





Walkbotr: Extracting ‘Walk Bouts’ from GPS and Accelerometry Data for Physical Activity Research

Lauren Blair Wilner¹ , Weipeng Zhou², Philip M Hurvitz³, Anne Vernez Moudon[†]⁴, Bumjoon Kang⁵, Brian E Saelens^{6,7}, Jimmy Phuong², Matthew Dekker¹, and Stephen J Mooney¹ 

¹ Department of Epidemiology, University of Washington, ² Biomedical Informatics and Medical Education, University of Washington, ³ Center for Studies in Demography and Ecology, University of Washington, ⁴ College of Built Environments, Urban Design and Planning, University of Washington (†deceased), ⁵ Department of Architecture and Architectural Engineering, Seoul National University, ⁶ Center for Child Health, Behavior and Development, Seattle Children's Research Institute, ⁷ Department of Pediatrics, University of Washington

Walking and measurement in public health research

Walking is the most common form of physical activity and a behavior of key interest for urban planners, health promotion researchers, and rehabilitation medicine practitioners. Data collected from monitoring devices, such as Global Positioning System (GPS) trackers and accelerometers, hold considerable public health research potential (Feng & Timmermans, 2013, Troped et al. (2008)). By analyzing patterns in individual energy expenditure and movement, these data can be used to objectively measure walking and its effects, unlocking a researcher's ability to identify encouragement for, and barriers to, this key cardioprotective behavior across geospatial contexts and populations (Kang et al., 2013, Jankowska et al. (2015)).

While GPS devices and accelerometers provide the gold standard measurement approach for walking, processing accelerometer traces and GPS device tracks to identify walking is a computational and algorithmic challenge. In their raw form – a series of timestamps, locations, and accelerometer counts – monitoring data are rarely of direct researcher interest and may also contain subject-identifying location data. Further, it can be challenging to process these data efficiently to identify behaviors of interest (e.g. periods of walking). Some methods to identify travel using personal monitoring data (e.g. PALMS, and its successor HABITUS) process data on secure servers (Carlson et al., 2015); however, researchers whose privacy agreements with study participants preclude storing data on 3rd party servers may prefer a local R package. To address this gap, we developed a package, walkbotr, that implements a previously validated algorithm to extract patterns in monitoring data consistent with walking (Kang et al., 2013) that may be safely shared across research teams. walkbotr allows researchers to (1) process raw personal time stamp-linked GPS and accelerometry data for identifying

THE DENOMINATOR



Citation:

Wilner, L. B., Zhou, W., Hurvitz, P. M., Moudon†, A. V., Kang, B., Saelens, B. E., Phuong, J., Dekker, M., & Mooney, S. J. (2026). Walkbotr: Extracting ‘Walk Bouts’ from GPS and Accelerometry Data for Physical Activity Research. <https://doi.org/DOI pending>

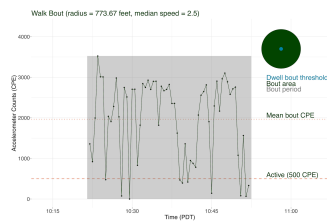
Published: 2026-03-16

License: CC-BY 2.0

Template based on [LaPreprint for Typst](#) by Rowan Cockett (MIT License).

periods and locations of walking and (2) create a de-identified summary of walking behavior that can be used in research and practice.

Related Download



[Walkbotr: Extracting 'Walk Bouts' from GPS and Accelerometry Data for Physical Activity Research](#) Walking is a common form of exercise and a popular intervention of interest for population health researchers. GPS devices and accelerometers, which can be worn by study participants to provide accurate information on walking duration, speed, and intensity, are an efficient and cost-effective method for measuring walking characteristics. The outcome of research on walking can inform policy... [read more](#)

Definition of a walk bout

A walk bout is defined as a period of activity in which (1) an accelerometer trace indicates movement consistent with walking (e.g., based on speed) and (2) a GPS track from the corresponding period indicates movement through space (e.g., based on distance traveled). The idea of a walk bout derives from the physical activity literature, in which monitored time is partitioned into activity bouts and inactivity bouts. **A walk bout is a physical activity bout in which both activity count range and GPS track is consistent with walking** (Kang et al., 2013).

To identify physical activity bouts, the package first classifies each non-overlapping discrete time window of accelerometry observations, or epochs, as active or inactive following Troiano et al.'s (2008) approach. Next, a physical activity bout is measured as any contiguous set of epochs that contains a cumulative sequence of active epochs that are preceded and followed by sequences of consecutive inactive epochs. These active bouts indicate periods of time in which the wearer appears to be physically active, though not necessarily walking (e.g., they could be playing a sport or working out).

Physical activity bouts are further classified as a walk bout or dwell bout based on GPS tracks. A dwell bout is a physical activity bout in which an individual did not leave a pre-specified radius (Kang et al., 2013). A walk bout is physical activity bout that is not a dwell bout, has a median speed consistent with that of walking, and has sufficient GPS coverage to indicate mobility rather than stasis.

Using the walkbotr package

I. Identifying physical activity bouts

Prior to use in the package, observed accelerometry measurements along each axis in three-dimensional space are processed to remove noise and gravitational effects, and the total count per epoch (CPE) is calculated as the sum of absolute values across all three axes. Once accelerometry measurements are broken down into total activity counts per time period (epochs), they are ready to be used in walkbotr. Epochs with a CPE above a selected threshold are considered active; for example, epochs 30 seconds in length with a CPE of 500 or higher are classified as active by the walkbotr package. While this threshold is relatively low compared with other physical activity research, it allows for measurement of slow walking. (Note: epoch length and CPE threshold can be set by the package user.)

A physical activity bout is identified as a sequence of consecutive epochs that is preceded and followed by a collection of consecutive inactive epochs. First, the data are divided into potential physical activity bouts and non-bouts using a run-length encoding algorithm to identify subsequences of epochs with at least 4 consecutive active epochs. Physical activity bouts are then identified from the potential bouts by filtering to those with 10 consecutive active epochs preceded and followed by at least 4

inactive epochs, as shown in Figure 1. These thresholds can be changed by the package user if desired. Note, walkbotr considers a physical activity bout to include the relevant active epochs along with the 3 inactive epochs immediately following the final active epoch.

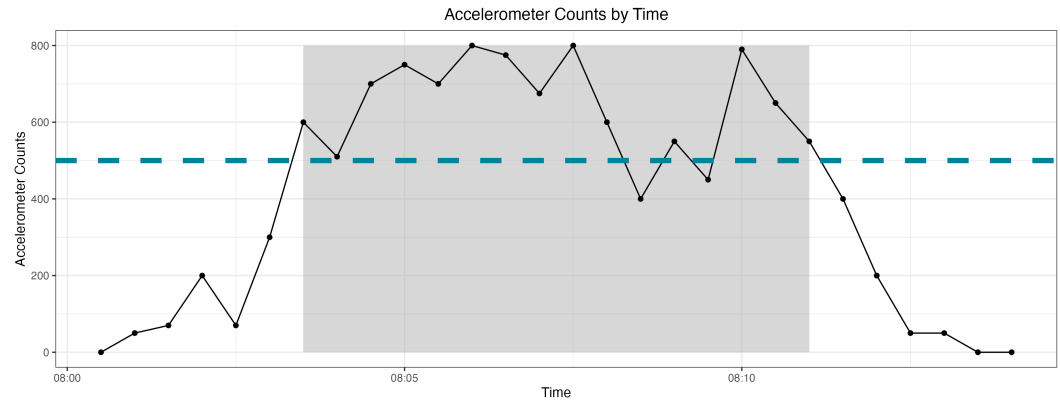


Figure 1: An accelerometry trace indicating a bout, where the bout period is indicated by the gray bounds and the epochs outside these bounds are not included in the bout. The blue line shows the threshold for an individual being considered active, within walkbotr the default is set to 500 CPE (but this can be changed by the user).

If no bouts are identified, walkbotr stops and returns a message indicating that there are no bouts. If there are physical activity bouts, walkbotr labels each bout with a numeric value. In a later step (see II.), these bouts will be labeled as either walk bouts or non-walk bouts. In addition to finding all bouts in the accelerometry data, walkbotr identifies non-wearing periods in the accelerometry data when there are 20 consecutive minutes in which activity counts are 0 CPE (Saelens et al., 2014). Finally, walkbotr flags whether the subject wore their accelerometer for a “complete day” or provided at least 8-hours of data (Saelens et al., 2014). As with other parameters, the package user may modify the non-thresholds for flagging non-wearing periods and complete days.

II. Merging GPS and accelerometer data

Once physical activity bouts are identified, walkbotr uses GPS data to classify bouts as walking or not walking. First, walkbotr merges GPS and accelerometry data by timestamp. This merge is complex for two reasons: (1) accelerometry data are typically recorded in local time whereas GPS devices record in UTC time, and (2) accelerometers record activity in consistent epochs from the time they are turned on, whereas GPS devices only record when they receive responses from GPS satellites. A fully launched GPS device pings satellites on a regular schedule, often aligned by design with accelerometer epochs. However, when devices re-establish connections with satellites (e.g., after a device restart or after time spent in a tunnel), timestamps may be off alignment. Accordingly, for each recorded GPS point, walkbotr identifies the accelerometer epoch synchronized to the local time zone of the GPS points (accounting for daylight saving time) before merging the datasets. When multiple GPS points fall

within one accelerometer epoch, walkbotr will assign the latest GPS point within the epoch to reconcile duplications.

Next, walkbotr determines whether each bout is a walk bout or a dwell bout. A circular boundary is constructed for each bout that is centered on the GPS point median location and includes 95% of all GPS points within its circumference. As shown in **Figure 2**, a bout is classified as a dwell bout (i.e., not walking) if the radius of the constructed circle is smaller than a prespecified threshold. If the bout's constructed circle is larger than the dwell bout threshold, that physical activity bout is a walk bout (Kang et al., 2013). By default, walkbotr prespecifies the dwell bout circle radius at 66 feet, but this parameter may be changed.

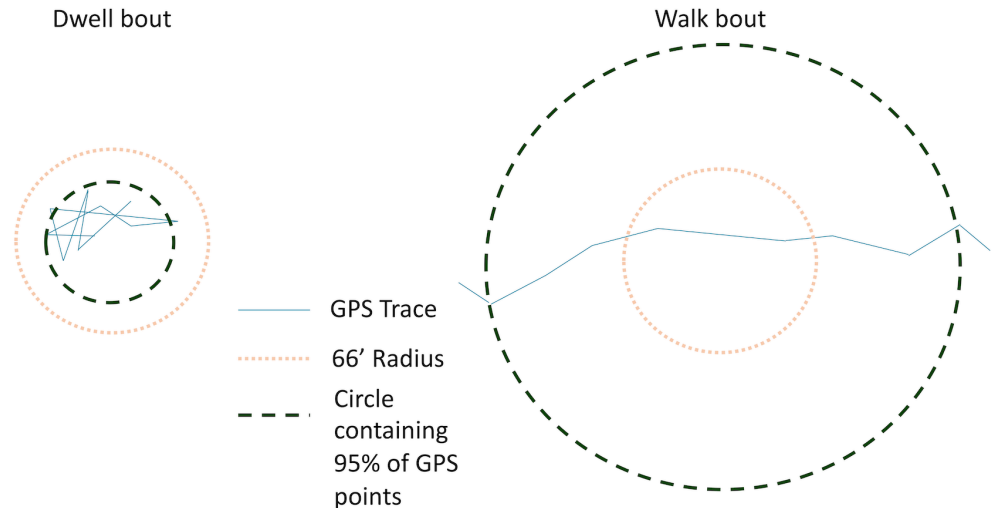


Figure 2: Dwell bout (left) and walk bout (right) show how an individual might remain within the the dwell bout threshold of 66 feet (shown in pink circle in both plots) on the left side and might leave the dwell bout threshold on a potential walk bout on the right side.

GPS data for each bout are evaluated for completeness, assessed both in terms of the number of GPS observations within the bout as well as the proportion of accelerometry observations that have GPS data. By default, a bout has sufficient GPS coverage if it has at least five GPS observations and at least 20% of the epochs have a paired GPS observation.

The final walkbotr output labels each physical activity bout as either a walk bout or a non-walk bout, with a specific label for five different types of non-walk bouts (Kang et al., 2013). Each physical activity bout can have just one category, making the order of labeling important. In order, walkbotr applies the following labels:

1. **non_walk_incomplete_gps**: all physical activity bouts without complete GPS data,
2. **non_walk_too_fast**: remaining physical activity bouts where the median speed exceeds the maximum walking speed (default: 6 km/h),
3. **non_walk_too_slow**: remaining physical activity bouts where the median speed falls below the minimum walking speed (default: 2 km/h),
4. **non_walk_too_vigorous**: remaining physical activity bouts whose mean CPE are too vigorous to be considered walking (default: > 2,863 CPE),
5. **dwel_bout**: remaining physical activity bouts whose GPS data do not exceed a circular dwell bout threshold radius (default: 66 ft), and, finally,
6. **walk_bout**: remaining physical activity bouts.

Producing walk bout datasets for analysis

From the processed GPS and accelerometry data, a complete, epoch-level dataset (containing epoch time as date-time in the UTC time zone, accelerometry counts per epoch, latitude, longitude, epoch speed, and wearing day complete flag) is used to create two different output datasets.

The first output is a complete, epoch-level dataset [Table 1](#) that returns all the original input data that the user provided and new variable columns created by the walkbotr package. This dataset could be merged back to the full original dataset, containing identifiable information (Dalmat et al., 2021).

Table 1: Variable dictionary for epoch-level dataset. The first column contains the dataset column names, the second column contains the object class of each dataset feature, and the final column provides a definition of each feature.

Column	Class	Definition
bout_category	Character	This column contains the category of the bout, which is described below.
activity_counts	Numeric	Accelerometer counts in counts per epoch (CPE).
bout_start	Date-time	This column contains date-time values in the UTC time zone.
non_wearing	Logical	Boolean flag for whether the user was wearing their device at the time (non_wearing = TRUE).

complete_day	Logical	This is a Boolean column indicating whether the calendar day of data was complete (assessed by determining whether the individual wore their accelerometer for greater than x hours, where x is passed in as a parameter min_wearing_hours_per_day or defaults to 8.)
latitude	Numeric	Latitude coordinate
longitude	Numeric	Longitude coordinate
median_speed	Numeric	This column contains the median speed, in km/h, of a given bout.
duration	Numeric	This column contains the length of a bout, in minutes.

The second output is a summary, bout-level dataset [Table 2](#) that has been collapsed to only include summary walk bout information and is completely de-identified. This dataset is intended to provide essential walking and physical activity metrics without any identifying information – thus serving as a product that can be easily shared. This dataset can be used in analyses where walking or walking frequency is the outcome, e.g. research studies identifying external factors or neighborhood features that do or do not increase walking in a population ([Mooney et al., 2020](#)).

Table 2: Variable dictionary for the bout-level dataset. The first column contains the dataset column names, the second column contains the object class of each dataset feature, and the final column provides a definition of each feature.

Column	Class	Definition
bout	Numeric	This column is a label for each walk bout – each bout is sequentially labeled with a number for easier identification purposes.
median_speed	Numeric	This column contains the median speed, in km/h, of a given bout.
bout_category	Character	This column contains the category of the bout, which is described below.
complete_day	Logical	This is a Boolean column indicating whether the calendar day of data was complete (assessed by determining whether the individual wore their accelerometer for greater than x hours, where x is passed in as a parameter min_wearing_hours_per_day or defaults to 8.)
bout_start	Date-time	This column contains date-time values in the UTC time zone.

duration	Numeric	This column contains the length of a bout, in minutes.
----------	---------	--

The walkbotr package also produces summary figures describing the walk bouts. [Figure 3](#) shows a walk bout where the accelerometry counts exceed the threshold for being considered active in combination with an image of the ratio of radii of the bout to the dwell bout threshold.

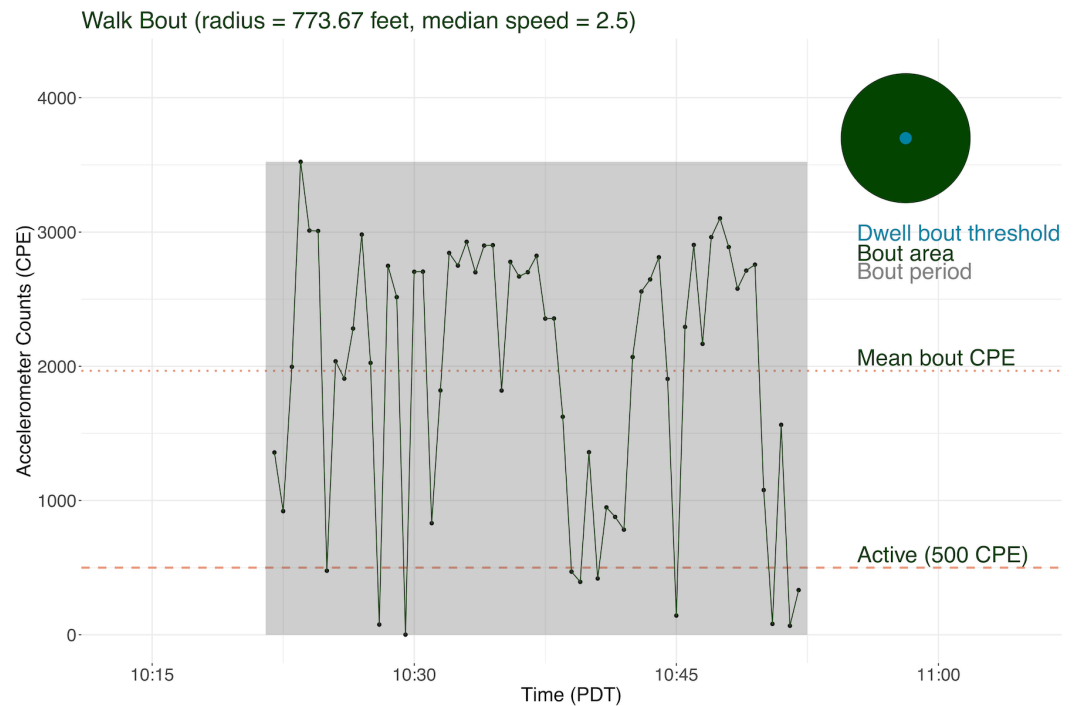


Figure 3: Example of a walk bout. The physical activity bout in the gray box is classified as a walk bout, because (1) the activity CPE is above the active threshold and is consistent with that of walking and (2) the bout area is larger than that of the dwell bout. The circles show the size of the walk bout (green) in relation to the dwell bout max circumference (blue). These two measures define walks and the application allows researchers to set the activity and bout area thresholds to define active walking for reproducibility and comparison to other research sites.

Conclusions and future directions

walkbotr enables researchers to leverage data collected from monitoring devices such as GPS trackers and accelerometers for improved objective measurement of walking behavior and its associated impacts on health across diverse geospatial contexts and populations. By implementing a previously validated algorithm (Kang et al., 2013), walkbotr facilitates not only the identification of walking periods in a consistent and

user-friendly manner, thereby bridging the gap between raw data and actionable insights, but also allows researchers to process complex and identifiable monitoring data into shareable insights.

As public health research increasingly embraces data-driven methodologies, the walkbotr package emerges as a tool for researchers to expand their collaborations and datasets. It may serve as an example of how public health researchers can begin to shift towards a new era of collaborative, privacy-conscious, and data-rich health research. By facilitating the extraction of valuable insights from monitoring data without compromising privacy or ethics, the package paves the way for a more informed understanding of human behavior, health, and well-being in the world of big data.

Acknowledgements: We thank Amy Youngbloom for her contributions and assistance in vetting these methods.

Computation & reproducibility

The original repository maintained by Lauren Wilner can be found here: <https://github.com/rwalkbotr/walkbotr>. Note: this repository is maintained by Lauren Wilner and may differ from that originally used to produce the results in this publication.

References

- Carlson, J. A., Jankowska, M. M., Meseck, K., Godbole, S., Natarajan, L., Raab, F., Demchak, B., Patrick, K., & Kerr, J. (2015). Validity of PALMS GPS scoring of active and passive travel compared to SenseCam. *Medicine and Science in Sports and Exercise*, 47(3), 662. <https://doi.org/10.1249/MSS.0000000000000446>
- Dalmat, R. R., Mooney, S. J., Hurvitz, P. M., Zhou, C., Moudon, A. V., & Saelens, B. E. (2021). Walkability measures to predict the likelihood of walking in a place: A classification and regression tree analysis. *Health and Place*, 72. <https://doi.org/10.1016/j.healthplace.2021.102700>
- Feng, T., & Timmermans, H. J. (2013). Transportation mode recognition using GPS and accelerometer data. *Transportation Research Part C: Emerging Technologies*, 37, 118–130. <https://doi.org/10.1016/j.trc.2013.09.014>
- Jankowska, M. M., Schipperijn, J., & Kerr, J. (2015). A framework for using GPS data in physical activity and sedentary behavior studies. *Exercise and Sport Sciences Reviews*, 43(1), 48. <https://doi.org/10.1249/JES.0000000000000035>
- Kang, B., Moudon, A. V., Hurvitz, P. M., Reichley, L., & Saelens, B. E. (2013). Walking objectively measured: classifying accelerometer data with GPS and travel diaries. *Medicine and Science in Sports and Exercise*, 45(7), 1419. <https://doi.org/10.1249/MSS.0b013e318285f202>
- Mooney, S. J., Hurvitz, P. M., Moudon, A. V., Zhou, C., Dalmat, R. R., & Saelens, B. E. (2020). Residential neighborhood features associated with objectively measured

-
- walking near home: Revisiting walkability using the Automatic Context Measurement Tool (ACMT). *Health Place*, 63. <https://doi.org/10.1016/j.trd.2017.09.026>
- Saelens, B. E., Moudon, A. V., Kang, B., Hurvitz, P. M., & Zhou, C. (2014). Relation between higher physical activity and public transit use. *American Journal of Public Health*, 104(5), 854–859. <https://doi.org/10.2105/AJPH.2013.301696>
- Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*, 40(1), 181–188. <https://doi.org/10.1249/mss.0b013e31815a51b3>
- Troped, P. J., Oliveira, M. S., Matthews, C. E., Cromley, E. K., Melly, S. J., & Craig, B. A. (2008). Prediction of activity mode with global positioning system and accelerometer data. *Medicine and Science in Sports and Exercise*, 40(5), 972–978. <https://doi.org/10.1249/MSS.0b013e318164c407>