Published in final edited form as:

Appetite. 2020 January 01; 144: 104451. doi:10.1016/j.appet.2019.104451.

Distance metrics optimized for clustering temporal dietary patterning among U.S. adults

Heather A. Eicher-Miller, PhD^a, Saul Gelfand, PhD^b, Youngha Hwang, PhD^b, Edward Delp, PhD^b, Anindya Bhadra, PhD^c, Jiaqi Guo^b

^aDepartment of Nutrition Science, Purdue University, 700 West State Street, West Lafayette, Indiana, 47907, USA

^bSchool of Electrical and Computer Engineering, 465 Northwestern Avenue, Purdue University, West Lafayette, Indiana, 47907, USA

^cDepartment of Statistics, Purdue University, 250 North University Street, West Lafayette, Indiana, 47907, USA

Abstract

Objective: Few attempts to determine dietary patterns have incorporated concepts of time, specifically time and proportion of energy intake consumed throughout a day. A type of modified dynamic time warping (MDTW) was previously developed using an appropriate distance metric for patterning these aspects to determine *temporal* dietary patterns (TDP). This study further explores dynamic time warping (DTW) distance metrics including unconstrained DTW (UDTW), constrained DTW (CDTW), and MDTW with modern spectral clustering methods to optimize TDP related to dietary quality. MDTW was expected to create TDP with the strongest relationships to dietary quality and distinct visualization among U.S. adults 20-65y of the National Health and Nutrition Examination Survey 1999-2004.

Methods: Proportional energy intake by time of day metrics were optimized to create TDP from complete day-one 24-hour dietary recalls using MDTW, UDTW with only a standard local constraint, and CDTW with standard local and global banding constraints, then clustered using spectral clustering. The association between each TDP distance metric clustering and mean dietary quality, as indicated by the 2005 Healthy Eating Index (HEI-2005), were determined using multiple linear regression controlled for potential confounders. Strength of association for each model was compared using adjusted R-squared. The results were also visualized to make qualitative comparisons.

Results: Four clusters representing distinct TDP for each distance metric by spectral clustering were generated among participants. MDTW exhibited TDP clusters with strongest associations to HEI compared with the TDP clusters generated from unconstrained and constrained DTW, and visualization of the TDP clusters from MDTW supported the association.

Corresponding Author Heather A. Eicher-Miller, (address listed above), 765-494-6815(phone); heicherm@purdue.edu.

Declarations of Interest

None

Implication: MDTW paired with spectral clustering is a useful tool for dimension reduction and uncovering temporal patterns with dietary data.

Keywords

temporal dietary patterns; time of eating; dietary patterns; dietary quality; patterning methods; energy intake

Introduction

Integration of the schedule and proportion of energy intake throughout a day, the temporal distribution of energy, is a novel perspective in dietary patterns research. Characterizing and evaluating the relationship of the temporal distribution of energy to health outcomes such as obesity, diabetes, and metabolic syndrome, holds promise to identify stronger associations compared to the relationships of the isolated components time of eating, energy, and number of eating occasions with these diet-related health outcomes. Investigation of temporal energy distribution is based on the compiled evidence supporting a more generalized concept, that the timing of energy intake as mainly early or late is linked with metabolism and long-term health (Eicher-Miller et al., 2015; Beccuti et al., 2017; Morgan et al., 2012). Current interest in the integration of multiple types of data to better understand lifestyles linked to health and disease is high due to the burgeoning data available from electronic platforms and the ability to self-monitor diet. Further investigation to the relationship of temporal energy distribution with health may provide new opportunities for interventions to reduce chronic disease.

Temporal distribution of energy in a sub-population or group may be discovered through temporal dietary patterning, or incorporating the distribution of time as a primary component of a dietary pattern. Temporal dietary patterns (TDP) integrating the timing, proportion of energy, and number of eating events throughout a 24-hour day were recently linked to dietary quality among US adults aged 20 to 65 years (Eicher-Miller et al., 2015; Khanna et al., 2011). Four clusters of adults representing distinct TDP were characterized. The cluster exhibiting proportionally equivalent energy intake during three evenly-spaced eating occasions exhibited higher daily dietary quality, quantified using the Healthy Eating Index-2005 (HEI-2005), compared with the three alternate TDP clusters. Thus, the predictive validity of TDP with dietary quality was demonstrated (Guenther et al., 2008).

This initial attempt at creating TDP utilized a modification of dynamic time warping or Modified Dynamic Time Warping (MDTW), where the distance metric is optimized based on the proportion of energy, time, and occurrence of eating occasions and paired with a heuristic variant of kernel k-means clustering. Further development of these methods may improve distinction of the TDP, increase dietary quality differences among the sub-population clusters, and improve predictive ability. For example, comparison of conventional DTW-type distance metrics, such as unconstrained and constrained DTW, with MDTW, and use of widely applied and well accepted clustering methods such as spectral clustering (Liao, 2005) are potential methodological improvements that may fulfill these goals and provide added practical application of the TDP.

Dynamic time warping computes a distance between dietary time series by elastically aligning them in time. This is frequently necessary for general time series comparisons, and in the present case when similar eating events do not occur at the exact same times. Unconstrained dynamic time warping (UDTW) allows the alignment of times which differ in value up to the full length of the data record (24 hours), whereas constrained dynamic time warping (CDTW) imposes a hard threshold on the maximum amount such times can differ. Modified dynamic time warping (MDTW) computes a distance between dietary time series by discretely aligning non-zero eating event times. Spectral clustering is a state-of-the-art method suitable for combining the resulting DTW-type metrics and k-means. K-means cannot be used directly due to the absence of a centroid after the time alignment. One distinct advantage of spectral clustering salient to improving TDP is the decomposition into principal components which explicitly allows for non-spherical clusters. Thus, the purpose of the current study is to: 1) explore additional DTW-type distance metrics including UDTW, CDTW, and MDTW, along with modern spectral clustering methods to determine and compare TDP among U.S. adults 20-65y of the National Health and Nutrition Examination Survey (NHANES) 1999-2004; 2) evaluate the new TDP clusters by determining their association with dietary quality using the HEI-2005; 3) visualize the patterns.

The hypothesis is that MDTW will classify the temporal dietary pattern with the strongest relationships to dietary quality compared with UDTW and CDTW. Visualization of patterns also informs suitable patterning selection because of the importance of directly observed interpretation of the patterns. The rationale to support this hypothesis is that MDTW solves a discrete matching problem, which aligns discrete events by adjusting their timing. It does not require every event be matched, and prohibits multiple events from being matched to a single event. Conventional DTW such as UDTW and CDTW solves a continuous matching problem, which aligns waveforms by stretching and compressing them. It requires every time to be matched, and allows for multiple times to be matched to a single time. MDTW is expected to be more suitable for dietary analysis because the eating events are inherently discrete events, occurring at certain isolated and specified times, as opposed to occurring at every time with smoothly varying intensity where DTW may perform better.

Methods

Study Design and Participants.

This secondary analysis comprised adult participants of the NHANES 1999-2004 using a cross-sectional study design, building on and maintaining consistency with previous studies (Eicher-Miller et al., 2015; Khanna et al., 2011). The NHANES is a program of the National Center for Health Statistics through the Centers for Disease Control and Prevention. Participants represented the civilian non-institutionalized US population and were recruited using a complex, multistage probability sampling method (Zipf et al., 2013). Subpopulations including non-Hispanic Black Americans, Mexican Americans, low-income White Americans, and individuals greater than 60 years old were oversampled in the 1999-2004 data in order to achieve precise and reliable estimation. Inclusion criteria for this analysis were adults ages 20 to 64 years, reporting as non-pregnant with exclusion for participants

who indicated that they were not sure of their pregnancy status, and complete 1-day 24-hour dietary recall ($n=8,421$). Children, teens and adults older than 65 years were not included in this analysis because these excluded age groups were expected to have distinct patterns from the non-pregnant adult sample included. Participant characteristics, including age, sex, race/ethnicity, and household poverty-income-ratio were surveyed in the participant's home. Body mass index (BMI) was determined using measures of height and weight that were collected during the health examination which was completed at the NHANES Mobile Examination Center (NCHS, 2018^a). Dietary data was also collected in the Mobile Examination Center using a 24-hour dietary recall, the United States Department of Agriculture (USDA) Automated Multiple-Pass Method (CDC, 2018). All NHANES protocols and content were reviewed by the National Center for Health Statistics Research Ethics Review Board who granted approval (NCHS, 2018^b). Consent was provided by all participants.

Dietary Assessment.

Detailed dietary information, including types and amounts of foods and beverages consumed during the 24-hr prior the interview, as well as time when each eating occasion occurred, was collected in a 5-step series of probes or passes comprising the Automated Multiple-Pass Method (USDA, 2018). Participants recalled as many foods and beverages as possible in the approximately 30 to 45 minutes it took to complete the assessment and were provided with a standard set of measuring guides to help participants estimate portion sizes. The duration of eating occasions was not available. Dietary components and energy content were calculated by linking the dietary data to the USDA Food and Nutrient Database for Dietary Studies for survey years 2001-2004 and the USDA 1994-1998 Survey Nutrient Database for years 1999-2000 (ARS, 2018).

The mean HEI-2005 was used as an outcome measure of dietary quality over one 24-hour day, maintaining consistency with the temporal aspect of the dietary patterns. HEI-2005 may be used as an indication of how well an individual or group adheres to the 2005 Dietary Guidelines for Americans such that the highest score on a scale of 0 to 100 indicates full conformance to the Dietary Guidelines for Americans (Guenther et al., 2008^a). The HEI-2005 is a density based metric comprised of numeric component scores for recommended servings of total fruit; whole fruit; total vegetables; dark green and orange vegetables and legumes; total grains; whole grains; milk; meat and beans; oils; saturated fat; sodium; and energy from solid fats, alcoholic beverages, and added sugars, that sum the total score. See Guenther et al. for detailed calculation of HEI-2005 (Guenther et al., 2008^b). Assumptions of normality were not rejected upon examination.

Temporal Dietary Patterns.

Individuals vary in the times that they choose to consume energy throughout a day and the proportion of energy they choose to consume at these times, such as consuming 10% of daily energy at 8:00 a.m., 10% at noon, and 80% at 6:00 p.m. as opposed to consuming 35% of daily energy at 7:00 a.m., 25% at 1:00 p.m. and 30% at 7:00 p.m. and 10% at 8:00 p.m. Unsupervised clustering based on DTW-type distance metrics was chosen as an appropriate tool to identify common patterns within the sample due to its ability to accommodate these

variations (Liao, 2005). Three versions of temporal dietary patterns were derived based on utilization of MDTW and two conventional DTW-type distance metrics followed by spectral clustering (Berndt and Clifford, 1994; Von Luxburg, 2007). All of these metrics were computed based on quantizing the consumed energy into hourly periods, normalized by the total consumed energy in the 24-hour day.

Modified dynamic time warping (MDTW) was created specifically to compare the proportion of energy intake and timing of intake in derivation of the metric. A detailed description of the steps to create the MDTW derived patterns is available from (Khanna et al., 2011; Khanna et al., 2017) and briefly described here. For MDTW, a series of the non-zero hourly eating periods was extracted from the 24-hour time series of proportional energy intakes, with each hour marked by its proportional energy intake and hourly time. These non-zero hourly eating periods are referred to as eating events in the sequel, and typically there were 2-5 such events in a 24-period. These series of eating events were then employed in the MDTW metric by separately weighting and then combining the difference between proportional energy intake and hourly time of eating events. A local constraint was also applied to allow an eating event of an individual to be matched to only one or no eating event of another individual.

Two conventional DTW metrics were also used for comparison with MDTW. These metrics were computed based directly on the 24-hour time series of proportional energy intake. The UDTW employed only standard local constraints, while the CDTW employed in addition a global banding constraint with band length equal to 3 hours (Sakoe and Chiba, 1978). CDTW is often used as a suboptimal method to reduce the complexity of unconstrained DTW, but here it was used to add a reasonable constraint on the different times that similar eating behaviors occur (times must lie within 3 hours) (Rakthanmanon et al., 2013).

The goal of clustering is to divide the data points (here time series) into several groups such that points in the same group are similar and points in different groups are dissimilar. Although there are many kernel methods which are suitable for clustering data points with DTW-type preprocessing (Von Luxburg, 2007), here the focus is on spectral clustering. Since only distances (or equivalently similarities) between data points are available, a similarity graph is a good candidate to represent the data. The data points are represented by vertices and their similarities by edge weights. Two vertices are connected if the similarity between the corresponding data points is positive. For clustering, a partition of the graph is found such that the edges between different groups have relatively low weights, and the edges within a group have relatively high weights. The edges are weighted by Gaussian kernel of the distances between the connected data points. The width or the standard deviation of the Gaussian kernel is computed as the average distance of its k-nearest neighbors.

The main tools for spectral clustering are graph Laplacian matrices. Here the normalized Laplacian by Shi and Malik (Shi and Malik, 2000) which scales the Laplacian by the inverse of the degree matrix is used. The eigenvectors from the normalized Laplacian with 20 smallest eigenvalues embed the time series. Then the standard k-means clustering is executed from the embedded space. One-hundred repetitions of k-means clustering, with

random initial centroids were performed, and the final clusters were selected based on the largest spread between their maximum and the minimum HEI-2005 averages. Four clusters were created for each DTW-type dietary patterning in order to maintain consistency with previous studies and to enhance focus on comparison of the DTW-type patterning (Eicher-Miller et al., 2015; Khanna et al., 2011).

Inferential Data Analysis.

Chi-square analysis was performed to compare the characteristics of the sub-population included in each cluster. These characteristics included survey year (1999-2000, 2001-2002, 2003-2004), age (20 to 34 years, 35 to 49 years, 50 to 65 years), sex (male or female), race/ethnicity (Mexican American, other Hispanic, non-Hispanic white, non-Hispanic black, or other), poverty-income-ratio (PIR), and weight status indicator BMI (weight (kg)/[height(m)]²). BMI was categorized as: underweight (<18.5), normal weight (18.5-24.9), overweight (25.0-29.9), and obese (>30.0) (WHOB, 2000). PIR, calculated as household income divided by the federal poverty guideline for household income, was divided to six categories: 0-0.99, 1-1.99, 2-2.99, 3-3.99, 4-4.99, and 5 or more; ratios below 1 indicate a PIR below the officially defined poverty level. Rao Scott F-adjusted Chi Square statistic was used to determine significant differences among clusters by characteristics when ($P < 0.05$).

Differences in the mean total HEI-2005 score for the 4 clusters of each DTW-type patterning were examined using analysis of variance. Cluster (independent variable) associations with HEI-2005 score (dependent variable) were also determined using a multiple linear regression model, with adjustment for covariates including sex, race/ethnicity, age group, BMI, survey year, and household PIR. Bonferroni-type correction was used to adjust the significance level for multiple comparisons between the clusters ($P < 0.05/6$). Model fit was evaluated based on the strength of association for each model determined using the adjusted R-squared. Predictive validity was indicated by significant differences between the cluster HEI-2005 means and aided discussion of method improvement. Proper sampling weights were constructed to account for non-response, non-coverage, planned oversampling, as well as unequal probability of selection due to clustering and stratification in NHANES design, so that the survey population could represent the non-institutionalized civilian U.S. population and provide unbiased estimates for survey year 1999-2004. Clustering analyses algorithms were carried out in MATLAB 2014a (The MathWorks, Inc, Natick, MA). SAS Survey procedures and related inferential analysis were performed in SAS 9.4 (SAS Institute Inc, Cary, NC).

Visualization.

Practical significance and application of the temporal dietary pattern clusters based on their visualization was also considered (Figures 1-4) in determining a suitable distance-metric based clustering. Time is shown on the x-axis, fractional amount of energy intake on the y-axis, and proportion of individuals in each cluster is indicated through shading in the heat maps for each of the 4 clusters of MDTW (Figure 1), UDTW (Figure 2), and CDTW (Figure 3) in order to capture more information compared with traditional graphs. Shading of the squares in each heat map directly corresponds to the proportion of individuals ranging from 0=0% to 0.1=10% of each cluster with fractional daily energy intake ranging from 0=0% to

1=100% of total daily energy at a particular time so that when shading is darker a greater proportion of that particular cluster reported energy intake at that particular time. Time ranged from 0=12:00 a.m. to time 24=12:00 a.m. the next day. Figure 4 maintains a similar structure but adds color in order to differentiate the 4 clusters of MDTW, UDTW, and CDTW which are each shown in a graph. The visual quality of the patterning prioritized showed a clear and constrained range of proportional energy intake and high conformity of individuals in a cluster to the pattern of energy intake throughout the day.

Results

This paper's hypothesis focus on the DTW-type metrics to differentiate the strongest relationships to dietary quality, guided attention to clusters of highest and lowest dietary quality, compared with the other clusters. Therefore, clusters of highest and lowest dietary quality are heavily featured in the results and discussion. Varying characteristics were prevalent in the 4 clusters of each DTW-type dietary patterning (Table 1).

The smallest cluster identified by all DTW-type patterning was prevalently male, non-Hispanic black, 20-34 years, very low-income (0-0.99 PIR) and also exhibited the lowest dietary quality as per the mean HEI-2005 score (Table 2) when compared with the other 3 clusters.

Similarly, all DTW-type patterning produced a cluster with the highest dietary quality that was prevalently male, non-Hispanic white, incomes 5 of the poverty-income-ratio and 35-65 years. The clusters with the largest percentage of participants for each DTW-type pattern had neither the highest nor lowest HEI scores among each of their DTW-type patterns. Among all DTW-type patterning, mean HEI cluster scores ranged from 35.7 to 53.7 with the widest range present among MDTW clusters (35.7 to 51.9).

Clusters determined using MDTW exhibited the strongest association with dietary quality yielding statistically significant differences of 14.5, 11.1, and 10.7 HEI points between various cluster means and the best model fit (0.086, Adjusted R-Square, Table 2). Clustered derived using UDTW and CDTW also exhibited significant differences; clusters of UDTW derived patterns showed statistically significant differences of 14.1, 10.0, and 9.1 HEI points which were also strong and model fit was relatively similar (0.085, Adjusted R-Square) while clusters of CDTW derived patterns exhibited statistically significant differences of 8.8, 6.4, and 4.4 HEI points and slightly lower model fit (0.083, Adjusted R-Square).

Visualization results (Figure 1) showed MDTW with the most distinct or constrained range of proportional energy intake among both the clusters with the highest and lowest dietary quality. MDTW cluster 4 with the highest dietary quality showed a clear and confined pattern of proportional energy intake between 0% and 40% for eating occasions throughout the day from 6:00 a.m. to 11:00 p.m. with a greater percentage of the cluster engaging in eating occasions from 8:00 a.m. to 11:00 a.m., noon to 4:00 p.m. and 6:00 p.m. to 9:00 p.m. compared with proportional energy and time patterns of respective high quality dietary clusters of UDTW (Figure 2) and CDTW patterning (Figure 3). A similar constrained pattern was visualized among the clusters exhibiting the lowest dietary quality, MDTW

Cluster 2 consumed 90-100% of energy in one eating occasion and less than 10% in others while these proportional energy ranges were wider in respective low dietary quality UDTW and CDTW patterning. Finally, both UDTW and MDTW exhibited clearer, distinct patterns of both time and amount of eating with regard to the eating occasion of highest proportional energy (Figure 4) compared with CDTW; MDTW exhibited the most distinct bounds for time and amount of eating for clusters of highest and lowest dietary quality.

Discussion

MDTW clustering exhibited the strongest relationships with dietary quality and the most distinct visualization compared with other TDP clustering methods UDTW and CDTW. However, inferential hypothesis testing and visualization-based differences among MDTW, UDTW, and CDTW and especially between MDTW and UDTW were slight. For example, the differences in strength of the relationship of the clusters to HEI exhibited by the 14.5 HEI point difference in MDTW clusters 2 and 4 and 14.1 point differences in UDTW clusters 3 and 4 may not translate to differences in health outcomes for which HEI unit changes of at least 2 can be meaningful (Berkowitz et al., 2014). Visualization of MDTW was also marginally better compared with UDTW. Yet taken together, even marginal superiority of MDTW distance metric compared with UDTW and CDTW reveals an improved method to determine temporal dietary patterns in the population.

Visualization of the MDTW temporal dietary pattern cluster 4 with the highest dietary quality exhibited clear confinement of proportional energy intake between 0 and 40% for eating occasions from 6:00 a.m. to 11:00 p.m. that translates in application to a dietary pattern where energy is consumed moderately throughout all eating occasions in the day. This pattern is supported in the Dietary Guidelines for Americans and gleaned from the compilation of evidence linking diet with the prevention of chronic disease. A pattern of moderation was observed more clearly in MDTW compared with CDTW and UDTW clusters exhibiting highest dietary quality. For example, proportional energy intake ranged from 0 to 60% in UDTW cluster 4 exhibiting the highest dietary quality among UDTW clusters with an evening eating occasion from 5:00 p.m. to 8:00 p.m. that was about 20% greater than other energy intake throughout the day. CDTW cluster 3 exhibiting highest dietary quality among CDTW clusters showed energy intake ranging from 0 to 80% throughout the day with a morning eating event ranging from 0 to 30% of energy, a mid-day eating event ranging from 0 to 50% and even wider-spread eating event in the evening ranging from 0 to 80%.

Observed ranges of proportional energy distribution were very different for the clusters of lowest compared with highest dietary quality among MDTW, CDTW, and UDTW but similarly showed MDTW as the patterning most clearly confining proportion of energy intake. MDTW cluster 2 with lowest dietary quality among MDTW clusters exhibited a dietary pattern where one main eating occasion constituted 90% to 100% of energy intake in the day and most of the participants in the cluster reported this meal between 3:00 p.m. and 11:00 p.m. UDTW cluster 3 and CDTW cluster 2 with lowest dietary quality among their respective clusters also exhibited patterns of most energy consumed at one eating occasion but the proportion of energy consumed at this meal was more variant between 80% to 100%.

Finally, visualization of the eating event of the largest or greatest proportional energy consumed throughout the day, again demonstrates the superiority of MDTW compared with UDTW and CDTW. MDTW cluster 4 exhibited a high percentage of participants with noon or 6:00 p.m. to 8:00 p.m. as the largest eating event at only 30-50% of energy, cluster 2 with a high percentage of participants with 3:00 p.m. to 6:00 p.m. as the eating event comprising 90% to 100% of energy, cluster 3 with noon to 5:00 p.m. as the largest meal comprising 35% to 50% of energy and cluster 1 exhibiting the largest eating event comprising 35% to 55% of energy intake from 6:00 to 9:00 p.m. In contrast, CDTW and UDTW clusters exhibited greater overlap in the times of largest eating occasions and range of energy intake at those times.

All DTW-type pattern clusters featured here exhibited clusters of highest dietary quality, where energy was more moderately consumed compared with the other clusters, and lowest dietary quality, where the majority of energy was consumed in one later eating occasion, consistent with the previous findings in Eicher-Miller et al., 2015, indicating the usefulness of DTW overall to the problem of temporal dietary clustering (Eicher-Miller et al., 2015). For example, the cluster that was prevalently young, non-Hispanic black, and male with incomes below the poverty line has previously been shown to have low dietary quality (Rehm et al., 2016) compared with other race/ethnic, age, and gender groups and higher dietary quality has been previously observed among higher income groups and older adults (Wang, et al., 2014). The widest HEI-2005 differences, detected among the MDTW cluster 2 and 4, of 14.5 as well as other differences of 11.1, and 10.7 are expected to be meaningful with regard to health (Kirkpatrick et al., 2018) although mean HEI of all of the clusters are still quite low and considered to be grades of “F” when translated to a traditional A-F grade model (Krebs-Smith et al., 2018). Several studies have used HEI to divide the population into segments (quartiles, quintiles) and found links of HEI to long-term risk (Schwingshackl et al., 2018); such population segments have greater HEI differences than the ones determined between the clusters of these DTW-type patterns, yet these studies have intentionally created the largest possible HEI differences. Integrating this evidence, it is apparent that the highest and lowest HEI segments in the population (quintiles, quartiles) may not exhibit uniform temporal dietary patterns as the temporal dietary patterns determined here had a more restricted range of mean HEI scores. Yet the prevalent characteristics of the temporal dietary patterns of lowest and highest dietary quality also characterize population sub-groups with higher risk for diet-related chronic disease and support the potential for temporal dietary patterns to link to long-term health.

Investigation and comparison of the DTW-type pattern clustering variable against important covariates further indicated the strength of temporal dietary patterns (data not shown) to explain variation in dietary quality. Poverty-income-ratio and age both have an established direct relationship with dietary quality in the U.S. population (Hiza et al, 2013). Separate simple modeling of MDTW, UDTW, and CDTW as the only independent variable linked with the dietary quality outcome resulted in the highest adjusted R-square values when compared with similar models including only poverty-income-ratio with dietary quality or only age with dietary quality. Such findings indicate the importance of temporal dietary patterns and need for future work to continue to refine measures to classify temporal dietary patterns.

The distance metrics compared in this study did not result in dramatic differences to mean dietary quality or in visualization. In addition to refinement of the distance metric, modification of the type of clustering is an additional characteristic of developing temporal dietary patterns that may potentially influence greater differences in HEI, and usefulness for prediction with regard to dietary quality and additional health outcomes. Modern spectral clustering was applied in all DTW-type patterning here but future studies should explore divergence measure based clustering (Banerjee et al., 2005). Further, future exploration of other temporal lifestyle patterns such as physical activity should be considered both independently and jointly with dietary data, for example, the timing and sequence of how and when individuals are physically active or when energy is consumed may have a relationship with the quality of dietary intake and other health outcomes such as BMI, waist circumference, fasting glucose, etc. Discovering more comprehensive temporal lifestyle patterns, holds promise to strengthen prediction of health outcomes and the potential for behavioral interventions.

The dietary quality as the outcome chosen in this study to validate the importance and meaningfulness of temporal dietary patterns is perhaps a conservative choice for an outcome as both quality and temporal patterns are limited to the 24-hour time frame. Relevance of the temporal dietary patterns to long-term outcomes such as BMI, diabetes, metabolic syndrome, and presence of other chronic disease or risk factors is likely based on these short-term results. For example, a single-day dietary quality may be less important to long-term health, however, repeated over time, the dietary quality differences observed here among MDTW clusters are likely to yield diverse chronic disease outcomes and should be investigated in future studies. Following significant results from investigation of meaningful differences in distal health outcomes, evidence may justify evaluation of temporal dietary patterns in experimental trials that could inform future population guidelines.

Strengths of this research are the inclusive nature of creating patterns integrating proportional energy, time, and frequency of eating with a thorough investigation and comparison of a modified approach to dynamic time warping developed specifically for dietary data with traditional constrained and unconstrained versions of dynamic time warping. Attention to refining methods is important because behavioral signals, such as TDP, with important links to health may be dulled by the errors inherent to patterning that compound when linked with the errors inherent to additional daily patterns. An important limitation to this study was the inclusion of only a 24-hour time period. Pattern robustness and participant membership to clusters over multiple days is unknown, nor how the relationship with dietary quality may also differ. All days of the week were represented in the dietary assessment data, further studies should explore how weekend may differ from weekday patterns and how other environmental cues may influence temporal patterns. Controlling for the presence of hunger is an additional need in future studies as it may be related to both TDP and dietary quality. Inherent measurement error including the under-reporting of energy intake is a known issue of self-reported energy intake among adults and self-reporting of the time of dietary intake may also entail error (Macdiarmid and Blundell, 1998). Finally, the relationship of dietary quality to TDP in this study representing the US population in 1999-2004 maintains consistency with development of TDP (Eicher-Miller et al., 2015; Khanna et al., 2011) but is expected to have changed as dietary quality in the US

population has improved (Wilson et al, 2016). Future studies should explore the relationship of TDP with dietary quality using more recent versions of NHANES and the HEI.

Conclusions

MDTW with spectral clustering exhibits superior links with dietary quality and more constrained temporal dietary patterns compared with UDTW and CDTW with spectral clustering. Investigation to refine the methodology to integrate time with dietary patterns is important to determine relationships of diet with dietary quality and potentially with long-term health.

Acknowledgements

The authors would like to acknowledge Ms. Yanling Zhao for analyzing preliminary data and drafting a previous version of this paper.

Funding

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number R21CA224764 and Purdue University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Prior to issuing a press release concerning the outcome of this research, please notify the NIH awarding IC in advance to allow for coordination

Abbreviations

TDP	Temporal Dietary Patterns
MDTW	Modified Dynamic Time Warping
UDTW	Unconstrained Dynamic Time Warping
CDTW	Constrained Dynamic Time Warping
HEI	Healthy Eating Index
NHANES	National Health and Nutrition Examination Survey
USDA	United States Department of Agriculture
BMI	Body Mass Index

References

- Agricultural Research Service (ARS), United States Department of Agriculture. Food and Nutrient Database for Dietary Studies Version 1.0 and 2.0. <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fhdds-download-databases/> Beltsville Human Nutrition Research Center, Food Surveys Research Group Beltsville, MD, Updated October 2018, Accessed December 2018.
- Banerjee A, Merugu S, Dhillon IS, Gosh J (2005). Clustering with Bregman Divergences. *Journal of Machine Learning Research*, 6, 1705–1749.
- Beccuti G, Monagheddu C, Evangelista A, Ciccone G, Broglio F, Soldaite L, Bo S (2017). Timing of food intake: Sounding the alarm about metabolic impairments? A systematic review. *Pharmacological Research*, 125, 132–141. [PubMed: 28928073]

- Berkowitz SA, Gao X, Tucker KL (2014). Food-insecure dietary patterns are associated with poor longitudinal glycemic control in diabetes: Results from the Boston Puerto Rican Health Study. *Diabetes Care*, 37(9), 2587–2592. [PubMed: 24969578]
- Berndt DJ, Clifford J (1994). Using dynamic time warping to find patterns in time series. In *Knowledge Discovery in Databases Workshop* vol. 10 (pp.359–370).
- Centers for Disease Control and Prevention (CDC), National Center for Health Statistics. NHANES dietary data. <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=dietary> Updated November 2007, Accessed December 2018.
- Eicher Miller HA, Khanna N, Boushey CJ, Gelfand SB, Delp EJ (2015). Temporal dietary patterns derived among the adult participants of the National Health and Nutrition Examination Survey 1999–2004 are associated with diet quality. *Journal of the Academy of Nutrition and Dietetics*, 116, 283–291. [PubMed: 26138502]
- Guenther PM, Reedy J, Krebs-Smith SM (2008a). Development of the Healthy Eating Index-2005. *Journal of the American Dietetic Association*, 108(11), 1896–1901. [PubMed: 18954580]
- Guenther PM, Reedy J, Krebs-Smith SM, Reeve BB (2008b). Evaluation of the healthy eating index-2005. *Journal of the American Dietetic Association*, 108(11), 1854–1864. [PubMed: 18954575]
- Hiza HA, Casavale KO, Guenther PM, Davis CA. (2013). Diet quality of Americans differs by age, sex, race/ethnicity, income, and education level. *Journal of the Academy of Nutrition and Dietetics*, 113(2), 297–306. [PubMed: 23168270]
- Khanna N, Eicher-Miller HA, Boushey CJ, Gelfand SB, Delp EJ (2011). Temporal dietary patterns using kernel k-means clustering. *Proceedings of the Institute of Electrical Electronics Engineers International Symposium on Multimedia*, 378–380.
- Khanna N, Eicher-Miller HA, Verma HK, Boushey CJ, Gelfand SB, Delp EJ (2017). Modified dynamic time warping (MDTW) for estimating temporal dietary patterns. *Proceedings of the Institute of Electrical and Electronics Engineers Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal, Quebec.
- Kirkpatrick SJ, Reedy J, Krebs-Smith SM, Pannucci TR, Subar AF, Wilson MM, Lerman JL, Tooze JA (2018). Applications of the Healthy Eating Index for surveillance, epidemiology, and intervention research: Considerations and caveats. *Journal of the Academy of Nutrition and Dietetics*, 118(9), 1603–1621. [PubMed: 30146072]
- Krebs-Smith SM, Pannucci TR, Subar AF, Kirkpatrick SI, Lerman JL, Tooze JA, Wilson MM, Reedy J (2018). Update of the Healthy Eating Index: HEI-2015. *Journal of the Academy of Nutrition and Dietetics*, 118(9), 1591–1602. [PubMed: 30146071]
- Liao TW (2005). Clustering of time series data – a survey. *Pattern Recognition*, 38(11), 1857–1874.
- Macdiarmid J and Blundell J (1998). Assessing dietary intake: Who, what and why of underreporting. *Nutrition Research Reviews* 11(2): 231–253. [PubMed: 19094249]
- Morgan LM, Shi JW, Hampton SM, Frost G (2012). Effect of meal timing and glycaemic index on glucose control and insulin secretion in healthy volunteers. *British Journal of Nutrition*, 108, 1286–1291. [PubMed: 22176632]
- National Center for Health Statistics (NCHS), National Health and Nutrition Examination Survey questionnaires, datasets, and related documentation. <https://wwwn.cdc.gov/nchs/nhanes/default.aspx>, Accessed December 2018.
- National Center for Health Statistics (NCHS), Research Ethics Review Board (ERB) approval. <https://www.cdc.gov/nchs/nhanes/irba98.htm> Updated November 2017, Accessed December 2018.
- Rakthanmanon T, Bilson J, Campana L, Mueen A, Batista G, Westover B, Zhu Q, Zakaria J, Keogh E (2013). Addressing big data time series: mining trillions of time series subsequences under dynamic time warping. *Association for Computing Machinery Transactions on Knowledge Discovery* 7(3), 10.
- Rehm CD, Penalvo JL, Afshin A, Mozaffarian D (2016). Dietary intakes among US adults, 1999–2012. *Journal of the American Medical Association*, 315(23), 2542–2553. [PubMed: 27327801]
- Sakoe H, Chiba S (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions*, 26(1), 43–49.

- Schwingshackl L, Bogensberger B, Hoffman G (2018). Diet quality as assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: An updated systematic review and meta-analysis of cohort studies. *Journal of the Academy of Nutrition and Dietetics*, 118(1), 74–100. [PubMed: 29111090]
- Shi J, Malik J (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8), 888–905.
- United States Department of Agriculture (USDA) automated multiple-pass method. <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/ampm-usda-automated-multiple-pass-method/> Updated September 2016, Accessed December 2018.
- Von Luxburg U (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395–416.
- Wang DD, Leung CW, Yyanping L, Ding EL, Chiuve SE, Hu FB, Willett WC (2014). Trends in dietary quality among adults in the United States, 1999 through 2010. *Journal of the American Medical Association Internal Medicine*, 174(10), 1587–1595. [PubMed: 25179639]
- Wilson MM, Reedy J, Krebs-Smith SM (2016). American diet quality: Where it is, where it is heading, and what it could be. *Journal of the Academy of Nutrition and Dietetics*, 116(2), 302–310.e1. [PubMed: 26612769]
- World Health Organization (WHO). (2000). World Health Organization Obesity: preventing and managing the global epidemic. Report of a WHO consultation WHO Technical Report Series 894. Geneva: Switzerland.
- Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J (2013) National Health and Nutrition Examination Survey: Plan and operations, 1999–2010. National Center for Health Statistics, *Vital Health Stat*, 1(56), 1–37.

Figure 1:

MDTW clustering depicting the fractional amount of energy intake for each US adult participant age 20 to 65 years as drawn from NHANES 1999-2004 in the cluster ranging from 0=0% to 1=100% over a 24-hour day from time 0=12:00 a.m. to time 24=12:00 a.m. the next day, by the proportion of the sample indicated by the inverse gray-scale sidebar legend with 0=0% of the cluster participants to 0.1=10% of the cluster participants.

Figure 2:

UDTW clustering depicting the fractional amount of energy intake for each US adult participant age 20 to 65 years as drawn from NHANES 1999-2004 in the cluster ranging from 0=0% to 1=100% over a 24-hour day from time 0=12:00 a.m. to time 24=12:00 a.m. the next day, by the proportion of the sample indicated by the inverse gray-scale sidebar legend with 0=0% of the cluster participants to 0.1=10% of the cluster participants.

Figure 3:

CDTW clustering depicting the fractional amount of energy intake for each US adult participant age 20 to 65 years as drawn from NHANES 1999-2004 in the cluster ranging from 0=0% to 1=100% over a 24-hour day from time 0=12:00 a.m. to time 24=12:00 a.m. the next day, by the proportion of the sample indicated by the inverse gray-scale sidebar legend with 0=0% of the cluster participants to 0.1=10% of the cluster participants.

Figure 4:

Distribution of the largest eating events for each US adult participant age 20 to 65 years as drawn from NHANES 1999-2004 in MDTW, UDTW, and CDTW clusters indicated by red, green, blue, and orange for clusters 1, 2, 3, and 4 respectively where the fractional amount of energy intake for each participant in the cluster ranging from 0=0% to 1=100% over a 24-hour day from time 0=12:00 a.m. to time 24=12:00 a.m. the next day, by the proportion of the sample indicated by the inverse color-scale legend 0=0% of the cluster participants to 0.12=12% of the cluster participants.

Characteristics of Temporal Dietary Pattern Clusters based on Dynamic Time Warping-type Spectral Clustering among US adults 20 to 65 years, NHANES 1999-2004^a

Characteristics	MDTW ^b				UDTW ^c				CDTW ^b				Total n
	MDTW 1 n (%) ^e	MDTW 2 n (%) ^e	MDTW 3 n (%) ^e	MDTW 4 n (%) ^e	UDTW 1 n (%) ^e	UDTW 2 n (%) ^e	UDTW 3 n (%) ^e	UDTW 4 n (%) ^e	CDTW 1 n (%) ^e	CDTW 2 n (%) ^e	CDTW 3 n (%) ^e	CDTW 4 n (%) ^e	
Total	5902 (70.1)	119 (1.4)	867 (10.3)	1533 (18.2)	2628 (31.2)	4924 (58.5)	142 (1.7)	727 (8.6)	1196 (14.2)	141 (1.7)	2918 (34.7)	4166 (49.5)	8421
Survey year													
1999-2000	1828 (31.0)	43 (36.1)	279 (32.2)	469 (30.6)	808 (30.7)	1529 (31.1)	50 (53.2)	232 (31.9)	397 (33.2)	51 (36.2)	887 (30.4)	1284 (30.8)	2619
2001-2002	2153 (36.5)	42 (35.3)	308 (35.5)	576 (37.6)	982 (37.4)	1789 (36.3)	47 (33.1)	261 (35.9)	416 (34.8)	45 (31.9)	1113 (38.1)	1505 (36.1)	3079
2003-2004	1921 (32.5)	34 (28.6)	280 (32.3)	488 (31.8)	838 (31.9)	1606 (32.6)	45 (31.7)	234 (32.2)	383 (32.0)	45 (31.9)	918 (31.5)	1377 (33.1)	2723
Gender													
Male	3303 (56.0)	73 (61.3)	465 (53.6)	841 (54.9)	1444 (54.9)	2768 (56.2)	86 (60.6)	384 (52.8)	645 (53.9)	86 (61.0)	1592 (54.6)	2359 (56.6)	4682
Female	2599 (44.0)	46 (38.7)	402 (46.4)	692 (45.1)	1184 (45.1)	2156 (43.8)	56 (39.4)	343 (47.2)	551 (46.1)	55 (39.0)	1326 (45.4)	1807 (43.4)	3739
Race/Ethnicity													
Mexican American	1356 (23.0)	23 (19.3)	278 (32.1)	329 (21.5)	621 (23.6)	1160 (23.6)	29 (20.4)	176 (24.2)	373 (31.2)	23 (16.3)	701 (24.0)	889 (21.3)	1986
Other Hispanic	296 (5.0)	6 (5.0)	49 (5.7)	70 (4.6)	128 (4.9)	258 (5.2)	9 (6.3)	26 (3.6)	56 (4.7)	10 (7.1)	133 (4.6)	222 (5.3)	421
Non-Hispanic White	2814 (47.7)	31 (26.1)	292 (33.7)	786 (51.3)	1286 (48.9)	2206 (44.8)	41 (28.9)	390 (53.6)	438 (36.6)	49 (34.8)	1455 (49.9)	1981 (47.6)	3923
Non-Hispanic Black	1203 (20.4)	53 (44.5)	214 (24.7)	304 (19.8)	500 (19.0)	1105 (22.4)	57 (40.1)	112 (15.4)	282 (23.6)	56 (39.7)	530 (18.2)	906 (21.7)	1774
Other	233 (3.9)	6 (5.0)	34 (3.9)	44 (2.9)	93 (3.5)	195 (4.0)	6 (4.2)	23 (3.2)	47 (3.9)	3 (2.1)	99 (3.4)	168 (4.0)	317
Age													
20-34 years	2050 (34.7)	57 (47.9)	323 (37.3)	511 (33.3)	873 (33.2)	1792 (36.4)	64 (45.1)	212 (29.2)	436 (36.5)	59 (41.8)	920 (31.5)	1526 (36.6)	2941
35-49 years	2240 (38.0)	36 (30.3)	333 (38.4)	595 (38.8)	1010 (38.4)	1854 (37.7)	44 (31.0)	296 (40.7)	429 (35.9)	53 (37.6)	1134 (38.9)	1588 (38.1)	3204
50-65 years	1612 (27.3)	26 (21.8)	211 (24.3)	427 (27.9)	745 (28.3)	1278 (26.0)	34 (23.9)	219 (30.1)	331 (27.7)	29 (20.6)	864 (29.6)	1052 (25.3)	2276
PIR ^f													
0-0.99	957 (17.5)	38 (35.2)	188 (23.8)	255 (18.2)	448 (18.5)	851 (18.7)	42 (32.8)	97 (14.3)	275 (24.7)	40 (31.3)	459 (17.0)	664 (17.3)	1438
1.00-1.99	1277 (23.4)	33 (30.6)	204 (25.8)	325 (23.2)	582 (24.0)	1081 (23.8)	34 (26.6)	142 (20.9)	267 (24.0)	29 (22.7)	641 (23.8)	902 (23.5)	1839
2.00-2.99	820 (15.0)	14 (13.0)	127 (16.1)	214 (15.3)	351 (14.5)	704 (15.5)	20 (15.6)	100 (14.7)	170 (15.3)	20 (15.6)	391 (14.5)	594 (15.5)	1175
3.00-3.99	710 (13.0)	10 (9.3)	90 (11.4)	161 (11.5)	292 (12.0)	583 (12.8)	13 (10.2)	83 (12.2)	122 (11.0)	14 (10.9)	338 (12.5)	497 (13.0)	971
4.00-4.99	572 (10.5)	6 (5.6)	62 (7.8)	144 (10.3)	253 (10.4)	433 (9.5)	6 (4.7)	92 (13.6)	103 (9.3)	7 (5.5)	310 (11.5)	364 (9.5)	784

Characteristics	MDTW ^b				UDTW ^c				CDTW ^b				Total n
	MDTW 1 n (%) ^e	MDTW 2 n (%) ^e	MDTW 3 n (%) ^e	MDTW 4 n (%) ^e	UDTW 1 n (%) ^e	UDTW 2 n (%) ^e	UDTW 3 n (%) ^e	UDTW 4 n (%) ^e	CDTW 1 n (%) ^e	CDTW 2 n (%) ^e	CDTW 3 n (%) ^e	CDTW 4 n (%) ^e	
5	1132 (20.7)	7 (6.5)	120 (15.2)	304 (21.7)	498 (20.5)	888 (19.6)	13 (10.2)	164 (24.2)	176 (15.8)	18 (14.0)	559 (20.7)	810 (21.1)	1563
BMI ^g													
Underweight	94 (1.6)	4 (3.5)	14 (1.7)	26 (1.7)	41 (1.6)	86 (1.8)	4 (2.9)	7 (1.0)	14 (1.2)	3 (2.2)	40 (1.4)	81 (2.0)	138
Normal	1738 (30.2)	39 (33.9)	253 (29.9)	511 (34.2)	858 (33.4)	1439 (30.0)	41 (29.7)	203 (28.7)	313 (26.8)	43 (31.2)	862 (30.5)	1323 (32.5)	2541
Overweight	2022 (35.2)	32 (27.8)	279 (33.0)	483 (32.3)	858 (33.4)	1653 (34.5)	45 (32.6)	260 (36.8)	410 (35.2)	45 (32.6)	982 (34.7)	1379 (33.8)	2816
Obese	1899 (33.0)	40 (34.8)	300 (35.5)	474 (31.7)	809 (31.5)	1619 (33.8)	48 (34.8)	237 (33.5)	429 (36.8)	47 (34.1)	946 (33.4)	1291 (31.7)	2713

^a Analyses were adjusted for clustering and stratification. Sample weights were appropriately constructed and applied to the analysis as directed by National Center for Health Statistics. Weights were rescaled so that the sum of the weights matched the survey population at the midpoint of the 6 years covering 1999 to 2004.

^b MDTW= modified dynamic time warping based clusters; Rao Scott F-adjusted χ^2 P-value is a goodness-of-fit, one-sided test; statistical significance is indicated when P<0.05 for gender, race/ethnicity (P<0.001), age (P<0.01), poverty-income ratio (P<0.001), and BMI

^c UDTW=unconstrained dynamic time warping based clusters; Rao Scott F-adjusted χ^2 P-value is a goodness-of-fit, one-sided test; statistical significance is indicated when P<0.05 for race/ethnicity, age (P<0.001), and poverty-income ratio

^d CDTW=constrained dynamic time warping based clusters; Rao Scott F-adjusted χ^2 P-value is a goodness-of-fit, one-sided test; statistical significance is indicated when P<0.05 for race/ethnicity (P<0.001), age (P<0.001), poverty-income ratio (P<0.001), and BMI

^e total numbers do not always add up to sample size due to missing values

^f PIR=poverty-income ratio

^g BMI=body mass index

Table 2.

Mean Healthy Eating Index 2005 Scores and Covariate Adjusted Regression Model Results for Temporal Dietary Pattern Clusters based on Dynamic Time Warping-type Spectral Clustering of US Adults 20 to 65 years as drawn from NHANES 1999-2004

Adjusted Models ^a	n(%)	Mean (SEM)	R-Square ^b	Adjusted R-Square ^c	β (SE) ^d compare to cluster 2	2005 Healthy Eating Index				
						95% CI ^e	β (SE) ^d compare to cluster 3	95% CI ^e	β (SE) ^d compare to cluster 4	95% CI ^e
MDTW ^f			0.020	0.086						
Cluster 1	5902(70.1)	48.7 (0.4)			11.1 (2.1)	6.9, 15.4*	0.5 (0.7)	-0.9, 1.8	-3.4 (0.5)	-4.5, -2.3*
Cluster 2	119 (1.4)	35.7 (2.1)					-10.7 (2.1)	-14.8, -6.5*	-14.5 (2.2)	-18.9, -10.1*
Cluster 3	867 (10.3)	48.2 (0.7)							-3.9 (0.8)	-5.5, -2.2*
Cluster 4	1533 (18.2)	51.9 (0.5)								
UDTW ^g			0.022	0.085						
Cluster 1	2628 (31.2)	49.7 (0.5)			0.9 (0.6)	-0.2, 2.0	10.0 (1.9)	6.1, 13.8*	-4.2 (0.8)	-5.8, -2.5*
Cluster 2	4924 (58.5)	48.4 (0.4)					9.1 (1.9)	5.2, 13.0*	-5.1 (0.8)	-6.7, -3.4*
Cluster 3	142 (1.7)	38.0 (1.9)							-14.1 (1.8)	-17.8, -10.1*
Cluster 4	727 (8.6)	53.7 (0.8)								
CDTW ^g			0.017	0.083						
Cluster 1	1196 (14.2)	47.2 (0.4)			4.4 (1.7)	1.0, 7.9	-4.4 (0.7)	-5.7, -3.0*	-2.0 (0.5)	-2.9, -1.0*
Cluster 2	141 (1.7)	41.6 (1.2)					-8.8 (1.7)	-12.2, -5.3*	-6.4 (1.7)	-9.8, -3.0*
Cluster 3	2918 (34.7)	50.8 (0.2)							2.4 (0.5)	1.3, 3.5*
Cluster 4	4166 (49.5)	48.7 (0.2)								

^aEstimates were adjusted for clustering and stratification. Sample weights were appropriately constructed and applied to the analysis as directed by the NCHS; Weights were rescaled so that the sum of the weights matched the survey population at the midpoint of the 6 years covering 1999-2004

^bSimple model not adjusted for covariates

^cModels were adjusted for survey year, gender, age, race/ethnicity, poverty-income-ratio, and BMI

^dbeta represents difference between mean HE-I2005 of cluster/cluster of reference

^estatistical significance is indicated by * when $P < 0.05/6$ using Bonferroni-type adjustment for multiple comparisons

^fMDTW=modified dynamic time warping

^gUDTW=unconstrained dynamic time warping

^hCDTW=constrained dynamic time warping