2022

# CRIB: A Novel Method for Device-Based Physical Behavior Analysis

Paul R. Hibbing
*Children's Mercy Hospital*

Seth A. Creasy

Jordan A. Carlson
*Children's Mercy Hospital*



## Recommended Citation

Hibbing PR, Creasy SA, Carlson JA. CRIB: A Novel Method for Device-Based Physical Behavior Analysis. Journal for the Measurement of Physical Behaviour. 2022;5(4):277-287. doi:10.1123/jmpb.2021-0059

***Journal for the Measurement of Physical Behaviour*, 2022, 5, 277-287**
https://doi.org/10.1123/jmpb.2021-0059

Human Kinetics
TECHNICAL NOTE

# CRIB: A Novel Method for Device-Based Physical Behavior Analysis

**Paul R. Hibbing,[1] Seth A. Creasy,[2] and Jordan A. Carlson[1,3]**

[1]Center for Children's Healthy Lifestyles & Nutrition, Children's Mercy Kansas City, Kansas City, MO, USA; [2]Division of Endocrinology, Metabolism, and Diabetes, and Anschutz Health and Wellness Center, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [3]Department of Pediatrics, University of Missouri-Kansas City, Kansas City, MO, USA

Physical behaviors (e.g., sleep, sedentary behavior, and physical activity) often occur in sustained bouts that are punctuated with brief interruptions. To detect and classify these interrupted bouts, researchers commonly use wearable devices and specialized algorithms. Most algorithms examine the data in chronological order, initiating and terminating bouts whenever specific criteria are met. Consequently, the bouts may encapsulate or overlap with later periods that also meet the activation and termination criteria (i.e., alternative bout solutions). In some cases, it is desirable to compare these alternative bout solutions before making a final classification. Thus, comparison-focused algorithms are needed, which can be used in isolation or in concert with their chronology-focused counterparts. In this technical note, we present a comparison-focused algorithm called CRIB (Clustered Recognition of Interrupted Bouts). It uses agglomerative hierarchical clustering to facilitate the comparison of different bout solutions, with the final classification being made in favor of the smallest number of bouts that comply with user-specified criteria (i.e., limits on the number, individual duration, and cumulative duration of interruptions). For demonstration, we use CRIB to assess bouts of moderate to vigorous physical activity in accelerometer data from the National Health and Nutrition Examination Survey, and we include a comparison against results from two established chronology-focused algorithms. Our discussion explores strengths and limitations of CRIB, as well as potential considerations and applications for using it in future studies. An online vignette (https://github.com/paulhibbing/PBpatterns/blob/main/vignettes/CRIB.pdf) is available to assist users with implementing CRIB in R.

***Keywords**:* wearable technology, interruptions, signal processing, data-driven algorithms, unsupervised data mining

Bout classification algorithms are often used to process data from wearable devices. The purpose of the algorithms is to assess sustained engagement in physical behaviors such as sleep, sedentary behavior (SB), or moderate to vigorous physical activity (MVPA). The specific details of bout classification are different for each behavior (Altenburg & Chinapaw, 2015; Winkler et al., 2016), but the fundamental similarity is a need to identify when each bout starts and ends, while accounting for brief interruptions (i.e., transient stoppages followed by prompt resumption of the behavior). For example, when quantifying bouts of MVPA, there is a need to account for the possibility that a person stops at a crosswalk while out for a walk or run.

Most bout classification algorithms examine the data in chronological order, as seen in prominent examples for assessing sleep (van der Berg et al., 2016), SB (Carson & Janssen, 2011), MVPA (Ostendorf et al., 2018; Troiano et al., 2008), and ambulation (Barry et al., 2015). In these algorithms, bouts start when predefined activation criteria are met and end when predefined termination criteria are met. If there are overlapping periods that meet the criteria for a bout (i.e., alternative bout solutions; see Figure 1), chronology-focused algorithms will select the first solution that occurs. This is a suitable approach for many research questions but in some cases it may be desirable to examine different solutions and identify bouts based on comparison rather than chronology. Such a comparison-focused approach could be conceptualized as looking at overlapping bout solutions (Figure 1) to address potential ambiguities regarding the number of bouts and when each bout starts and ends. Fewer existing methods reflect this type of comparison-focused approach, with the leading example being a combined sleep and nonwear detection algorithm for activPAL devices (Winkler et al., 2016). Thus, there is warrant for developing new comparison-focused bout classification algorithms that apply to other behaviors and devices.

In this technical note, we build on a concept from Twaites (2019) and propose a novel comparison-focused bout classification method called CRIB (Clustered Recognition of Interrupted Bouts), which can be applied to analysis of any physical behavior. The premise of CRIB is that bout identification can be viewed as an unsupervised clustering problem in which the goal is to group separate occurrences of the target behavior into common bins, based on their temporal proximity to one another. Below, we present the inner workings of the CRIB technique, followed by an example analysis of accelerometer data from the National Health and Nutrition Examination Survey (NHANES). We conclude with a discussion exploring the strengths and limitations of CRIB, the importance of time resolution (i.e., epoch length) when using CRIB and other bout analysis methods, and potential novel applications for CRIB in future research.

## Technique

The CRIB technique is implemented in three phases, namely preprocessing, iterative clustering, and postprocessing. Table 1 describes the key components for each phase, and below we present the step-by-step procedures. For illustration, we will focus on using CRIB to identify bouts of MVPA in an excerpt of NHANES accelerometer

Creasy https://orcid.org/0000-0002-4176-0917
Carlson https://orcid.org/0000-0002-6008-7983
Hibbing (prhibbing@cmh.edu) is corresponding author, https://orcid.org/0000-0002-1469-5160

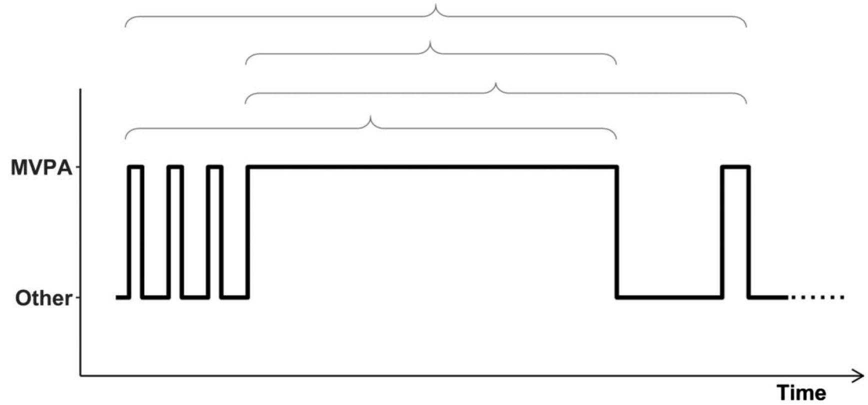



**Figure 1** — Depiction of overlapping bout solutions when analyzing a period of interrupted MVPA. Each gray bracket represents a potential grouping of events into a bout. The depicted options are illustrative, not exhaustive. MVPA = moderate to vigorous physical activity.

**Table 1 Components of CRIB**

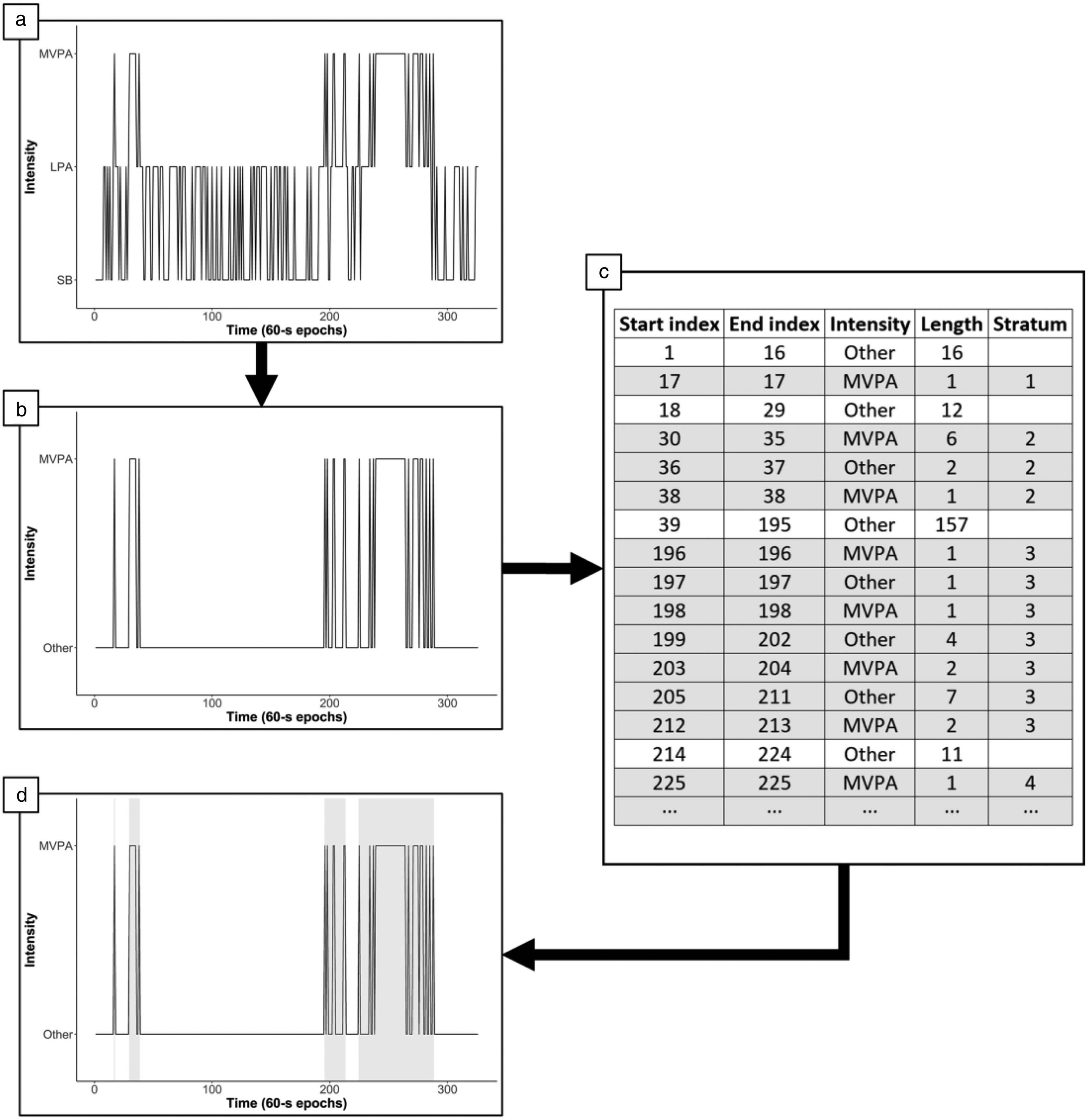
| Component | Description/summary |
|---|---|
| Phase 1: Preprocessing | Data are encoded, compressed, and stratified in preparation for clustering. |
| Input data stream | A time series of predetermined behavior classifications, e.g., minute-by-minute indications of SB/LPA/MVPA. |
| Target behavior[a] | The one specific value (e.g., MVPA) to analyze from *input data stream*. A search is made for occurrences of this value, while any others (e.g., SB, or LPA) are recategorized as "Other." |
| Target buffer[a,b] | A stratifying threshold of $n$ epochs. Before clustering, the data set is split whenever two occurrences of *target behavior* (e.g., MVPA) are separated by more than this amount (e.g., 10 min). Occurrences in different strata are prohibited from being grouped together in the same bout, helping to limit runtime. |
| Run-length encoding | A standard compression algorithm, used here to identify the location and duration of each behavioral occurrence in the data sequence (e.g., Other-MVPA-Other-MVPA . . . ). |
| Phase 2: Iterative clustering | Within each stratum, occurrences of the target behavior are grouped together in a range of ways to find the option that aligns best with user-specified settings. |
| Agglomerative hierarchical clustering (McQuitty, 1957) | A standard clustering algorithm, used here to establish groupings of events within each stratum. The groupings (i.e., potential bouts) are nested in a dendrogram that is then examined more closely to determine the smallest number that meets the below criteria. |
| Maximum number of interruptions[a,c] | The maximum number of interruptions allowed within a bout (e.g., $n = 3$). |
| Longest allowable interruption[a,c] | The maximum length allowed for any single interruption within a bout (e.g., 2 min). |
| Required percent engagement[a,d] | A threshold defining the minimum percentage of the full bout duration that must be spent in a noninterrupted state (e.g., 80% of bout duration must be MVPA). |
| Phase 3: Postprocessing | The cluster outputs are pooled across strata, then filtered if necessary. |
| Minimum bout duration[a] | A filtering criterion; after running the clustering algorithm, bouts will be removed from the output if the total *target behavior* engagement is not last at least this long (e.g., to exclude bouts with <10 min of MVPA). Can be set to 0 if no restriction is desired. |

*Note*. SB = sedentary behavior; LPA = light physical activity; MVPA = moderate to vigorous physical activity; CRIB = Clustered Recognition of Interrupted Bouts. [a]User-specified setting. [b]Can be set to 0 for continuous bouts, or to infinity for exhaustive consideration of all occurrences (not recommended due to the nonlinear time complexity of the clustering algorithm). [c]Can be set to 0 for continuous bouts, or to infinity if no restriction is desired. [d]Can be set to 100 for continuous bouts, or to 0 if no restriction is desired.

data. However, it should be noted that CRIB can also be applied for classifying bouts of other behaviors such as SB or sleep.

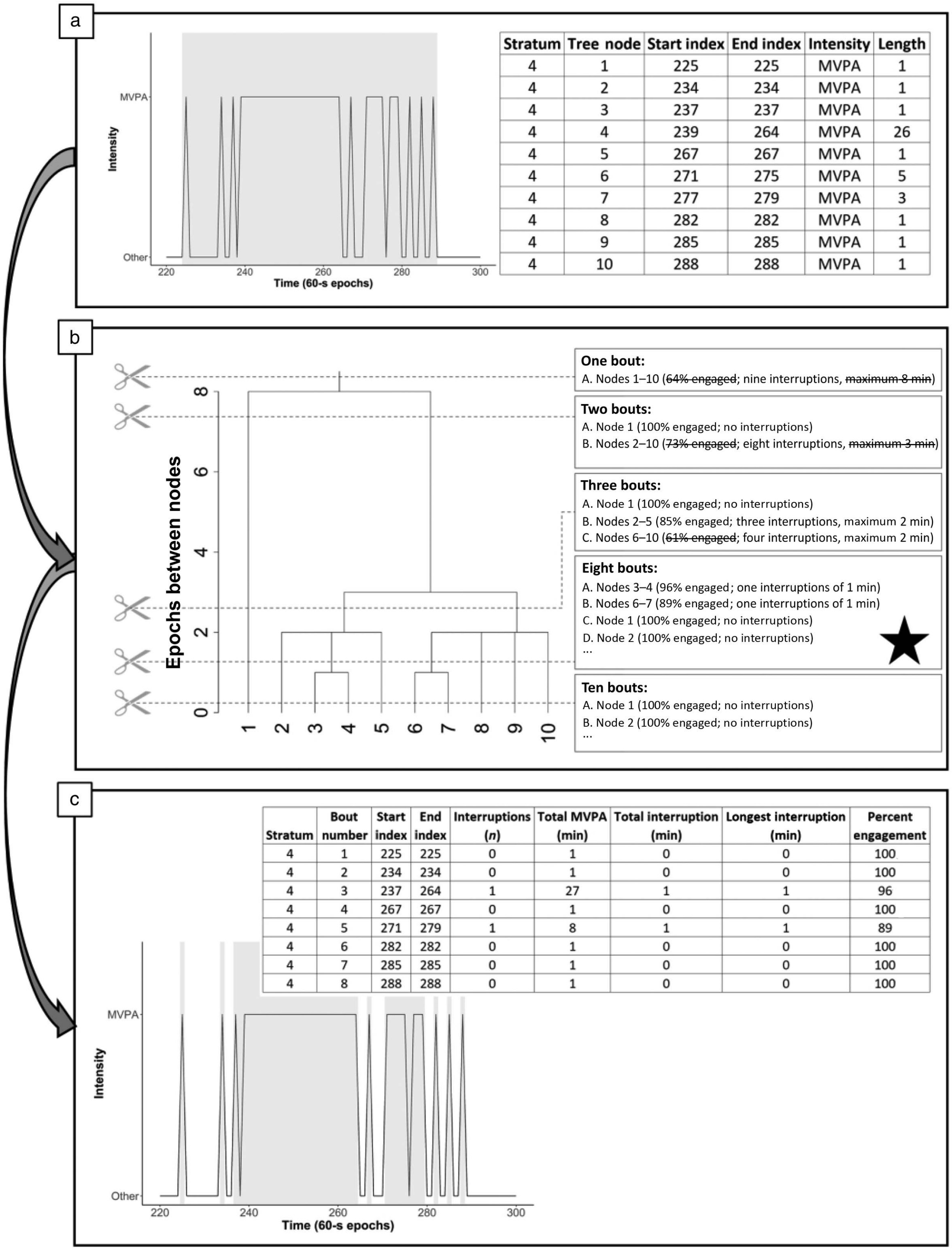
## Phase 1: Preprocessing

Figure 2 illustrates the steps of the preprocessing phase. First, each epoch receives a dichotomous classification as either the target behavior (defined in Table 1) or "other" (i.e., nontarget behaviors). This ensures that consecutive nontarget behaviors will be treated as a single interruption, consistent with traditional approaches to bout classification. Once each epoch has been dichotomized, run-length encoding is applied to extract the start and stop times for each event (i.e., each run of consecutive epochs in a single category), resulting in a sequence that alternates between the target behavior and "other." The final step of preprocessing is stratifying the results whenever an "other" event meets or exceeds a user-specified duration threshold called the target buffer (defined in Table 1 and set to 10 min for the current example). The purpose of data stratification is to improve runtime, as discussed later. The preprocessing phase is complete once the data have been run-length encoded and stratified.




| Start index | End index | Intensity | Length | Stratum |
|---|---|---|---|---|
| 1 | 16 | Other | 16 | |
| 17 | 17 | MVPA | 1 | 1 |
| 18 | 29 | Other | 12 | |
| 30 | 35 | MVPA | 6 | 2 |
| 36 | 37 | Other | 2 | 2 |
| 38 | 38 | MVPA | 1 | 2 |
| 39 | 195 | Other | 157 | |
| 196 | 196 | MVPA | 1 | 3 |
| 197 | 197 | Other | 1 | 3 |
| 198 | 198 | MVPA | 1 | 3 |
| 199 | 202 | Other | 4 | 3 |
| 203 | 204 | MVPA | 2 | 3 |
| 205 | 211 | Other | 7 | 3 |
| 212 | 213 | MVPA | 2 | 3 |
| 214 | 224 | Other | 11 | |
| 225 | 225 | MVPA | 1 | 4 |
| ... | ... | ... | ... | ... |

**Figure 2** — Steps of the preprocessing phase for a sample analysis where the target behavior is MVPA. (a) The input stream has raw labels of SB, LPA, or MVPA. (b) The raw labels are recoded to binary, that is, MVPA or "Other." (c) Run-length encoding is applied to identify the sequence of engagement in MVPA and "Other." Data are stratified by splitting the data set whenever an "Other" event lasts ≥10 min. (d) For visualization only, representing how stratification would appear if run-length encoding were undone. MVPA = moderate to vigorous physical activity; SB = sedentary behavior; LPA = light physical activity.




| Stratum | Tree node | Start index | End index | Intensity | Length |
|---|---|---|---|---|---|
| 4 | 1 | 225 | 225 | MVPA | 1 |
| 4 | 2 | 234 | 234 | MVPA | 1 |
| 4 | 3 | 237 | 237 | MVPA | 1 |
| 4 | 4 | 239 | 264 | MVPA | 26 |
| 4 | 5 | 267 | 267 | MVPA | 1 |
| 4 | 6 | 271 | 275 | MVPA | 5 |
| 4 | 7 | 277 | 279 | MVPA | 3 |
| 4 | 8 | 282 | 282 | MVPA | 1 |
| 4 | 9 | 285 | 285 | MVPA | 1 |
| 4 | 10 | 288 | 288 | MVPA | 1 |

| Stratum | Bout number | Start index | End index | Interruptions (*n*) | Total MVPA (min) | Total interruption (min) | Longest interruption (min) | Percent engagement |
|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 225 | 225 | 0 | 1 | 0 | 0 | 100 |
| 4 | 2 | 234 | 234 | 0 | 1 | 0 | 0 | 100 |
| 4 | 3 | 237 | 264 | 1 | 27 | 1 | 1 | 96 |
| 4 | 4 | 267 | 267 | 0 | 1 | 0 | 0 | 100 |
| 4 | 5 | 271 | 279 | 1 | 8 | 1 | 1 | 89 |
| 4 | 6 | 282 | 282 | 0 | 1 | 0 | 0 | 100 |
| 4 | 7 | 285 | 285 | 0 | 1 | 0 | 0 | 100 |
| 4 | 8 | 288 | 288 | 0 | 1 | 0 | 0 | 100 |

**Figure 3** — Steps of the iterative clustering phase for a single stratum of data. (a) The MVPA events are passed into the hierarchical agglomerative clustering algorithm. (b) The algorithm yields a tree called a dendrogram. The height of the dendrogram represents the temporal distance between each node, and agglomeration occurs from the bottom up, recursively grouping the two branches that are closest together until all events are in a single bout. The dendrogram is then cut at each branching point to examine all the possible bout solutions, from the top of the tree (all leaf nodes in one bout) to the bottom (all leaf nodes in separate bouts). The resulting bouts are tested to see whether they all meet the user-specified criteria (see Table 1). Here, the criteria allowed unlimited interruptions as long as they comprised <20% of the overall duration, with no single interruption lasting >2 min. Strikethrough text indicates the requirement is not met. Notably, equidistant nodes (e.g., 8–10) are agglomerated together in a single iteration. The final bout selection (black star) is defined as the smallest number of bouts that meet all criteria. (c) For visualization only, representing the results of iterative clustering for the stratum in question. MVPA = moderate to vigorous physical activity.

### Phase 2: Iterative Clustering

In the iterative clustering phase, a systematic clustering and revision process is applied separately for each of the strata from the preprocessing phase. This broadly involves computing different bout solutions (e.g., as shown in Figure 1), then comparing the solutions, and making a selection based on simple criteria. The process for one stratum is described below and illustrated in Figure 3.

First, the input data (Figure 3a) are analyzed using agglomerative hierarchical clustering with single linkage (McQuitty, 1957). This involves an iterative three-step process, as follows: (1) combine the pair of target behavior events that are temporally closest to each other, (2) store the results for later analysis, and (3) repeat the process until all events have been combined into a single cluster (i.e., a single bout solution). If there are ties in Step 1, the step is recursed for all qualifying pairs before moving to Step 2.

Once the clustering iterations are finished, the results can be visualized as a tree called a dendrogram, which groups the events into nested clusters (Figure 3b). Each cluster represents a potential bout solution, and the dendrogram presents a wide range of possibilities, from the top of the tree (all events belong to a single bout) to the bottom (all events belong to separate bouts). CRIB then examines the dendrogram layer-by-layer to determine which bout clustering arrangement aligns best with user-specified criteria (Figure 3b). The best alignment is defined as the smallest possible number of bouts for which each bout complies with the user-specified criteria defined in Table 1 (i.e., maximum number of interruptions, longest allowable interruption, and required percent engagement). For the current example, these criteria were set to allow an unlimited number of interruptions during the bout as long as no consecutive interruptions spanned >2 min and there was ≥80% engagement in MVPA over the course of the full bout. The iterative clustering phase is complete once a final bout selection (e.g., Figure 3c) has been made for each stratum of data.

There are several important considerations to note for the iterative clustering phase. One is the linkage method used for agglomerative hierarchical clustering. Single linkage was selected because its "chaining" effect is advantageous in certain applications, of which bout classification is one (Hartigan, 1981; Kuiper & Fisher, 1975). Another crucial consideration is the need to ensure a feasible runtime because the clustering algorithm makes pairwise comparisons among all the data points. For large time series data (e.g., a week or more of accelerometry), there could easily be millions of comparisons to make, most of which would be unnecessary (e.g., evaluating the distance between MVPA instances on different days),therefore stratification was done in the preprocessing phase. By running the algorithm separately on each stratum (i.e., prohibiting bouts from spanning multiple strata), it is possible to limit the number of computations, thereby improving runtime.

### Phase 3: Postprocessing

Figure 4 illustrates the steps for the postprocessing phase. First, the clustering results are pooled across all strata, and then bouts are filtered out if they do not meet the minimum bout duration requirement (defined in Table 1). For the current example, bouts with <10 min of MVPA were filtered out. Completion of the postprocessing phase yields final bouts for subsequent analysis.

## Example Analysis

For illustration, we provide an analysis of MVPA bouts using accelerometer data from adults (≥18 years old) in the 2003–2004 and 2005–2006 cycles of the NHANES. We use CRIB to obtain a range of bout metrics, with comparisons against values obtained from the chronology-focused MVPA algorithms of Troiano et al. (2008) and Ostendorf et al. (2018).

### Protocol and Data Preprocessing

The NHANES protocol required participants to wear an ActiGraph AM-7164 accelerometer (ActiGraph, LLC) on the hip for up to 7 days. Of 9,601 adults with accelerometer data, we excluded those whose data had been flagged by NHANES as being unreliable ($n = 116$) or out of calibration ($n = 358$). We screened for nonwear using the algorithm of Choi et al. (2011) and defined valid days as those with at least 10 hr of wear time. We then excluded 2,280 participants who had <4 valid days of data. MVPA was defined as ≥1,952 counts per min during wear time on a valid day (Freedson et al., 1998). As a final cleaning step, we excluded data from 4,114 participants for whom there were no MVPA bouts using any of the three algorithms. Thus, the final analytic sample included 2,733 individuals (43.3% females; mean ± *SD*: age 42.9 ± 17.8 years). All bout classification algorithms, including CRIB, were implemented in R using the "analyze_bouts" function in the PBpatterns package (version 0.3.0, see https://github.com/paulhibbing/PBpatterns/releases/tag/v0.3.0).

### Bout Classification Algorithms

The algorithm of Troiano et al. (2008) operates using a 10-min sliding window that scans the data from start to end. Bouts are preliminarily activated when the window encounters a span with ≥8 min of MVPA, then confirmed when ≥10 min of MVPA have accrued. Bouts are terminated when the window reaches a span with ≥3 consecutive minutes of non-MVPA. If termination occurs after preliminary activation yet prior to confirmation, the bout is excluded.

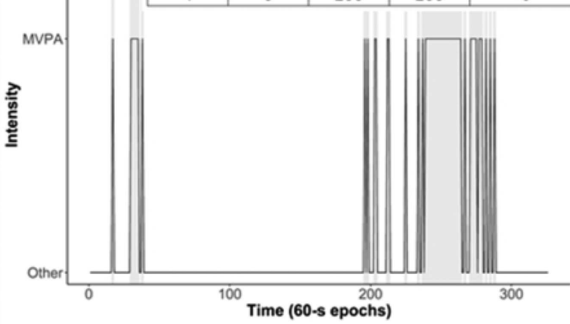
The algorithm of Ostendorf et al. (2018) allows a bout to be preliminarily activated by a single minute of MVPA. The bout then continues until there are ≥3 consecutive minutes of non-MVPA. To be confirmed, the bout must contain ≥10 min of MVPA, and ≥ 80% of all epochs must also be classified as MVPA. Like the prior algorithm, preliminarily activated bouts are discarded if the confirmation criteria are not met.

For this sample analysis, CRIB settings were selected to align with the general concepts of the algorithms of Troiano et al. (2008) and Ostendorf et al. (2018). Specifically, there was no limit on the number of interruptions during the bout, but all interruptions (i.e., periods of consecutive non-MVPA epochs) had to last < 3 min, and at least 80% of the epochs in the final bout had to be classified as MVPA. The minimum bout duration was set to 10 min, as was the target buffer. The purpose of aligning the CRIB settings with those from the other algorithms was to maximize the interpretability of the findings. If similar output was obtained from each algorithm, it would suggest limited occurrence (or limited impact) of alternative bout solutions like what is shown in Figure 1. If discrepant outputs were obtained, it would suggest the opposite.

### Statistical Methods and Results

For all three algorithms, we analyzed the following: (a) bouted MVPA time, (b) interruption time within the bouts, and (c) the a

| Stratum | Bout number | Start index | End index | Interruptions (*n*) | Total MVPA (min) | Total interruption (min) | Longest interruption (min) | Percent engagement |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 17 | 17 | 0 | 1 | 0 | 0 | 100 |
| 2 | 1 | 30 | 35 | 0 | 6 | 0 | 0 | 100 |
| 2 | 2 | 38 | 38 | 0 | 1 | 0 | 0 | 100 |
| 3 | 1 | 196 | 196 | 0 | 1 | 0 | 0 | 100 |
| 3 | 2 | 198 | 198 | 0 | 1 | 0 | 0 | 100 |
| 3 | 3 | 203 | 204 | 0 | 2 | 0 | 0 | 100 |
| 3 | 4 | 212 | 213 | 0 | 2 | 0 | 0 | 100 |
| 4 | 1 | 225 | 225 | 0 | 1 | 0 | 0 | 100 |
| 4 | 2 | 234 | 234 | 0 | 1 | 0 | 0 | 100 |
| 4 | 3 | 237 | 264 | 1 | 27 | 1 | 1 | 96 |
| 4 | 4 | 267 | 267 | 0 | 1 | 0 | 0 | 100 |
| 4 | 5 | 271 | 279 | 1 | 8 | 1 | 1 | 89 |
| 4 | 6 | 282 | 282 | 0 | 1 | 0 | 0 | 100 |
| 4 | 7 | 285 | 285 | 0 | 1 | 0 | 0 | 100 |
| 4 | 8 | 288 | 288 | 0 | 1 | 0 | 0 | 100 |




b

| Stratum | Bout number | Start index | End index | Interruptions (*n*) | Total MVPA (min) | Total interruption (min) | Longest interruption (min) | Percent engagement |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 17 | 17 | 0 | 1 | 0 | 0 | 100% |
| 2 | 1 | 30 | 35 | 0 | 6 | 0 | 0 | 100% |
| 2 | 2 | 38 | 38 | 0 | 1 | 0 | 0 | 100% |
| 3 | 1 | 196 | 196 | 0 | 1 | 0 | 0 | 100% |
| 3 | 2 | 198 | 198 | 0 | 1 | 0 | 0 | 100% |
| 3 | 3 | 203 | 204 | 0 | 2 | 0 | 0 | 100% |
| 3 | 4 | 212 | 213 | 0 | 2 | 0 | 0 | 100% |
| 4 | 1 | 225 | 225 | 0 | 1 | 0 | 0 | 100% |
| 4 | 2 | 234 | 234 | 0 | 1 | 0 | 0 | 100% |
| 4 | 3 | 237 | 264 | 1 | 27 | 1 | 1 | 96% |
| 4 | 4 | 267 | 267 | 0 | 1 | 0 | 0 | 100% |
| 4 | 5 | 271 | 279 | 1 | 8 | 1 | 1 | 89% |
| 4 | 6 | 282 | 282 | 0 | 1 | 0 | 0 | 100% |
| 4 | 7 | 285 | 285 | 0 | 1 | 0 | 0 | 100% |
| 4 | 8 | 288 | 288 | 0 | 1 | 0 | 0 | 100% |

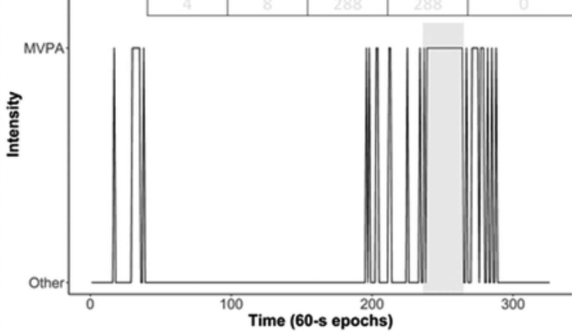



**Figure 4** — Steps of the postprocessing phase when analyzing bouts of MVPA. (a) Cluster results are pooled across strata. (b) Bouts are filtered out if they do not meet the minimum duration criteria (10 min in this example). MVPA = moderate to vigorous physical activity.

number of bouts. Day-level totals were first calculated for each participant, then averaged to obtain participant-level daily summaries for the analyses. We tested group-level agreement among the methods using pairwise paired *t* tests for each metric ($\alpha = .05$), with *p* values corrected using the false discovery rate correction (Benjamini & Hochberg, 1995). We also performed pairwise Bland–Altman analyses of mean bias and limits of agreement, including regression summaries (slope and $R^2$) to reflect the presence of systematic differences (Bland & Altman, 1999). Since none of the algorithms was a criterion measure, the pairwise models were fitted by regressing bias (i.e., Method A − Method B for each participant) against the mean of both methods (i.e., $\frac{\text{Method A}+\text{Method B}}{2}$ for each participant).

Results for all analyses are shown in Table 2. Group-level agreement was strongest between the Ostendorf et al. (2018) and CRIB algorithms, with small mean differences for bouted MVPA time (within ± 0.3 min/day of one another; $p = .51$), interruption time (within ± 0.3 min/day; $p < .001$), and number of bouts (within ± 0.1 bouts/day; $p = .23$). Slightly larger mean differences were seen for comparisons involving the algorithm of Troiano et al. (2008), across bouted MVPA time (within ± 3.7 min/day of the other methods; both $p < .001$), interruption time (within ± 1.4 min/day; both $p < .001$), and number of bouts (within ± 0.3 bouts/day; both $p < .001$).

Individual-level agreement followed similar patterns to what was shown at the group level. Specifically, limits of agreement were 1.5–2.8 times as wide for comparisons involving the algorithm of Troiano et al. (2008) than they were when comparing the algorithm of Ostendorf et al. (2018) with CRIB. Across all pairwise comparisons, there was limited evidence of systematic differences for bouted MVPA time or number of bouts (all slope magnitudes ≤0.04 with $R^2 \leq 1.2\%$), whereas interruption time differed more systematically (slopes of 0.21–0.98; $R^2$ of 9.5%–46.0%).

Overall, the three algorithms were highly concordant at the group and individual levels, with the strongest agreement between the Ostendorf and CRIB algorithms. Interruption time was the metric for which the greatest differences were seen, which is consistent with the design areas in which the algorithms are most distinct from one another.

## Implications

The similar output from each algorithm provides some indication that participants rarely engaged in edge-case behavior patterns (e.g., few overlapping bout solutions like what is shown in Figure 1) that would be captured differently by the three algorithms. However, there were certainly individual cases where a unique pattern led to disagreement among the algorithms. For example, despite the overall strong agreement between the Ostendorf et al. (2018) and CRIB algorithms, there were nevertheless individual participant-level differences up to 36 MVPA min/day, 20 interruption min/day, and four bouts/day. Figure 5 (drawn from Stratum 4 of the example data in Figures 2–4) shows a specific excerpt of data where the three algorithms yielded disparate results, with the algorithm of Troiano et al. (2008) identifying two bouts (39 min of MVPA) and

**Table 2 Pairwise Comparisons of Bout Analysis Outcomes for Three Methods of Assessing MVPA Bouts (*N* = 2,733)**

| | Troiano (A) vs. Ostendorf (B) | Troiano (A) vs. CRIB (B) | Ostendorf (A) vs. CRIB (B) |
|---|---|---|---|
| Bouted MVPA (min/day) | | | |
| Method A (mean ± *SD*) | 28.4 ± 18.9 | 28.4 ± 18.9 | 25.0 ± 18.6 |
| Method B (mean ± *SD*) | 25.0 ± 18.6* | 24.7 ± 18.2* | 24.7 ± 18.2 |
| Mean bias (95% LOA) | 3.4 (−12.2, 19.0) | 3.7 (−12.7, 20.2) | 0.3 (−7.9, 8.6) |
| LOA width | 31.3 | 32.9 | 16.5 |
| Bland–Altman slope | 0.02 | 0.04 | 0.02 |
| Bland–Altman $R^2$ | .1% | .8% | 1.1% |
| Within-bout interruptions[a] (min/day) | | | |
| Method A (mean ± *SD*) | 2.8 ± 3.5 | 2.8 ± 3.5 | 1.7 ± 1.9 |
| Method B (mean ± *SD*) | 1.7 ± 1.9* | 1.4 ± 1.6* | 1.4 ± 1.6* |
| Mean bias (95% LOA) | 1.1 (−4.5, 6.6) | 1.3 (−4.9, 7.5) | 0.2 (−2.0, 2.5) |
| LOA width | 11.1 | 12.4 | 4.4 |
| Bland–Altman slope | 0.7 | 0.98 | 0.21 |
| Bland–Altman $R^2$ | 35.9% | 46.0% | 9.5% |
| MVPA bouts (*n*/day) | | | |
| Method A (mean ± *SD*) | 1.5 ± 0.7 | 1.5 ± 0.7 | 1.2 ± 0.7 |
| Method B (mean ± *SD*) | 1.2 ± 0.7* | 1.3 ± 0.8* | 1.3 ± 0.8 |
| Mean bias (95% LOA) | 0.2 (−0.7, 1.1) | 0.2 (−0.7, 1.1) | 0.0 (−0.6, 0.6) |
| LOA width | 1.8 | 1.8 | 1.2 |
| Bland–Altman slope | 0.02 | −0.03 | −0.04 |
| Bland–Altman $R^2$ | .1% | .2% | 1.2% |

*Note*. CRIB = Clustered Recognition of Interrupted Bouts; LOA = limits of agreement; MVPA = moderate to vigorous physical activity.
[a]Minutes of non-MVPA occurring within valid MVPA bouts, as defined by the algorithm in question.
*Significant difference (adjusted $p < .001$) between methods.

six interruptions (totaling 10 min), the algorithm of Ostendorf et al. (2018) identifying one bout (29 min of MVPA) and three interruptions (totaling 5 min), and CRIB identifying one bout (27 min of MVPA) and one interruption (totaling 1 min). It is important to consider the potential for such differences when interpreting bout-focused data and drawing comparisons across methods and studies. Notably, it is not always clear which algorithm (if any) gives the correct estimate, and thus algorithms should be selected based on how suitable their design is for the research question being investigated. Additionally, it may be a useful practice to compare results from several algorithms, as a form of sensitivity analysis that shows whether overlapping bout solutions are influential in the data. In the following section, we provide a more generalized discussion of CRIB and its applicability for ongoing research into bouts of different physical behaviors.

# Discussion

The CRIB technique provides a comparison-focused approach to classifying interrupted bouts of physical behavior from wearable device data. Users can implement the method in R via the *PBpatterns* package (see https://www.github.com/paulhibbing/PBpatterns), and specific guidance and sample code are available in a vignette (https://www.github.com/paulhibbing/PBpatterns/blob/main/vignettes/CRIB.pdf). Below, we discuss the novelty and strengths of CRIB, its limitations, the influence of time resolution (i.e., epoch length) in bout analysis, and potential novel applications of CRIB for future research.

## Novelty and Strengths of the CRIB Technique

The main novelty of CRIB is the framework it provides for systematically comparing different bout solutions in a way that addresses the number, timing, individual duration, and cumulative duration of interruptions. CRIB is also designed as a general purpose tool for use with any behavior, device, or population. This may promote standardization across studies, although the degree of standardization is limited by the need to use different settings for each application.

Another novel aspect of CRIB is its data-driven flow, which arises from the combination of event-based analysis (via run-length encoding) and unsupervised data mining (via hierarchical agglomerative clustering with single linkage). These design

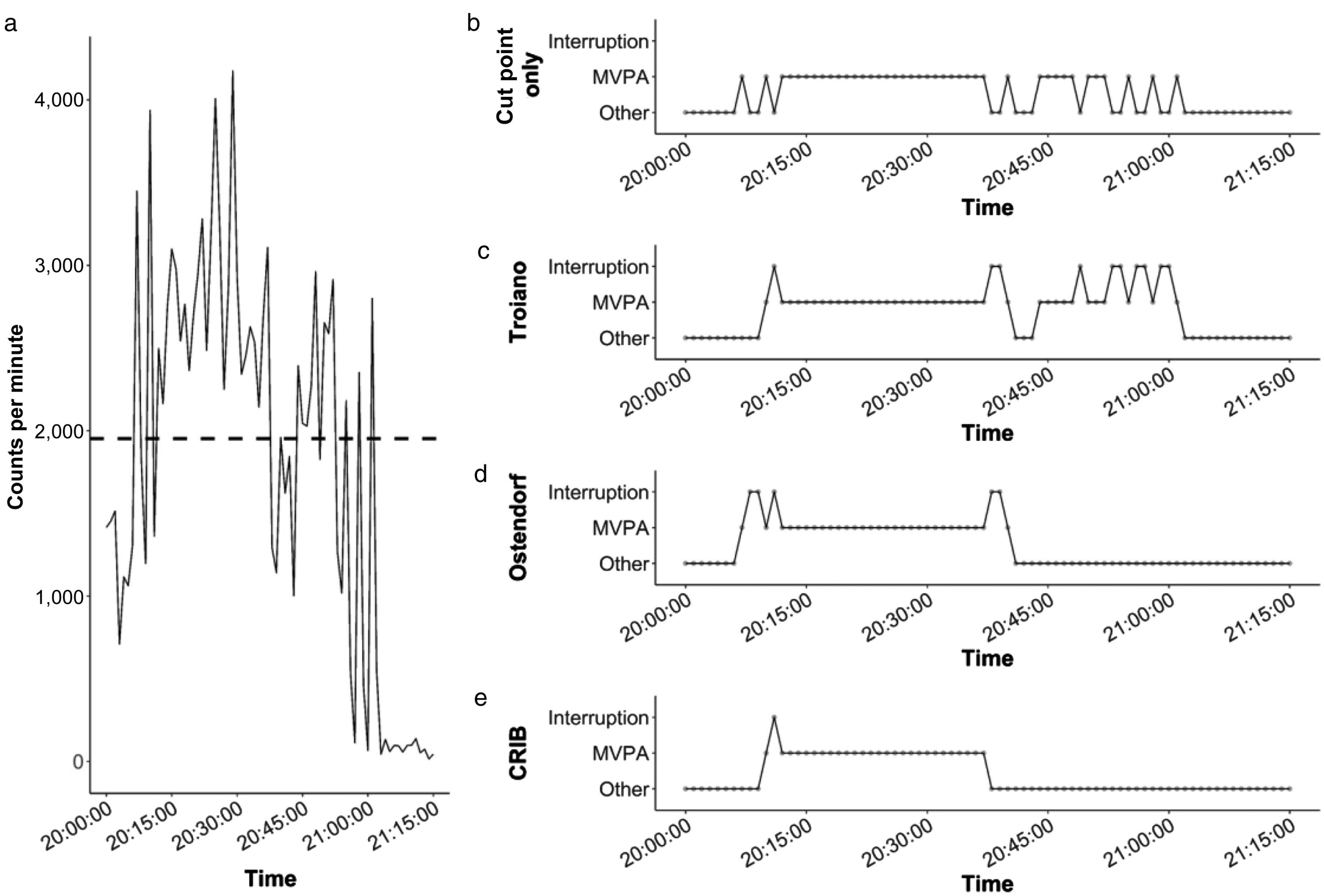



**Figure 5** — Depiction of a special case when analyzing bouts of MVPA for one participant in the NHANES. The left-hand panel (a) shows the accelerometer time series in 60-s epochs, with the horizontal line indicating the MVPA cut point of 1,952 counts/min. The right-hand column shows bout classifications when using only the cut point, that is, continuous bouts (b), versus the algorithm of Troiano et al. (2008) (c), the algorithm of Ostendorf et al. (2018) (d), or the CRIB algorithm (e). NHANES = National Health and Nutrition Examination Survey; MVPA = moderate to vigorous physical activity; CRIB = Clustered Recognition of Interrupted Bouts.

features place CRIB in strong alignment with general concepts of event-based analysis (e.g., those overviewed by Granat, 2012) as well as specific concepts of bout classification presented by Twaites (2019).

Overall, the comparison-focused design is a considerable strength that makes CRIB complementary to a range of existing bout classification algorithms. CRIB may be especially useful in cases where comparison-related information is desired in addition to chronology-related information. As noted previously, comparing output from multiple algorithms may help to show whether the data are strongly influenced by patterns that one method detects while the others do not. These capabilities position CRIB to make a valuable contribution to the science of bout analysis.

## Limitations of the CRIB Technique

While the comparison-focused and general purpose features of CRIB enhance its novelty, they are also accompanied by limitations that are important to note. One such limitation is the need to select different settings when applying CRIB to different behaviors, devices, and populations. Prior work by Barry et al. (2015) has shown that bout analysis algorithms can yield much different estimates when changing the allowable duration of individual interruptions within each bout. Additional differences are likely to occur when changing the allowable number, timing, and cumulative duration of interruptions (Altenburg & Chinapaw, 2015). Therefore, settings must be carefully selected when implementing CRIB, which may require an assortment of tailored validation studies.

Another potential limitation of CRIB is that it does not account for helpful assumptions that can sometimes be made for specific behaviors. For example, other algorithms have been enhanced by factoring in assumptions about the times of day when sleep is most likely (van der Berg et al., 2016) or which behavior types are most likely to generate the longest event each day (Winkler et al., 2016). Future work could potentially create adapted CRIB methods to account for similar assumptions, but it is unclear how much benefit this would provide.

Lastly, CRIB is computationally intensive, making runtime another potential limitation. This issue is complicated by the fact that runtime is dependent on many factors including the user-specified settings (especially the target buffer), the size of the data, the amount of target behavior engagement, the machine on which the analysis is run, and the number of other tasks the machine is performing in addition to running the algorithm. While these are important considerations, the practical impact may be trivial in some situations. For example, we timed 10 iterations of the Troiano et al. (2008), Ostendorf et al. (2018), and CRIB algorithms on the NHANES MVPA data when processing the five smallest participant files (49.1–49.7 KB in compressed R-native format) and the five largest files (71.0–73.3 KB). The full files were used (i.e., not screened for wear time or valid days), and the tests were run on a Lenovo machine with 32 GB of random-access memory and a quad-core processer (2.11 GHz base speed). Results are shown in Figure 6. When processing small files, the median runtime was ≤0.05 s per file for the Troiano et al. (2008) and Ostendorf et al. (2018) algorithms, whereas it was 0.30 s per file for CRIB. When processing large files, the Troiano algorithm ran faster (median 0.04 s per file, a twofold increase compared to the small files) than the Ostendorf algorithm (0.24 s per file, a 4.5-fold increase) or the CRIB algorithm (1.79 s per file, a sixfold increase). Although these are stark differences, it is important to note that the runtimes remained reasonably fast and were applied to a considerable volume of data for each participant (i.e., 1 week of 60-s epochs). Thus, for small and midsized data sets, runtime may pose only a minor barrier when analyzing bouts of MVPA (e.g., an extra 4 s per 100 participants with the Troiano method vs. an extra 3 min per 100 participants with CRIB).

## Additional Consideration of Time Resolution (Epoch Length)

With all bout classification algorithms, it is important to consider the impact of epoch length on the analysis. The shorter the epoch length, the shorter the interruptions that can be detected within a bout, which can potentially be an advantage in some cases (i.e., when focused on granular behavior patterns) and a disadvantage in others (i.e., when focused on higher level behavior patterns). Thus, epoch lengths must be selected with careful consideration of specific study objectives, and algorithm output must be interpreted accordingly.

For our example analysis, the NHANES data were available in 60-s epochs. Thus, very short interruptions (e.g., ≤30 s) may have been missed if the rest of the epoch contained enough movement to elicit an MVPA classification. Furthermore, these partial-epoch interruptions could have been part of longer interruptions that spanned several epochs. For example, a continuous 2-min interruption could have been spread across three 60-s epochs, with only the middle epoch being labeled an interruption. As noted above, this type of issue would be a definite limitation for research questions aimed at analyzing granular behavior patterns, whereas the concern would likely be smaller for research questions aimed at assessing higher level activity.

Epoch length is also important to consider because of its computational implications. Specifically, shorter epoch lengths lead to longer runtimes because of the direct relationship between the number of epochs and the required number of computations. As before, this affects all bout classification algorithms. However, unlike before, the consequences may be especially great for CRIB, since the number of computations will not necessarily change on a linear scale as epoch length is manipulated. To illustrate this, we randomly selected five participants from our sample analysis and reintegrated the data from 60- to 120-s epochs. We then timed 10 iterations when processing the data from each participant with all three algorithms at both time resolutions (again using the full files without screening for wear time or valid days). As expected, all algorithms had longer runtime for 60-s than 120-s epochs, but the discrepancy was less for the Troiano et al. (2008) algorithm (medians differing by a factor of 2.0) and the Ostendorf et al. (2018) algorithm (a factor of 2.3) than for CRIB (a factor of 2.4). While the practical implications were again quite small (median runtimes < 3.4 s per file for all algorithms), 120-s epochs are rarely seen in the literature, and greater consequences are likely when adjusting the epoch length downward to more common settings (i.e., <60 s). Future studies should consider the above factors when selecting an epoch length and deciding whether to use CRIB instead of (or in addition to) another algorithm.

## Potential Novel Applications for CRIB

While CRIB is applicable to common behaviors of interest in physical behavior research (e.g., sleep, SB, MVPA, and ambulation), there are also potential applications in less common areas.

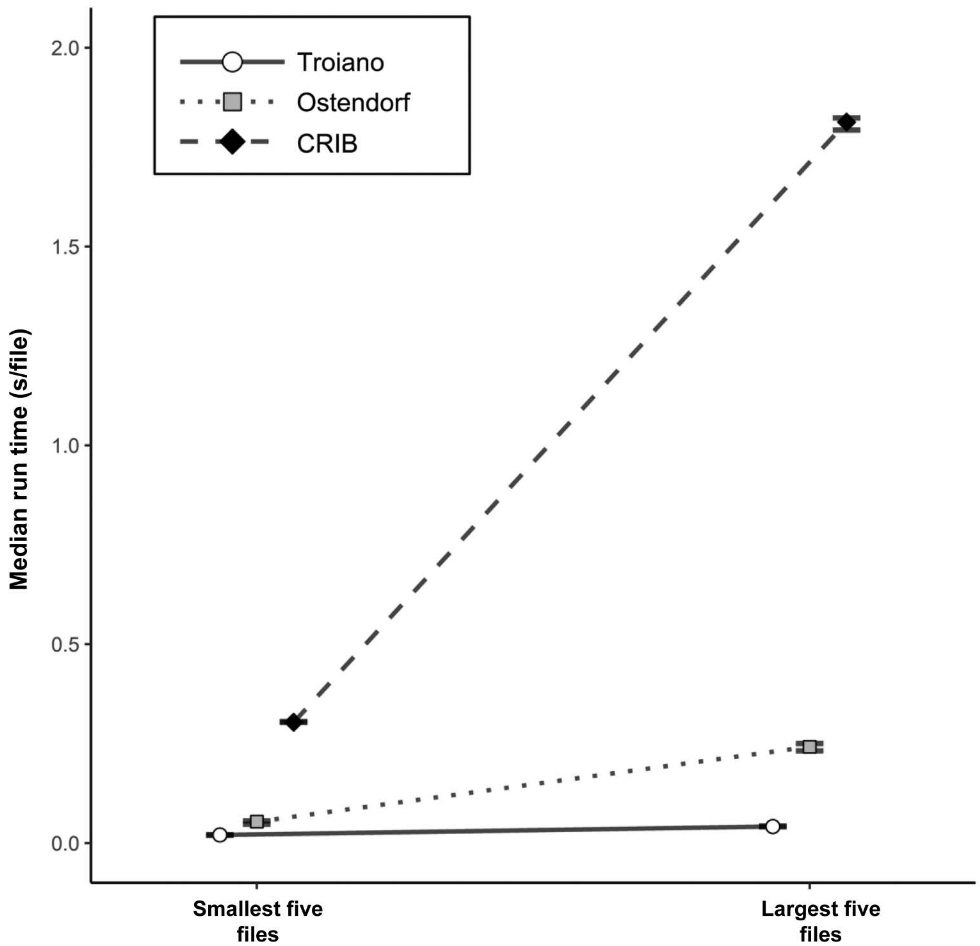



**Figure 6** — Runtimes when processing the five smallest and largest accelerometer data files (compressed R-native format extracted from the original.xpt files) for the participants in the sample analysis (overall $N = 2{,}733$). Each file was processed 10 times, and runtime values (seconds per file) are shown as median ± interquartile range. CRIB = Clustered Recognition of Interrupted Bouts.

For example, its utility can potentially be extended to spatial data from global positioning systems and geographic information systems. An example would be identifying bouts of engagement in certain activity spaces (e.g., within a certain buffer around the home or school address). Once the global positioning systems and geographic information systems data have been processed into a time stamped sequence of activity space labels, CRIB can be applied to classify the bouts while accounting for brief excursions outside the buffer. This may position CRIB to address certain previously noted challenges associated with defining buffers in the first place (Jankowska et al., 2015).

Nonwear analysis is another potential novel application for CRIB. Although nonwear analysis is not normally considered a type of bout analysis, it can nevertheless be conceptualized as a stable pattern of accelerometer signal (traditionally 0 counts per min), with the possibility of brief interruptions (e.g., being moved from one location to another). To apply CRIB in this setting, a starting point could be to use settings similar to the algorithm of Choi et al. (2011), for example, by setting the target buffer to 30 min and allowing ≤1 interruption per motionless bout, with a maximum interruption length of 2 min and a minimum bout length of 60 min. Alternatively, an extension could be implemented by adding a setting for the required percent engagement within the bout.

## Conclusion

The comparison-focused design of CRIB allows it to complement a range of chronology-focused algorithms. The additional information may help to better understand behavioral patterns and their health implications. Future studies can refine CRIB by experimenting with design tweaks, for example, using different clustering approaches, linkage techniques, or distance metrics. Validation work is also necessary to develop a catalog of recommended settings when using CRIB for different applications (e.g., for different behaviors, devices, and populations). Within this work, it will be important to investigate the implications for health, in addition to looking at alignment of the predicted bouts with criterion data (e.g., from direct observation). Additionally, as CRIB is applied in these diverse settings, more insights may emerge to suggest how else bout classification can be standardized and enriched, particularly through the use of unsupervised data mining (Twaites, 2019).

## References

Altenburg, T.M., & Chinapaw, M.J.M. (2015). Bouts and breaks in children's sedentary time: Currently used operational definitions and recommendations for future research. *Preventive Medicine, 77*, 1–3. https://doi.org/10.1016/j.ypmed.2015.04.019

Barry, G., Galna, B., Lord, S., Rochester, L., & Godfrey, A. (2015). Defining ambulatory bouts in free-living activity: Impact of brief stationary periods on bout metrics. *Gait & Posture, 42*(4), 594–597. https://doi.org/10.1016/j.gaitpost.2015.07.062

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, Methodological, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bland, J.M., & Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Carson, V., & Janssen, I. (2011). Volume, patterns, and types of sedentary behavior and cardio-metabolic health in children and adolescents: A cross-sectional study. *BMC Public Health, 11*(1), 274. https://doi.org/10.1186/1471-2458-11-274

Choi, L., Liu, Z., Matthews, C.E., & Buchowski, M.S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine & Science in Sports & Exercise, 43*(2), 357–364. https://doi.org/10.1249/MSS.0b013e3181ed61a3

Freedson, P.S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine & Science in Sports & Exercise, 30*(5), 777–781. https://doi.org/10.1097/00005768-199805000-00021

Granat, M.H. (2012). Event-based analysis of free-living behaviour. *Physiological Measurement, 33*(11), 1785–1800. https://doi.org/10.1088/0967-3334/33/11/1785

Hartigan, J.A. (1981). Consistency of single linkage for high-density clusters. *Journal of the American Statistical Association, 76*(374), 388–394. https://doi.org/10.1080/01621459.1981.10477658

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), 48–56. https://doi.org/10.1249/JES.0000000000000035

Kuiper, F.K., & Fisher, L. (1975). A Monte Carlo comparison of six clustering procedures. *Biometrics, 31*(3), 777–783. https://doi.org/10.2307/2529565

McQuitty, L.L. (1957). Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement, 17*(2), 207–229. https://doi.org/10.1177/001316445701700204

Ostendorf, D.M., Lyden, K., Pan, Z., Wyatt, H.R., Hill, J.O., Melanson, E.L., & Catenacci, V.A. (2018). Objectively measured physical activity and sedentary behavior in successful weight loss maintainers. *Obesity, 26*(1), 53–60. https://doi.org/10.1002/oby.22052

Troiano, R.P., Berrigan, D., Dodd, K.W., Mâsse, L.C., Tilert, T., & Mcdowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise, 40*(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3

Twaites, J. (2019). *Data dependent identification and characterisation of physical activity bouts. Symposium IV: Doubts about bouts: Time in Bouts of MVPA fails to characterise patterns of physical activity* [Conference presentation June 26-28, 2019, Maastricht, The Netherlands]. International Conference on Ambulatory Monitoring of Physical Activity and Movement. https://ismpb.org/wp-content/uploads/2019/06/ICAMPAM-2019-program-107-web.pdf

van der Berg, J.D., Willems, P.J.B., van der Velde, J.H.P.M., Savelberg, H.H.C.M., Schaper, N.C., Schram, M.T., Sep, S.J.S., Dagnelie, P.C., Bosma, H., Stehouwer, C.D.A., & Koster, A. (2016). Identifying waking time in 24-h accelerometry data in adults using an automated algorithm. *Journal of Sports Sciences, 34*(19), 1867–1873. https://doi.org/10.1080/02640414.2016.1140908

Winkler, E.A.H., Bodicoat, D.H., Healy, G.N., Bakrania, K., Yates, T., Owen, N., Dunstan, D.W., & Edwardson, C.L. (2016). Identifying adults' valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol. *Physiological Measurement, 37*(10), 1653–1668. https://doi.org/10.1088/0967-3334/37/10/1653