Received: 23 August 2021 | Accepted: 29 September 2022

DOI: 10.1111/biom.13773

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Pattern-based clustering of daily weigh-in trajectories using dynamic time warping

Samantha Bothwell[1] | Alex Kaizer[1] | Ryan Peterson[1] | Danielle Ostendorf[2] | Victoria Catenacci[2] | Julia Wrobel[1]

[1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

[2]Department of Medicine, Division of Endocrinology, Metabolism, and Diabetes, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA

Correspondence
Samantha Bothwell, Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA.
Email: samantha.bothwell@cuanschutz.edu

Abstract

"Smart"-scales are a new tool for frequent monitoring of weight change as well as weigh-in behavior. These scales give researchers the opportunity to discover patterns in the frequency that individuals weigh themselves over time, and how these patterns are associated with overall weight loss. Our motivating data come from an 18-month behavioral weight loss study of 55 adults classified as overweight or obese who were instructed to weigh themselves daily. Adherence to daily weigh-in routines produces a binary times series for each subject, indicating whether a participant weighed in on a given day. To characterize weigh-in by time-invariant patterns rather than overall adherence, we propose using hierarchical clustering with dynamic time warping (DTW). We perform an extensive simulation study to evaluate the performance of DTW compared to Euclidean and Jaccard distances to recover underlying patterns in adherence time series. In addition, we compare cluster performance using cluster validation indices (CVIs) under the single, average, complete, and Ward linkages and evaluate how internal and external CVIs compare for clustering binary time series. We apply conclusions from the simulation to cluster our real data and summarize observed weigh-in patterns. Our analysis finds that the adherence trajectory pattern is significantly associated with weight loss.

KEYWORDS
binary time series, dynamic time warping, hierarchical clustering, "smart"-scales

1 | INTRODUCTION

Obesity is an epidemic in the United States, affecting an estimated 42.5% of American adults over the age of 20 as of 2018 (CDC - National Center for Health Statistics, 2021). In weight loss studies, despite initial weight loss success, participants often slowly regain the weight lost over time (Ross et al., 2018). For example, Anderson et al. (2001) found that participants who enrolled in weight loss studies regained about 50% of weight lost within two years. There is substantial evidence that frequent weighing is associated with weight loss maintenance (Linde et al., 2005; Steinberg et al., 2013, 2015; Wilkinson et al., 2017; Wing & Phelan, 2005) and a decrease in self-weighing is associated with weight regain in the National Weight Control Registry (NWCR) (Thomas et al., 2014).

New "smart"-scales are a remote tool for clinicians and weight-loss researchers to electronically collect at-home weight measurements (Krukowski & Ross, 2020; Thomas & Bond, 2014). Participants are typically asked to weigh themselves daily, which allows for monitoring of near-continuous weight change and weigh-in behavior in a

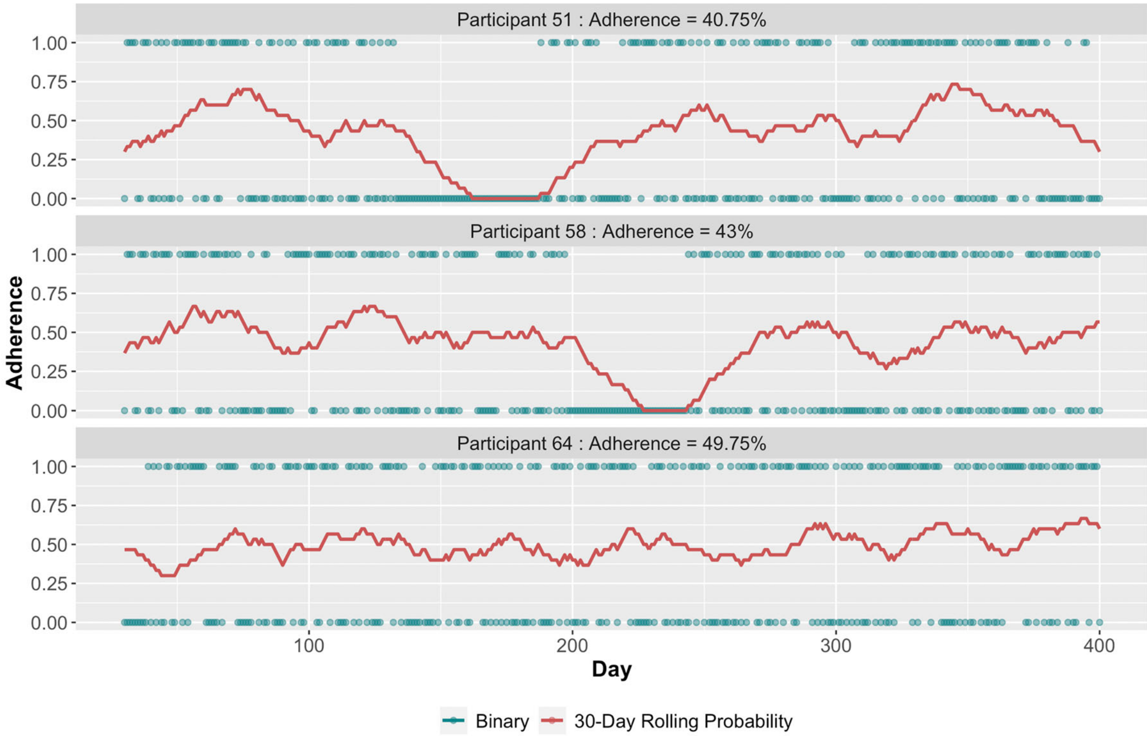
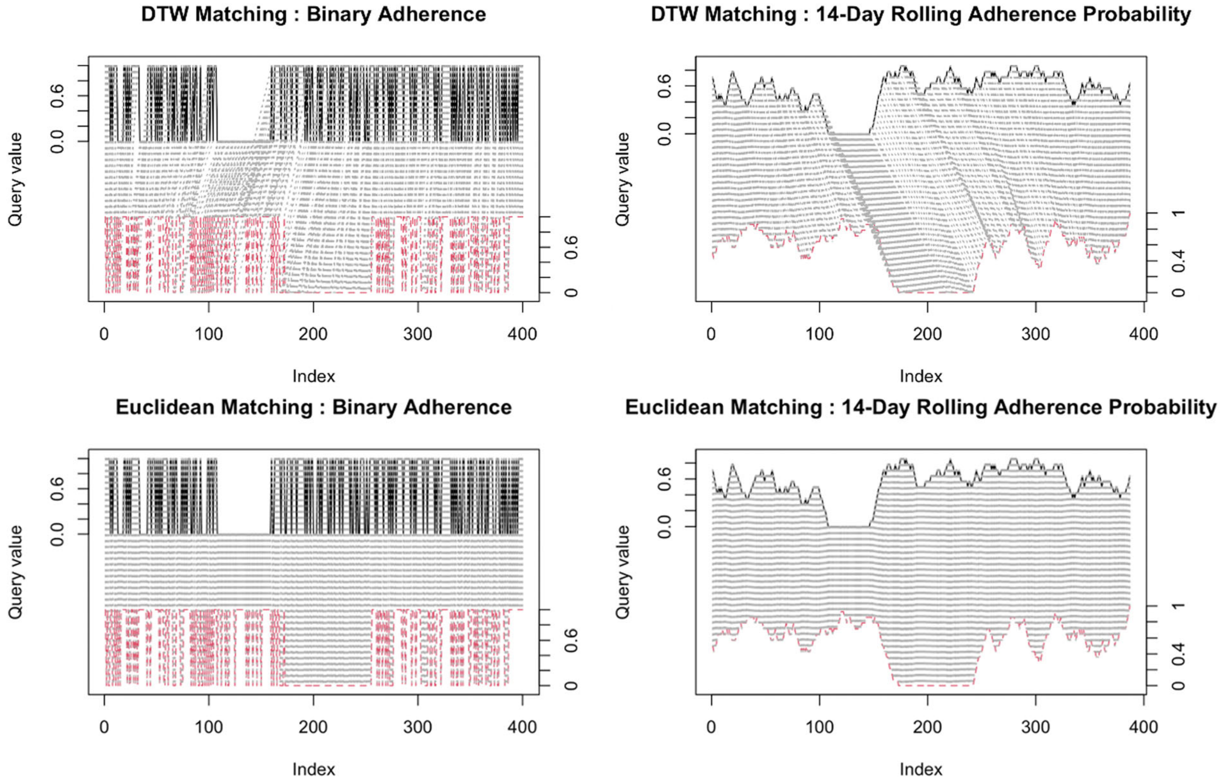


FIGURE 1 Example adherence trajectories $Y_i(t)$ for three simulated participants over 400 days, with the same level of overall weigh-in adherence. Probabilities $\pi_i(t)$ are calculated using a rolling probability with a 30-day window. Participants 51 and 58 have a "vacation" break from weighing in. Participant 64 maintains consistent adherence throughout the 400 days. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

free-living environment, potentially over long periods of time. As a by-product of daily weight, the "smart"-scales also implicitly record adherence to daily weigh-in instructions, which offers an opportunity to uncover patterns in self-weighing behavior, which have recently been shown to associate with weight-loss in a latent class trajectory analysis (Zheng et al., 2019).

Measurements from these "smart"-scales can be used to construct a binary time series for each study participant that indicates whether they stepped on their "smart"-scale each day. Let $Y_i(t)$, $i \in (1, \ldots, N)$, represent the daily weigh-in behavior over many months for participant i at time t and $\pi_i(t) = \Pr(Y_i(t) = 1)$ represent a participant-specific probability of weigh-in at time t. Example binary time series and their underlying probability of weigh-in, the combination of which we refer to as *adherence trajectories*, are shown in Figure 1. All three participants have an average weigh-in adherence of 40–50% over 400 days, but with different patterns of weigh-in behavior over time. Participant 64 maintains relatively consistent adherence throughout the study, whereas participants 51 and 58 have a gap where they do not weigh-in. Patterns in weigh-in behavior revealed by these adherence trajectories may offer additional insight into long-term weight loss and weight maintenance. Our goal is to cluster these trajectories to understand time-varying weigh-in adherence patterns, and how they correlate with overall weight loss.

Clustering is a popular class of machine learning algorithms for characterizing trends and grouping data based on user-defined similarity. Despite the immense popularity of clustering, applications focused on identifying patterns in binary time series are limited. Our approach is to use agglomerative hierarchical clustering with a dynamic time warping (DTW) distance to define similarity between series from different participants. DTW is a popular approach for creating clusters based on patterns over a time series, and has been used to recognize speech patterns, predict stock price, monitor crop dynamics for industrial farms, and monitor wear time of medical interventions in sleep apnea (Belgiu et al., 2020; Berndt & Clifford, 1994; Bottaz-Bosson et al., 2021; Juang, 1984). In Figure 1, participants 51 and 58 have similar patterns of breaks, or "vacations," during the study. However, their total adherence percentage is close to that of participant 64. Euclidean matching maps points at the same time between two series. With this strategy, patterns that occur at different points in time will not be identified. DTW, on the other hand, performs an optimal matching that considers a shift in time series. This strategy thus identifies patterns that happen at different time points and would be



DTW Cost Matrix : Binary Data

1	2	2	2	2	2	2	1	1
0	2	2	2	2	1	1	1	0
1	1	1	1	1	2	3	0	1
1	1	1	1	1	2	3	0	1
1	1	1	1	1	2	2	0	1
1	1	1	1	1	1	1	0	1
0	1	1	1	1	0	0	1	1
1	0	0	0	0	1	1	2	3
	1	**1**	**1**	**1**	**0**	**0**	**1**	**0**

DTW Cost Matrix : Continuous Data

6	18	9	9	9	9	7	5	7
10	16	8	8	8	8	5	9	9
5	10	5	5	5	3	6	7	10
7	9	3	3	3	5	6	7	8
7	6	3	3	3	5	6	7	8
4	3	6	9	10	9	11	13	15
3	3	6	7	8	7	11	11	13
6	2	3	4	5	6	8	8	10
	4	**7**	**7**	**7**	**5**	**8**	**6**	**8**

FIGURE 2 The top four graphs represent matching strategies for the DTW and Euclidean distances under binary and continuous conditions. The two series represent simulated data for participants who have a "vacation" break from weighing in. The bottom two matrices represent DTW cost matrices for binary and continuous data, respectively, where the bolded values on the x- and y-axes represent distinct time series. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

more likely to place participants 51 and 58 together in a different cluster than participant 64 because they exhibit similar behavior.

Figure 2 visualizes the matching strategy for the DTW and Euclidean distances under binary and continuous conditions of two similar series. Both series experience a "vacation," where they stop weighing in, but at different times. Euclidean matching is one-to-one, so, as we see in the figure, the vacations are not matched for either the binary or continuous cases. Since DTW is sensitive to the time shift, we see that vacations are matched in the continuous case but, because DTW is not optimized for binary series, these are not as easily picked up in the binary case.

We implement a DTW distance-based hierarchical clustering approach for dense time series of weigh-in adherence data arising from "smart"-scale data. These scales provide the opportunity to assess free-living weigh-in behavior and its relation to weight-loss. Because DTW-based clustering has primarily been developed for continuous time series and applications of DTW for binary data are limited, we conduct extensive simulations to investigate the properties of DTW-based clustering on our binary adherence data and establish guidelines for selecting optimal clustering and data processing strategies including linkage, number of clusters, and probability window for smoothing binary series. The simulation framework and binary clustering guidelines we provide can be generalized to other binary time series applications. Briefly, Section 2 reviews the literature on hierarchical clustering pertinent to our application, Section 3 describes our weigh-in study data and our proposed clustering and analysis approach, Section 4 describes our simulation design and results, and Section 5 extends the simulation findings to our motivating data. We conclude in Section 6.

2 | LITERATURE REVIEW

2.1 | Hierarchical clustering

Common clustering approaches include hierarchical clustering and partition-based clustering, the most popular of which is K-means (Saxena et al., 2017). Partition-based approaches like K-means minimize the squared-Euclidean distance between observations and have been found to be suboptimal for performing DTW-based clustering (Niennattrakul & Ratanamahatana, 2007). In contrast, hierarchical clustering defines similarity between pairs of observations based on a user-specified distance method. Because hierarchical clustering can seamlessly incorporate the DTW distance, it is better suited for our application.

Hierarchical clustering can be performed in an agglomerative or divisive manner (Maimon & Rokach, 2006). Agglomerative clustering is a "bottom-up" approach where individual observations begin as their own cluster and are merged in pairs at each hierarchy. Divisive clustering is a "top-down" approach where all observations begin in one cluster and are split at each hierarchy. We focus on an agglomerative approach. In addition, hierarchical clustering operates under a user-specified distance method and linkage. With these specifications, the algorithm combines observations into clusters one-by-one, creating a hierarchical tree, or dendrogram, which visualizes the order in which clustering occurred. The dendrogram is then split to form the desired number of clusters.

2.2 | Distance methods

The distance method defines the relationship between each pair of time series. Careful selection of distance method is critical to achieve desired clustering behavior. Euclidean and Jaccard distances are common choices, with the Jaccard distance particularly well suited for binary data. While both distances are used regularly in practice, neither accounts for shifts in the time scale between series (Keogh & Pazzani, 2000). DTW creates clusters based on patterns over a time series (Berndt & Clifford, 1994) and has been shown to outperform extensions of Euclidean distance for this purpose (Keogh & Pazzani, 2000). However, DTW distance has not been widely used for binary data.

DTW was first proposed by Sakoe and Chiba (1970) to recognize audio signatures of specific words given different speaking patterns. The authors proposed adjusting the time differences between two speech patterns to attain maximum matching. Three criteria were specified for DTW. First, every index from both series must be matched with one or more indices from the other series. Second, the first and last indices from both series must be matched. Finally, the mapping of indices must be monotonically increasing.

2.2.1 | Time series clustering using dynamic time warping

DTW has been used extensively to cluster continuous time series data. Keogh and Pazzani (2000) compared the performance of DTW to Euclidean distance on an Australian Sign Language dataset that measured the x-axis position of a participant's right hand when signing 95 words. The study found that Euclidean distance, while quick to compute, performed only slightly better than random guessing due to its inability to adjust for the time axis distortion. Giusti and Batista (2013) compared the performance of 48 distance methods including DTW, Euclidean, and Jaccard on 42 continuous time series datasets. Jaccard distance performed slightly, but not significantly, better than Euclidean. DTW, however, performed significantly better than Euclidean and performed best when parameters were optimally tuned.

DTW is known to be computationally expensive, and work has been done to improve its accuracy and efficiency for clustering continuous time series (Iwana et al., 2020; Luczak, 2016; Li et al., 2020; Mueen et al., 2016). AWarp (Mueen et al., 2016), developed for sparse time series, and DTW-NN (Iwana et al., 2020), a neural-network based adaptation, have both been shown to be more computationally efficient alternatives to standard DTW for pattern-based clustering. In addition, Luczak (2016) found that applying a parametric combination of DTW on the raw and numeric first derivative of the data outperformed strict DTW. Adaptive constrained DTW (Li et al., 2020) incorporates adaptive penalty functions to overcome overstretching and overcompression problems that come with DTW.

Further adaptations to DTW have been made to identify and cluster by specific time-localized features in time series data. Guijo-Rubio et al. (2021) explored a segmentation approach, TS3C, which clusters segments of different time series to determine common patterns. Zakaria et al. (2012) and Shah et al. (2016) propose "shapelet," or "subsequence," approaches to extract localized shapes for clustering. Liang et al. (2021) extended DTW to two-dimensional vessel spatial trajectories using a convolutional autoencoder.

These extensions have been shown to improve the performance of DTW for clustering continuous time series, but little work has been done to assess the performance of DTW for clustering dense binary time series. While our adherence trajectories can be smoothed to produce $\pi_i(t)$, continuous trajectories that represent probability of adherence over time, the smoothing window must be carefully selected to ensure optimal clustering performance on our data. In Section 3, we discuss our strategy for selecting optimal probability window, linkage, and number of clusters, and through simulation establish guidelines for selecting these parameter values in our real data.

2.3 | Linkages

Linkage defines the strategy for combining pairs of observations into clusters. Common linkages are single, complete, average, and Ward. In each linkage strategy, every point begins as its own cluster, and then the first cluster is formed between the points with the minimum distance. Clusters are further combined using the following definitions (Ackerman & Ben-David, 2016; Miyamoto et al., 2015). With *single linkage*, the minimum distance between points in different clusters is minimized over all clusters. With *complete linkage*, the maximum distance between points in different clusters is minimized over all clusters. With *average linkage*, the average distance between points in different clusters is minimized. Finally, with *Ward linkage*, the within cluster variance after combining clusters is minimized.

Vijaya and Batra (2019) compared the performance of the single, complete, and Ward linkages to cluster groups of individuals based on income and shopping expenditure. Single linkage created long chains of points, making it difficult to distinguish clusters. Complete and Ward linkages performed similarly but the Ward linkage performed better for poorly separated clusters.

2.4 | Number of clusters

Hierarchical clustering does not require the number of clusters, K, to be prespecified. Rather, the dendrogram resulting from a specific combination of distance method and linkage strategy is split into K clusters after the algorithm has run. Determining the optimal K for a unique dataset is challenging and has motivated research into methods that guide the decision, but it remains an open problem (Salvador & Chan, 2004). Internal cluster validation, which we discuss in Section 2.5, can assist with the selection of optimal K when the true number of clusters is unknown (Islam et al., 2015).

2.5 | Validation indices

Cluster validation indices (CVIs) are used to evaluate the quality of a clustering algorithm. Internal CVIs are useful for estimating the optimal number of clusters when the true cluster assignment is unknown, whereas external CVIs are used to assess clustering accuracy when true cluster labels are available. Internal CVIs are based on compactness, separation, and connectivity (Liu et al., 2010). *Compactness* measures how close objects within the same cluster are. A lower within-cluster variance of the measures indicates better compactness (Liu et al., 2010). *Separation* measures how far apart clusters are from each other (Liu et al., 2010). *Connectivity* measures the extent to which items are placed in the same cluster as their nearest neighbors in the data space (Liu et al., 2010). Common internal CVIs are the Silhoutte, Davies-Bouldin (DB), Dunn, and COP indices (Ansari et al., 2015; Arbelaitz et al., 2013). Arbelaitz et al. (2013) conducted a comparative study of internal CVIs using real and simulated continuous data. The results of their study found that the Silhouette and Davies-Bouldin indices selected the correct number of clusters most often.

Common external CVIs are the Rand, Adjusted Rand, Jaccard, and Fowlkes-Mallows indices (Arbelaitz et al., 2013). Brun et al. (2007) compared external CVIs with 600 simulated datasets under various clustering algorithms, and found that the Rand index outperformed the Jaccard and Fowlkes-Mallows indices in their setting.

Through simulation in which ground truth is known, a comparison of best performing external and internal CVIs can be used to guide clustering parameter choices for real data analyses in which only internal CVIs can be used because cluster assignment is unknown. However, these previous studies only compared CVIs in continuous data scenarios.

3 | METHODS

Through simulation, we delineate a pragmatic approach to pattern-based clustering under combinations of distance methods, including Euclidean, Jaccard, and DTW, and linkage strategies, including single, complete, average, and Ward. Specifically, it is necessary to examine clustering performance for binary time series. We expect that our smoothed adherence probabilities will cluster more naturally using DTW than the raw binary series. This requires optimizing an additional parameter, the probability window, for smoothing the binary trajectories into continuous trajectories representing probability of adherence. The goal of our simulation study is to establish how internal and external CVIs can be used to guide parameter choices for analysis of our motivating daily weigh-in data.

3.1 | Notation

In Section 1, we introduced a binary weigh-in indicator $Y_i(t), i \in (1, \ldots, N)$, for participant i at time t; $t \in (1, \ldots, T)$. On a discrete scale, the binary series can be represented as $\boldsymbol{Y}_i$ where $\{Y_{i1}, Y_{i2}, \ldots, Y_{iT}\} \in \boldsymbol{Y}_i$. We then define a discrete participant-specific probability of weigh-in at time t to be $\pi_{it} = \Pr(Y_{it} = 1)$. We further define π_{it}^{w} to be a rolling probability of adherence obtained by smoothing $\boldsymbol{Y}_i$ over window $w \in \{7, 14, 30\}$.

For each pair of time series, $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$, we calculate the Euclidean, Jaccard, and DTW distances, denoted as $d_{Euclidean}$, $d_{Jaccard}$, and d_{DTW}, respectively. We evaluate four linkages, L, denoted as L_{single}, $L_{complete}$, $L_{average}$, and L_{Ward}. We calculate the linkages in Section 3.4.2 using example clusters A and B. Cluster performance is evaluated using two external CVIs, Adjusted Rand (AR) and Jaccard (J), and two internal CVIs, Dunn (D) and Silhouette (Sil).

3.2 | Data collection

Data were collected as part of an ongoing randomized weight loss trial at the University of Colorado Anschutz Health and Wellness Center (Ostendorf et al., 2022). Participants were men and women aged 18–55 years with BMI in 27–45 kg/m^2. The primary outcome of the weight loss trial is weight change at the end of a 12-month weight loss intervention that consisted of an energy-restricted diet, increased physical activity, and weekly group-based weight-loss support; with a follow-up measure 18 months after initial enrollment.

Participants were asked to step onto a BodyTrace "smart"-scale (BodyTrace, Inc., 2021) once a day to record their weight. Scale data were converted into binary trajectories of weigh-in adherence, where 1 indicates the participant weighed in on that day and 0 indicates they did not weigh in, and truncated to 400 days for all participants.

3.3 | Data processing

Data cleaning was required to ensure "smart"-scale weight measurements were obtained from the study participant and not other household members. Researcher-verified clinic weights were measured at CU-AHWC every week for the first 15 weeks and then biweekly for the remainder of the study. "Smart"-scale weights were removed if they differed substantially from the mean of the researcher-verified weights for a given period. Additionally, same day observations were removed and the weight closest to the researcher verified weight for that period was selected. When specified, binary time series were transformed into to trajectories of rolling probabilities for each participant, π_{it}^{w}, defined as $\pi_{it}^{w} = \frac{\sum_{s=t-w+1}^{t} Y_{is}}{w}$ for probability window w and time t; $t \in \{1, \ldots, 400\}$. We evaluate $w \in \{7, 14, 30\}$ days.

3.4 | Clustering approach

Here, we describe simulation, clustering, and analysis of our adherence trajectories.

3.4.1 | Distance methods

We compare the Euclidean, Jaccard, and DTW distances for clustering the binary series. Given two time series $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$, Euclidean distance is:

$$d_{Euclidean} = \sqrt{\sum_{s=1}^{T} (Y_{is} - Y_{js})^2}.$$

The Jaccard distance, commonly used for binary data, is calculated as the ratio of the size of the intersection between the two series and the size of the union (Irani et al., 2016):

$$d_{Jaccard} = 1 - \frac{|\boldsymbol{Y}_i \cap \boldsymbol{Y}_j|}{|\boldsymbol{Y}_i \cup \boldsymbol{Y}_j|}.$$

The DTW distance between two series is calculated by creating a $T \times T$ cost matrix, M, with one series on the x-axis and the other on the y-axis (see Figure 2). Let $s \in (1, \ldots, T)$ represent the time index for series $\boldsymbol{Y}_i$ and $r \in (1, \ldots, T)$ the time index for $\boldsymbol{Y}_j$. Then, $M(s, r)$ corresponds to the specific element of the cost matrix, M, that maps Y_{si} to Y_{rj}. The value of $M(s, r)$ is calculated as:

$$M(s, r) = |Y_{is} - Y_{jr}| + \min[M(s-1, r-1), M(s, r-1), M(s-1, r)].$$

Using the cost matrix , M, an optimal matching series of length l is formed such that $d_{DTW} = \frac{\sum_{n=1}^{l} m_n}{l}$ is minimized, where m_n is the nth element in matched series (Sakoe & Chiba, 1970). The third row of Figure 2 shows example DTW cost matrices for binary and continuous data with the optimal matching series highlighted. The x- and y-axes represent two time series. For the binary example (left column), the first four columns are identical. This can make it different to determine a uniquely optimal path, compared to continuous data (right column).

3.4.2 | Linkage strategies

Our simulation evaluates each distance method combined with complete, average, single, and Ward linkages. Let A and B represent two clusters, then the linkages are defined as (Ackerman & Ben-David, 2016; Miyamoto et al., 2015):

- Single Linkage : $L_{single}(A, B, d) = \min_{Y_i \in A, Y_j \in B} d(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$
- Complete Linkage : $L_{complete}(A, B, d) = \max_{Y_i \in A, Y_j \in B} d(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$
- Average Linkage : $L_{average}(A, B, d) = \frac{\sum_{\boldsymbol{Y}_i \in A, \boldsymbol{Y}_j \in B} d(\boldsymbol{Y}_i, \boldsymbol{Y}_j)}{|A| \cdot |B|}$
- Ward Linkage : After combining clusters, $L_{Ward}(A) = \sum_{Y_i \in A} ||\boldsymbol{Y}_i - \bar{A}||^2$ is minimized

3.4.3 | Validation indices

In our simulation, we use external CVIs to determine which distance method is most appropriate for recovering true clusters in our data. Both internal and external CVIs are evaluated to determine the optimal linkage strategy and probability window for application to our real data. The external CVIs that we consider are the adjusted Rand and Jaccard indices. The internal CVIs that we consider are the Dunn and Silhouette indices. Equations for these validation indices are in Supporting Information Section 1.

3.4.4 | Simulation design

Our simulation study evaluates pattern-based clustering on the raw binary trajectories $\mathbf{Y}$ as well as three different smoothed adherence probability trajectories $\boldsymbol{\pi}^7, \boldsymbol{\pi}^{14}$, and $\boldsymbol{\pi}^{30}$ representing 7-day, 14-day, and 30-day rolling probability windows, respectively. For each type of series, for example, the rolling 14-day adherence trajectory, $\boldsymbol{\pi}^{14}$, we evaluate performance of both the Euclidean and DTW distances in combination with each of the four linkages (single, average, complete, and Ward). For the raw binary trajectories only, we also evaluate the performance of the Jaccard distance.

We generate clusters from $Bernoulli(p_{adherence})$ distributions under five simulation scenarios, where $p_{adherence} \in [0, 1]$ is the probability of adherence and varies by cluster. Unless specified otherwise, data across time points and participants were generated independently. The first three clusters are the low, medium, and high adherence clusters with $p_{adherence} = 0.32, 0.655$, and 0.80, respectively. These values of $p_{adherence}$ were chosen because they align with the lower 25%, median, and upper 75% quantiles of adherence values from our real data. The fourth cluster represents "Drop-outs" who have an initial adherence of $p_{adherence} = 0.655$, but fall to $p_{adherence} = 0.05$ for the rest of the study after a randomly selected drop-out start day generated from a $N(200, (50)^2)$ distribution. The final cluster represents "Vacationers" who have an initial adherence of $p_{adherence} = 0.655$, but fall to $p_{adherence} = 0$ during a "vacation" period, with a start day that is randomly generated from a $Uniform[30, 300]$ distribution and a vacation length generated from a $Uniform[30, 90]$ distribution.

We simulate 500 datasets where each dataset contained 100 participants, 20 from each cluster. For each simulated dataset, we cluster under each distance, linkage, and probability window and calculate external and internal CVIs.

3.4.5 | Data analysis

In our simulation, we use agreement between external and internal CVIs to determine optimal rolling probability windows and linkage strategies. We also determine which internal CVI(s) best recover the true number of clusters. We then use these findings to guide analysis of our real adherence trajectory data to uncover weigh-in patterns. Optimal number of clusters will be determined using the internal validation metric chosen in the simulation study.

We test the association between percent weight change from baseline and cluster assignment using a linear model. Both "smart"-scale and clinic weights were available for each participant, and the "final" weight measurement for each participant was taken as the most recent "smart"-scale or clinic weight measurement. We run a second linear model replacing cluster assignment with overall percent adherence and compare these models using the Akaike information criterion (AIC). Both models also control for baseline weight and age.

TABLE 1 Validation indices to compare distances using a 14-day rolling window for DTW and Euclidean distances and binary for the Jaccard distance. AR, adjusted rand. CVI summaries are reported as Median [IQR]. Bold values indicate most optimal results within each linkage and CVI combination

	AR index	Jaccard index
Complete linkage		
DTW	**0.78 [0.78, 0.78]**	**0.70 [0.70, 0.70]**
Euclidean	0.74 [0.69, 0.78]	0.68 [0.64, 0.72]
Jaccard	0.74 [0.69, 0.78]	0.68 [0.64, 0.72]
Average linkage		
DTW	**0.78 [0.78, 0.78]**	**0.70 [0.70, 0.70]**
Euclidean	0.69 [0.63, 0.74]	0.64 [0.60, 0.69]
Jaccard	0.02 [0.02, 0.03]	0.28 [0.28, 0.28]
Single linkage		
DTW	**0.99 [0.79, 0.99]**	**0.99 [0.75, 0.99]**
Euclidean	0.67 [0.56, 0.68]	0.63 [0.55, 0.64]
Jaccard	0.02 [0.02, 0.02]	0.28 [0.28, 0.28]
Ward linkage		
DTW	**0.78 [0.78, 0.78]**	**0.70 [0.70, 0.70]**
Euclidean	0.57 [0.53, 0.61]	0.51 [0.48, 0.54]
Jaccard	0.73 [0.68, 0.82]	0.68 [0.63, 0.76]

3.4.6 | Implementation details

All analyses were performed in R, version 4.0.1 (R Core Team, 2020). Euclidean and Jaccard distances were calculated using the `distance` function from the `ecodist` package (Goslee & Urban, 2007), and DTW distance was calculating using `dist` with the `dtwclust` package (Sardá-Espinosa, 2019). The linkage and cluster number were specified in the `hclust` and `cutree` functions, respectively.

4 | SIMULATION RESULTS

DTW is more computationally costly than Euclidean or Jaccard distance, so slower run times were expected. Across all simulation scenarios, the average running time for DTW was 42.5 s compared to 0.27 and 0.89 s for Euclidean and Jaccard, respectively (Table S1 in the Supporting Information). Table 1 summarizes external CVI values for DTW and Euclidean distances calculated on pairs of adherence trajectories with a 14-day probability window, $\boldsymbol{\pi}^{14}$, and Jaccard distance calculated on pairs of raw binary trajectories $\mathbf{Y}$. These were found to be the optimal transformations for each distance method, see Table S2 in the Supporting Information for CVI results using other probability windows. External CVIs in Table 1 are displayed for these distances in combination with each linkage. Across all linkages, DTW outperforms both the Euclidean and Jaccard distances as measured by maximum score for both the AR and Jaccard indices. It is expected that DTW would outperform Euclidean distance for accurately clustering the "Drop-out" and "Vacationer" clusters because it is designed to recognize patterns that occur at different times between series. Note, however, that DTW even performs well in accurately clustering the low, medium, and high adherence groups, which do not contain shape-based patterns and are expected to be adequately identified using Euclidean distance.

Table S3 in the Supporting Information summarizes external and internal CVIs for data transformations $\mathbf{Y}$, π^7, π^{14}, and π^{30} (representing raw binary trajectories and smoothed adherence probability trajectories with 7-, 14-, or 30-day rolling probability windows) in combination with each linkage strategy. Results are shown only using DTW distance, with the goal of using these results to determine which data transformation and linkage strategy is best for DTW-based clustering. As expected, since DTW is designed for clustering continuous trajectories, the rolling probability windows provide better clustering performance than the unprocessed binary data. The Dunn index and external CVIs are maximized across simulation scenarios for the single linkage with a 14-day rolling probability. Under Ward linkage, we also find that a 14- or 30-day rolling probability window is optimal. While CVIs are lower for the Ward linkage than single linkage, Ward is more robust to different selections for the rolling probability window. In addition, there is less variability within external CVIs for the Ward linkage compared to all other linkages. With the exception of the Ward 30-day window, the Dunn index agrees with external CVIs results across all simulation scenarios, whereas the Silhouette index is consistently highest for the 30-day window. This indicates that the Dunn index may be a better evaluation metric for our real data analysis.

Recall that there are five true clusters in the simulated data. For each simulation, we calculated internal CVI values for three, four, five, and six clusters when using both the Ward and single linkage applied to 14- and 30-day data transformations. Table S4 in the Supporting Information shows the percentage of simulated datasets in which each cluster number was chosen under these linkages and data transformations. The Dunn index is able to uncover the true number of clusters more often than the Silhouette index. Using the Dunn index for selecting number of clusters, the correct number of clusters was chosen most often under Ward linkage with a 30-day probability window and single linkage with a 30-day probability window.

To cluster weigh-in adherence for our study, we will use DTW distance with Ward linkage as it was the most robust, least variable linkage across window sizes, and led to correctly identifying the number of clusters most often. We will cluster using both the 14- and 30-day rolling probability windows as both the Dunn and Silhouette indices generally agreed with the external CVIs in these scenarios. The Dunn index was able to recover the true number of clusters more often than the Silhouette index and will therefore be used to choose the number of clusters for our real weigh-in adherence trajectory data.

5 | DATA APPLICATION

Of the 55 included participants in the weight loss trial, 50% were between 33 and 50 years old, and 50% had a baseline BMI between 29.44 and 35.33. The majority of the cohort is white (87.3%) and female (74.5%); see Table S7 in the Supporting Information.

5.1 | Cluster evaluation

Guided by our simulation results, we clustered 14- and 30-day probability windows of our weigh-in adherence trajectories using DTW distance with Ward linkage. Internal CVIs were calculated for three, four, five, and six clusters. Dunn index was used to select the best clustering strategy among the combinations of two data transformations and four cluster sizes. This resulted in a final clustering strategy that used a 14-day probability window data transformation, DTW distance with Ward linkage, and six clusters.

Figure 3 visualizes the individual binary trajectories (top row) and 14-day probability trajectories (bottom row) by cluster. In the bottom row, the bold green line in each panel shows the average trajectory for that cluster. Cluster 1 represents the lowest adhering cluster, with a mean adherence of 20.5% and cluster 6 represents the highest adhering cluster, with a mean adherence of 96.7%. Clusters 1, 3, and 4 experience a noticeable drop-off in adherence with the steepest decline in adherence occurring in cluster 3. Clusters 2, 5, and 6 generally maintain a consistent average adherence throughout the study.

5.2 | Model analysis of cluster as a covariate

Our final model of the effect of weigh-in patterns on overall weight loss contains cluster assignments from the clusters depicted in Figure 3 as the primary predictor, and controls for age and baseline weight. For comparison, we also ran a model with overall percent adherence as the primary predictor (Table S6 of the Supporting Information). Percent adherence was significant ($p = 0.0022$), which is consistent with prior findings that frequent weighing is associated with weight loss (Steinberg et al., 2013, 2015). However, our percent adherence model (AIC = 368) had a higher AIC than the cluster assignment model (AIC = 344), indicating that the cluster assignment model is a better fit to the data and that the clusters offer additional information about weigh-in behavior beyond weigh-in frequency.

TABLE 2 Model performance with cluster assignment as the primary covariate. Cluster 1, the lowest adhering cluster, is the reference cluster. Significance level = 0.05. AIC = 344.8

Effect	Estimate	Standard error	t-Value	Pr > $\|t\|$
Intercept	−5.4274	7.7700	−0.699	0.4884
Baseline Weight	0.0283	0.0286	0.991	0.3271
Baseline Age	−0.0685	0.1084	−0.632	0.5305
Cluster 2	−5.2430	1.1942	−1.215	0.2307
Cluster 3	−0.7176	4.2828	−0.168	0.8677
Cluster 4	−8.0560	3.2602	−2.471	**0.0172**
Cluster 5	−7.4890	2.9341	−2.552	**0.0141**
Cluster 6	−13.7177	4.2778	−3.207	**0.0024**

Tables 2 and 3 provide the cluster assignment model summary and cluster pairwise comparisons, respectively. There was a significant overall effect of cluster assignment on weight change ($p = 0.0198$). At an unadjusted significance level of 0.05, Clusters 4, 5, and 6 lost significantly more weight than cluster 1 ($p = 0.0172$, $p = 0.0141$, and $p = 0.0024$, respectively) holding baseline age and weight constant. In addition, participants in cluster 3 lost significantly less weight than cluster 6 ($p = 0.0149$). There were no significant overall demographic differences between clusters; see Table S7 of the Supporting Information.

6 | DISCUSSION

We provide a pragmatic approach to pattern-based clustering of daily weigh-in adherence trajectories. Through simulation, we develop a framework for selecting cluster parameters for analysis of "smart"-scale data from a weight-loss study, and uncover meaningful adherence trajectory clusters that represent self-maintained weigh-in behavior and correlate with overall weight loss. Specifically, we discover interpretable clusters of time-varying weigh-in patterns, which are consistent with temporal patterns found in a recent latent class analysis of "smart"-scale data (Zheng et al., 2019), which provide additional information regarding overall weigh-in adherence, which is known to correlate with weight loss and weight maintenance (Linde et al., 2005; Steinberg et al., 2013, 2015; Wilkinson et al., 2017; Wing & Phelan, 2005). As the interventional trial is ongoing, future analyses will include more participants and may better inform our weigh-in clustering as well as increase cluster sizes. Once the study

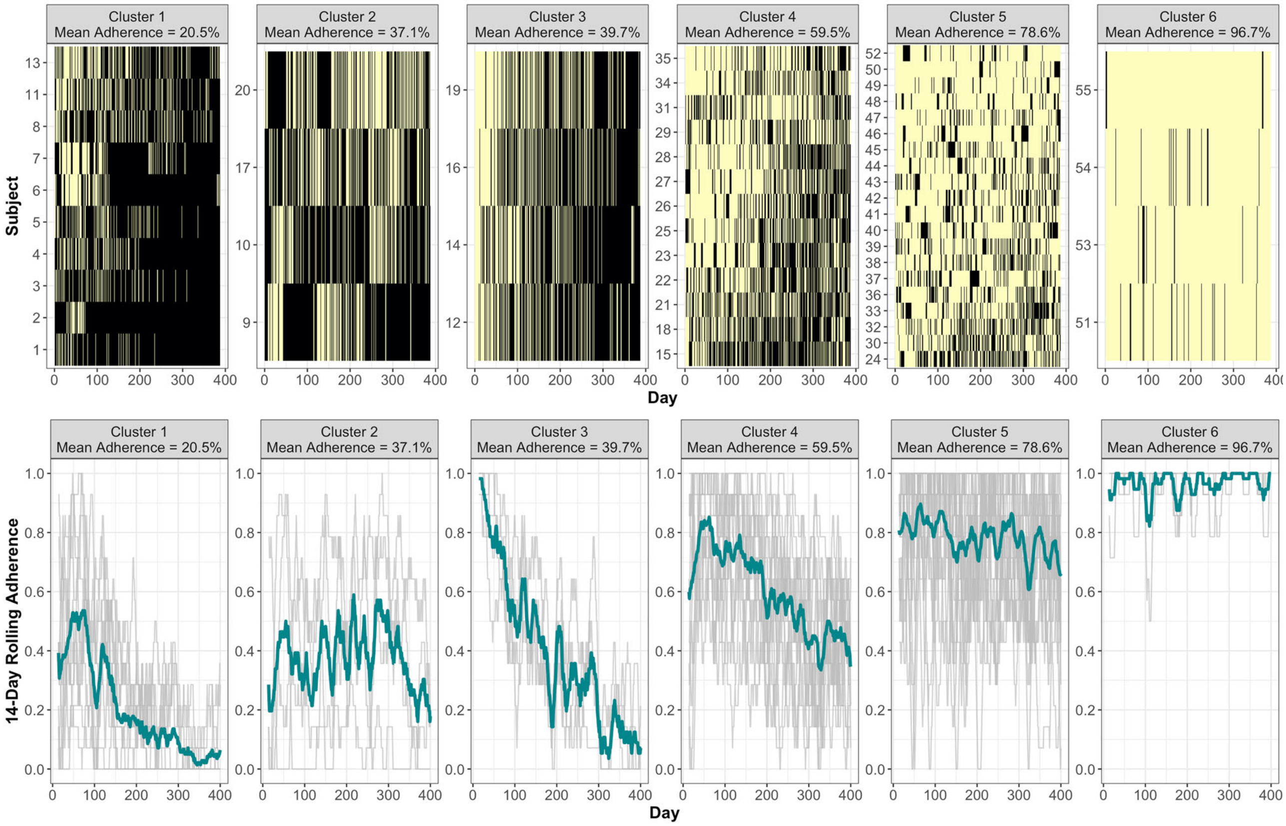


FIGURE 3 Study data cluster results using DTW distance. Top row: Individual binary trajectories by cluster assignment. Bottom row: Individual (gray) and cluster average (green) 14-day probability trajectories. Participants are numbered in order of adherence, where 1 indicates the lowest adhering participant and 55 indicates the highest adhering participant. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

TABLE 3 Model estimated pairwise differences in percent weight change from baseline between clusters. Adjusted p-values are presented using a Tukey's adjustment

Cluster	Cluster	Estimated difference	Standard error	t-Value	Unadjusted p-value	Adjusted p-value
Cluster 1	Cluster 2	5.243	4.32	1.215	0.231	0.819
Cluster 1	Cluster 3	0.718	4.28	0.168	0.868	1.000
Cluster 1	Cluster 4	8.056	3.26	2.471	**0.017**	0.146
Cluster 1	Cluster 5	7.489	2.93	2.552	**0.014**	0.123
Cluster 1	Cluster 6	13.718	4.28	3.207	**0.002**	**0.026**
Cluster 2	Cluster 3	−4.525	5.07	−0.893	0.377	0.943
Cluster 2	Cluster 4	2.813	4.13	0.680	0.500	0.982
Cluster 2	Cluster 5	2.246	3.93	0.572	0.570	0.992
Cluster 2	Cluster 6	8.475	5.11	1.659	0.104	0.551
Cluster 3	Cluster 4	7.338	4.12	1.780	0.082	0.474
Cluster 3	Cluster 5	6.771	3.95	1.714	0.093	0.515
Cluster 3	Cluster 6	13.00	5.14	2.531	**0.015**	0.129
Cluster 4	Cluster 5	−0.567	2.64	−0.215	0.831	1.000
Cluster 4	Cluster 6	5.662	4.32	1.312	0.196	0.766
Cluster 5	Cluster 6	6.229	4.01	1.552	0.128	0.620

has concluded, we will be able to incorporate additional data on participant exercise patterns, nutritional choices, and randomized interventional group assignment.

Our analysis found that subjects in clusters 4, 5, and 6 lost significantly more weight than subjects in cluster 1. These findings indicate that participants with low overall weigh-in adherence and a steep pattern of drop-off exhibit significantly less weight loss, on average, than participants with consistently high overall adherence (over 59.5%). In addition, there is a significant difference between clusters 3 and 6, which indicates that participants who steeply decline in adherence exhibit significantly less weight loss than participants who consistently maintain high levels of adherence. However, there were no significant differences in weight loss between clusters that held a constant self-weighing adherence (clusters 2, 5, and 6). This finding indicates that individuals who form a consistent weigh-in habit, regardless of frequency, are more adept at losing weight.

Our analysis clusters whole time series, rather than subsetting localized features, to understand long-term weigh-in behaviors. Shapelet methodology identifies and matches subpatterns of time series (Shah et al., 2016; Zakaria et al., 2012), and can be incorporated in future analyses to reveal additional insight into more localized weigh-in behavior. Neamtu et al. (2018) created a generalized DTW method that extends the warping methodology to general "point-to-point" distances. Adapting this generalized DTW approach to our binary time series using the Jaccard distance is an exciting potential future direction for our study. After determining an appropriate continuous transformation, an additional further analysis could implement a wavelet approach (D'Urso & Maharaj, 2012) to evaluate the effect of scale on clustering performance and understand local as well as sustained behaviors.

A drawback to the standard DTW approach is the computational cost. Because our application contained a small sample size of 55 subjects, cluster assignments using standard DTW were obtained in just under 11 s. However, given a larger dataset, more computationally efficient alternatives, including AWarp and DTW-NN (Iwana et al., 2020; Mueen et al., 2016), should be considered. In addition, to improve simulation time, we recommend running DTW in parallel.

To determine our optimal clustering, we provided a novel evaluation of how internal and external CVIs compare for binary clustering. However, to our knowledge, there remains no optimal method to cluster patterns of binary data without a continuous transformation. Our study used three windows to create a rolling probability of adherence and simulations to guide the tuning of these windows in our application. Future studies might consider model-based or latent variable approaches to identify an optimal latent probability trajectory. Because the meaning of resultant clusters can be sensitive to how data are transformed, we recommend that future studies also use a similar simulation-based tactic to guide parameter selection.

ACKNOWLEDGMENTS

We would like to thank Kevin Hall, PhD, for providing financial support to cover the cost of the BodyTrace smart-scale devices, used in the present study. We thank Liza Wayland and Kristen Bing for data curation and management, and Benjamin Steinhart for data cleaning.

DATA AVAILABILITY STATEMENT

The data that support the findings of this paper are available in the Supporting Information at the Biometrics website on Wiley Online Library (Ostendorf et al., 2022).

ORCID

Samantha Bothwell https://orcid.org/0000-0002-1616-8137

REFERENCES

Ackerman, M. & Ben-David, S. (2016) A characterization of linkage-based hierarchical clustering. *Journal of Machine Learning Research*, 17, 1–17.

Anderson, J.W., Konz, E.C., Frederich, R.C. & Wood, C.L. (2001) Long-term weight-loss maintenance: a meta-analysis of US studies. *The American Journal of Clinical Nutrition*, 74(5), 579–584.

Ansari, Z., Azeem, M.F., Ahmed, W. & Babu, A.V. (2015) Quantitative evaluation of performance and validity indices for clustering the web navigational sessions. *CoRR*. arXiv: 1507.03340.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M. & Perona, I. (2013) An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), 243–256.

Belgiu, M., Zhou, Y., Marshall, M. & Stein, A. (2020) Dynamic time warping for crops mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2020, 947–951.

Berndt, D.J. & Clifford, J. (1994) Using dynamic time warping to find patterns in time series. In: *AAAIWS'94: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, p. 12.

BodyTrace, Inc. (2021) BodyTrace scale: frequently asked questions, available from [http://www.bodytrace.com/medical/faq.html].

Bottaz-Bosson, G., Hamon, A., Pépin, J.L., Bailly, S. & Samson, A. (2021) Continuous positive airway pressure adherence trajectories in sleep apnea: clustering with summed discrete fréchet and dynamic time warping dissimilarities. *Statistics in Medicine*, 40, 5373–5396.

Brun, M., Sima, C., Hua, J., Lowey, J., Carroll, B., Suh, E. & Dougherty, E.R. (2007) Model-based evaluation of clustering validation measures. *Pattern Recognition*, 40(3), 807–824.

CDC - National Center for Health Statistics (2021) FastStats: obesity and overweight, available from [https://www.cdc.gov/nchs/fastats/obesity-overweight.htm].

D'Urso, P. & Maharaj, E.A. (2012) Wavelets-based clustering of multivariate time series. *Fuzzy Sets and Systems*, 193, 33–61.

Giusti, R. & Batista, G.E. (2013) An empirical comparison of dissimilarity measures for time series classification. In: *2013 Brazilian Conference on Intelligent Systems*, pp. 82–88.

Goslee, S.C. & Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22, 1–19.

Guijo-Rubio, D., Durán-Rosal, A.M., Gutiérrez, P.A., Troncoso, A. & Hervás-Martínez, C. (2021) Time-series clustering based on the characterization of segment typologies. *IEEE Transactions on Cybernetics*, 51(11), 5409–5422. Conference Name: IEEE Transactions on Cybernetics.

Irani, J., Pise, N. & Phatak, M. (2016) Clustering techniques and the similarity measures used in clustering: a survey. *International Journal of Computer Applications*, 134, 9–14.

Islam, M.A., Alizadeh, B.Z., Heuvel, E.R. v. d., Bruggeman, R., Cahn, W., Haan, L.d., Kahn, R.S., Meijer, C., Myin-Germeys, I., Os, J.v. & Wiersma, D. (2015) A comparison of indices for identifying the number of clusters in hierarchical clustering: a study on cognition in schizophrenia patients. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1(2), 98–113.

Iwana, B.K., Frinken, V. & Uchida, S. (2020) Dtw-nn: a novel neural network for time series recognition using dynamic alignment between inputs and weights. *Knowledge-Based Systems*, 188, 104971.

Juang, B.H. (1984) On the hidden Markov model and dynamic time warping for speech recognition—a unified view. *AT&T Bell Laboratories Technical Journal*, 63(7), 1213–1243.

Keogh, E.J. & Pazzani, M.J. (2000) Scaling up dynamic time warping for datamining applications. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '00. New York, NY: Association for Computing Machinery, pp. 285–289.

Krukowski, R.A. & Ross, K.M. (2020) Measuring weight with electronic scales in clinical and research settings during the coronavirus disease 2019 pandemic. *Obesity (Silver Spring, Md.)*, 28(7), 1182–1183.

Li, H., Liu, J., Yang, Z., Liu, R.W., Wu, K. & Wan, Y. (2020) Adaptively constrained dynamic time warping for time series classification and clustering. *Information Sciences*, 534, 97–116.

Liang, M., Wen Liu, R., Li, S., Xiao, Z., Liu, X. & Lu, F. (2021) An unsupervised learning method with convolutional auto-encoder for vessel trajectory similarity computation. *Ocean Engineering*, 225, 1-16.

Linde, J.A., Jeffery, R.W., French, S.A., Pronk, N.P. & Boyle, R.G. (2005) Self-weighing in weight gain prevention and weight loss trials. *Annals of Behavioral Medicine*, 30(3), 210–216.

Liu, Y., Li, Z., Xiong, H., Gao, X. & Wu, J. (2010) Understanding of internal clustering validation measures. In: *2010 IEEE International Conference on Data Mining*, pp. 911–916. ISSN: 2374-8486.

Luczak, M. (2016) Hierarchical clustering of time series data with parametric derivative dynamic time warping. *Expert Systems with Applications*, 62, 116–130.

Maimon, O. & Rokach, L. (2006) *Clustering methods*. Data Mining and Knowledge Discovery Handbook, Springer, Berlin, pp. 321–352.

Miyamoto, S., Abe, R., Endo, Y. & Takeshita, J.i. (2015) Ward method of hierarchical clustering for non-Euclidean similarity measures. In: *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 60–63.

Mueen, A., Chavoshi, N., Abu-El-Rub, N., Hamooni, H. & Minnich, A. (2016) AWarp: fast warping distance for sparse time series. In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*. Barcelona, Spain: IEEE, pp. 350–359.

Neamtu, R., Ahsan, R., Rundensteiner, E.A., Sarkozy, G., Keogh, E., Dau, H.A., Nguyen, C. & Lovering, C. (2018) Generalized dynamic time warping: unleashing the warping power hidden in point-wise distances. In: *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pp. 521–532.

Niennattrakul, V. & Ratanamahatana, C.A. (2007) On clustering multimedia time series data using k-means and dynamic time warping. In: *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, pp. 733–738.

Ostendorf, D., Caldwell, A., Zaman, A., Pan, Z., Bing, K., Wayland, L., Creasy, S., Bessesen, D., MacLean, P., Melanson, E. & Catenacci, V. (2022) Comparison of weight loss induced by daily caloric restriction versus intermittent fasting (drift) in individuals with obesity: study protocol for a 52-week randomized clinical trial. *Trials*, 23, 718.

R Core Team. (2020) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ross, K.M., Qiu, P., You, L. & Wing, R.R. (2018) Characterizing the pattern of weight loss and regain in adults enrolled in a 12-week Internet-based weight management program. *Obesity (Silver Spring, Md.)*, 26(2), 318–323.

Sakoe, H. & Chiba, S. (1970) A similarity evaluation of speech patterns by dynamic programming. In: *Dig. 1970 National Meeting*. Japan: Institute of Electronics, Information and Communication Engineers, p. 136.

Salvador, S. & Chan, P. (2004) Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: *16th IEEE International Conference on Tools with Artificial Intelligence*, pp. 576–584. ISSN: 1082-3409.

Sardá-Espinosa, A. (2019) Time-Series Clustering in R Using the dtwclust Package.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O., Tiwari, A., Er, M., Ding, W. & Lin, C.T. (2017) A review of clustering techniques and developments. *Neurocomputing*, 267, 664–681.

Shah, M., Grabocka, J., Schilling, N., Wistuba, M. & Schmidt-Thieme, L. (2016) Learning dtw-shapelets for time-series classification. In: *Proceedings of the 3rd IKDD Conference on Data Science, 2016*, CODS '16. New York, NY: Association for Computing Machinery, pp. 1–8.

Steinberg, D.M., Bennett, G.G., Askew, S. & Tate, D.F. (2015) Weighing everyday matters: daily weighing improves weight loss and adoption of weight control behaviors. *Journal of the Academy of Nutrition and Dietetics*, 115(4), 511–518.

Steinberg, D.M., Tate, D.F., Bennett, G.G., Ennett, S., Samuel-Hodge, C. & Ward, D.S. (2013) The efficacy of a daily self-weighing weight loss intervention using smart scales and email. *Obesity (Silver Spring, Md.)*, 21(9), 1789–1797.

Thomas, J.G. & Bond, D.S. (2014) Review of innovations in digital health technology to promote weight control. *Current Diabetes Reports*, 14(5), 485.

Thomas, J.G., Bond, D.S., Phelan, S., Hill, J.O. & Wing, R.R. (2014) Weight-loss maintenance for 10 years in the national weight control registry. *American Journal of Preventive Medicine*, 46(1), 17–23.

Vijaya, S. & Batra, N. (2019) Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering. In: *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*.

Wilkinson, L., Pacanowski, C.R. & Levitsky, D. (2017) Three-year follow-up of participants from a self-weighing randomized controlled trial. *Journal of Obesity*, 2017, 1–7.

Wing, R.R. & Phelan, S. (2005) Long-term weight loss maintenance. *The American Journal of Clinical Nutrition*, 82, 222–225.

Zakaria, J., Mueen, A. & Keogh, E.J. (2012) Clustering time series using unsupervised-shapelets. In: *2012 IEEE 12th International Conference on Data Mining*.

Zheng, Y., Sereika, S.M., Burke, L.E., Olgin, J.E., Marcus, G.M., Aschbacher, K., Tison, G.H. & Pletcher, M.J. (2019) Temporal patterns of self-weighing behavior and weight changes assessed by consumer purchased scales in the health eheart study. *Journal of Behavioral Medicine*, 42(5), 873–882.

SUPPORTING INFORMATION

Web Appendices and Tables referenced in Sections 3.4.3, 4, and 5 are available with this paper at the Biometrics website on Wiley Online Library. Code to replicate our simulation and data analysis is publicly available on GitHub at https://github.com/sjbothwell/BinaryShapeClust and is available with this paper at the Biometrics website on Wiley Online Library.

Table Web.App.1. Computation time (in seconds) for each distance (median [IQR])

Table Web.App.2 Validation indices to evaluate distance methods across probability windows and linkages.

Table Web.App.3. Validation indices to optimize rolling window under DTW.

Table Web.App.4. Evaluation of internal CVI performance to uncover the true number of clusters with DTW.

Table Web.App.5. Model estimated pairwise differences in percent weight change from baseline between clusters.

Table Web.App.6. Model performance with total percent adherence as the primary covariate.

Table Web.App.7. Cohort demographics stratified by assigned cluster.

Data S1

How to cite this article: Bothwell, S., Kaizer, A., Peterson, R., Ostendorf, D., Catenacci, V. & Wrobel, J. (2022) Pattern-based clustering of daily weigh-in trajectories using dynamic time warping. *Biometrics*, 1–13. https://doi.org/10.1111/biom.13773