

Najafi, Bijan, O'Driscoll, R. ORCID

logoORCID: <https://orcid.org/0000-0003-3995-0073>, Turicchi, J., Duarte, Cristiana ORCID logoORCID: <https://orcid.org/0000-0002-6566-273X>, Michalowska, J., Larsen, S. C., Palmeira, A. L., Heitmann, B. L., Horgan, G. W. and Stubbs, R. J. (2020) A novel scaling methodology to reduce the biases associated with missing data from commercial activity monitors. PLOS ONE, 15 (6). e0235144.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/5742/>

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version: <http://dx.doi.org/10.1371/journal.pone.0235144>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repository Policy Statement](#)

# RaY

Research at the University of York St John

For more information please contact RaY at [ray@yorks.ac.uk](mailto:ray@yorks.ac.uk)

RESEARCH ARTICLE

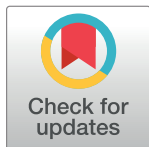
# A novel scaling methodology to reduce the biases associated with missing data from commercial activity monitors

R. O'Driscoll<sup>1</sup>\*, J. Turicchi<sup>1</sup>, C. Duarte<sup>1</sup>, J. Michalowska<sup>2</sup>, S. C. Larsen<sup>3</sup>, A. L. Palmeira<sup>4,5</sup>, B. L. Heitmann<sup>3,6,7</sup>, G. W. Horgan<sup>8</sup>, R. J. Stubbs<sup>1</sup>

**1** Appetite Control and Energy Balance Group, School of Psychology, University of Leeds, Leeds, United Kingdom, **2** Department of Treatment of Obesity, Metabolic Disorders and Clinical Dietetics, Medical Faculty, Poznan University of Medical Sciences, Poznan, Poland, **3** Research Unit for Dietary Studies, The Parker Institute, Bispebjerg and Frederiksberg Hospital, The Capital Region, Denmark, **4** Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Portugal, **5** Universidade Lusófona, Lisbon, Portugal, **6** Department of Public Health, Section for General Medicine, Copenhagen University, Copenhagen, Denmark, **7** Charles Perkins Centre, The Boden Institute, University of Sydney, Sydney, Australia, **8** Biomathematics & Statistics Scotland, Aberdeen, United Kingdom

\* These authors contributed equally to this work.

\* [psrod@leeds.ac.uk](mailto:psrod@leeds.ac.uk)



## OPEN ACCESS

**Citation:** O'Driscoll R, Turicchi J, Duarte C, Michalowska J, Larsen SC, Palmeira AL, et al. (2020) A novel scaling methodology to reduce the biases associated with missing data from commercial activity monitors. PLoS ONE 15(6): e0235144. <https://doi.org/10.1371/journal.pone.0235144>

**Editor:** Bijan Najafi, Baylor College of Medicine, UNITED STATES

**Received:** January 24, 2020

**Accepted:** June 9, 2020

**Published:** June 24, 2020

**Copyright:** © 2020 O'Driscoll et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** There are legal restrictions on sharing data these data because they contain potentially identifying or sensitive person information. The restrictions are imposed by The Danish Data Protection Agency (<https://www.datatilsynet.dk/english/>). Data used in the current study will be made available upon request after application to the NoHoW data controller (The James Hutton Institute: <https://www.hutton.ac.uk/>). The application procedure can be obtained from The James Hutton Institute ([DPO@hutton.ac.uk](mailto:DPO@hutton.ac.uk)).

## Abstract

### Background

Commercial physical activity monitors have wide utility in the assessment of physical activity in research and clinical settings, however, the removal of devices results in missing data and has the potential to bias study conclusions. This study aimed to evaluate methods to address missingness in data collected from commercial activity monitors.

### Methods

This study utilised 1526 days of near complete data from 109 adults participating in a European weight loss maintenance study (NoHoW). We conducted simulation experiments to test a novel scaling methodology (NoHoW method) and alternative imputation strategies (overall/individual mean imputation, overall/individual multiple imputation, Kalman imputation and random forest imputation). Methods were compared for hourly, daily and 14-day physical activity estimates for steps, total daily energy expenditure (TDEE) and time in physical activity categories. In a second simulation study, individual multiple imputation, Kalman imputation and the NoHoW method were tested at different positions and quantities of missingness. Equivalence testing and root mean squared error (RMSE) were used to evaluate the ability of each of the strategies relative to the true data.

### Results

The NoHoW method, Kalman imputation and multiple imputation methods remained statistically equivalent ( $p < 0.05$ ) for all physical activity metrics at the 14-day level. In the second simulation study, RMSE tended to increase with increased missingness. Multiple imputation

The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland UK Tel: +44 (0)344 928 5428.

**Funding:** The NoHoW study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 643309. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

showed the smallest RMSE for Steps and TDEE at lower levels of missingness (<19%) and the Kalman and NoHoW methods were generally superior for imputing time in physical activity categories.

## Conclusion

Individual centred imputation approaches (NoHoW method, Kalman imputation and individual Multiple imputation) offer an effective means to reduce the biases associated with missing data from activity monitors and maximise data retention.

## Introduction

Participation in physical activity and limiting sedentary behaviours is associated with increased total energy expenditure and potentially beneficial homeostatic matching of energy intake to energy expenditure [1]. As such, more active lifestyles are associated with a reduced risk of obesity [2], weight loss and prevention of weight regain following weight loss [3–5], as evidence suggests that weight maintenance is more readily achieved at higher degrees of energy flux [6]. Thus, the accurate and precise quantification of physical activity behaviours is critical to the study of overweight, obesity and associated comorbidities.

Accelerometry-based measures of physical activity have been available for a number of years [7]. Their objective nature offers a significant advantage over questionnaire-based assessments, which are biased by misreporting [8]. In current activity monitors, tri-axial piezoelectric sensors detect acceleration in anteroposterior, mediolateral and vertical axes and are used to objectively quantify human movement [9]. Technological advances in terms of size, data aggregation/storage capabilities and the associated fall in cost facilitates the use of tri-axial accelerometers in most new devices [9], as opposed to the uni-axial [10], bi-axial accelerometers [11] and burdensome battery packs required for earlier devices [12]. Taken together, these advances mean that it is increasingly feasible to objectively and continuously monitor the intra-day physical activity patterns of large groups of participants.

A well-recognised phenomenon in accelerometer research is missing data [13] attributable to behavioural (removal for aesthetic reasons) and non-behavioural reasons (device technical failures, charging). Non-wear time in accelerometers has previously been detected by defining periods in which the signal of acceleration in each axis falls below a threshold for some period of time, often a predefined period between 10–120 minutes [14,15]. Researchers then permit a maximum amount of non-wear time per day, which may be up to 14 hours [16]. The aim of defining such a period is to determine the amount of missing data which minimally influences the inferences of the study [17]. It is also common to define a minimum number of valid days within a measurement period and if these criteria are met, an average or total value for physical activity metrics can be estimated [18,19].

Missing accelerometer data may detrimentally influence the conclusions of a study in a number of ways. If physical activity summaries are calculated from incomplete data, true physical activity may be under-estimated (depending on the assumptions made about missing data). If missing periods occur in individuals that differ behaviourally or demographically from those with more complete data then the generalisability of the study's conclusions may be compromised [20]. A range of strategies have been developed with the aim of limiting the bias introduced by missing accelerometer data [21]. These methods make use of the observed (non-missing) data to build predictive models of missing data points and have utilised mean

imputation [22], combined multivariate strategies [23,24] or normalisation by the amount of wear-time [25,26].

Commercial activity monitors are increasingly prevalent in research environments and may be utilised in large cohorts and over long durations for assessment of physical activity. Commercial activity monitors are cloud-connected, facilitating the assessment of physical activity for longer time periods than research-grade equivalents (i.e. Actigraph GT3-x), which typically measure physical activity maximally over a single week [27]. Commercial activity monitors are also increasingly equipped with heart rate monitoring devices [28], which can facilitate the estimation of the relative intensity of physical activity or energy expenditure, through heart rate reserve (HRR) or flex methodologies [29–32] but also creates different patterns of missingness. For example, missing data may be identified through loss of contact with the wrist (and therefore no measured heart rate), inferring that the device has most likely been removed. This results in the detection of smaller windows of removal, compared to longer periods used when accelerometer signal is the determinant of missingness [14,15]. These differences highlight an important need to develop methods to limit the bias associated with missing data from these devices. There has been no attempt to develop or apply imputation methodologies to commercially available multisensory activity monitors (i.e. Fitbit charge 2; FC2).

The purpose of the present study is to propose and evaluate a methodology designed to minimise the bias introduced by missing data collected from a commercial activity monitor (FC2). Firstly, we conducted a series of intra-class correlation analyses to investigate the minimum data required to achieve a reasonably non-biased aggregation of physical activity data collected by a FC2. Next, the results of autocorrelation analyses are presented, which serve as the rationale for the development of a method which scales temporally proximate data to produce summaries over a given measurement period. Lastly, in a series of simulation experiments using real datasets with simulated missingness, we compared the performance of the proposed methodology to alternative imputation strategies.

## Materials and methods

### Participants

Data were collected as part of the NoHoW trial (ISRCTN88405328), an 18-month randomised 2x2 controlled trial testing the efficacy of an ICT based toolkit for weight loss maintenance across three European centres: United Kingdom, (Leeds), Denmark (Copenhagen), and Portugal (Lisbon). The NoHoW study received funding from the European Union's Horizon 2020 research and innovation programme (grant agreement number: 643309). The study was conducted in accordance with the Helsinki Declaration and ethical approval has been granted by local institutional ethics committees at the Universities of Leeds (17–0082; 27-Feb-2017), Lisbon (17/2016; 20-Feb-2017) and the Capital Region of Denmark (H-16030495; 8-Mar-2017) and all participants provided informed consent to have their data used for research purposes by this research team. Full details of the trial protocol have been published previously [33]. The NoHoW trial recruited 1,627 participants and some of the observational work reported in this study utilised the entire sample of NoHoW participants and when this is the case, this is specified in the manuscript.

For the simulation experiments conducted in this study, FC2 data from 109 participants each wearing a FC2 for 14 days (minutes = 2,197,440, hours = 36,624, days = 1526) were used. This sample was selected based on the quantity of non-wear time (<2.5% data missing within the first 14 days). Utilising a sample with minimal degrees of missingness allows 'true', near-complete data to be held back for comparison with imputation methods.

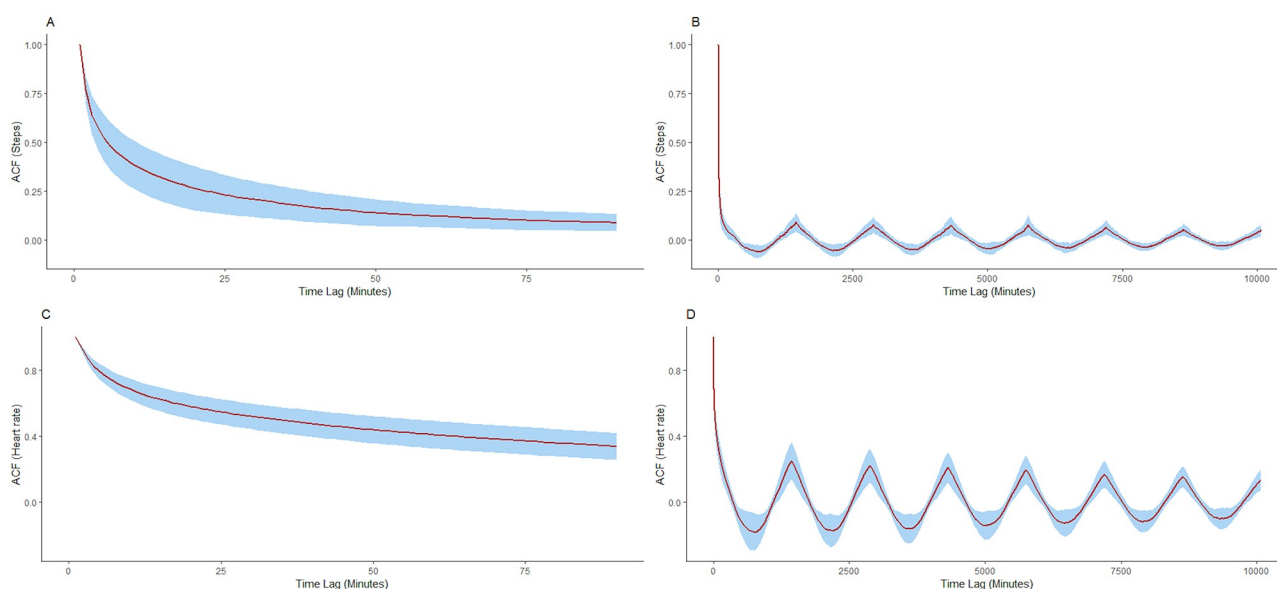
## Fitbit Charge 2 (FC2)

All participants enrolled in the NoHoW trial were provided with a FC2 (FC2; Fitbit Inc, San Francisco, CA, USA). The FC2 is a wrist-worn activity monitor which derives estimates of energy expenditure and physical activity based on data obtained from incorporated sensors and proprietary algorithms. The FC2 estimates of heart rate are obtained through a patented technology called 'PurePulse', which uses light-emitting diodes to monitor blood volume [28]. Data are aggregated to the minute-level and synced via the Fitbit mobile application to Fitbit servers through an application programming interface. In the present study, non-wear time is defined by the absence of a heart rate measure and all devices were set to 'auto' mode by default, which ensured that no heart rate reading was transmitted when the device was not on the wrist.

## Autocorrelation analyses

The algorithm proposed in this study was initially based on a series of autocorrelation analyses which are presented below. In autocorrelation analyses, the correlation between values in the time series are computed as a function of the time lag between them, defined in minutes in this case. For these analyses we calculated the autocorrelation value for all time lags of up to 7 days (10080 minutes) for each participant individually, thus indicating time points within a week with the highest correlation. Fig 1 illustrates the autocorrelation for steps and heart rate for 90 minutes and 10081 minutes, respectively.

The average of the autocorrelation values (ACF) reached within 60 minutes for steps are: 15 mins: ACF = 0.31, 30 mins: ACF = 0.21, 45 mins: ACF = 0.15, 60 mins: ACF = 0.12, comparatively, heart rate values are higher: 15 mins: ACF = 0.62, 30 mins: ACF = 0.52, 45 mins: ACF = 0.46, 60 mins: ACF = 0.41. Although there is evidence of periodic patterns on subsequent days, the value does not exceed ACF = 0.09 for steps, which is observed at a lag of 1441 minutes and ACF = 0.25 is observed for heart rate at 1440 minutes, the differences in these values are likely attributable to the stochastic nature of steps when compared to heart rate.



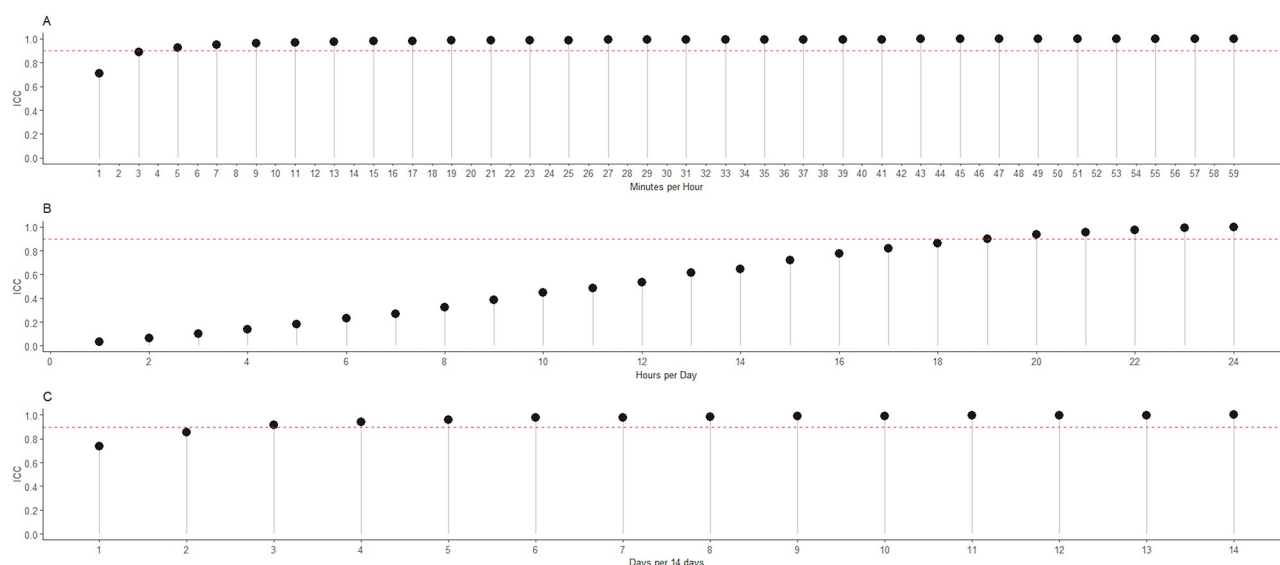
**Fig 1. Autocorrelation (ACF) values for steps with time lags of 90 minutes (A), 10,080 minutes (B) and heart rate with time lags of 90 minutes (C) and 10,080 minutes (D).** Average ACF values are shown in red and the blue ribbon represents  $\pm 1$  standard deviation.

<https://doi.org/10.1371/journal.pone.0235144.g001>

Notably, the value at 10081 mins (7 days) is  $ACF = 0.05$  for steps and  $ACF = 0.13$  for heart rate. Thus, the greatest autocorrelation values are observed locally for both steps and heart rate.

### Wear time requirements

In order to investigate the minimum amount of wear-time required for a valid hour, day or 14-day period, intraclass correlation (ICC) analyses were conducted, as ICC is a widely used and accepted means of determining measurement agreement [34]. In each of these experiments, data were deleted incrementally and at random and the ICC was calculated between the partially deleted data and the 'true' steps at each increment. An ICC threshold of 0.9 was used as the selection criterion to represent 10% similarity of true values [18]. We first investigated the minimum time required within a single hour with adjustment for wear time, and thus the remaining data was divided by the proportion of the wear time and this adjusted value was used for ICC analyses. In the daily and 14-day analyses, adjustments for wear time were not made. For all analyses, two-way mixed-effects agreement models were used [34] and this was conducted with the 'icc' function from the 'rel' package in R. Fig 2a demonstrates that if 5 minutes of data are present and scaled to 60 minutes, the ICC threshold of 0.9 is reached. In the daily analysis, the ICC threshold was met at 18–19 hours per day (Fig 2b). It is important to note that our ICC comparisons for each day include non-scaled data despite using scaled data in our algorithm (outlined below). When scaling by the proportion of wear time per day, the number of hours required will be lower. We utilise 18 hours to ensure that true data are available from different parts of the day (i.e. morning, afternoon, evening) and this is a conservative requirement in line with previous research [35]. To establish minimum 14-day requirements, the ICC threshold was met at 3 days (Fig 2c). For the final algorithm, we required 4 days including at least one weekend day as the minimum criteria for inclusion, owing to the potential for differential patterns of physical activity between weekdays and weekend days [36].



**Fig 2. Intraclass correlations (ICC) for incrementally deleted data and 'true' data.** Data are presented for scaled minutes per hour (A), for hours per day (B) and for number of days per 14 days (C).

<https://doi.org/10.1371/journal.pone.0235144.g002>



## NoHoW algorithm

Based on these analyses we propose a scaling algorithm, referred to from hereon as the ‘NoHoW algorithm’ as follows:

1. *If non-missing minutes per hour < 5 then remove hour from dataset else sum available minutes to provide hourly total*
2. Divide the number of available minutes per hour by 60 to give the proportion of wear time per hour
3. Divide hourly total by the proportion of wear time per hour to provide a scaled hourly total
4. *If available hours per day < 18 then remove day from dataset else sum all available hours to give daily total*
5. Divide the number of available hours by 24 to give proportion of wear time per day
6. Divide daily total by the proportion of wear time per day to provide a scaled daily total
7. *If available days per 14 days < 4 or < 1 weekend day then remove 14-day period from dataset else average all valid days*

## Simulation experiments

In order to test the algorithm, we performed two simulation experiments. In the first experiment, we tested traditional imputation methods as well as the proposed algorithm. This was achieved by creating datasets with simulated missingness from each of the included participant’s true data and holding back this true data to be compared to the imputed datasets. The time point at which the data were removed was random and the length of each deleted period was uniformly sampled between one and 120 minutes in duration. The decision to insert missing data at random positions was informed by observing the proportion of missing FC2 data for each hour in the first 14 days of the NoHoW study, on average 22.83% was missing with a range of 21.1% at 13:00–13:59 to 25.96% at 23:00–23:59 (S1 Fig). To determine the length of missing periods in this study, we quantified the length of each missing period in the first 14 days of the NoHoW study, where the length was less than an entire day (1440 minutes). Of the 146,165 missing periods, 139,213 (95.24%) were less than 60 minutes and 3882 (2.7%) were greater than 120 minutes (S2 Fig), thus we set 120 minutes as the upper limit for the length of insertions. The final parameter in the missing data algorithm was the number of missing periods, which was set to 40. This resulted in the amount of missing data per day being 13.7% (11.76% inserted) on average and ranging up to 44.4% (36.81% inserted) in simulation study 1.

Utilising the same simulated missing datasets, our first simulation study tested the methodologies below for dealing with missing data.

## Removal

The effect of no imputation or adjustment strategy was demonstrated by simply reporting the physical activity summaries for the simulated missing datasets.

**Mean imputation.** Missing data were imputed with the i) mean of all the remaining data and ii) with the mean of the individuals remaining data. This was conducted with the *Hmisc* package in R.

**Random forest imputation.** We performed random forest imputation, utilising the ‘*missForest*’ package in R. This is a non-parametric imputation method, which implements the

original random forest algorithm [37]. We performed random forest imputation to predict the missing values for steps, heart rate and calories on each participants data using weekday and hour as observed, non-missing variables. Hyperparameters were selected with consideration of computational feasibility; We utilised 100 trees in each forest, the number of randomly sampled variables at each split was set to the square root of the number of variables and the maximum number of iterations was set to 5.

### Multiple imputation

We tested multiple imputation with the use of bootstrapping and predictive mean matching utilising i) the entire sample and ii) individual-level data. In the case of the overall model, we utilised age, gender and day of the week as covariates, as they have previously been shown to be associated with differential patterns of physical activity [18,38]. In the individual models, hour of the day was used as an additional covariate. An advantage of multiple imputation is the repetition of the imputation process thus attempting to address the uncertainty associated with a single imputation. We utilised 5 imputations in the overall model, and in the individual level model we utilised 7 imputations. Multiple imputation was implemented with the *Hmisc* package in R.

### Kalman imputation

Lastly, we tested Kalman smoothing imputation using a structural time series model. Kalman imputation was implemented with the *imputeTS* package in R to impute caloric expenditure, steps and heart rate.

### Simulation study 2

In simulation study 2, we investigated how the bias introduced by the NoHoW algorithm, Kalman imputation and individual level multiple imputation may vary depending of the quantity and position of missing data. We chose to include these individual centred approaches as they were the only individualised approaches that were statistically equivalent to the true data across all activity types in simulation study 1. As in the first simulation study, we utilised 14-days (20160 minutes) of data for each participant. We simulated missingness randomly throughout the day and in all iterations, the maximum length of each insertion was set to 120 minutes. The simulations were split in to 10 windows of missingness, where the number of missing periods inserted for each participant increased incrementally with each simulation window. In the first window, the number of missing periods per participant was sampled from a uniform distribution between 0–10, the second between 10–20 up to the tenth which inserted 90–100 missing periods in each iteration. Within each window 20 simulations were conducted per participant, for a total of 21,800 iterations of each algorithm overall.

### Physical activity metrics

Each of the imputation methods tested in both simulation studies were used to address a number of distinct physical activity metrics including total steps, total daily energy expenditure (TDEE) and minutes of sedentary, light, moderate and vigorous physical activity. Both steps and TDEE for a given interval are extracted from the FC2 and time in each of sedentary, light, moderate and vigorous are defined by the heart rate reserve (HRR) method which is computed for each minute in the dataset. To facilitate this method, we estimated maximum heart rate for each participant using the Tanaka method;  $(208 - 0.7 \times \text{age})$  [39]. To define resting heart rate, we first determined sleeping heart rate, which was defined as the mean of the lowest 20



consecutive minutes observed between 00:00 and 08:00 am, when steps/min were  $< 5$ . After sleeping heart rate was defined, an 8% increase was used to approximate resting heart rate as this represents a typical difference between resting and sleeping heart rate [40]. Relative intensity of each minute was then calculated:

$$\%HRR = \frac{(HR - HR_{REST})}{(HR_{MAX} - HR_{REST})} * 100 \quad (1)$$

The following cut points for were applied: Sedentary ( $<20\%$  HRR), light ( $20\text{--}40\%$  HRR), moderate ( $40\text{--}60\%$  HRR), and vigorous ( $\geq 60\%$  HRR) [32]. For each missing minute in the dataset, each of the imputation methods described above were used to impute or scale steps, caloric expenditure and heart rate to produce hourly, daily and average physical activity estimates.

### Statistical analysis

All data are presented as means and standard deviations unless otherwise stated and a flow-chart detailing both simulation studies is available in S2 Fig. To evaluate the performance of each method, root mean squared error (RMSE) was calculated for all physical activity metrics for hourly, daily and 14-day averages, relative to the observed data. Where RMSE is defined as:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (2)$$

Where  $\hat{y}_i$  refers to predicted values,  $y_i$  refers to the true values and  $n$  refers to the number of observations. Equivalence tests were performed to investigate whether the models were statistically equivalent to the true data. To be considered equivalent, the 90% confidence interval of the estimate must fall within  $\pm 10\%$  of the criterion mean. Simulation study 1 was conducted on an intel i7-8750H with 32GB RAM and 12 logical processors. Simulation study 2 was undertaken on ARC3, part of the High-Performance Computing cluster at the University of Leeds, UK. Statistical analyses were conducted with R version 3.6.3 using a p-value of  $< 0.05$  to determine statistical significance.

### Results

The participants meeting the minimum criteria were predominantly female ( $n = 93$ , male = 16) and were primarily from the Danish centre (DK = 69, UK = 23, Portugal = 17), Table 1 presents the demographic and physical activity results for the included sample.

The computation time for each of the included algorithms in the first simulation were as follows: Overall mean imputation: 18.23 Minutes, Individual mean imputation: 1.27 Minutes, Overall multiple imputation: 17.61 Hours, Individual multiple imputation: 17.04 Minutes, Random forest imputation: 4.36 Hours, Kalman imputation: 2.16 Minutes, NoHoW method: 2.12 Seconds.

Table 2 illustrates the results of the first simulation study for 14-day, daily and hourly comparisons and Table 3 presents the results of equivalence tests for each of the methods. For TDEE, Individual multiple imputation had the smallest RMSE for 14-day (36.32 kcal), followed by the NoHoW method (39.51 kcal), and for the hourly comparison, Kalman imputation was superior (14.11 kcal). In the daily comparison the smallest RMSE was observed for the NoHoW method (115.86 kcal). All methods except removal (mean difference: -343.44 kcal) were statistically equivalent to the true data, with the smallest mean difference observed for Individual multiple imputation. For steps, the lowest RMSE was observed for the NoHoW

**Table 1. Demographic data and physical activity averages for the included sample (n = 109).** Total daily energy expenditure (TDEE) is presented in kcal/day, sedentary, light, moderate and vigorous are presented in minutes/day.

	Mean $\pm$ SD	Minimum	Maximum
Age	47.46 $\pm$ 9.62	22	75
Height	1.69 $\pm$ 0.08	1.54	1.87
Weight	84.76 $\pm$ 15.59	50.5	148.4
BMI	29.64 $\pm$ 5	20.2	44.8
TDEE	2626.59 $\pm$ 504.66	1754.24	4492.25
Steps	10570.34 $\pm$ 3208.67	3202.50	19941.07
Sedentary	1087.76 $\pm$ 112.72	847.21	1284.64
Light	266.77 $\pm$ 94.83	102.29	484.14
Moderate	50.24 $\pm$ 31.6	6.43	132.86
Vigorous	7.29 $\pm$ 9.09	0.00	47.07

<https://doi.org/10.1371/journal.pone.0235144.t001>

method for 14-day (397.83 steps) and daily comparison (1366.92 steps) and Kalman imputation for hourly comparison (173.78 steps). All methods except removal (mean difference: -1320.74 steps,  $p$ -value  $>0.05$ ), were statistically equivalent to the true data. In the HRR analysis, multiple imputation methods, Kalman imputation and the NoHoW algorithm were statistically equivalent for all sedentary, light, moderate and vigorous comparisons.

In the second simulation study, which is visually represented as boxplots in Fig 3, the aggregated RMSE for each of the tested approaches tended to increase with the proportion of missing data. For the TDEE estimation (Fig 3A), the first iteration (1% missingness added) resulted in a mean RMSE of 31.14 kcal/day for the NoHoW method (range 28.82–33.12 kcal/day) compared to multiple imputation: 21.30 kcal/day (range 19.20–23.11 kcal/day) and Kalman imputation: 37.44 kcal/day (range 35.49–39.90 kcal/day). Comparatively, at the 10th insertion of missingness (~28% missingness added) a maximum RMSE of 68.89 kcal/day, 68.05 kcal/day and 72.55 kcal/day was observed for NoHoW, multiple imputation and Kalman imputation, respectively. For steps (Fig 3B), evidence of slightly superior performance was observed for multiple imputation at the lower levels of missingness ( $<19\%$ ). However, mean RMSE values for each of the methods remained similar and did not differ by more than 86 steps/day. In the HRR analysis, differences were the greatest in the sedentary comparison (Fig 3C), with the NoHoW and Kalman methods having a lower mean RMSE than multiple imputation at each window. The largest difference was observed at 28% missingness, where the mean RMSE values were 24.87 mins/day (range: 23.15–26.39 mins/day) for the NoHoW method, 55.56 (range 53.69–57.76) mins/day for multiple imputation and 23.73 mins/day (range 21.46–26.89 mins/day) for Kalman imputation. For light (Fig 3D) and moderate (Fig 3E) the NoHoW method showed the lowest mean RMSE values after 13% missingness. Its largest mean RMSE of 15.19 mins/day (range 12.81–17.42 mins/day) for light activity and 5.38 mins/day (range 4.72–6.26 mins/day) for moderate activity were observed at 28% missingness. Lastly, in the vigorous activity simulation (Fig 3F), multiple imputation had the lowest mean RMSE with  $<7\%$  added missingness but Kalman and NoHoW methods were superior at higher levels of missingness. In the 28% missingness window, NoHoW reached a mean RMSE of 2.25 mins/day (range 1.84–3.03 mins/day) mins/day and Kalman reached 2.28 mins/day (range 1.85–2.95 mins/day). Results of the second simulation study are available in S1 Table.

## Discussion

The use of commercial activity monitors in research environments is proliferating, creating new research opportunities, however, it is critical to take steps to ensure the integrity of these

**Table 2. Mean  $\pm$  standard deviation estimates and Root Mean Squared Error (RMSE) for each of the imputation methods tested in simulation study 1. Total daily energy expenditure (TDEE) is presented in kcals, sedentary, light, moderate and vigorous are presented in minutes.**

Comparison	TRUE	Removal	Overall mean	Individual mean	Overall Multiple	Individual Multiple	Random Forest	Kalman	NoHoW
TDEE	2626.59 $\pm$ 504.66	228.15 $\pm$ 445.78	2645.66 $\pm$ 443.93	2645.09 $\pm$ 515.87	2649.59 $\pm$ 457.11	2638.06 $\pm$ 513.24	2658.48 $\pm$ 571.72	2660.96 $\pm$ 518.63	2653.61 $\pm$ 515.8
	RMSE	351.27	71.30	40.97	67.52	36.32	117.37	49.57	39.51
Day	2626.59 $\pm$ 607.05	228.3.15 $\pm$ 583.3	2645.66 $\pm$ 545.14	2645.49 $\pm$ 602.91	2649.59 $\pm$ 555.65	2638.06 $\pm$ 600.29	2658.48 $\pm$ 654.09	2660.96 $\pm$ 632.27	2653.61 $\pm$ 627.21
	RMSE	413.74	134.48	117.53	132.73	116.36	172.07	134.71	115.86
hour	109.7 $\pm$ 65.24	100.12 $\pm$ 65.22	110.23 $\pm$ 61.53	110.25 $\pm$ 62.03	110.36 $\pm$ 61.7	110.01 $\pm$ 62.31	110.61 $\pm$ 63.2	110.69 $\pm$ 66.86	110.61 $\pm$ 67.38
	RMSE	28.00	15.03	14.51	15.21	14.66	16.55	14.11	15.13
Steps	10570.34 $\pm$ 3208.67	9249.6 $\pm$ 2867.46	10718.22 $\pm$ 3860.93	10716.67 $\pm$ 3309.78	10741.09 $\pm$ 2817.02	10593.71 $\pm$ 3274.98	10049.5 $\pm$ 3472.95	10755.97 $\pm$ 3249.79	10791.34 $\pm$ 3300.09
	RMSE	1412.90	519.53	426.36	591.06	398.18	864.48	432.66	397.83
Day	10570.34 $\pm$ 4775.05	9249.6 $\pm$ 4447	10718.22 $\pm$ 4560.78	10716.67 $\pm$ 4657.6	10741.09 $\pm$ 4343.52	10593.71 $\pm$ 4637.29	10049.5 $\pm$ 4804.18	10755.97 $\pm$ 5013.85	10791.34 $\pm$ 5009.86
	RMSE	2049.36	1407.11	1388.44	1453.05	1387.16	1603.70	1588.95	1366.92
hour	442.55 $\pm$ 738.52	405.61 $\pm$ 699.45	446.55 $\pm$ 694.48	446.58 $\pm$ 696.27	447.17 $\pm$ 695.47	442.93 $\pm$ 697.7	428.19 $\pm$ 700.15	448.23 $\pm$ 751.38	449.94 $\pm$ 763.31
	RMSE	198.65	176.66	175.54	180.37	179.96	184.75	173.78	188.66
Sedentary	1087.76 $\pm$ 112.72	956.05 $\pm$ 101.01	1139.11 $\pm$ 109.98	1151.53 $\pm$ 105.13	1118.75 $\pm$ 100.63	1120.96 $\pm$ 110.35	1138.12 $\pm$ 112.52	1101.93 $\pm$ 117.73	1105.57 $\pm$ 116.03
	RMSE	133.22	74.14	68.34	39.84	35.50	58.43	19.72	20.50
Day	1087.76 $\pm$ 170.63	956.05 $\pm$ 174.15	1139.11 $\pm$ 160.95	1151.53 $\pm$ 156.73	1118.75 $\pm$ 152.3	1120.96 $\pm$ 157.97	1138.12 $\pm$ 165.97	1101.93 $\pm$ 175.74	1105.57 $\pm$ 173.68
	RMSE	158.92	93.09	86.87	61.30	56.77	78.45	47.30	42.33
hour	45.51 $\pm$ 17.82	41.93 $\pm$ 19.1	47.03 $\pm$ 16.83	47.38 $\pm$ 16.61	46.46 $\pm$ 16.44	46.54 $\pm$ 16.52	47.01 $\pm$ 17	45.99 $\pm$ 18.04	46.07 $\pm$ 17.89
	RMSE	10.57	7.05	6.71	5.68	5.55	6.59	4.79	4.59
Light	266.77 $\pm$ 94.83	235.5 $\pm$ 84.53	247.25 $\pm$ 93.15	236.96 $\pm$ 86.51	264.11 $\pm$ 84.32	261.72 $\pm$ 91.8	247.05 $\pm$ 94.01	269.32 $\pm$ 96.42	274.66 $\pm$ 96.81
	RMSE	34.02	48.07	35.76	20.66	11.20	30.39	10.63	10.98
Day	266.77 $\pm$ 139.25	235.5 $\pm$ 127.82	247.25 $\pm$ 134.98	236.96 $\pm$ 129.6	264.11 $\pm$ 126.65	261.72 $\pm$ 131.03	247.05 $\pm$ 137.41	269.32 $\pm$ 143.71	274.66 $\pm$ 144.45
	RMSE	46.30	61.21	48.15	35.59	29.24	45.61	32.25	27.43
hour	11.17 $\pm$ 14.15	10.33 $\pm$ 13.59	10.65 $\pm$ 13.86	10.36 $\pm$ 13.62	11.11 $\pm$ 13.47	11.03 $\pm$ 13.54	10.63 $\pm$ 13.92	11.2 $\pm$ 14.54	11.44 $\pm$ 14.69
	RMSE	3.94	4.77	4.02	3.58	3.38	4.23	3.32	3.51
Moderate	50.24 $\pm$ 31.6	44.63 $\pm$ 28.42	44.63 $\pm$ 28.42	44.63 $\pm$ 28.42	48.81 $\pm$ 28.63	48.25 $\pm$ 31.18	44.65 $\pm$ 28.43	48.76 $\pm$ 31.17	52.22 $\pm$ 33.37
	RMSE	6.83	6.83	6.83	4.90	3.90	6.80	3.81	3.79
Day	50.24 $\pm$ 47.85	44.63 $\pm$ 43.95	44.63 $\pm$ 43.95	44.63 $\pm$ 43.95	48.81 $\pm$ 44.07	48.25 $\pm$ 45.54	44.65 $\pm$ 43.95	48.76 $\pm$ 49.15	52.22 $\pm$ 51.02
	RMSE	11.38	11.38	11.38	10.20	9.68	11.38	12.73	10.29
hour	2.11 $\pm$ 5.53	1.96 $\pm$ 5.29	1.96 $\pm$ 5.29	1.96 $\pm$ 5.29	2.07 $\pm$ 5.29	2.06 $\pm$ 5.3	1.96 $\pm$ 5.29	2.07 $\pm$ 5.67	2.18 $\pm$ 5.88
	RMSE	1.31	1.31	1.31	1.29	1.28	1.31	1.49	1.59
Vigorous	7.29 $\pm$ 9.09	6.51 $\pm$ 8.38	6.51 $\pm$ 8.38	6.51 $\pm$ 8.38	6.92 $\pm$ 8.35	6.86 $\pm$ 9.13	6.51 $\pm$ 8.38	7.17 $\pm$ 9.03	7.55 $\pm$ 9.57
	RMSE	1.44	1.44	1.44	1.29	1.28	1.44	1.00	1.16
Day	7.29 $\pm$ 15.48	6.51 $\pm$ 14.61	6.51 $\pm$ 14.61	6.51 $\pm$ 14.61	6.92 $\pm$ 14.59	6.86 $\pm$ 14.99	6.51 $\pm$ 14.61	7.17 $\pm$ 15.66	7.55 $\pm$ 16.62
	RMSE	4.05	4.05	4.05	4.03	4.05	4.05	4.03	4.62
hour	0.3 $\pm$ 2.45	0.29 $\pm$ 2.36	0.29 $\pm$ 2.36	0.29 $\pm$ 2.36	0.3 $\pm$ 2.36	0.3 $\pm$ 2.37	0.29 $\pm$ 2.36	0.31 $\pm$ 2.52	0.32 $\pm$ 2.6
	RMSE	0.47	0.47	0.47	0.48	0.48	0.47	0.52	0.60

<https://doi.org/10.1371/journal.pone.0235144.t002>

**Table 3. Mean  $\pm$  standard deviation estimates and equivalence test results for each of the imputation methods tested in simulation study 1.** Total daily energy expenditure (TDEE) is presented in kcal, sedentary, light, moderate and vigorous are presented in minutes. Bounds refers to the equivalence boundaries and p-value upper and lower refers to equivalence tests at the upper and lower equivalence bounds.

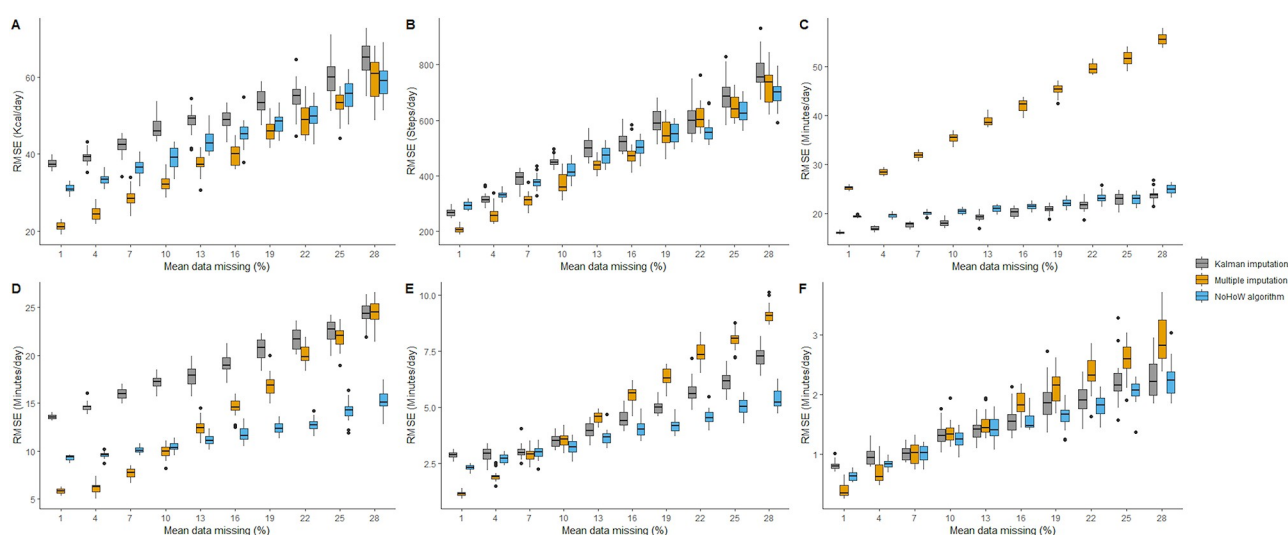
		TRUE	Imputed	Mean difference	Bounds	P-value lower	P-value upper
<b>TDEE</b>	Removal	2626.59 $\pm$ 504.66	2283.15 $\pm$ 445.78	-343.44	$\pm$ 262.66	1	0
	Overall mean	2626.59 $\pm$ 504.66	2645.66 $\pm$ 443.93	19.08	$\pm$ 262.66	0	0
	Individual mean	2626.59 $\pm$ 504.66	2645.49 $\pm$ 515.87	18.9	$\pm$ 262.66	0	0
	Overall Multiple	2626.59 $\pm$ 504.66	2649.59 $\pm$ 457.11	23.01	$\pm$ 262.66	0	0
	Individual Multiple	2626.59 $\pm$ 504.66	2638.06 $\pm$ 513.24	11.48	$\pm$ 262.66	0	0
	Random Forest	2626.59 $\pm$ 504.66	2658.48 $\pm$ 571.72	31.89	$\pm$ 262.66	0	0
	Kalman	2626.59 $\pm$ 504.66	2660.96 $\pm$ 518.63	34.37	$\pm$ 262.66	0	0
	NoHoW	2626.59 $\pm$ 504.66	2653.61 $\pm$ 515.8	27.02	$\pm$ 262.66	0	0
	Removal	10570.34 $\pm$ 3208.67	9249.6 $\pm$ 2867.46	-1320.74	$\pm$ 1057.03	1	0
<b>Steps</b>	Overall mean	10570.34 $\pm$ 3208.67	10718.22 $\pm$ 2860.93	147.88	$\pm$ 1057.03	0	0
	Individual mean	10570.34 $\pm$ 3208.67	10716.67 $\pm$ 3309.78	146.33	$\pm$ 1057.03	0	0
	Overall Multiple	10570.34 $\pm$ 3208.67	10741.09 $\pm$ 2817.02	170.75	$\pm$ 1057.03	0	0
	Individual Multiple	10570.34 $\pm$ 3208.67	10593.71 $\pm$ 3274.98	23.37	$\pm$ 1057.03	0	0
	Random Forest	10570.34 $\pm$ 3208.67	10049.5 $\pm$ 3472.95	-520.84	$\pm$ 1057.03	0	0
	Kalman	10570.34 $\pm$ 3208.67	10755.97 $\pm$ 3249.79	185.63	$\pm$ 1057.03	0	0
	NoHoW	10570.34 $\pm$ 3208.67	10791.34 $\pm$ 3309.09	221	$\pm$ 1057.03	0	0
	Removal	1087.76 $\pm$ 112.72	956.05 $\pm$ 101.01	-131.71	$\pm$ 108.78	1	0
	Overall mean	1087.76 $\pm$ 112.72	1139.11 $\pm$ 109.98	51.36	$\pm$ 108.78	0	0
<b>Sedentary</b>	Individual mean	1087.76 $\pm$ 112.72	1151.53 $\pm$ 105.13	63.78	$\pm$ 108.78	0	0
	Overall Multiple	1087.76 $\pm$ 112.72	1118.75 $\pm$ 100.63	30.99	$\pm$ 108.78	0	0
	Individual Multiple	1087.76 $\pm$ 112.72	1120.96 $\pm$ 110.35	33.2	$\pm$ 108.78	0	0
	Random Forest	1087.76 $\pm$ 112.72	1138.12 $\pm$ 112.52	50.36	$\pm$ 108.78	0	0
	Kalman	1087.76 $\pm$ 112.72	1101.93 $\pm$ 117.73	14.17	$\pm$ 108.78	0	0
	NoHoW	1087.76 $\pm$ 112.72	1105.57 $\pm$ 116.03	17.81	$\pm$ 108.78	0	0
	Removal	266.77 $\pm$ 94.83	235.5 $\pm$ 84.53	-31.27	$\pm$ 26.68	1	0
	Overall mean	266.77 $\pm$ 94.83	247.25 $\pm$ 93.15	-19.51	$\pm$ 26.68	0.046	0
	Individual mean	266.77 $\pm$ 94.83	236.96 $\pm$ 86.51	-29.8	$\pm$ 26.68	0.949	0
<b>Light</b>	Overall Multiple	266.77 $\pm$ 94.83	264.11 $\pm$ 84.32	-2.65	$\pm$ 26.68	0	0
	Individual Multiple	266.77 $\pm$ 94.83	261.72 $\pm$ 91.8	-5.05	$\pm$ 26.68	0	0

(Continued)

Table 3. (Continued)

		TRUE	Imputed	Mean difference	Bounds	P-value lower	P-value upper
	Random Forest	266.77 ± 94.83	247.05 ± 94.01	-19.72	± 26.68	0.001	0
	Kalman	266.77 ± 94.83	269.32 ± 96.42	2.55	± 26.68	0	0
	NoHoW	266.77 ± 94.83	274.66 ± 96.81	7.89	± 26.68	0	0
<b>Moderate</b>	Removal	50.24 ± 31.6	44.63 ± 28.42	-5.61	± 5.02	0.938	0
	Overall mean	50.24 ± 31.6	44.63 ± 28.42	-5.61	± 5.02	0.938	0
	Individual mean	50.24 ± 31.6	44.63 ± 28.42	-5.61	± 5.02	0.938	0
	Overall Multiple	50.24 ± 31.6	48.81 ± 28.63	-1.44	± 5.02	0	0
	Individual Multiple	50.24 ± 31.6	48.25 ± 31.18	-1.99	± 5.02	0	0
<b>Vigorous</b>	Random Forest	50.24 ± 31.6	44.65 ± 28.43	-5.59	± 5.02	0.933	0
	Kalman	50.24 ± 31.6	48.76 ± 31.17	-1.48	± 5.02	0	0
	NoHoW	50.24 ± 31.6	52.22 ± 33.37	1.98	± 5.02	0	0
	Removal	7.29 ± 9.09	6.51 ± 8.38	-0.78	± 0.73	0.672	0
	Overall mean	7.29 ± 9.09	6.51 ± 8.38	-0.78	± 0.73	0.672	0
	Individual mean	7.29 ± 9.09	6.51 ± 8.38	-0.78	± 0.73	0.672	0
	Overall Multiple	7.29 ± 9.09	6.92 ± 8.35	-0.37	± 0.73	0.002	0
	Individual Multiple	7.29 ± 9.09	6.86 ± 9.13	-0.43	± 0.73	0.005	0
	Random Forest	7.29 ± 9.09	6.51 ± 8.38	-0.78	± 0.73	0.672	0
	Kalman	7.29 ± 9.09	7.17 ± 9.03	-0.12	± 0.73	0	0
	NoHoW	7.29 ± 9.09	7.55 ± 9.57	0.26	± 0.73	0	0

<https://doi.org/10.1371/journal.pone.0235144.t003>



**Fig 3. Boxplots detailing Root Mean Squared Error (RMSE) values from simulation study 2 for each window of missingness.** Data are presented for TDEE (A), Steps (B), Sedentary (C), Light (D), Moderate (E), Vigorous (F). Mean missing data refers to the additional data added in the simulations.

<https://doi.org/10.1371/journal.pone.0235144.g003>

data is not challenged by missing data. The purpose of the present study was to develop and test a methodology to account for missingness in physical activity data collected with a commercial activity monitor in a free-living environment. In our initial experiments, we utilised ICC analyses to show that if data are scaled within an hour, the relative data requirements to meet an ICC threshold of 0.9 are minimal (~5 minutes). This relates to the relative similarity between 'local' data points, as confirmed by our autocorrelation analyses. We also show that if the data are not scaled by wear time the relative requirements for a day equates to approximately 18 hours per day. This is in contrast to a previous study, which showed that relative to a 14 hours/day criterion, at least 13 hours/day of accelerometer data are required [41]. This slight discrepancy in the proportion of the day required may relate to the inclusion of night hours in our sample. Given the likelihood that this is a highly sedentary period, missing data at night is likely to be less influential on daily totals.

In simulation study 1, we used each of the tested methods to impute metrics that are likely to be of importance depending on the specific research aims. Our results suggest differential outcomes depending on the metric selected, for instance, random forest imputation, overall mean and individual mean methods did not impute vigorous or moderate minutes regularly, as reflected in the non-significant equivalent results (indicating these methods are not statistically equivalent). This is likely due to the low proportion of the day in which these activities are performed. In the first simulation study, we observed a slight tendency for the NoHoW method to overestimate minutes of moderate and vigorous activity. This may relate to the position of the missing data in simulation 1; For example, if missing data occurs in the sedentary period after an exercise bout then this period will be overestimated. As exercise is infrequent in non-athlete populations this is unlikely to result in a large error in mean differences. Indeed, the estimates for moderate and vigorous differed by < 2 minutes/day in the 14-day comparison. Researchers should consider imputation strategies based on observed activity data from their sample or should select methodologies which are statistically equivalent in the specific activities of interest.

We have also shown that all tested methods for all comparisons resulted in a RMSE which was lower than no imputation (i.e. removal). Making no attempt to adjust for missingness effectively assumes that activity was 0 and our results demonstrate the potential implications of this. In our first study, ~14% of the day was missing on average with ~12% inserted, equating to a wear time of 20–22 hours, which falls within the acceptable levels of missingness for most accelerometer research [14,15] and therefore evidences the importance of using one of these methods even in the case of relatively small quantities of missing data. Of the imputation methods tested, an advantage of individual-centred methods was observed, specifically Kalman imputation, individual multiple imputation and the NoHoW algorithm. Indeed, in our second simulation study, in which the maximal missingness approached double the quantity of our first simulation study the RMSE for TDEE was lower than the values observed for removal, overall mean and random forest imputation in simulation study 1, indicating the efficacy of these methods.

We simulated missingness evenly throughout the entire 24-hour period in relation to the observed patterns of missingness in the NoHoW trial. This is contrary to a previous study observing that missing data patterns more frequently occur at the beginning and end of the day [42]. It is of note that we utilised wrist-worn devices compared to the aforementioned study, which utilised hip worn accelerometers. Unlike wrist-worn monitors, hip-worn accelerometers are generally removed with changing of clothes. This may encourage compliance [43] and contribute to a more uniform distribution of missingness throughout the day.

We consider the relative computational simplicity of the NoHoW method to be a significant advantage. Accelerometer data of this kind can be extremely high volume and researchers



must select their imputation strategy with consideration of both error reduction and computational feasibility. It may be possible to utilise advanced machine learning techniques to impute missing data, but these methods are computationally expensive and may be technically inaccessible to many researchers. In addition, more information (e.g. physiological, psychological or behavioural factors) may allow for more accurate multivariate imputation techniques but in free-living widescale settings this information is likely to be limited, thus our method is likely to be widely applicable. A further advantage of the present study is the testing of numerous activity metrics in addition to steps. Steps are a highly interpretable and relatable metric produced by wearable devices and some evidence suggests that estimates of steps from Fitbit devices are more valid and reliable than other derived variables, i.e. TDEE [44–46] although machine learning techniques may facilitate the refinement of energy expenditure estimates [47]. Nevertheless, the metric of interest to researchers will vary depending on the aims and hypotheses of a study and we demonstrate that the NoHoW method, Kalman imputation and individual level multiple imputation perform particularly well across a variety of physical activity metrics.

Key limitations of the present study are the utilisation of participants with a high proportion of wear time (>97.5%). Whilst highly adherent participants were required in order to have a near-complete dataset to validate against, we cannot rule out the possibility that the included participants are in some way behaviourally different from the participants that remove the FC2 more frequently. Second, we inserted missing data at random positions, and it remains uncertain how representative this is of free-living data in other studies. Participants may remove devices for comfort, aesthetic reasons, charging or under conditions where they would not wish to have measurements made (e.g. extreme sedentariness) and thus, it is possible that missingness is not completely at random [48] and may differ between populations and research studies. Unfortunately, no definitive method exists to test if data are missing at random [49] and many imputation strategies have limited capabilities to overcome this. However, our second simulation study simulates a wide variety of missing patterns in an attempt to identify such biases and worst-case scenarios in the selected methods.

Incorporation of activity monitoring devices is a necessary step in improving physical activity and energy balance tracking in research and clinical settings. We have proposed a simple and accessible methodology which effectively reduces the bias introduced to physical activity estimates by non-wear time and may improve the validity of research conclusions. Other imputation strategies (i.e. multiple imputation and Kalman imputation) performed comparatively well and importantly, all the methods tested in this study are superior to data removal. Researchers and clinicians utilising commercial activity monitors to monitor physical activity longitudinally should account for missingness in datasets and the algorithm presented in this study offers an approach to this.

## Supporting information

**S1 Table. Aggregated results for each window of missingness for each physical activity metric in simulation study 2.** Difference refers to the difference in means between the imputed and true value. Abbreviations: SD: standard deviations, RMSE: Root mean squared error.

(DOCX)

**S1 Fig. The percentage of missing data for each hour of the day in the NoHoW trial.**

(DOCX)

**S2 Fig. A density plot detailing the lengths of missing data (<1440 minutes in length) in the NoHoW trial.** The mean is represented by the red dashed line and the median is represented by the blue dashed line.

(DOCX)

**S1 Data. A flowchart detailing the simulation procedures conducted in this study.**

(PDF)

## Author Contributions

**Conceptualization:** R. O'Driscoll, J. Turicchi, J. Michalowska, B. L. Heitmann, G. W. Horgan.

**Data curation:** R. O'Driscoll.

**Formal analysis:** R. O'Driscoll.

**Funding acquisition:** A. L. Palmeira, B. L. Heitmann, R. J. Stubbs.

**Investigation:** R. O'Driscoll, C. Duarte, J. Michalowska, S. C. Larsen, A. L. Palmeira, G. W. Horgan.

**Methodology:** R. O'Driscoll, J. Turicchi, A. L. Palmeira, G. W. Horgan, R. J. Stubbs.

**Project administration:** S. C. Larsen, A. L. Palmeira, B. L. Heitmann, R. J. Stubbs.

**Resources:** J. Michalowska.

**Software:** R. O'Driscoll, G. W. Horgan.

**Supervision:** C. Duarte, B. L. Heitmann.

**Validation:** R. J. Stubbs.

**Visualization:** R. O'Driscoll.

**Writing – original draft:** R. O'Driscoll, J. Turicchi, R. J. Stubbs.

**Writing – review & editing:** R. O'Driscoll, J. Turicchi, C. Duarte, J. Michalowska, S. C. Larsen, B. L. Heitmann, G. W. Horgan, R. J. Stubbs.

## References

1. Beaulieu K, Hopkins M, Blundell J, Finlayson G. Impact of physical activity level and dietary fat content on passive overconsumption of energy in non-obese adults. *Int J Behav Nutr Phys Act.* 2017; 14: 14. <https://doi.org/10.1186/s12966-017-0473-3> PMID: 28166797
2. Swift DL, McGee JE, Earnest CP, Carlisle E, Nygard M, Johannsen NM. The Effects of Exercise and Physical Activity on Weight Loss and Maintenance. *Prog Cardiovasc Dis.* 2018; 61: 206–213. <https://doi.org/10.1016/j.pcad.2018.07.014> PMID: 30003901
3. Kerns JC, Guo J, Fothergill E, Howard L, Knuth ND, Brychta R, et al. Increased Physical Activity Associated with Less Weight Regain Six Years After “The Biggest Loser” Competition. *Obesity.* 2017; 25: 1838–1843. <https://doi.org/10.1002/oby.21986> PMID: 29086499
4. Wadden TA, Neiberg RH, Wing RR, Clark JM, Delahanty LM, Hill JO, et al. Four-year weight losses in the look AHEAD study: Factors associated with long-term success. *Obesity.* 2011; 19: 1987–1998. <https://doi.org/10.1038/oby.2011.230> PMID: 21779086
5. Schoeller DA, Shay K, Kushner RF. How much physical activity is needed to minimize weight gain in previously obese women? *Am J Clin Nutr.* 1997; 66: 551–556. <https://doi.org/10.1093/ajcn/66.3.551> PMID: 9280172
6. Drenowatz C, Greier K. The Role of Energy Flux in Weight Management. *Exerc Med.* 2017; 1: 4. <https://doi.org/10.26644/em.2017.004>

7. Sirard JR, Melanson EL, Li L, Freedson PS. Field evaluation of the Computer Science and Applications, Inc. physical activity monitor. *Med Sci Sports Exerc.* 2000; 32: 695–700. <https://doi.org/10.1097/00005768-200003000-00022> PMID: 10731015
8. Helmerhorst HJF, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *Int J Behav Nutr Phys Act.* 2012; 9: 103. <https://doi.org/10.1186/1479-5868-9-103> PMID: 22938557
9. Hills AP, Mokhtar N, Byrne NM. Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures. *Front Nutr.* 2014; 1: 1–16. <https://doi.org/10.3389/fnut.2014.00001> PMID: 25988106
10. Swartz AM, Strath SJ, Bassett DRJ, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc.* 2000; 32: S450–6. <https://doi.org/10.1097/00005768-200009001-00003> PMID: 10993414
11. Whybrow S, Ritz P, Horgan GW, Stubbs RJ. An evaluation of the IDEEATM activity monitor for estimating energy expenditure. *Br J Nutr.* 2013; 109: 173–183. <https://dx.doi.org/10.1017/S0007114512000645> PMID: 22464547
12. Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: New insights and validation studies. *Obes Rev.* 2013; 14: 451–462. <https://doi.org/10.1111/obr.12021> PMID: 23398786
13. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008; 40: 181–8. <https://doi.org/10.1249/mss.0b013e31815a51b3> PMID: 18091006
14. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc.* 2011; 43: 357–364. <https://doi.org/10.1249/MSS.0b013e3181ed61a3> PMID: 20581716
15. Ridgers ND, Fairclough S. Assessing free-living physical activity using accelerometry: Practical issues for researchers and practitioners. *Eur J Sport Sci.* 2011; 11: 205–213. <https://doi.org/10.1080/17461391.2010.501116>
16. Tudor-Locke C, Camhi SM, Troiano RP. A catalog of rules, variables, and definitions applied to accelerometer data in the national health and nutrition examination Survey, 2003–2006. *Prev Chronic Dis.* 2012; 9: E113. <https://doi.org/10.5888/pcd9.110332> PMID: 22698174
17. Liu B, Yu M, Graubard BI, Troiano RP, Schenker N. Multiple imputation of completely missing repeated measures data within person from a complex sample: application to accelerometer data in the National Health and Nutrition Examination Survey. *Stat Med.* 2016; 35: 5170–5188. <https://doi.org/10.1002/sim.7049> PMID: 27488606
18. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. Buchowski M, editor. *PLoS One.* 2017; 12: e0169649. <https://doi.org/10.1371/journal.pone.0169649> PMID: 28146576
19. Kapteyn A, Banks J, Hamer M, Smith JP, Steptoe A, Van Soest A, et al. What they say and what they do: Comparing physical activity across the USA, England and the Netherlands. *J Epidemiol Community Health.* 2018; 72: 471–476. <https://doi.org/10.1136/jech-2017-209703> PMID: 29643112
20. Loprinzi PD, Cardinal BJ, Crespo CJ, Brodowicz GR, Andersen RE, Smit E. Differences in demographic, behavioral, and biological variables between those with valid and invalid accelerometry data: Implications for generalizability. *J Phys Act Heal.* 2013; 10: 79–84. <https://doi.org/10.1123/jpah.10.1.79>
21. Stephens S, Beyene J, Tremblay MS, Faulkner G, Pullnayegum E, Feldman BM. Strategies for Dealing with Missing Accelerometer Data. *Rheum Dis Clin North Am.* 2018; 44: 317–326. <https://doi.org/10.1016/j.rdc.2018.01.012> PMID: 29622298
22. Meng Y, Speier W, Shufelt C, Joung S, E Van Eyk J, Bairey Merz CN, et al. A Machine Learning Approach to Classifying Self-Reported Health Status in a Cohort of Patients with Heart Disease Using Activity Tracker Data. *IEEE J Biomed Heal Informatics.* 2020; 24: 878–884. <https://doi.org/10.1109/JBHI.2019.2922178> PMID: 31199276
23. Lee PH. Data imputation for accelerometer-measured physical activity: The combined approach. *Am J Clin Nutr.* 2013; 97: 965–971. <https://doi.org/10.3945/ajcn.112.052738> PMID: 23553165
24. Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc.* 2012; 44: S61–S67. <https://doi.org/10.1249/MSS.0b013e3182399e0f> PMID: 22157776
25. Katapally TR, Muhajarine N. Towards uniform accelerometry analysis: a standardization methodology to minimize measurement bias due to systematic accelerometer wear-time variation. *J Sports Sci Med.* 2014; 13: 379–86. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24790493> PMID: 24790493

26. Chen C, Jerome GJ, Laferriere D, Young DR, Vollmer WM. Procedures used to standardize data collected by RT3 triaxial accelerometers in a large-scale weight-loss trial. *J Phys Act Health*. 2009; 6: 354–9. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19564665> PMID: 19564665
27. Thraen-Borowski KM, Gennuso KP, Cadmus-Bertram L. Accelerometer-derived physical activity and sedentary time by cancer type in the United States. *PLoS One*. 2017; 12. <https://doi.org/10.1371/journal.pone.0182554> PMID: 28806753
28. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the fitbit charge 2 for monitoring heart rate. *PLoS One*. 2018; 13: e0192691. <https://doi.org/10.1371/journal.pone.0192691> PMID: 29489850
29. Rennie KL, Hennings SJ, Mitchell J, Wareham NJ. Estimating energy expenditure by heart-rate monitoring without individual calibration. *Med Sci Sports Exerc*. 2001; 33: 939–945. <https://doi.org/10.1097/00005768-200106000-00013> PMID: 11404659
30. Silva AM, Santos DA, Matias CN, Júdice PB, Magalhães JP, Ekelund U, et al. Accuracy of a combined heart rate and motion sensor for assessing energy expenditure in free-living adults during a double-blind crossover caffeine trial using doubly labeled water as the reference method. Buchowski M, editor. *Eur J Clin Nutr*. 2015; 69: 20–27. <https://doi.org/10.1038/ejcn.2014.51> PMID: 24690589
31. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc*. 2012; 44: S32–S38. <https://doi.org/10.1249/MSS.0b013e3182399cf7> PMID: 22157772
32. Schrack JA, Leroux A, Fleg JL, Zipunnikov V, Simonsick EM, Studenski SA, et al. Using Heart Rate and Accelerometry to Define Quantity and Intensity of Physical Activity in Older Adults. *Journals Gerontol—Ser A Biol Sci Med Sci*. 2018; 73: 668–675. <https://doi.org/10.1093/gerona/gly029> PMID: 29509832
33. Scott SE, Duarte C, Encantado J, Evans EH, Harjuma M, Heitmann BL, et al. The NoHoW protocol: A multicentre 2×2 factorial randomised controlled trial investigating an evidence-based digital toolkit for weight loss maintenance in European adults. *BMJ Open*. 2019; 9: e029425. <https://doi.org/10.1136/bmjopen-2019-029425> PMID: 31575569
34. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016; 15: 155–63. <https://doi.org/10.1016/j.jcm.2016.02.012> PMID: 27330520
35. Shook RP, Hand GA, O'Connor DP, Thomas DM, Hurley TG, Hébert JR, et al. Energy Intake Derived from an Energy Balance Equation, Validated Activity Monitors, and Dual X-Ray Absorptiometry Can Provide Acceptable Caloric Intake Data among Young Adults. *J Nutr*. 2018; 148: 490–496. <https://doi.org/10.1093/jn/nxx029> PMID: 29546294
36. Shiroma EJ, Lee IM, Schepps MA, Kamada M, Harris TB. Physical Activity Patterns and Mortality: The Weekend Warrior and Activity Bouts. *Med Sci Sports Exerc*. 2019; 51: 35–40. <https://doi.org/10.1249/MSS.0000000000001762> PMID: 30138219
37. Breiman L. Random Forests. *Mach Learn*. 2001; 45: 5–32. <https://doi.org/10.1023/A:1010933404324>
38. Berkemeyer K, Wijndaele K, White T, Cooper AJM, Luben R, Westgate K, et al. The descriptive epidemiology of accelerometer-measured physical activity in older adults. *Int J Behav Nutr Phys Act*. 2016; 13: 1–10. <https://doi.org/10.1186/s12966-015-0325-y> PMID: 26733186
39. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol*. 2001; 37: 153–156. [https://doi.org/10.1016/s0735-1097\(00\)01054-8](https://doi.org/10.1016/s0735-1097(00)01054-8) PMID: 11153730
40. Kräuchi K, Wirz-Justice A. Circadian Clues to Sleep Onset Mechanisms. *Neuropsychopharmacology*. 2001; 25: S92–S96. [https://doi.org/10.1016/S0893-133X\(01\)00315-3](https://doi.org/10.1016/S0893-133X(01)00315-3) PMID: 11682282
41. Herrmann SD, Barreira TV., Kang M, Ainsworth BE. How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates. *J Phys Act Heal*. 2013.
42. Yue Xu S, Nelson S, Kerr J, Godbole S, Patterson R, Merchant G, et al. Statistical approaches to account for missing values in accelerometer data: Applications to modeling physical activity. *Stat Methods Med Res*. 2018; 27: 1168–1186. <https://doi.org/10.1177/0962280216657119> PMID: 27405327
43. Diaz KM, Krupka DJ, Chang MJ, Shaffer JA, Ma Y, Goldsmith J, et al. Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. *BMC Res Notes*. 2016; 9: 213. <https://doi.org/10.1186/s13104-016-2020-8> PMID: 27068022
44. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Young Yoo J, et al. Accuracy of fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR mHealth uHealth*. 2018; 6: e10527. <https://doi.org/10.2196/10527> PMID: 30093371
45. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *Br J Sports Med*. 2020; 54: 332–340. <https://doi.org/10.1136/bjsports-2018-099643> PMID: 30194221

46. O'Driscoll R, Turicchi J, Hopkins M, Gibbons C, Larsen SC, Palmeira AL, et al. The validity of two widely used commercial and research-grade activity monitors, during resting, household and activity behaviours. *Health Technol (Berl)*. 2020; 10: 637–648. <https://doi.org/10.1007/s12553-019-00392-7>
47. O'Driscoll R, Turicchi J, Hopkins M, Horgan GW, Finlayson G, Stubbs JR. Improving energy expenditure estimates from wearable devices: A machine learning approach. *J Sports Sci*. 2020; 00: 1–10. <https://doi.org/10.1080/02640414.2020.1746088> PMID: 32252598
48. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*. 2009; 339: 157–160. <https://doi.org/10.1136/bmj.b2393> PMID: 19564179
49. Ae Lee J, Gill J. Missing value imputation for physical activity data measured by accelerometer. *Stat Methods Med Res*. 2018; 27: 490–506. <https://doi.org/10.1177/0962280216633248> PMID: 26994215