



HHS Public Access

Author manuscript

Softw Impacts. Author manuscript; available in PMC 2024 June 13.

Published in final edited form as:

Softw Impacts. 2023 November ; 18: . doi:10.1016/j.simpa.2023.100577.

acc: An R package to process, visualize, and analyze accelerometer data

Jaejoon Song^{a,*}, Michael D. Swartz^b, Karen Basen-Engquist^c

^aCenter for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

^bDepartment of Biostatistics and Data Science, The University of Texas, Houston, TX, USA

^cDepartment of Behavioral Science, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

Wearable activity monitors are now widely used in behavioral and epidemiological studies to measure physical activity in free-living conditions. Despite the widespread use in research, the development of software to explore the data collected from these devices has been limited. We present *acc*, a comprehensive, free, and open-source R package to provide a seamless environment for exploring accelerometer data (<https://cran.r-project.org/web/packages/acc>). In this article, we demonstrate the software for processing, visualizing, and analyzing accelerometer data using real and simulated datasets.

Keywords

Physical activity monitoring; Accelerometer; Wearable sensor

Code metadata

Current code version	1.3.3
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2023-319

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Corresponding author. jaejoon.song@fda.hhs.gov (J. Song).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.simpa.2023.100577>.

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

Disclaimer

This manuscript reflects the views of the authors and should not be construed to represent the views or policies of the FDA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Permanent link to reproducible capsule	https://codeocean.com/capsule/4784632/tree/v1
Legal code license	<i>GPL-2,GPL-3</i>
Code versioning system used	<i>Git</i>
Software code languages, tools and services used	<i>R, C++</i>
Compilation requirements, operating environments and dependencies	<i>R >= 2.10</i>
If available, link to developer documentation/manual	
Support email for questions	jaejoonsong@gmail.com

1. Introduction

Accelerometers are wearable devices used to collect real-time physical activity data in free-living conditions. The devices have been used in clinical and observational studies to complement the traditional survey-based measures of physical activity [1–3]. The R package *acc* is designed to be a comprehensive software for (i) importing and processing, (ii) visualizing, (iii) simulating, and (iv) summarizing both uni-axial and tri-axial accelerometer data. The following code can be used to install the R package *acc*.

```
R> install.packages("acc")
```

2. The structure of accelerometer data

In accelerometers, activity intensities are recorded in a continuous manner with a corresponding time stamp. Two types of data can be extracted from accelerometers.

2.1. Activity counts data

The activity count is a measure of activity intensity that can be extracted from Actigraph (ActiGraph, LLC, Pensacola, FL, USA) devices. The activity counts quantify activity intensities over short periods of time (e.g., one minute). The value of activity counts ranges from zero to infinity and can be translated into meaningful physical activity categories.

2.2. Raw acceleration data

More recently, physical activity monitors include a high-resolution tri-axial acceleration sensor. These sensors can log human motion in three orthogonal directions in sub-second resolution. For example, the UK Biobank [10] collected 7-day physical activity for 100,000 individuals in real-life conditions using a tri-axial accelerometer (AX3; Axivity LTD, Newcastle, UK) which logged activity in units of gravitational constant g (9.8 m/s^2) every hundredth of a second.

3. Simulating epoch-level data

We provide two functions to simulate accelerometer data. First, the function *simAcc* generates activity intensity data for any epoch size (e.g., per minute). Second, the function *simPA* generates daily minutes of physical activity with realistic missing data patterns. We highlight the usage of the two functions in the following sections.

3.1. Simulating epoch-level data

Epoch level data can be generated from a hidden Markov model. Activity intensities are simulated from multiple latent states (e.g., (i) non-wear, (ii) sedentary activity, and (iii) moderate to vigorous physical activity (MVPA)). Each of the underlying latent states are assumed to have a normal distribution with mean μ_j and standard deviation σ_j . The user provides the means and the standard deviations for the underlying activity states, as well as a transition matrix for probability of change from one state to the other. Below is an example using the function `simAcc` to simulate activity counts data with three activity states: (i) non-wear, (ii) sedentary, and (iii) MVPA.

```
R> library(acc)

R> # Specifying the transition probability matrix (tpm) for change from one state to the
other

R> tpm <- matrix(c(0.95,0.04,0.01,0.09,0.9,0.01,0.1,0.2,0.7), byrow='TRUE', nrow=3)

R> # Simulating minute-level data for one day (24 hrs.) using the tpm specified above

R> sim <- simAcc(timelength=(60*24), tpm=tpm,
```

```
+   epoch="1 min", seedset=33,
+   startDate=ISOdate(2017,1,1,hour=0),
+   endDate=ISOdate(2017,1,2,hour=0),
+   mu=c(0,30,2500), sigma=c(0,30,1000))
```

We also provide three default settings to generate epoch level data. The user can choose between MVPA levels of 'low', 'moderate', or 'high', to conveniently simulate data using pre-specified parameters (see Appendix 1 or the package manual for details).

3.2. Simulating day level data

Day level data, such as minutes of physical activity per day, can be generated from a non-homogeneous Poisson process. This formulation allows us to simulate realistic physical activity monitoring data over multiple days with missing data. Using the function `simPA`, the user provides the maximum and minimum number of observation days (i.e., using options `minday` and `maxday`). Data can be generated from two scenarios when: (i) missing data patterns are independent of the physical activity process (`type='noninf'`), and (ii) missing data patterns are dependent on the physical activity process (`type='inf'`). This is indicated by the argument `type`. Below is an example of generating data from scenario (i) for a single individual (`n=1`), over 7 scheduled observation days.

```
R> # Example generating data for one individual with non-informative missing data

R> simDays <- simPA(n=1, type='noninf', minday=6, maxday=7)
```

Details of using functions `simAcc` and `simPA` to simulate epoch and day level physical activity data can be found in Appendix 1 and in the package manual.

4. Processing accelerometer data

The R package `acc` provides functions to import and process both uni-axial and tri-axial accelerometer data. We highlight the usage of the functions in the following sections.

4.1. Functions to read activity counts data

The R package `acc` can read activity counts from various filetypes (e.g., `.agd`, `.csv`, or `.dat`) using the function `readCounts`. The user specifies the file name and path to the function.

```
R> accData1 <- readCounts("C:/mydata.agd")
```

In addition, the import function can be performed in a batch mode to read multiple files of a certain file type in a folder using the function `readCountsBatch` (see Appendix 2).

4.2. Data processing procedure for activity counts data

The procedure for accelerometer data processing can be outlined as the following [2–5].

1. First, observations suspected to be non-human signals (i.e., ‘spurious observations’) are identified and replaced with missing data indicators. Spurious observations are defined as a single non-zero epoch bordered before and after consecutive zero epochs (e.g., 60 min) [6].
2. Second, non-spurious observations are classified as either wear or non-wear time [6–9]. Common definitions of non-wear time classify extensive consecutive 1-minute zero epochs (e.g., at least 60 min) to be considered as non-wear time.
3. Third, types of physical activity are identified in pre-defined bouts, among the wear time. For instance, moderate to vigorous physical activities are defined based on published cut-points (e.g., 1952 for moderate-vigorous physical activity by [10]). In other words, consecutive 1-minute epochs that persist (e.g., 1952) for at least the pre-specified bout size (e.g., ten minutes) and allowance (e.g., allowance of up to two epochs below the cut-point inside the pre-specified bout size), are considered to be bouts of moderate to vigorous physical activity.
4. Fourth, total wear time and minutes of activity (sum of the number of epochs in bouts) are summarized by day [2–4].

The data processing rules can be applied using the function `accSummary`. Data processing can be also performed in a batch mode, to obtain physical activity summaries for multiple individuals using function `accBatch`. Details of the input parameters as well as the sample output from our functions `accSummary` and `accBatch` can be found in Appendix 2. We provide a working example with simulated data. Data are simulated using the function `simAcc` to generate minute-level activity counts for seven consecutive days. Then we use function `acc` to process the data.

```
R> # Generating data for three activity states: (i) non-wear, (ii) sedentary, (iii) MVPA
```

```
R> tpm <- matrix(c(0.95,0.04,0.01, 0.09,0.9,0.01, 0.1,0.2,0.7), byrow='TRUE', nrow=3)
R> mu <- c(0, 30, 2500)
R> sigma <- c(0, 30, 1000)
R> simData <- simAcc(timelength=60*24*7, tpm=tpm, mu=mu, sigma=sigma)
R> summary1 <- acc(data=simData, spuriousDef=60,
```

```
+ patype=c('Sedentary','MVPA'),
+ pacut=c(c(0,99),c(1952,Inf)),
+ boutsize=c(10,10),
+ tolerance=c('FALSE','TRUE'))
```

5. Visualizing accelerometer data

Data visualization is an important component for exploratory analysis and data cleaning. Accelerometer data can be visualized in the R package `acc`: using function `plotAcc` to visualize physical activity levels across the entire day (Fig. 1, panel a), using function `racePlot` to visualize daily activity minutes (Fig. 1, panel b), and using function `visGroup` to generate physical activity summaries using boxplots (Fig. 1, panel c) or heatmap (Fig. 1, panel d). Full details of the function usage and input parameters are provided in Appendix 3 and in the package manual.

6. Impact overview

Despite the widespread use of accelerometers in research to measure physical activity in free living conditions, development of software to explore the data collected from these devices has been limited. Unique contributions of our R package include: (i) functions to process various types of accelerometry data (e.g., activity counts, raw accelerometer data), (ii) functions to simulate realistic accelerometer data, and (iii) extensive data visualization options.

The functions in R package `acc` has been used extensively in a publication [11] to demonstrate the performance of a new methodology to account for irregular accelerometer wear time [6–9]. Details of the methodology is described in Appendix 4 and in the publication [11].

Furthermore, some of the key functions in the `acc` package were written in the C++ language for efficiency. The computational speed gain by writing function in the C++ language was substantial. Data processing speed for our package `acc` is illustrated in Appendix 5, with sample code to further expedite the implementation using parallel processing. Such efficiency would benefit any users tasked with processing a large accelerometer data.

We believe that our open-source R package `acc` will be useful to: (i) researchers who have collected accelerometer data for their own study, (ii) researchers working with public data sources to study physical activity behavior in the population, or (iii) anyone who is interested in developing methodology or software for wearable sensor data.

7. Limitations and future research directions

Recently, there has been increased interest in utilizing wearable sensors in clinical trials to derive clinical endpoints [12]. One of the major challenges in deriving clinical endpoints from wearable physical activity monitors has been associated with addressing missing data arising from the nature of collecting data in free-living conditions [12]. Several studies have investigated the methods to handle missingness in physical activity data measured using wearable sensors, for the purpose of deriving clinical endpoints [13,14].

Currently, the R package `acc` does not include functionalities to perform missing data imputation. Future version of the R package `acc` can consider including a comprehensive functionality to address missing data in accelerometer measured physical activity data to support development of valid and reliable clinical endpoints.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This study was supported by the National Institutes of Health/ National Cancer Institute, USA through R01 CA 109919, R25T CA057730, R25E CA056452, P30 CA016672 (M. D. Anderson's Cancer Center Support Grant and PROSPR Shared Resource), and the Center for Energy Balance in Cancer Prevention and Survivorship, Duncan Family Institute for Cancer Prevention and Risk Assessment, USA.

References

- [1]. Sudlow C, Gallacher J, Allen N, et al. , UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age, *PLoS Med.* 12 (3) (2015) e1001779.
- [2]. Troiano RP, Berrigan D, Dodd KW, et al. , Physical activity in the united states measured by accelerometer, *Med. Sci. Sports Exerc* 40 (1) (2008) 181–188. [PubMed: 18091006]
- [3]. Troiano RP, McClain JJ, Brychta RJ, Chen KY, Evolution of accelerometer methods for physical activity research, *Br. J. Sports Med* 48 (1019) (2014) 1023.
- [4]. Evenson KR, Terry JW, Assessment of differing definitions of accelerometer nonwear time, *Res. Q. Exerc. Sport* 80 (2) (2009) 355–362. [PubMed: 19650401]
- [5]. Choi L, Liu Z, Matthews CE, Buchowski MS, Validation of accelerometer wear and nonwear time classification algorithm, *Med. Sci. Sports Exerc* 3 (2) (2011) 357–364.
- [6]. Evenson KR, Buchner DM, Morland KM, Objective measurement of physical activity and sedentary behavior among us adults aged 60 years or older, *Prevent. Chronic Dis* 9 (2012) 110109.
- [7]. Robertson W, Stewart-Brown S, Wilcock E, et al. , Utility of accelerometers to measure physical activity in children attending an obesity treatment intervention, *J. Obesity* 2011 (2011) 398918.
- [8]. Perry MA, Hendrick PA, Hale L, Utility of the rt3 triaxial accelerometer in free living: an investigation of adherence and data loss, *Applied Ergon.* 41 (3) (2010) 469–476.

- [9]. Kirby J, Tibbins C, Callens C, et al. , Young peoples views on accelerometer use in physical activity research: Findings from a user involvement investigation, *Obesity* 2012 (2012) 948504.
- [10]. Freedson P, Melanson E, Sirard J, Calibration of the computer sciences and applications, inc. accelerometer, *Med. Sci. Sports Exerc* 30 (1998) 77781.
- [11]. Song J, Swartz MD, Gabriel KP, et al. , A semiparametric model for wearable sensor-based physical activity monitoring data with informative device wear, *Biostatistics* 20 (2) (2019) 287–298. [PubMed: 29415194]
- [12]. Di J, Demanuele C, Ketterman A, et al. , Considerations to address missing data when deriving clinical trial endpoints from digital health technologies, *Contemp. Clin. Trials* 113 (2022) (2022) 106661.
- [13]. Tachney MS, Cook DG, Stahl D, et al. , A framework for handling missing accelerometer outcome data in trials, *Trials* 22 (379) (2021).
- [14]. Lee JA, Gill J, Missing value imputation for physical activity data measured by accelerometer, *Stat. Methods Med. Res* 27 (2) (2018) 490–506. [PubMed: 26994215]

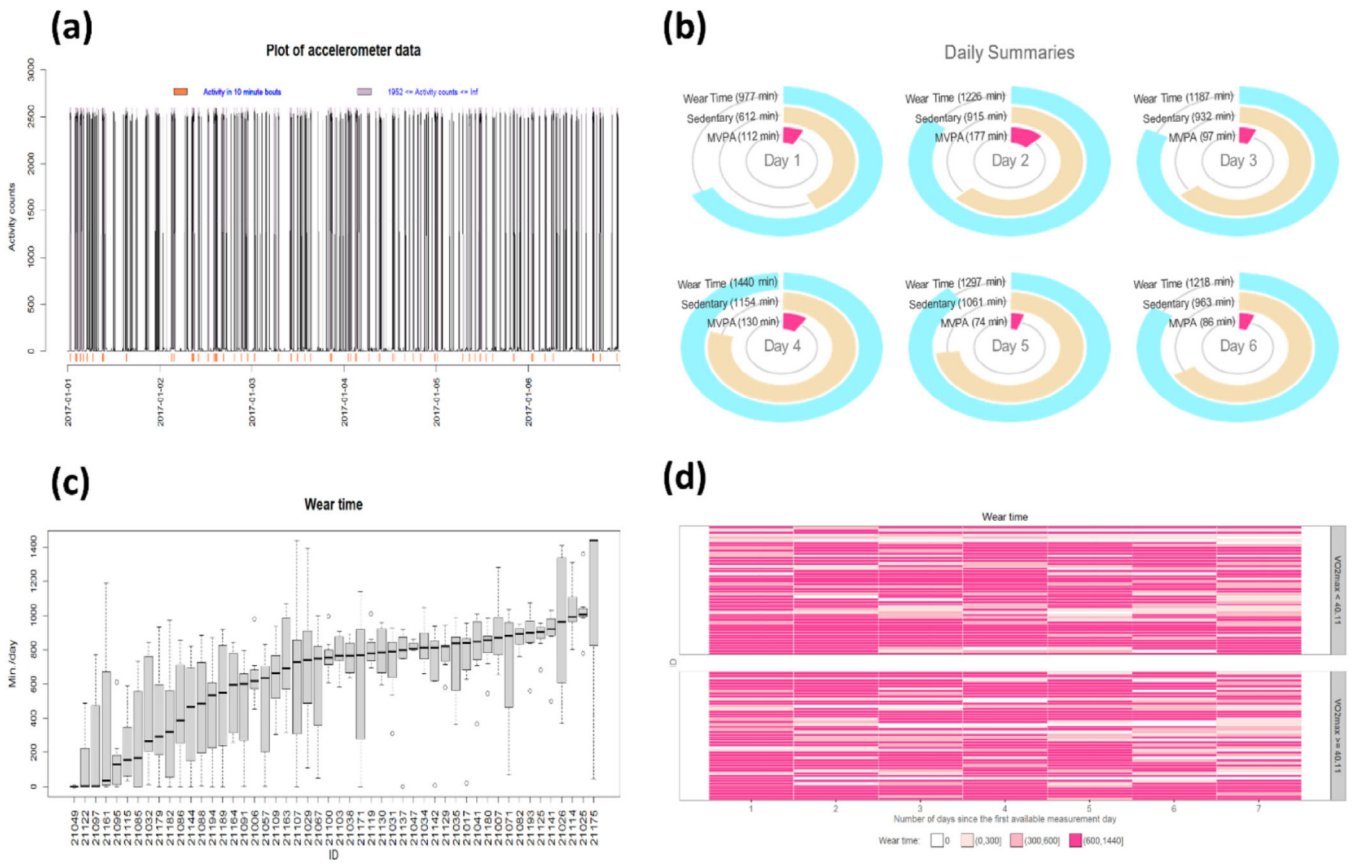


Fig. 1. Accelerometer data visualization available in the R package acc.