BEHAVIORAL FAILURE IN THE PROCESS OF WEIGHT REGAIN: A DATA DRIVEN PROTOCOL

by

Qianheng Ma

BS, Sun Yat-Sen University, China, 2009

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Qianheng Ma

It was defended on

June 9, 2016

and approved by

Committee Chair
Gary M. Marsh, PhD, FACE, Professor, Biostatistics and Epidemiology, Graduate School of Public Health; Clinical and Translational Science, School of Medicine, University of Pittsburgh

Lora E. Burke, PhD, FAHA, FAAN, Professor, Health and Community Systems, School of Nursing; Epidemiology, Graduate School of Public Health; Clinical Translation Science, School of Medicine, University of Pittsburgh

Susan M. Sereika, PhD, Professor, Health and Community Systems, School of Nursing; Biostatistics and Epidemiology, Graduate School of Public Health; Clinical Translation Science, School of Medicine, University of Pittsburgh

Gary M. Marsh, PhD

BEHAVIORAL FAILURE IN THE PROCESS OF WEIGHT REGAIN: A DATA DRIVEN PROTOCOL

Qianheng Ma, MS

University of Pittsburgh, 2016

ABSTRACT

Background: The prevalence of obesity is still an issue of high public health significance. Dietary self-monitoring (DSM) has been identified as the key component in standard behavioral treatment (SBT) for obesity that supports weight loss maintenance. However, little is known about the process of the weight regain in interventions using SBT. Previous research showed the temporal trend of adherence to DSM which preceded weight regain. We hypothesized that participants experienced the failure in adherence to DSM before the onset of weight regain. We then hypothesized that the adherence to daily time-contingent surveys in ecological momentary assessments (EMA) would protect people from behavioral failure and weight regain.

Methods: In this study, we provided a data-driven protocol to define and analyze the weight regain and behavioral failure. With the self-weighing data, we used piecewise linear model to detect the onset of weight regain and classified participants as *maintainers* and *regainers*. We used Bai-Perron's test to detect the failure in DSM adherence before weight regain and classified participants as *collapsers* and *sustainers*. Group-based trajectory modeling was used to cluster the longitudinal patterns of adherence to the time-contingent EMA surveys into two groups (*the consistent group* and *the decline group*). We constructed a three-state Markov transition model for the process of weight regain via a behavioral failure and used Cox models to explore the group effect on the transition intensities among states.

Results: According to the self-weighing trajectories, 148 participants were classified as *regainers* (66.89%) and *maintainers* (25.68%). Among the *regainers* and *maintainers* (N=137), 62.04% was classified as *collapsers* versus *sustainers* (37.96%) for adherence to DSM. of the participants were organized as *the consistent group* (73.8%) versus *the decline group* for adherence to EMA surveys. Being consistently adherent to EMA surveys significantly was related to: (1) greater amount of percent weight loss before weight regain; (2) longer duration of weight loss and maintenance before weight regain; (3) longer duration without behavioral failure and weight regain; and (4) lower hazard of behavioral failure in DSM adherence.

Conclusions: Failure in the adherence of DSM was a more hazardous state for weight regain. Consistent adherence to time-contingent EMA surveys was associated with lower hazard of failure in the adherence to DSM.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First, I would like to express my sincere gratitude to Dr. Gary M. Marsh, my academic advisor who always helps me, and supports me in my life choice. I would also like to thank Dr. Yuting Zhang, my first mentor of my first project at Pitt, excellent economist and I learned a lot from her acumen to numbers and quick mind. I also would like to express my respect and my genuine gratitude to Dr. Lora E Burke, my mentor in EMPOWER research team. I have learnt not just a great deal from her insights in research, her attitude for work and her kindness and elegance. Without her, I would never know the real research world and never know what kind of role statistics can play in fit in other novel hot topics. And my gratitude would also go to Dr. Susan M. Sereika, amazing professional statistician but not just the simple statistician, who can always raise critical points and interesting ideas that inspired me so much to improve my works. She will be the role model for my future career. I would like to thank the EMPOWER team, while working here, I am so proud to be a member in this excellent and efficient research team. You are like my family members!

Last but not the least, I would also like to thanks my family and all of my friends, especially my parents, who supported me financially and encouraged me to pursue my dream.

Hail to Pitt!

Acknowledgements:

R01HL107370 & UL1RR024153

Real Time Data Collection With Adaptive Sampling And Innovative Technologies

1.0 INTRODUCTION

1.1 STANDARD BEHAVIORAL TREATMENT

Standard behavioral treatment (SBT) is an efficacious non-medical approach to help obese (body mass index (BMI) $\geq$ 30 kg/m^2) or over-weight (30 kg/m^2 > BMI> 25 kg/m^2) individuals (Main, Rao, & O'Keefe, 2010) develop a more healthy lifestyle that will support weight loss and maintenance (Ryan & Kushner, 2010). Key components of SBT includes self-monitoring, goal settings for reduced dietary intake and increased energy expenditure, and behavioral strategies such as development of problem-solving skills and enhancement of self-efficacy (Wing, 2004). In traditional SBT, self-monitoring of dietary intake and physical activity has been central to the intervention (Burke, Wang, & Sevick, 2011). However, weight regain is still prevalent in behavioral interventions for weight with approximately only 20% of the participants being successful on long-term maintenance of weight loss (Wing & Phelan, 2005).

1.2 SELF-MONITORING OF DIETARY INTAKE

Self-monitoring of dietary intake (DSM) is a key component in most behavioral weight loss interventions. Paying deliberate attention to dietary intake and recording the details of eating behaviors helps people develop self-regulatory skills that will support successful weight loss, reduce relapse incidents and support long-term weight loss maintenance (Burke, Wang, et al., 2011). In other words, low consistency in dietary self-monitoring could result in greater likelihood of weight loss relapse (Butryn, Phelan, Hill, & Wing, 2007). Given the positive association between weight loss maintenance and adherence to DSM, it is hypothesized that people experienced a change or even failure in the DSM behavior before the onset of weight regain.

1.3 STUDIES ON WEIGHT REGAIN

To date previous studies have focused more on the endpoint status of weight maintenance rather than the underlying process of regain. Most behavioral weight loss interventions defined the term “weight regain” and “weight loss and the maintenance” based on the amount of weight change regarding the pre-specified duration. For example, *maintainers* have been defined as participants who maintained 5% (Crawford, Jeffery, & French, 2000) or 10% (Byrne, Cooper, & Fairburn, 2003; Wing & Hill, 2001) for more than 1 year from baseline (Wing & Hill, 2001) or 2 years (Crawford et al., 2000). Whereas, *regainers* have been defined as individuals who regained their weight to certain level following intentional weight loss, for example, weight regain to within 3.6kg of the baseline weight (Byrne et al., 2003). The endpoint results of these weight loss treatments for the following studies were reported in several publications. However, the process of weight change and the onset of weight regain have been seldom explored. Moreover, measurement on weights were usually too sparse in these studies to capture the timing of weight regain, e.g., every 6 month (Franz et al., 2007). In order to show the temporal trend of weight change, a reliable source of intensive weight data is warranted.

1.4 FEASIBILITY OF USING SELF-WEIGHING DATA

Recently, adherence to self-weighing was shown to be significantly associated with greater weight loss and more successful weight loss maintenance (Zheng et al., 2015). Moreover, no significant association was found between daily self-weighing and negative psychological effects (Zheng et al., 2015). Furthermore, with the more prevalent use of wireless scales and Wi-Fi spots, people are able to weigh themselves without location restriction and transfer weight data to server in real time. By linking such smart scales to self-monitoring applications on smartphones, people can self-monitor their weights with less burden.

Although the reliability of self-reported weights has been questioned (Ambwani & Chmielewski, 2013; Gunnare, Silliman, & Morris, 2013; Stewart, 1982), the trend of weight change can still be clearly observed with the highly intensive daily self-weighing data. Thus, we are able to estimate the time to onset of weight regain with less variance and uncertainty. In addition, the real time data transmission minimizes the self-report bias and recall bias. Therefore, the utilization of the daily self-weighing data is feasible for the exploration of the process of weight regain

1.5 MOBILE-BASED ECOLOGICAL MOMENTARY ASSESSMENTS

Previous research has showed that frequent prompting may encourage adherence to physical activity (Lombard, Lombard, & Winett, 1995; op den Akker, Jones, & Hermens, 2014) and enhance other healthful behaviors (Neff & Fry, 2009). Mobile-based ecological momentary interventions for behavior change have been proved to provide support in participants' everyday life by delivering treatments with higher frequency and less burden (Heron & Smyth, 2010). In the usual case, mobile-based ecological momentary assessments (EMA) were for monitoring behaviors and recording environmental risks for behavioral relapse. Although the primary aim of EMA in behavioral treatments was data collection, we still wondered whether completing the EMA surveys provide support and served as an aid for participants to maintain a healthy lifestyle. We were also interested in whether the completion of mobile EMA surveys was able to reflect the temporal trend of the motivation change or to forecast the behavioral lapses. Therefore, in this study, we hypothesized that, the completion or EMA surveys could be associated with lower hazard of failure in dietary self-monitoring behaviors and the onset of weight regain. We also hypothesized that the adherence to EMA surveys was related to a lowered hazard of weight regain and behavioral failure (see Figure 1 in 2.0).

2.0 AIMS AND HYPOTHESE

Aim I: we aimed to detect the natural process of weight regain and behavioral failure using daily self-weighing data.

Hypothesis I: participants may experience the collapse of beneficial behaviors (dietary self-monitoring) before the onset of weight regain

Aim II: we aimed to evaluate the impact of adherence to daily EMA surveys on behavioral failure and weight regain.

Hypothesis II: adherence to daily EMA surveys can protect participants from behavioral collapse and weight regain

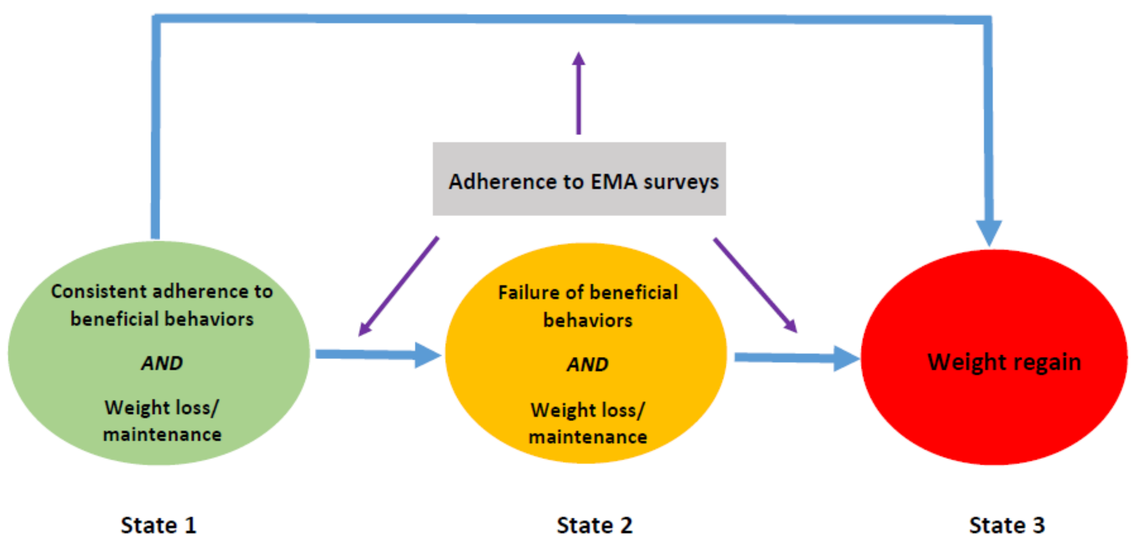


Figure 1. The hypothesized behavioral failure model for weight regain

3.0 DATA

3.1 PARENT STUDY

3.1.1 EMPOWER study overview

This study was a secondary analysis of daily self-weighing data, daily dietary self-monitoring data and ecological momentary assessments data from the EMPOWER study (PI: Lora E. Burke, R01HL107370). EMPOWER was a 12-month standard behavioral trial for obese or overweight individuals that employed EMA data collection. The EMA data included time-contingent assessments at the beginning of the day (BOD) and the end of the day (EOD), self-initiated event contingent assessments and random assessments of participants' behavioral (e.g., quality of sleepiness, lapses to eat, dietary or exercise plan for the day) and psychological status (e.g., mood, energy state) and environments (location, context, temptation to eat). All participants received the standard behavioral intervention as 24 group treatment sessions that included self-monitoring of dietary intake (DSM) and exercise behaviors and goal setting for daily dietary intake and weekly exercising. Participants were instructed to use a smartphone application (*LoseIt!*, *FitNow*, Inc., Boston, MA) to self-monitor their dietary intake and minutes of physical activity. In addition, participants were provided with the a WI-FI scale (*Withings*, Inc., IssylesMoulineaux, France) and instructed to weigh themselves daily at the beginning of the day or at the same time every day.

3.1.2 Sample recruitment and eligibility criteria

Eligibility criteria for EMPOWER participants were: (1) age $\geq$ 18 years, (2) 27 $\leq$ BMI $\leq$ 44, (3) no participation in another weight loss intervention in the past 3 months. Exclusion criteria included: (1) medical conditions such as diabetes, pregnancy, post bariatric surgery, (2) planned to be pregnant in the following year, (3) plans frequent travel, extended vacations or relocation in the following year, (4) received current treatment for a serious mental illness, (5) reports of alcohol intake $\geq$ 4 drinks/day, or (6) unable or unwilling to use a smartphone. At the completion of the baseline assessment, 151 individuals were enrolled in the study. At six months, 142 participants (94%) completed the assessment.

3.1.3 Protocol of EMA data collection

3.1.3.1 Event-contingent EMA surveys

The primary aim of the event-contingent surveys were to assess the trigger of participants' lapse and desire to eat beyond what is consistent with their daily goal, which were initiated by the participants when they encountered a temptation. It also recorded whether the participant gave in to the temptation, the social setting and location where the participant was, their mood and self-efficacy in real time.

3.1.3.2 Random-prompted assessments

This type of assessments were prompted at random times from 1to 5 times per day to assess mood, self-efficacy, the social settings and locations in which the participant was to provide background information.

3.1.3.3 Time-contingent EMA surveys

The time-contingent EMA surveys were prompted at fixed time intervals such as the beginning of the day (BOD) right after the participant awoke and end of the day (EOD) shortly before the participant went to bed. The BOD surveys assessed participants' mood, self-efficacy, quality of sleepiness and the plan for dietary intake and physical activities. The EOD surveys assesses participant's motivation and self-efficacy to adhere to their healthy plan the next day, the burden of the random assessments and the completion of self-initiated surveys during the days and how typical the day was. Although BOD and EOD surveys did not assess participants all along the day, we assumed that time-contingent surveys served as a reliable reminder or the participant's summary of the day. In this secondary analysis, we only used the time-contingent EMA surveys.

3.2 MEASURES

3.2.1 Socio-demographics

We used the investigator-developed self-administered Socio-demographic and Life-style surveys to collect the socio-demographic data. This surveys has 25 primary questions to assess socio-demographic information such as age, gender, ethnicity, marital status, employment status, years of formal education completed.

3.2.2 Self-weighing weight

The self-weighing data were transmitted to the server in real time and participants were able to view their weights every day in *LoseIt!*. Therefore, each participant in the study had a trajectory of self-weighing weights (lbs.) over the 51-week (357 days) period. If the participant did not complete the self-weighing, the system automatically imputed the missing self-weighing data using the most recent self-weighing weight.

3.2.3 Calorie goals and adherence to self-monitoring on dietary intake

Calorie goals for dietary intake were assigned to participants based on their baseline weights and gender. For males with a baseline weight >200 lbs, the participant was assigned an 1,800 kcal dietary goal; for a baseline weight ≤200 lbs, the participant was assigned a 1,500 kcal of dietary goal. For females with a baseline weight >200 lbs, the participant was assigned a calorie goal of 1,500 kcal; for a baseline weight ≤200 lbs, the participant was assigned a calorie goal of 1,200 kcal per day

The dietary self-monitoring data of *LoseIt!* were transmitted to the server every night permitting a 24-hour lag for the participant who did not complete the dietary self-monitoring in the evening. "Non-adherent to dietary self-monitoring" was defined as dietary entry of <50% the calorie goal including the missing entry and otherwise, "adherent to dietary self-monitoring was established by the research team and used or more than ten years ago.

3.2.4 Adherence to the time-contingent surveys (BOD and EOD)

The sample questions from BOD and EOD surveys were listed in Appendice A and B. The completion of all questions in BOD and EOD surveys was defined as adherence for the day over 51 weeks (357 days).

4.0 STATISTICAL METHODS AND ANALYSES

4.1 THE ONSET OF WEIGHT REGAIN

The first “lowest point” of the weight change trajectory was assumed as the onset of weight regain. To avoid accidental fluctuations in self-weighing data, we used subject-specific piecewise linear model (PLM) with maximum likelihood estimation rather than the observed lowest weight to define the onset of weight regain.

A PLM consists of more than one linear segments each with different slopes and continues everywhere in certain domain. In this thesis, daily self-weighing weighs were treated as independent variables with constant variance over time within each participant. We sequentially compared models with $j+1$ segments to j segments ($j = 1, 2, 3, \ldots$) using Bayesian Information Criterions (BICs). Better fit was indicated by lower BIC values. The sequential comparison process stopped when the decrease in BIC was terminated as we increased the number of segments in the model. From this, the optimal number of segments was determined and the slope of each segment, and breakpoints between each pair of adjacent segments (j vs. $j+1$, $j = 1, 2, 3, \ldots$) were estimated. For example, a two-piece linear model was described by the following equations, with b_{j0} and b_{j1} as the intercept and slope for the j^{th} segment,

$$y_t = b_{10} + b_{11} * t + \varepsilon_1, (t < T),$$

$$y_2 = b_{20} + b_{21} * t + \varepsilon_2, (t \geq T).$$

We defined segments as weight gain, weight maintenance, and weight loss with slopes to time and the significance: (1) a weight gain was defined as segment with significantly (p-value<.05) positive slope to time; (2) a weight loss was defined as significantly negative slope to time; (3) weight maintenance was defined as p-value $\geq$ 0.05. Self-weighing weights were analyzed by PROC NLIN of SAS version 9.4 (SAS Institute, Cary, NC).

Next, we classified participants with different weight change patterns by scanning the segments in the chronological order. The classification process would be terminated once a weight gain or the last segment was encountered: (1) if the process experienced at least one weight loss or weight maintenance segments and ended by one weight gain segment, then the participant was classified as a *regainer*; (2) if a weight gain segment never emerged till the end, then the participant was classified as a *maintainer*; (3) if the process terminated without encountering a weight loss segment, then the participant was defined as one who never experienced weight loss during the study.

With three different classes of weight change pattern identified, we performed chi-square tests of independence (or Fisher's exact tests if expected cell counts < 5 were encountered in the contingency table) for categorical characteristics and Analysis of Variance (ANOVA) for continuous characteristics to examine the difference in demographic characteristics among classes. The mean (SD) number of days of continuous weight loss/maintenance and mean (SD) of weight change right before the onset of weight regain were computed for *regainers*.

The primary aim of this study was to explore the process of weight regain following weight loss so we excluded 11 participants who gained weight initially from the following analyses.

4.2 FAILURE IN THE ADHERENCE TO DIETARY SELF-MONITORING

The second outcome we were interested in was, when the participant started to fail in the adherence to DSM. Previous research has shown the temporal decline of adherence to DSM(Burke, Conroy, et al., 2011; Burke et al., 2008). Thus, participants who experienced the failure in DSM were defined as the *collapsers* in self-monitoring behavior as once they gave up self-monitoring, they were not able to resume the consistent adherence to self-monitoring.

We used the structural break estimation method (Gregory & Hansen, 1996) in time series analysis to define the failure in the adherence to DSM, which is also a popular methodology in public policy and macroeconomics (Hansen, 2001). A structural break means an unexpected change in a sequence of observations over time, for example, the participant 5042 (see Figure 2, the red line signified the estimated structural break by Bai-Perron's test). Arbitrary definitions of a change point in a longitudinal data trend were also considered, such as the first episode of non-adherence to dietary self-monitoring. However, by using time series approach, we were able to take advantage of the highly intensive daily data to estimate the most likely change point for each participant and avoid the disturbance of singular non-adherence.

A structural break can be in means, variance or both. For example, in our case, after a certain break day, the trend of adherence to self-monitoring may be with widerfluctuation and participants were less likely to self-monitor their calorie intake afterwards, which may imply the existence of a break in both variances and means. We assumed people started to failure in adherence to DSM at a certain time point. Thus, we statistically tested the null hypothesis of zero structural break versus a single break. (see Figure 2)

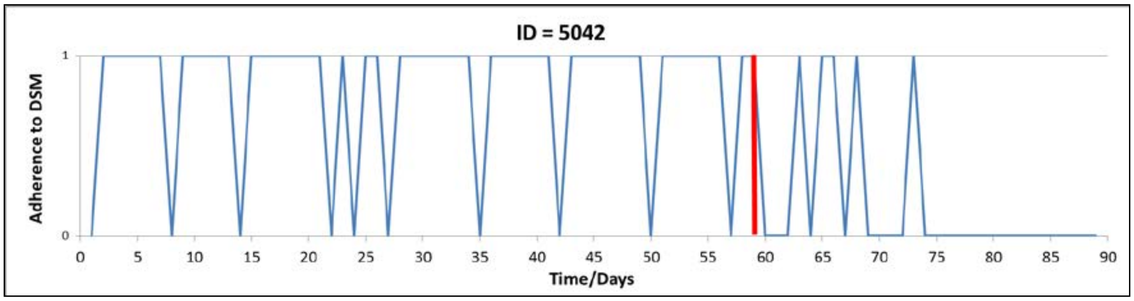


Figure 2. Participant 5042, "Failure" in dietary self-monitoring

Chow's test (Chow, 1960) and Bai-Perron's test (BP test) (Bai & Perron, 2003) are two popular tests to assess the structural break in time-series study. Chow's test can be used in the circumstance when the hypothesis of certain break time such as the time of the release of a policy, is well justified (Chow, 1960). However, in our case, the time of failure in the adherence to DSM was not known in advance and whether the particular participant would fail. In addition, Chow's test can only test the structural break in mean whereas BP test consider structural break in both the mean and variance (Bai & Perron, 2003). Therefore, for this investigation we used BP test to explore whether and when there was a structural break in each participant's time-series of daily adherence to DSM. We performed the BP test for each participant using PROC AUTOREG in SAS version 9.4 (SAS Institute, Cary, NC). Participant who had a significant (p-value<0.05) break in the adherence to DSM would considered to have experienced the DSM failure and he/she was defined as "*collapser*" and otherwise, "*sustainer*"

Bai and Perron(Bai & Perron, 2003) proposed the least-square estimation of the breakpoints. In the multi-variate regression system, m breakpoints were hypothesized,

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \overline{\boldsymbol{Z}}\delta + \boldsymbol{U},$$

Here $\overline{\boldsymbol{Z}} = diag(\boldsymbol{Z_1}, \dots, \boldsymbol{Z_{m+1}})$, $\boldsymbol{Z_i}$ denotes the partition-specific covariates from $T_{i-1} + 1$ to T_i. According to the least square method, the objective function for minimization is

$\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{U}=\sum_{i=1}^{m+1}\sum_{t=T_{i-1}+1}^{T_i}[y_t - x_t\beta - z_i\delta]^2$, therefore the minimization is taken over all partitions and the breakpoints estimated were the global estimator (Bai & Perron, 2003).

We used generalized estimation equations (GEE)(Zeger & Liang, 1986) for logistic regression to analyze the post-failure effect, group effect (identified in 4.3), and their interaction. Assumption of missing completely at random (MCAR) was naturally met because missing values of daily adherence to DSM were treated as non-adherence (in 3.2.3) so that there were no missing values. We considered different structures for working correlation matrix (compound symmetry, first order autoregressive (AR(1)), exchangeable) for within-subject correlation among time points. Although parameter estimates from GEE are consistent even with misspecified working correlation matrix (Liang & Zeger, 1986), the AR(1) structure demonstrated to have the best fit with lowest Quasi Information Criterion (QIC). We used PROC GENMOD in SAS version 9.4 to perform this analysis.

4.3 LONGITUDINAL PATTERNS OF ADHERENCE TO TIME-CONTINGENT EMA SURVEYS

For *regainers*, we retained the data before weight regain defined in 4.1 and for *maintainers*, we used their data until the end of the study.

The percentage of days adherence to the time-contingent surveys (BOD and EOD) was computed on a weekly basis to avoid the daily fluctuation of adherence. We calculated the percent days adherent by the number of days of completing all the questions in both BOD and EOD surveys divided by 7 days for each week and expressed as percentage.

The Group-based trajectory modeling (GBTM), PROC TRAJ in SAS version 9.4 (SAS Institute, Cary, NC) was used to identify different longitudinal patterns of adherence into groups (Jones & Nagin, 2007). We assumed the percent days adherent followed a censored normal distribution. In GBTM approach, the more groups and the higher order of polynomial time effects were pre-specified, the better fit of the data would be obtained, which at the same time added to the complexity of the model with more parameters and smaller group size. Thus, we required that, each resulted group should have >10 % of the total sample size to ensure enough statistical power for later analyses. Then we compared BICs of models of different number of groups determined and different orders of polynomial effects added according to the improvement in goodness of fit judged by Jeffrey's Scale (Jones, Nagin, & Roeder, 2001). To achieve higher efficiency, we first compared models with different numbers of groups and we only assumed linear time effect for each group. Once the optimal number of groups was found, we built a sequence of nested models by adding higher orders of polynomial time effect one at a time until no significant increment in goodness of fit or until the model no longer met the 10% group size restriction.

The chi-square tests of independence were performed to test whether the time-invariant demographic variables were related to the resulted group membership.

4.4 THE THREE-STATE MODEL FOR WEIGHT REGAIN PROCESS

Using definition of the failure in adherence to DSM in 4.2 and presence of weight regain in 4.1, we defined an integrated three-state model here (see Figure 1), which showed paths of transition among the three states, pre-failure in the adherence to DSM (state 1), failure in adherence to

DSM (state 2), and weight regain (state 3). In state 1 and state 2, participants were still losing weights or maintaining their weight losses. We assumed there was no recovery for participant who failed in the adherence to DSM. However, participants could progress to weight regain without experiencing the failure in DSM. Therefore, we used discrete time multi-state model (MSM) to explore the process of state transition (see Figure 1). The MSM model is frequently used in analyzing disease progression such as commonly used *illness-death model*.

In a MSM, the transition probabilities between states *r* and states *s* were expressed for $t \leq u$ as $P(S(t) = s|S(u) = r)$ The transition intensity was defined as the instantaneous risk of moving from state *r* to state *s*,

$$q_{\mathrm{rs}} = \lim_{\delta t \to 0} P(S(t + \delta t) = s|S(t) = r)/\delta t,$$

$$q_{\mathrm{rr}} = -\sum_{s \neq r} qrs,\ 1 \leq r < s \leq 3.$$

So that the transition intensity matrix Q is,

$$\boldsymbol{Q} = \begin{bmatrix} -q_{12} - q_{13} & q_{12} & q_{13} \\ 0 & -q_{23} & q_{23} \\ 0 & 0 & 0 \end{bmatrix}.$$

$P(t)$ can be computed by taking the matrix exponential of the scaled transition intensity maxtrix. $\boldsymbol{P}(\boldsymbol{t}) = exp(t\boldsymbol{Q})$, analytically, $exp(t\boldsymbol{Q}) = 1 + t\boldsymbol{Q} + t^2\boldsymbol{Q}^2/2! + t^3\boldsymbol{Q}^3/3! + \cdots$.

By assuming the time-homogeneous property, the corresponding time *t* transition probabilities are,

$$p_{11}(t) = e^{-(q_{12}+q_{13})t},$$

$$p_{12}(t) = \frac{q_{12}}{q_{12} + q_{13} - q_{23}}(e^{-q_{23}t} - e^{-(q_{12}+q_{13})t}),$$

$$p_{13}(t) = 1 - e^{-(q_{12}+q_{13})t} - \frac{q_{12}}{q_{12} + q_{13} - q_{23}}(e^{-q_{23}t} - e^{-(q_{12}+q_{13})t}),$$

$$p_{21}(t) = 0,$$

$$p_{22}(t) = e^{-q_{23}t},$$

$$p_{23}(t) = 1 - e^{-q_{23}t},$$

$$\mathrm{p}_{31}(\mathrm{t}) = 0,$$

$$\mathrm{p}_{32}(\mathrm{t}) = 0,$$

$$\mathrm{p}_{33}(\mathrm{t}) = 1.$$

We used the intermittently-observed processes to construct the likelihood and then maximize it to estimate the transition intensities. The times ($t_{i1},\ldots,t_{ini}$) were the exact time for transitions. Then for exact transition times, the likelihood function was given by,

$$L_{i,j} = exp\left(q_{S(t_j)S(t_j)}(t_{j+1} - t_j)\right) q_{S(t_j)S(t_{j+1})}.$$

The exponential component of the likelihood is the probability of staying in state *S* at time $t_{j.}$

Maintainers were censored cases in this analysis. For censored(C) cases,

$$L_{i,j} = \sum_{m \in C} p_{S(t_j)m(t_{j+1}-t_j)}.$$

The estimation approach was performed by *R* package '*msm*' version 1.6.1 by Jackson, 2016 (Jackson, 2011).

In order the explore the covariate effects on transitions between states, the primary strategy to model the transition intensities was to decouple the MSM model into several survival models and using the semi-parametric proportional hazard methods to estimate the covariate effects (Cox model). Thus, the transition intensities q_{rs} can be modeled using Cox models in the form of :

$$q_{rs}(t\,; \boldsymbol{X}) = q_{rs}(t\,; \boldsymbol{X} = 0) exp({\boldsymbol{\beta}_{rs}}^T \boldsymbol{X}), 1 \leq r < s \leq 3.$$

Except for our covariate of primary interest, the group membership of the completion trajectory in time-contingent EMA surveys as the fixed factor, we also controlled for percent weight loss between states and the percent weight change per day between states. Percent weight change from state *r* and state *s* were calculated by the weight change between the time points that participant entered state *r* and state *s*, divided by weight of the entry time into state *r* and expressed in 100% percentage. Thus, the percent weight change per day was computed by the percent weight change from state r to state *s* divided by the time (days) spent in state *r* until the entry to state *s*. Here, the entry time to state 1 is 1 for each participant.

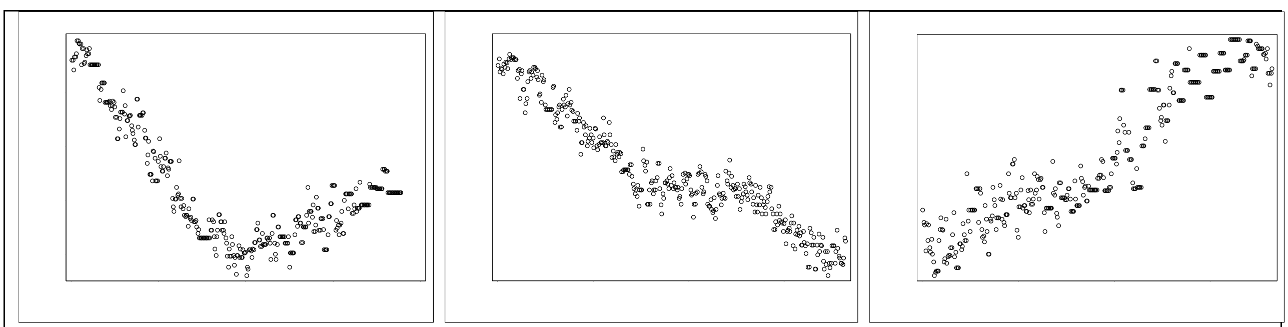
The above Cox models were performed with *R* package *"p3state.msm"* (Meira-Machado & Roca-Pardiñas, 2011) for it allowed the estimation of continuous covariates in the model.

5.0 RESULTS

5.1 SAMPLE DESCRIPTION

The sample consisted of 148 participants with two participants excluded due to pregnancy during the first 2~3 months of participation and one who withdrew on the first day of study. The sample was predominantly female (90.54%) and white (81.08%), full-time employed (82.43%) and currently married (56.46%) with a mean BMI of 34.08 ± 4.58 kg/m^2, and a mean age of 51.27 ± 10.13 years. (Table 1 in 5.2, the first column).

5.2 PATTERNS OF WEIGHT CHANGE AND THE ONSET OF WEIGHT REGAIN



Pattern 1: weight regain	Pattern 2: no weight regain throughout the study	Pattern 3: weight gain from the start

Figure 3. Weight change patterns. Participant 3053, 1072 and 1069

The longitudinal weight change trajectories of the 148 participants were classified into three patterns using subject-specific PLM: pattern 1 was "*weight regain following weight loss*", pattern 2 was "*no weight regain after weight loss throughout the study*" and pattern 3 as "*weight gain from the start*". Participants (n=99, 66.89%) were classified as *regainers* (pattern 1); 38 participants (25.68%) were classified as *maintainers* (pattern 2); and 11 (7.43%) participants gained weight right infially (pattern 3). For *regainers* (pattern 1), the average duration of weight loss was 186.29 ± 78.59 days after the entry of the study and mean weight loss as 20.01 ± 12.17 lbs. and the percent weight loss before the onset of weight regain was 9.88% ± 5.59%. The socio-demographic characteristics were not significantly different over participants of three weight change patterns. (Table 1)

Table 1. Socio-demographics of three weight change patterns

Characteristic	**Total** (n=148)	**Pattern 1** (n=99)	**Pattern 2** (n=38)	**Pattern 3** (n=11)
Age(year), mean ± SD	51.27±10.13	50.86±10.17	52.47±10.43	50.82±9.29
BMI(kg/m^2), mean ± SD	34.08±4.58	33.80±4.60	34.75±4.72	34.21±3.98
Education (year), mean ± SD	16.39±2.81	16.30±2.67	16.74±3.32	16.00±2.24
Gender (female), n (%)	134(90.54)	88(88.89)	36(94.74)	10(90.91)
Race (white), n (%)	120(81.08)	85(85.86)	28(73.68)	7 (63.64)
Marital status n (%)				
Never married	25(17.01)	20(20.41)	4(10.53)	1(9.09)
Currently married	83(56.46)	51(52.04)	25(65.79)	7(63.64)
Divorced/separated/widowed	30(20.41)	22(22.45)	5(13.16)	3(27.27)
Fulltime, n (%)	122(82.43)	83(83.84)	32(84.21)	7(63.64)

5.3 TRAJECTORY ANALYSIS FOR ADHERENCE TO EMA SURVEYS

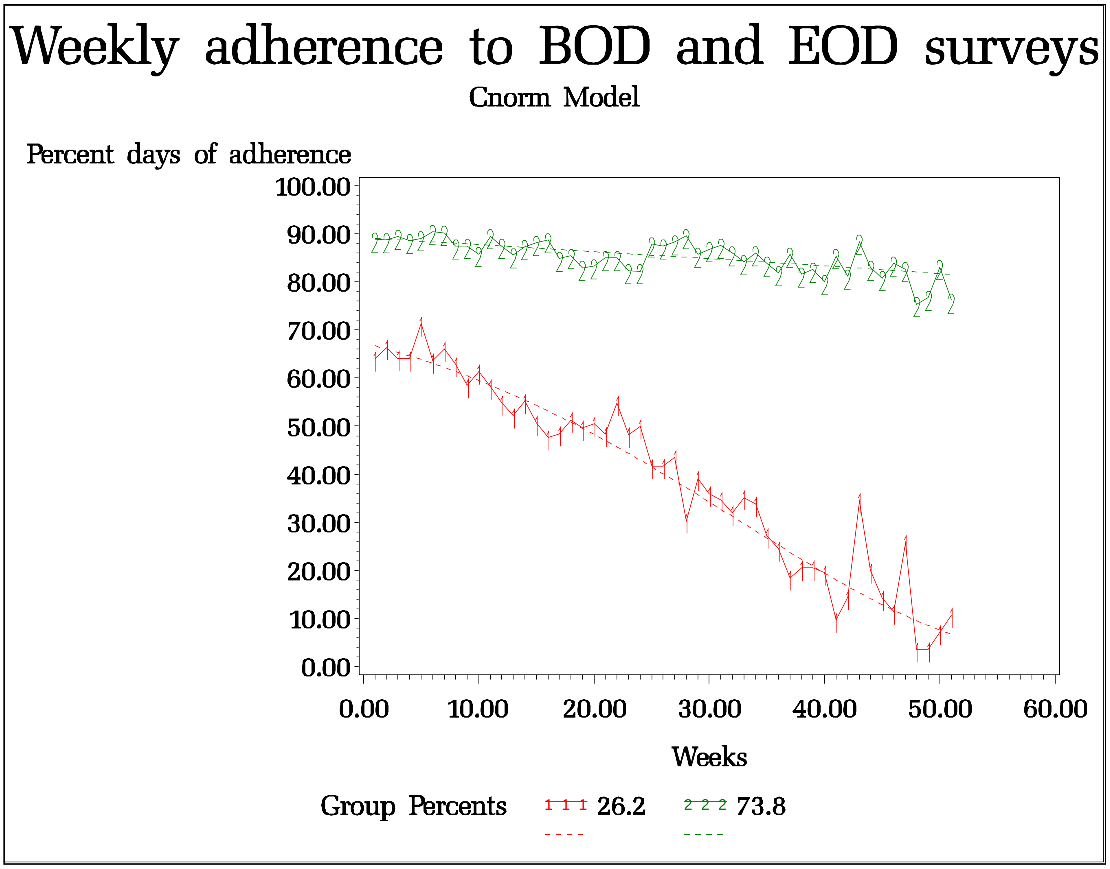


Figure 4. Group-based trajectory modeling for adherence to time-contingent EMA surveys

With GBTM method, we identified two groups for percentage of days adherent to time-contingent EMA surveys per week of 51 weeks for *maintainers* and *regainers* (n=137). (see Figure 4) Adherence to EMA surveys in group 2 (*the consistent group*) was predominantly higher than in group 1 (*the decline group*) over the 51-week period. Group 1 consisted of 26.21% of the sample. For group 1, the adherence to the time-contingent EMA surveys displayed a significant quadratic time effect and declined rapidly over time. Group 2 consisted of 73.79% of the sample, in which the adherence to the time-contingent EMA surveys minimally declined over time in a linear trend. (Table 2)

Table 2. Results of GBTM for each group

Group	Parameter	Estimate	Standard Error	T(H0=0)	Prob > |T|
1	Intercept	69.45	2.48	28.05	0.00
	Linear	-0.73	0.27	-2.74	0.01
	Quadratic	-0.02	0.01	-2.81	0.01
2	Intercept	101.65	1.06	96.01	0.00
	Linear	-0.26	0.04	-6.46	0.00

There was no significant difference in socio-demographic characteristics between these two groups. (Table 3)

Table 3. Socio-demographics of two resulted groups from GBTM

Characteristic	**Total** (n=137)	***The decline group*** (n=36)	***The consistent group*** (n=101)
Age(year), mean ± SD	51.31 ± 10.23	49.00 ± 12.25	51.13 ± 9.34
BMI(kg/m^2), mean ± SD	34.07 ± 4.64	34.16 ± 5.34	34.04 ± 4.39
Education (year), mean ± SD	16.42 ± 2.86	16.14 ± 2.46	16.52 ± 2.99
Gender (female), n (%)	124 (90.51)	33 (91.67)	91 (90.10)
Race (white), n (%)	113 (82.48)	28 (73.68)	85 (85.86)
Marital status n (%)			
Never married	24 (17.65)	9 (25.00)	15 (15.00)*
Currently married	76 (55.88)	16 (44.44)	60 (60.00)*
Divorced/separated/widowed	27 (19.85)	6 (16.67)	21 (21.00)*
Fulltime, n (%)	115 (83.94)	31 (86.11)	84 (83.17)

*1 observation missing.

The proportion of *regainers* was higher in *the decline group* (77.78%) than in *the consistent group* (70.30%). Larger proportion of *maintainers* was found in *the consistent group* (29.70%) than *the decline group* (22.22%). However, there was no significant difference in weight change patterns between these two groups (p = .52).

For *regainers*, (participants of pattern 1 in 5.2), the percent weight loss right before the onset of regain was significant (p = .03) greater in *the consistent group* (10.64% ± 5.81%, 21.57

± 12.79 lbs.) than in *the decline group* (7.95% ± 4.55%, 16.09 ± 9.60 lbs.). However, there was no significant difference (p = .51) in percent weight loss per day between group 1 (0.05% ± 0.04%, 0.11 ± 0.08 lbs. per day) and group 2 (0.06% ± 0.03%, 0.12 ± 0.06 lbs. per day). Moreover, there was no significant difference (p = .25) in average days of maintaining weight loss before weight regain between group 1 (171.82 ± 76.28 days) and group 2 (192.00 ± 79.28 days), though group 2 had a more lasting weight loss than group 1.

For *maintainers* (participants of pattern 2 in 5.2), the weight loss still end of the study was not significantly (p = .07) greater in group 2 (15.20% ± 6.61%, 29.61 ± 14.39 lbs.) than in group 1 (8.53% ± 5.49%, 18.05 ± 10.50 lbs.). The average percent weight loss per day was 0.02% ± 0.02% per day (0.05 ± 0.03 lbs. per day) for group 1 and 0.04% ± 0.02% per day (0.08 ± 0.04 lbs. per day) for group 2. However, the difference of percent weight loss per day is not significant (p = .07).

5.4 FAILURE IN ADHERENCE TO DIETARY SELF-MONITORING

Figure 5 showed the percentage of adherence to DSM over the 357-day period. Percentage of days adherent to DSM in *the consistent group* was predominantly higher than *the decline group*, in which the adherence to DSM decreased dramatically over time.

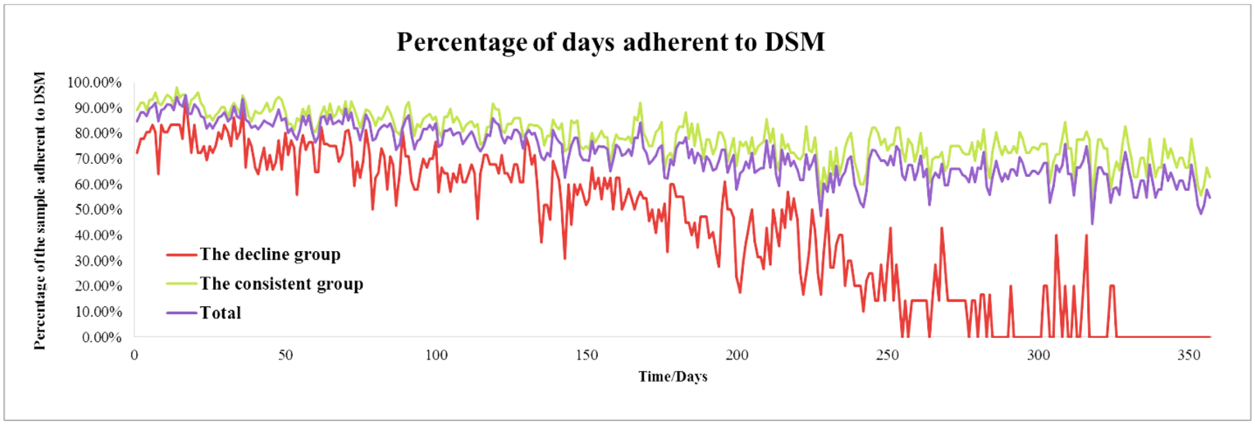


Figure 5. Percentage of of days adherent to DSM in resulted groups from GBTM

Participants who experienced the failure in adherence to DSM (*collapser*) consisted of *Collapsers* consisted of 62.04% of the sample and participants who did not experience the failure (*sustainer*) consisted of 37.96% of the sample. There was no significant difference in socio-demographic characteristics between *collapsers* and *sustainers*. (Table 4)

Table 4. Socio-demographics of *collapsers* and *sustainers*

Characteristic	**Total** (n=137)	**Sustainer** (n=52)	**Collapser** (n=85)
Age(year), mean ± SD	51.31±10.23	52.67±8.57	50.47±11.09
BMI(kg/m^2), mean ± SD	34.07±4.64	32.79±3.68	34.85±4.99
Education (year), mean ± SD	16.42±2.86	16.88±3.18	16.14±2.61
Gender (female), n (%)	124(90.51)	46(88.46)	78(91.76)
Race (white), n (%)	113(82.48)	47(90.38)	66(77.65)
Marital status n (%)			
Never married	24(17.65)	7(13.46)	17(20.24)*
Currently married	76(55.88)	33(63.46)	43(51.19)*
Divorced/separated/widowed	27(19.85)	10(19.23)	17(20.24)*
Fulltime, n (%)	115(83.94)	38(73.08)	77(90.59)

*1 observation missing.

The average time to fail in adherence to DSM for *collapsers* was 125.87 ± 70.76 days after the entry of the study. The time to fail in DSM was significantly different ($p < .01$) between *the decline group* (92.03 ± 62.34 days) and *the consistent group* (143.39 ± 68.94 days).

Post-failure effect ($p < 0.01$) and group effect ($p < 0.01$) on adherence to DSM were significant but their interaction ($p = 0.11$) was not significant. The odds ratio of non-adherence post versus pre the failure in *the decline group* was 6.25 ($p < .01$, 95% CI = (3.85, 10.00)) while the odds ratio of non-adherence post the failure versus pre the failure in *the consistent group* was 3.70 ($p < .01$, 95% CI = (2.50, 5.26)). In both group 1 and group 2, the risk of being non-adherent was higher after the failure than before the failure. The odds ratios of non-adherence in group 1 versus group 2 before ($p < .01$, OR = 2.19, 95% CI = (1.32, 3.63)) and after ($p < .01$, OR = 3.71, 95% CI = (1.92, 7.24)) the failure were both significant, which suggested that consistent adherence to EMA surveys may protect participants from being non-adherent to DSM.

5.5 THE BEHAVIORAL FAILURE MODEL FOR WEIGHT REGAIN

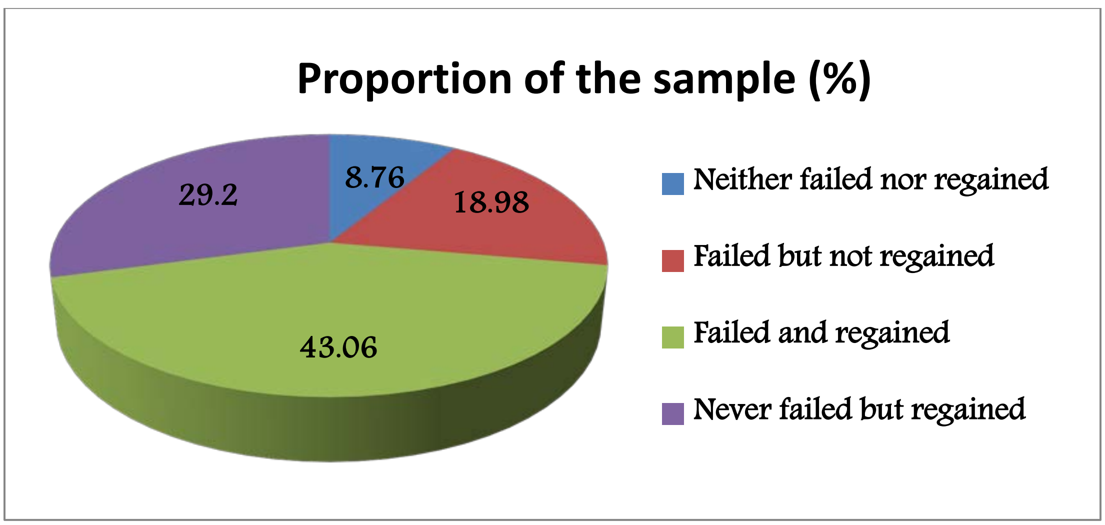


Figure 6. Participants of four classes in the three-state model

According to the states we defined in 3.6, we first categorized participants into 4 classes: (1) never left state 1 (weight loss/maintenance and consistently adherent to DSM); (2) moved to state 2 (failure in adherence to DSM) and stayed in state 2; (3) experienced state 2 and then moved to state 3 (weight regain); (4) without experiencing state 2 but directly moved to state 3. (see Figure 6). Interestingly, participants who never experienced the failure in adherence to DSM and maintained their weight losses were all from group 2 (*the consistent group*). (Table 5) Also, the proportion of participants who failed in adherence to DSM and then regained the weights were larger in group 1 than group 2. However, larger percentage of participants who never ailed in adherence to DSM but regained weights at the end was found in group 2 and larger percentage of participants who failed in adherence to DSM but not regained was found in group 1.

Table 5. Proportion of participants' path in two groups

Path	**Group 1** (*the decline group*) (n=36)	**Group 2** (*the consistent group*) (n=101)
Neither failed nor weight regained, n(%)	0(0)	12(11.88)
Failed but not regained, n(%)	8(22.22)	18(17.82)
Failed and regained, n(%)	21(58.33)	38(37.62)
Never failed but regained. , n(%)	7(19.44)	33(32.67)

We assumed the transition intensities between states were constant over time (time-homogeneous assumption). The hazard (transition intensity) from state 1 to state 2 was 0.003891 (95% CI = (0.003146, 0.004813)), the hazard from state 1 to state 3 was 0.001831 (95% CI = (0.001343, 0.002496)). If the participant had experienced the failure in adherence to DSM, the hazard from failure in adherence to DSM to weight regain was 0.005884 (95% CI = (0.004559, 0.007594)). Therefore, the hazard from state 2 to state 3 was significantly higher than from state 1 to state 2 and state 1 to state 3, which meant that *collapsers* may have higher risk for weight regain than *sustainers*. *Sustainers* had a predominantly higher model predicted probability of maintaining their weight loss over time than the *collapsers* (see Figure 7).

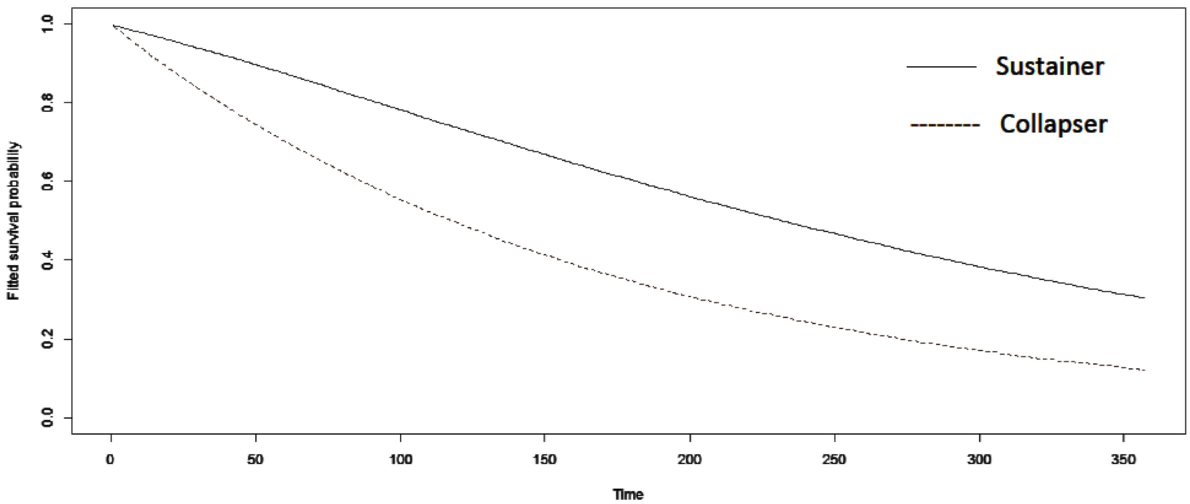


Figure 7. The survival curves: black solid line of *sustainers* and dash line for *collapsers*

The percentage of participants who stayed in state 1 decreased over time but the percentage of participants who entered state 3 (weight regain) increased over time. As the intermediate state (behavioral failure before weight regain), percentage of the sample reached the peak around 195~201 days (27-29 weeks, around the 6th month) after the entry of the study. (see Figure 8)

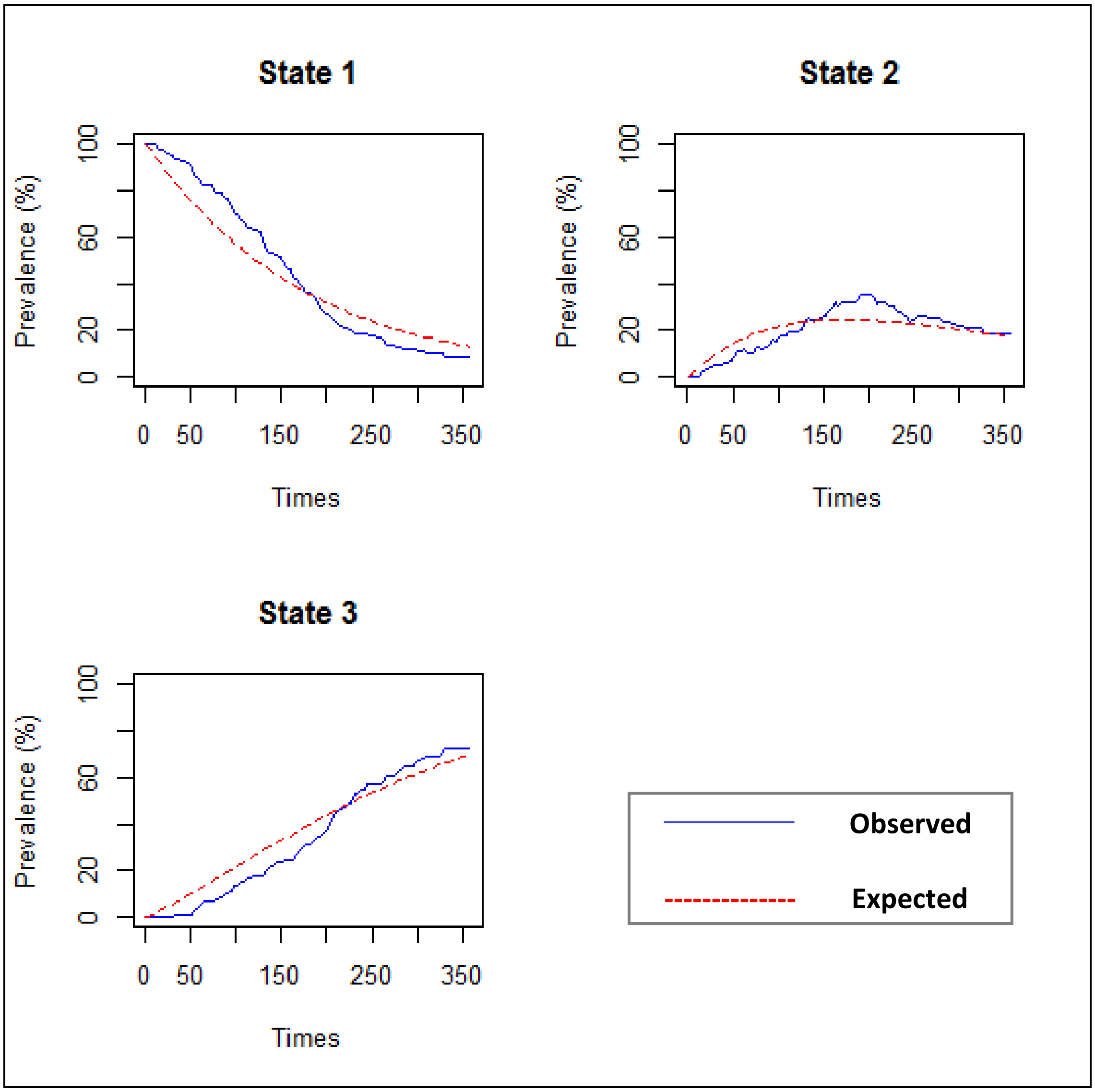


Figure 8. Prevalence of three states

The mean length of time staying in state 1 (weight loss/maintenance without failure in the adherence to DSM) was 174.76 ± 15.63 days. Participants in group 2 (202.84 ± 21.50 days, 95% CI = (164.79, 249.68)) had a significant longer duration in state 1 than group 1 (105.33 ± 17.56 days, 95% CI = (79.98, 146.03)). For *collapsers*, the mean length of time staying in state 2 before moving to state 3 was 169.95 ± 22.13 days and there was no significant difference in length of stay between group 1 (182.81 ± 39.89 days, 95% CI = (119.19, 280.38)) and group 2 (162.84 ± 26.42 days, 95% CI = (118.49, 223.80)).

For participants who lost the same amount of weight (MODEL A), being consistently adherent to EMA surveys (group 2) was significantly associated with lower hazards to fail in adherence to DSM (q_{12}, p = .03). Although possibly the group effect was protective for participants from weight regain, the effect was not significant. (p = .45 for q_{13}, p = .09 for q_{23}).

For participants who lost their weight at the same rate (MODEL B), being in group 2 was significantly associated with lower hazards to fail in adherence to DSM (q_{12}, p < .01). Although possibly the group effect was protective for participants from weight regain, the effect was not significant. (p = .10 for q_{13}, p = .26 for q_{23}). (Table 9)

The effect of percent weight loss and the effect of percent weight loss per day (the rate of losing weight) could be opposite. Greater percent weight loss was significantly associated with less hazards of transition from state 1 to state 2 and state 1 to state 3 (p's < .01) but not with transition intensity from state 2 to state 3 (p = .45). However, greater rate of losing weight was always significantly associated with higher risk of failure and regain (p's < .01).

Table 6. The Cox Markov models for transition intensities among the three states

	Failure (q_{12})		**Regain without failure (q_{13})**		**Regain after failure (q_{23})**	
Model	A	B	A	B	A	B
Group	-0.52* (0.25)	-1.11* (0.24)	-0.37 (0.43)	-0.72 (0.43)	-0.46 (0.27)	-0.33 (0.29)
Weight loss (%)	-0.17* (0.02)		-0.12* (0.03)		0.02 (0.02)	
Weight loss rate, %/day		9.31* (2.77)		18.27* (6.26)		13.10* (2.24)

*p-value < 0.05

6.0 DISCUSSION

This data-driven study utilized the daily self-weighing data, EMA data and daily self-monitoring data to look at the process of weight regain following the weight loss in a 1- year period. We assumed that, all the participants at the beginning were consistently adherent to certain beneficial weight loss behaviors (such as dietary self-monitoring in our study) and kept losing weights or maintained their weight losses (state 1); but then the adherence to the key beneficial behaviors failed (state 2) and later participants started their weight regain (state 3). With this three-state model, we were able to look into every state and the transitions between states.

In our sample of 148 participants, over two thirds of participants (66.89%) experienced the weight regain after the weight loss was defined as the *regainers*. In this one-year period, weight regain was still prevalent even with the efficacious standard behavioral treatment. One fourth (25.68%) of the participants who did not regain their weights throughout the study were defined as the *maintainers*. Other than these two categories, people who increased their weights from the beginning consisted of 7.43% of the sample. *Regainers* kept losing weights or maintained their weight losses until the 6th month (186.29 ± 78.59 days), when they started their weight regain. From this respect, the formal assessment at the first 6th month on weights, BMIs and other obesity related variables which were adopted by many behavioral weight loss interventions is still necessary. Also, the average percent weight loss before the weight regain for the *regainers* was 9.88% ± 5.59%, which coincides with the clinically significant percent weight loss of 10% (Wing & Hill, 2001).

Almost three-fourths (73.99%) of the participants were highly consistent to completing both the time-contingent EMA surveys. The proportion of *regainers* (or *maintainers*) was not significantly different between *the consistent group* and *the decline group*, which meant that being consistently adherent to the EMA surveys was not related to the long-term maintenance. However, for both *regainers* and *maintainers*, being highly consistent in completing the EMA surveys was associated with greater percent weight loss and longer duration of losing or maintaining weights.

Eight-five participants (62.04%) in the sample of 137 experienced the significant structural break in their self-monitoring trajectories and they were defined as *collapsers* and 52 participants had never failed in adherence to DSM were defined as the *sustainers*. *Collapsers* started to fail in DSM after 125.87 ± 70.76 days (the 4^{th} month) from the baseline. However, being consistently adherent to time-contingent EMA surveys would postpone the time to fail for almost 2 months (51 days, increased from average 92 days to average 143 days). Moreover, being consistently adherent to the EMA surveys can protect participants from giving up DSM no matter before the failure or after the failure.

In our proposed three-state transition model, we made two assumptions to simplify the analysis. The first assumption was no recovery from failure in adherence to DSM back to highly consistent adherence to DSM since many of the previous studies showed that adherence to DSM declined over time (Burke, Wang, et al., 2011). The second assumption was the Markov assumption that the transition probability from state 2 to state 3 was not affected by from which state participants moved to state 2. It was naturally met if the assumption of no-recovery was satisfied. For the transition intensities, similar to the hazards in the survival model, q_{23} was significantly greater than q_{13}. Participants who had experienced the failure in the adherence to DSM would have higher hazard for weight regain than the participants who never experienced the failure. Other than this, participants who were adherent to EMA surveys (202.84 ± 21.50 days) would have almost twice longer stay at state 1 than participants whose adherence to EMA surveys dramatically declined (105.33 ± 17.56 days). At around 6-7 months after the entry of the study (195~201 days), the number of participants at state 2 reached the peak. It could be the critical time interval to perform the enhanced intervention to prevent participants, both *collapsers* and *sustainers*, from weight regain and which again supported the 6th month on-site assessment.

The Cox-Markov models showed that, being high consistent to EMA surveys was related to lower hazards to fail in the adherence to DSM but was not related to lower hazard of regain. Another interesting finding was that, the effect of percent weight loss and the effect of average percent weight loss per day were opposite. Greater percent weight loss was related to lower hazards of weight regain and failure of DSM adherence. However, higher rate of losing weight had a negative effect on weight regain prevention. According to our findings, we supposed that if the participant changed his or her behaviors excessively or even to an extreme, e.g., harshly controlled the diet or did much more intensive physical activities so that his body could not endure this intensity, then the participant may give up the healthful behaviors earlier and start his/her weight regain, which would not benefit the long-term weight loss and maintenance.

Although consistent adherence to time-contingent EMA surveys was associated with lower hazards of state progression and greater amount of weight loss, we were not able to conclude on a casual relationship between adherence to EMA surveys and duration of weight loss and the maintenance of the beneficial behaviors. For example, participants who completed the EMA surveys usually had higher motivation and higher self-efficacy, which may be the true cause for more consistent adherence to dietary self-monitoring and the longer duration of weight loss and maintenance.

7.0 STRENGTHS AND LIMITATIONS

This is the first studyto use adata-driven approach to look into the process of how obese or overweight participants in standard behavioral weight study progressed from a healthy life style to the failure of the healthy life style and to weight regain. It provided information for the length of time particpants remained in each state before the onset of weight regain. We also identified factors such as adherence to time-contingent EMA surveys that might have impact on the progression or movement from one state to anothe. This study can be exemplified as an innovative protocol for other researchers to study the change (failure) in beneficial behaviors and to explore the cumulative factors associated weight regain.

We used a piecewise linear model with lowest BICs to fit the weight change trajectory of each participant to estimate the time to weight regain. We also applied self-referential time-series method in detecting the failure in beneficial behavior for each participant, which was based only on the participant's routine behaviors patterns and bypassed the impact from the individual differences such as different marital status and professions. For example, 3 days a week of adherence to DSM may be sufficient to be defined as high adherence for participants of particular stressful professions, but not for other participants with a relatively normal schedule. We used the structural break method in time-series analysis to process the daily binary dietary self-monitoring data. With a significant post-break effect, we showed that this method was able to detect the failure in adherence to dietary self-monitoring. The second advantage was the use of Bai-Perron's test which was able to detect the break in both means and variances. For the binary outcome of adherence or non-adherence to self-monitoring, either means or variances were sufficient for a meaningful structural break. However, if for other obesity related behaviors with a continuous scale of measurement, we would still need to detect the change in the average scale and in the fluctuation of the scale. Therefore in this study, we applied the BP test in case that we would need to estimate the failure of other healthful behaviors in the future.

There were several limitations in this study. We focused only on the period before the first weight regain following the weight loss. So we may ignore the possibility that the participants regained their weight for a short while and then restarted the weight loss, which can be another type of *maintainer*, e.g., participant 2020 in (see Figure 9). In addition, time effect may not be linear, e.g., for participant 3028, the best fit of the data was probably the higher order time effect or the non-polynomial time effect, e.g. the logarithm form or the exponential form. The piecewise linear function can approximate to any continuous function when the trajectory was separated into enough number of pieces (Baire's Theorum). However, we only applied BICs to control for the number of pieces which may cause low efficiency, e.g. for participant 3079, we fitted 7 linear pieces to obtain the best fit of data (see Figure 9).

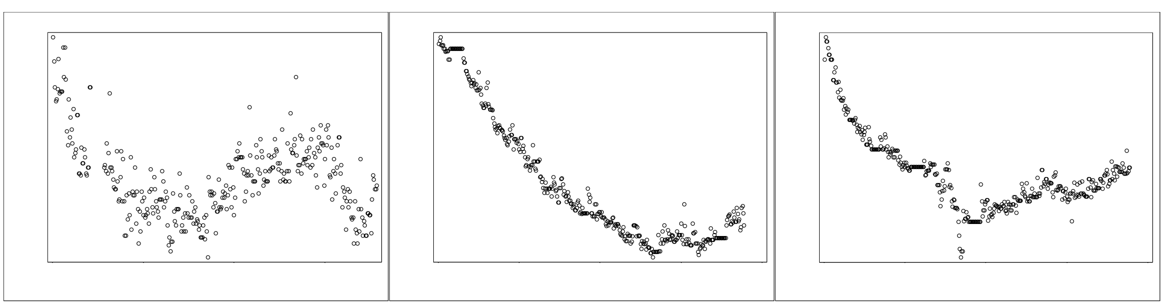


Figure 9. Special cases that our protocol failed to account for (participant 2020, 3028 and 3079)

Second, we assumed only a single structural break in the adherence to DSM. In reality, it may take some time for a participant after study entry to establish the healthful behaviors such as consistent self-monitoring of dietary intake. Therefore, we were not able to rule out the probability that the single structural break we estimated was actually the end of the learning curve rather than the end of the sustained healthful behavior. Thus, multiple stages of failure in adherence to DSM might exist in reality. Moreover, similar to the piecewise modelling, for potential structural break happening near the beginning or the end of the interval, with a decreased number of data points, it can be more difficult to detect the significant structural break for lack of statistical power.

The third limitation relied in the non-recover three-state transition model. We first ignored the possibility that people might recover from failure of adherence to self-monitoring so we skipped exploring factors that may promote or obstruct the recovery. Moreover, we only focused on one intermediate state so that this approach may not be suitable for a multi-state model of behavior failures. Moreover, the strict time-homogeneous assumption and the proportional hazards assumption are sometimes difficult to meet in reality.

8.0 CONCLUSIONS AND FUTURE DIRECTIONS

This thesis has presented an automized data-driven protocol which focus on the arising practical issues of combining and processing highly intensive longitudinal mobile Health (mHealth) data, e.g., daily self-weighing data, ecological momentary assessments data. We have considered the multiple structural-breaks method to detect potential failure in the highly intensive behavioral time-series and piece-wise linear model to detect the onset of weight regain.

In the traditional approaches, participants may need to wait until the next formal assessment (e.g., the next 6^{th} month assessment) to know and get advices on their health status. However, if this data-driven protocol is programmed into the mHealth related smartphone apps to deliver automatic treatments for participants, the cost of behavioral intervention will be much reduced without any location and time restriction and the adverse health outcomes can be prevented in time since participants no longer need to wait for the next on-site assessment. For example, once the significant behavioral failure is detected, the smartphone app. can automatically send a reminder or warning to the participants to control their diet and enhance physical activities in order to prevent weight regain. Or with this protocol, we can identified participants whose onset of weight regain have been detected and performed booster treatments to support them to resume the weight loss.

Based on the limitations we have discussed, in the future, we will not just focus on the first time of weight regain but also allow participants to resume their weight losses. We will also allow participants to have several stages of behavioral failure and the repeatedly recoveries from the failure, which makes more sense in reality. In this thesis, we only used the time-contingent EMA data so we can expand to other components of EMA data, such as the random assessments and self-initiated assessments.

Since we have proved the association between the adherence to EMA surveys and behavioral failure, we can even forecast the behavioral failure via the adherence to EMA surveys, e.g. using hidden Markov model, we can observe the participant's adherence to EMA surveys to speculate which hidden state (behavioral failure or weight regain) the participant might be in. This information will support our decision-making process to perform booster treatments.

APPENDIX A: BOD SURVEY

1. Did you have trouble falling asleep last night? (0=no trouble at all, 10=a lot of trouble)
2. How many hours of sleep did you get?
3. Number of awakenings?
4. Rate how well you slept last night? (0=poor, 10=excellent)
5. Do you feel tired? (y/n)
6. What is your general mood? (Very Good/Good/OK/Bad/Very Bad)
7. Are you on track to meet the weekly physical activity goal ? (y/n/?)
8. Do you intend to be physically active today? (y/n)
   (if yes) Do you have <u>a plan</u> to fit physical activity into your day today? (y/n)
9. Do you intend to meet your calorie and fat goals today? (y/n)
   (if yes) Do you have <u>a plan</u> for meeting your calorie and fat goals today? (y/n)
10. How confident are you that you will be able to stick to your healthy lifestyle plan today? (1-10)

APPENDIX B: EOD SURVEYS

1. Did you have any lapses for which you did not self-initiate a survey today? No – go to 3
    a. 1 missed
    b. 2 missed
    c. 3 missed
    d. >3 missed
    e. Did any of these lead to altered eating? (Y/N)
2. How was your eating today?
    a. Exceeded my goals (did better than planned)
    b. Met my goals
    c. Met a portion of my goals
    d. Did not meet my goals
3. How was your physical activity today?
    a. Did more than intended
    b. Did what I intended to do
    c. Did less than intended but on track for weekly goal
    d. Did < intended, at risk of not meeting weekly goal
4. Were you ill?
5. What is your general mood? (Very Good/Good/OK/Bad/Very Bad)
6. How motivated are you to stay on this healthy lifestyle plan? (0-10)
7. Rate the overall burden of the prompts received today? (0-10)
8. How typical was today (0=not at all typical, 10=very typical)? (If 0-5, GO TO Q10)
9. Was there a significant event that occurred?
    a. Yes – continue
    b. No - End
10. Was it negative/positive?
11. Did this affect your eating?
    a. Yes – continue
    b. No - End
12. If yes, Did this increase/decrease your food intake?

APPENDIX C: SAS CODES

```
/*macro optimal pieces selection*/
%macro optimal_number;
ods _all_ close;
data BIC;
run;
data p;
run;
proc sql noprint;
  select distinct id into : id_list separated by ' ' from work.loseit;
quit;

%local i next_id;
%let i=1;

%do %while (%scan(&id_list, &i) ne );
%let next_id = %scan(&id_list, &i);

/*1 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0;
ll=b0+b1*date;
model weight~normal(ll,s1u);
where id=&next_id.;

ods OUTPUT fitstatistics=BIC&next_id._1;
ods OUTPUT ParameterEstimates=p&next_id._1;
run;
```

```
/*2 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0 b2=0 cutpoint1=179;
if date<cutpoint1 then weightpart=b0+b1*date;
else if date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._2;
ods OUTPUT ParameterEstimates=p&next_id._2;
run;
/*3 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0 b2=0  b3=0 cutpoint1=119  cutpoint2=238  ;
if date<cutpoint1 then weightpart=b0+b1*date;
if cutpoint2>date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
if date>=cutpoint2 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+b3*date;
model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._3;
ods OUTPUT ParameterEstimates=p&next_id._3;
run;
/*4 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0  b2=0  b3=0 b4=0 cutpoint1=110 cutpoint2=200 cutpoint3=300 ;
if date<cutpoint1 then weightpart=b0+b1*date;
if cutpoint2>date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
if cutpoint3>date>=cutpoint2 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+b3*date;
if date>=cutpoint3 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+b4*date;
model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._4;
ods OUTPUT ParameterEstimates=p&next_id._4;
run;
/*5 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
```

```
PARMS  b0=199.34 b1=0 b2=0 b3=0 b4=0 b5=0  cutpoint1=71.4 cutpoint2=142.8  cutpoint3=214.2 cutpoint4=285.6;
if date<cutpoint1 then weightpart=b0+b1*date;
if cutpoint2>date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
if cutpoint3>date>=cutpoint2 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+b3*date;
if cutpoint4>date>=cutpoint3 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+b4*date;
if date>=cutpoint4 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+b5*date;
model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._5;
ods OUTPUT ParameterEstimates=p&next_id._5;
run;

/*6 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0 b2=0 b3=0 b4=0 b5=0 b6=0 cutpoint1=59.5 cutpoint2=119  cutpoint3=178.5 cutpoint4=238 cutpoint5=297.5;
if date<cutpoint1 then weightpart=b0+b1*date;
if cutpoint2>date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
if cutpoint3>date>=cutpoint2 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+b3*date;
if cutpoint4>date>=cutpoint3 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+b4*date;
if cutpoint5>date>=cutpoint4 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+b5*date;
if date>=cutpoint5 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+cutpoint5*(b5-b6)+b6*date;
model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._6;
ods OUTPUT ParameterEstimates=p&next_id._6;
run;
/*7 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0 b2=0 b3=0 b4=0 b5=0 b6=0 b7=0 cutpoint1=51 cutpoint2=102  cutpoint3=153 cutpoint4=204 cutpoint5=255 cutpoint6=306;
if date<cutpoint1 then weightpart=b0+b1*date;
if cutpoint2>date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
if cutpoint3>date>=cutpoint2 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+b3*date;
if cutpoint4>date>=cutpoint3 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+b4*date;
```

```
if cutpoint5>date>=cutpoint4 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+b5*date;
if cutpoint6>date>=cutpoint5 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+cutpoint5*(b5-
b6)+b6*date;
if date>=cutpoint6 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+cutpoint5*(b5-
b6)+cutpoint6*(b6-
b7)+b7*date;
model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._7;
ods OUTPUT ParameterEstimates=p&next_id._7;
run;



/*8 pieces*/
PROC nlmixed data=loseit method=gauss qpoints=24 maxfunc=3000 maxiter=2000;
PARMS  b0=199.34 b1=0 b2=0 b3=0 b4=0 b5=0 b6=0 b7=0 b8=0 cutpoint1=45 cutpoint2=89  cutpoint3=134 cutpoint4=179 cutpoint5=223
cutpoint6=268
cutpoint7=312;
if date<cutpoint1 then weightpart=b0+b1*date;
if cutpoint2>date>=cutpoint1 then weightpart= b0 + cutpoint1*(b1-b2) + b2*date;
if cutpoint3>date>=cutpoint2 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+b3*date;
if cutpoint4>date>=cutpoint3 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+b4*date;
if cutpoint5>date>=cutpoint4 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+b5*date;
if cutpoint6>date>=cutpoint5 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+cutpoint5*(b5-
b6)+b6*date;
if cutpoint7>date>=cutpoint6 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+cutpoint5*(b5-
b6)+cutpoint6*(b6-b7)+b7*date;
if date>=cutpoint7 then weightpart= b0+cutpoint1*(b1-b2)+cutpoint2*(b2-b3)+cutpoint3*(b3-b4)+cutpoint4*(b4-b5)+cutpoint5*(b5-
b6)+cutpoint6*(b6-
b7)+cutpoint7*(b7-b8)+b8*date;

model weight~normal(weightpart,s1u);
where id=&next_id.;
ods OUTPUT fitstatistics=BIC&next_id._8;
```

```
ods OUTPUT ParameterEstimates=p&next_id._8;
run;

data BIC&next_id._1;
set   BIC&next_id._1;
where Descr="BIC (smaller is better)";
id=&next_id.;
model=1;
run;
data BIC&next_id._2;
set   BIC&next_id._2;
where Descr="BIC (smaller is better)";
model=2;
id=&next_id.;
run;
data BIC&next_id._3;
set   BIC&next_id._3;
where Descr="BIC (smaller is better)";
id=&next_id.;
model=3;
run;
data BIC&next_id._4;
set   BIC&next_id._4;
where Descr="BIC (smaller is better)";
id=&next_id.;
model=4;
run;
data BIC&next_id._5;
set   BIC&next_id._5;
where Descr="BIC (smaller is better)";
id=&next_id.;
model=5;
run;
data BIC&next_id._6;
set   BIC&next_id._6;
```

```
where Descr="BIC (smaller is better)";
id=&next_id.;
model=6;
run;
data BIC&next_id._7;
set   BIC&next_id._7;
where Descr="BIC (smaller is better)";
id=&next_id.;
model=7;
run;
data BIC&next_id._8;
set   BIC&next_id._8;
where Descr="BIC (smaller is better)";
id=&next_id.;
model=8;
run;

proc sort data=BIC&next_id._1;
by id model;
run;

proc sort data=BIC&next_id._2;
by id model;
run;
proc sort data=BIC&next_id._3;
by id model;
run;
proc sort data=BIC&next_id._4;
by id model;
run;
proc sort data=BIC&next_id._5;
by id model;
run;
proc sort data=BIC&next_id._6;
by id model;
```

```
run;
proc sort data=BIC&next_id._7;
by id model;
run;
proc sort data=BIC&next_id._8;
by id model;
run;
data BIC&next_id.;
merge BIC&next_id._1 BIC&next_id._2 BIC&next_id._3 BIC&next_id._4 BIC&next_id._5 BIC&next_id._6 BIC&next_id._7 BIC&next_id._8 ;
by id model;
run;
proc datasets library=work;
  delete  BIC&next_id._1  BIC&next_id._2  BIC&next_id._3  BIC&next_id._4   BIC&next_id._5  BIC&next_id._6  BIC&next_id._7
BIC&next_id._8;
quit;
data BIC;
set BIC&next_id. BIC ;
run;
data p&next_id._1;
set  p&next_id._1;
retain polynomial 0 id model;
id=&next_id.;
model=1;
keep id model Parameter Estimate Probt;
run;
data p&next_id._2;
set  p&next_id._2;
retain polynomial 0 id model;
id=&next_id.;
model=2;
keep id model Parameter Estimate Probt;
run;
data p&next_id._3;
set  p&next_id._3;
retain polynomial 0 id model;
```

```
id=&next_id.;
model=3
keep id model Parameter Estimate Probt;
run;
data p&next_id._4;
set   p&next_id._4;
retain polynomial 0 id model;
id=&next_id.;
model=4;
keep id model Parameter Estimate Probt;
run;
data p&next_id._5;
set   p&next_id._5;
retain polynomial 0 id model;
id=&next_id.;
model=5;
keep id model Parameter Estimate Probt;
run;
data p&next_id._6;
set   p&next_id._6;
retain polynomial 0 id model;
id=&next_id.;
model=6;
keep id model Parameter Estimate Probt;
run;
data p&next_id._7;
set   p&next_id._7;
retain polynomial 0 id model;
id=&next_id.;
model=7;
keep id model Parameter Estimate Probt;
run;
data p&next_id._8;
set   p&next_id._8;
retain polynomial 0 id model;
```

```
id=&next_id.;
model=8;
keep id model Parameter Estimate Probt;
run;
proc sort data=p&next_id._1;
by  id model ;
run;
proc sort data=p&next_id._2;
by  id model ;
run;
proc sort data=p&next_id._3;
by  id model;
run;
proc sort data=p&next_id._4;
by  id model ;
run;
proc sort data=p&next_id._5;
by  id model ;
run;
proc sort data=p&next_id._6;
by  id model ;
run;
proc sort data=p&next_id._7;
by  id model ;
run;
proc sort data=p&next_id._8;
by  id model ;
run;
data p&next_id.;
set p&next_id._1 p&next_id._2 p&next_id._3 p&next_id._4 p&next_id._5 p&next_id._6  p&next_id._7 p&next_id._8;
run;
proc datasets library=work;
   delete p&next_id._1 p&next_id._2 p&next_id._3 p&next_id._4 p&next_id._5 p&next_id._6 p&next_id._7 p&next_id._8;
quit;
data p;
```

```
set p&next_id. p ;
run;
proc datasets library=work;
   delete p&next_id. BIC&next_id.;
quit;
   %let i = %eval(&i + 1);
%end;
%mend;
/*Clean the data again for EOD+BOD and do the trajectory analysis*/
PROC IMPORT OUT= WORK.bod
            DATAFILE= "C:\Users\samsung\Desktop\Thesis\Adherence\bod2"
            DBMS=SPSS REPLACE;
RUN;
PROC IMPORT OUT= WORK.eod
            DATAFILE= "C:\Users\samsung\Desktop\Thesis\Adherence\eod(cleaned_newscore).sav"
            DBMS=SPSS REPLACE;
RUN;
PROC IMPORT OUT= WORK.pfinal
            DATAFILE= " C:\Users\samsung\Desktop\Thesis\Adherence\empower_pfinal.sav"
            DBMS=SPSS REPLACE;
RUN;
data eod;
set eod;
keep id date complete;
run;
data eod;
set eod;
rename complete=complete_e;
run;
data bod;
set bod;
keep id date complete;
run;
data bod;
set bod;
```

```
rename complete=complete_b;
run;
proc sort data=eod;
by id date;
proc sort data=bod;
by id date;
data all;
merge bod eod;
by id date;
run;
/*all data points complete both bod and eod*/
data all;
set all;
if complete_b=1 and complete_e=1 then complete=1;
else complete=0;
run;
data maintenance;
set pfinal;
keep id maintanence;
run;
proc sort data=maintenance;
by id;
run;
proc sort data=all;
by id;
run;
data all1;
merge maintenance(in=a) all;
by id;
if a;
run;
data all1;
set all1;
where  date<=maintAnence and maintanence ne 1;
run;
```

```
data all1;
set all1;
w=ceil(date/7);
run;
proc tabulate data=all1 out=des_all1;
class id w ;
var complete;
table id*w,complete=""*(n sum);
run;
data des_all1;
set des_all1;
percent=complete_sum/complete_n*100;
run;
proc transpose data=des_all1 out=call;
   by id;
   id w;
   var percent;
run;
data call;
set call;
array  w(52) w1-w52;
do i=1 to 52;
w[i]=i;
end;
run;
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                    MODEL cnorm; MAX 100; NGROUPS 2 ; ORDER  1 1   ;
RUN;
/*  BIC=-14124.16 (N=4405)  BIC=-14113.74 (N=137)  AIC=-14104.98  L=-14098.98
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;

    MODEL cnorm; MAX 100; NGROUPS 2 ; ORDER  2 1   ;
```

```
RUN;

/*  BIC=-14124.38 (N=4405)  BIC=-14112.23 (N=137)  AIC=-14102.01  L=-14095.01

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

                    MODEL cnorm; MAX 100; NGROUPS 2 ; ORDER  3 1   ;

RUN;

/*BIC=-14128.56 (N=4405)  BIC=-14114.67 (N=137)  AIC=-14102.99  L=-14094.99

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

                    MODEL cnorm; MAX 100; NGROUPS 2 ; ORDER  2 2   ;

RUN;

/*BIC=-14128.47 (N=4405)  BIC=-14114.59 (N=137)  AIC=-14102.91  L=-14094.91

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

                    MODEL cnorm; MAX 100; NGROUPS 4 ; ORDER 1 1 1 1   ;

RUN;

/*no 5%*/


PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

                    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  1 1 1   ;

RUN;

/*

 BIC=-13897.97 (N=4416)  BIC=-13882.69 (N=148)  AIC=-13869.20

L=-13860.20

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  2 1 1   ;

RUN;

/*

 BIC=-13895.48 (N=4416)  BIC=-13878.50 (N=148)  AIC=-13863.51
```

```
L=-13853.51
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  1 2 1   ;
RUN;
/*
 BIC=-13912.99 (N=4416)  BIC=-13896.01 (N=148)  AIC=-13881.02
L=-13871.02
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  1 1 2   ;
RUN;
/*
 BIC=-13911.46 (N=4416)  BIC=-13894.48 (N=148)  AIC=-13879.49
L=-13869.49
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  2 2 1   ;
RUN;
/* BIC=-13897.93 (N=4416)  BIC=-13879.25 (N=148)  AIC=-13862.77
L=-13851.77
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  2 1 2   ;
RUN;
/* BIC=-13900.87 (N=4416)  BIC=-13882.19 (N=148)  AIC=-13865.71
L=-13854.71*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  2 2 2   ;
RUN;
```

```
/* BIC=-13900.25 (N=4416)  BIC=-13879.88 (N=148)  AIC=-13861.89
L=-13849.89
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  1 2 2   ;
RUN;
/* BIC=-13902.96 (N=4416)  BIC=-13884.28 (N=148)  AIC=-13867.80
L=-13856.80*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                  MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 1 1   ;
RUN;
/*
 BIC=-13896.59 (N=4416)  BIC=-13877.92 (N=148)  AIC=-13861.43
L=-13850.43
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                  MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  1 3 1   ;
RUN;
/*
 BIC=-13904.73 (N=4416)  BIC=-13886.05 (N=148)  AIC=-13869.56
L=-13858.56
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                  MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  1 1 3   ;
RUN;
/* BIC=-13901.61 (N=4416)  BIC=-13882.93 (N=148)  AIC=-13866.45
L=-13855.45
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                   MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 2 1   ;
```

```
RUN;
/*
 BIC=-13899.14 (N=4416)  BIC=-13878.76 (N=148)  AIC=-13860.78
L=-13848.78
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 1 2   ;
RUN;
/*BIC=-13904.89 (N=4416)  BIC=-13884.51 (N=148)  AIC=-13866.53
L=-13854.53
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 1 0   ;
RUN;
 /*BIC=-13906.27 (N=4416)  BIC=-13889.29 (N=148)  AIC=-13874.31
L=-13864.31*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 2 2   ;
RUN;
/*
 BIC=-13901.46 (N=4416)  BIC=-13879.38 (N=148)  AIC=-13859.90
L=-13846.90
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 1 1   ;
RUN;
/*
 BIC=-13896.59 (N=4416)  BIC=-13877.92 (N=148)  AIC=-13861.43
L=-13850.43
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
```

```
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 3 3   ;
RUN;
/* BIC=-13909.58 (N=4416)  BIC=-13884.11 (N=148)  AIC=-13861.64
L=-13846.64
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 3 1   ;
RUN;
/* BIC=-13906.41 (N=4416)  BIC=-13884.33 (N=148)  AIC=-13864.85
L=-13851.85
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
                MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  3 1 3   ;
RUN;
/* BIC=-13905.96 (N=4416)  BIC=-13883.89 (N=148)  AIC=-13864.41
L=-13851.41
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 1 1   ;
RUN;
/*
 BIC=-13894.54 (N=4416)  BIC=-13874.17 (N=148)  AIC=-13856.18
L=-13844.18
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 2 1   ;
RUN;
/* BIC=-13895.96 (N=4416)  BIC=-13873.89 (N=148)  AIC=-13854.41
L=-13841.41
*/
```

```
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;

    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 1 2   ;
RUN;
/* BIC=-13895.80 (N=4416)  BIC=-13873.72 (N=148)  AIC=-13854.24
L=-13841.24
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 2 2   ;
RUN;
/*
 BIC=-13898.28 (N=4416)  BIC=-13874.51 (N=148)  AIC=-13853.53
L=-13839.53
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
   ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 3 1   ;
RUN;
/* BIC=-13899.76 (N=4416)  BIC=-13875.99 (N=148)  AIC=-13855.01
L=-13841.01
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 1 3   ;
RUN;
/* BIC=-13909.93 (N=4416)  BIC=-13886.16 (N=148)  AIC=-13865.18
L=-13851.18
*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 3 2   ;
RUN;
/*
```

```
 BIC=-13902.08 (N=4416)  BIC=-13876.61 (N=148)  AIC=-13854.13

L=-13839.13

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 3 3   ;

RUN;

/* BIC=-13910.18 (N=4416)  BIC=-13883.02 (N=148)  AIC=-13859.04

L=-13843.04

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 4 1   ;

RUN;

/* BIC=-13904.65 (N=4416)  BIC=-13879.18 (N=148)  AIC=-13856.70

L=-13841.70

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;


    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 4 0   ;

RUN;

/*

 BIC=-13910.24 (N=4416)  BIC=-13886.47 (N=148)  AIC=-13865.49

L=-13851.49

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;

    ID ID; VAR _1-_51; INDEP w1-w51;

    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 4 2   ;

RUN;

/*

 BIC=-13912.15 (N=4416)  BIC=-13884.98 (N=148)  AIC=-13861.01

L=-13845.01

*/

PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
```

```
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER 5 1 1   ;
RUN;
/*
 BIC=-13905.39 (N=4416)  BIC=-13883.32 (N=148)  AIC=-13863.84
L=-13850.84
*/
/*Best model*/
PROC TRAJ DATA=call OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
    ID ID; VAR _1-_51; INDEP w1-w51;
    MODEL cnorm; MAX 100; NGROUPS 3 ; ORDER  4 2 1   ;
RUN;
/* BIC=-13895.96 (N=4416)  BIC=-13873.89 (N=148)  AIC=-13854.41
L=-13841.41
DATA MEMBERSHIP;
SET of;
KEEP ID GROUP;
RUN;
DATA PFINAL;
SET PFINAL;
DROP GROUP;
RUN;
proc sort data=pfinal;
by id;
run;
proc sort data=membership;
by id;
run;
data pfinal;
merge membership(in=a) pfinal;
by id;
if a;
run;
/* clean data for structural change in variance*/
/*BP test for single structural break*/
```

```
ods graphics off;
PROC IMPORT OUT= WORK.LOSEIT
        DATAFILE= "E:\bimodal\loseit.sav"
        DBMS=SPSS REPLACE;
RUN;

/**??? ????? ???*/
data empower;
set empower;
gender=ssf001;
run;
/*1 male 2 female*/
/*
For people with weight<=200 lbs,
the cal goal is 1200 for female and 1500 for male
; for people with weight>200 lbs, the cal goal is 1500 for female and 1800 for male
*/
data empower;
set empower;
if weight0<=200 and gender=1 then calgoal=1500;
if weight0<=200 and gender=2 then calgoal=1200;
if weight0>200 and gender=2 then calgoal=1500;
if weight0>200 and gender=1 then calgoal=1800;
keep id calgoal;
run;
proc sort data=empower;
by id ;
proc sort data=loseit;
by id;
run;
/*missing--will or will not*/
data loseit1;
merge loseit(in=a) empower;
by id;
if a;
```

```
run;
data loseit1;
set loseit1;
if foodcals>=0.5*calgoal then sm=1;
else sm=0;
run;
data wlr;
set pfinal;
keep id maintanence;
run;
proc sort data=loseit1;
by id;
proc sort data=wlr;
by id;
run;
data loseit_wlr;
merge wlr(in=a) loseit1;
by id;
if a;
run;
data loseit_wlr;
set loseit_wlr;
where date<=maintanence ;
run;
/*write a macro to find the break in variance*/
ods graphics off;
proc sort data=loseit_wlr;
by id date;
run;
data bptest;
run;
%macro bp;
proc sql noprint;
  select distinct id into : id_list separated by ' ' from work.loseit_wlr;
quit;
```

```
%local i next_id;
%let i=1;
%do %while (%scan(&id_list, &i) ne );
%let next_id = %scan(&id_list, &i);
proc autoreg data=loseit_wlr;
        model sm=/bp=(M=1) ;
        where id=&next_id.;
        ods output SeqFSCBP=bptest1;
run;
data bptest1;
set bptest1;
if _N_=1;
run;
data bptest1;
set bptest1;
id=&next_id.;
run;
data bptest;
set bptest bptest1;
run;
 %let i = %eval(&i + 1);
 %end;
%mend;
%bp;
/*assume sm have change*/
data bptest;
set bptest;
if Prob<=0.05 then changesm=1;
else changesm=0;
run;
data bptest;
set bptest;
if newbreak=0 then changesm=0;
run;
```

APPENDIX D: R CODES

```
library("p3state.msm",lib.loc('?C:/Users/samsung/Desktop/Complete dietary active/msm'))
relapse=read.table("C:/Users/samsung/Desktop/Complete dietary active/p3state4.csv",sep=',',header=TRUE)
coxdata=data.creation.reg(relapse)

cmm13=coxph(Surv(start,stop,event)~group1+group3+aweightloss13p,data=coxdata,subset=(treat==0))
cmm13=coxph(Surv(start,stop,event)~factor(GROUP)+aweightloss13p,data=coxdata,subset=(treat==0))
cmm12=coxph(Surv(start,stop,1-event-aux)~group1+group3+aweightloss12p,data=coxdata,subset=(treat==0))
cmm12=coxph(Surv(start,stop,1-event-aux)~factor(GROUP)+aweightloss12p,data=coxdata,subset=(treat==0))


cmm23=coxph(Surv(start,stop,event)~factor(GROUP)+aweightloss23p,data=coxdata)
cmm23=coxph(Surv(start,stop,event)~factor(GROUP)+aweightloss23p,data=coxdata)
cmm23=coxph(Surv(start,stop,event)~group1+group3+aweightloss232p,data=coxdata)
cmm23=coxph(Surv(start,stop,event)~group1+group3+aweightloss23p,data=coxdata)
library("msm",lib.loc('?C:/Users/samsung/Desktop/Complete dietary active/msm'))
Q=rbind(c(0,0.1,0.3),c(0,0,0.3),c(0,0,0))
relapse=read.table("C:/Users/samsung/Desktop/Complete dietary active/msmnew2.csv",sep=',',header=TRUE)
relapse.msm=msm(state~time,subject=id,data=relapse,qmatrix=Q,exacttimes=TRUE)
relapse.msm
sojourn.msm(relapse.msm)
```

BIBLIOGRAPHY

Ambwani, S., & Chmielewski, J. F. (2013). Weighing the evidence: Social desirability, eating disorder symptomatology, and accuracy of self-reported body weight among men and women. *Sex roles, 68*(7-8), 474-483.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics, 18*(1), 1-22.

Burke, L. E., Conroy, M. B., Sereika, S. M., Elci, O. U., Styn, M. A., Acharya, S. D., . . . Glanz, K. (2011). The Effect of Electronic Self‐Monitoring on Weight Loss and Dietary Intake: A Randomized Behavioral Weight Loss Trial. *Obesity, 19*(2), 338-344.

Burke, L. E., Sereika, S. M., Music, E., Warziski, M., Styn, M. A., & Stone, A. (2008). Using instrumented paper diaries to document self-monitoring patterns in weight loss. *Contemporary clinical trials, 29*(2), 182-193.

Burke, L. E., Wang, J., & Sevick, M. A. (2011). Self-monitoring in weight loss: a systematic review of the literature. *Journal of the American Dietetic Association, 111*(1), 92-102.

Butryn, M. L., Phelan, S., Hill, J. O., & Wing, R. R. (2007). Consistent Self-monitoring of Weight: A Key Component of Successful Weight Loss Maintenance. *Obesity, 15*(12), 3091-3096. doi: 10.1038/oby.2007.368

Byrne, S., Cooper, Z., & Fairburn, C. (2003). Weight maintenance and relapse in obesity: a qualitative study. *International journal of obesity, 27*(8), 955-962.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 591-605.

Crawford, D., Jeffery, R. W., & French, S. A. (2000). Can anyone successfully control their weight? Findings of a three year community-based study of men and women. *International Journal of Obesity & Related Metabolic Disorders, 24*(9).

Franz, M. J., VanWormer, J. J., Crain, A. L., Boucher, J. L., Histon, T., Caplan, W., . . . Pronk, N. P. (2007). Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. *Journal of the American Dietetic Association, 107*(10), 1755-1767.

Gregory, A. W., & Hansen, B. E. (1996). Residual-based tests for cointegration in models with regime shifts. *Journal of econometrics, 70*(1), 99-126.

Gunnare, N. A., Silliman, K., & Morris, M. N. (2013). Accuracy of self-reported weight and role of gender, body mass index, weight satisfaction, weighing behavior, and physical activity among rural college students. *Body Image, 10*(3), 406-410.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in US labor productivity. *The Journal of Economic Perspectives, 15*(4), 117-128.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology, 15*(1), 1-39.

Jackson, C. H. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software, 38*(8), 1-29.

Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological methods & research, 35*(4), 542-571.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological methods & research, 29*(3), 374-393.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22.

Lombard, D. N., Lombard, T. N., & Winett, R. A. (1995). Walking to meet health guidelines: the effect of prompting frequency and prompt structure. *Health Psychology, 14*(2), 164.

Main, M. L., Rao, S. C., & O'Keefe, J. H. (2010). TRends in obesity and extreme obesity among us adults. *JAMA, 303*(17), 1695-1696. doi: 10.1001/jama.2010.517

Meira-Machado, L., & Roca-Pardiñas, J. (2011). p3state. msm: Analyzing survival data from an illness-death model. *Journal of Statistical Software, 38*(3), 1-18.

Neff, R., & Fry, J. (2009). Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. *Journal of medical Internet research, 11*(2), e16.

op den Akker, H., Jones, V. M., & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: a literature survey and model. *User modeling and user-adapted interaction, 24*(5), 351-392.

Ryan, D. H., & Kushner, R. (2010). The state of obesity and obesity research. *JAMA, 304*(16), 1835-1836.

Stewart, A. L. (1982). The reliability and validity of self-reported weight and height. *Journal of chronic diseases, 35*(4), 295-309.

Wing, R. R. (2004). Behavioral approaches to the treatment of obesity. *Handbook of obesity: Clinical applications, 2*, 147-167.

Wing, R. R., & Hill, J. O. (2001). Successful weight loss maintenance. *Annual review of nutrition, 21*(1), 323-341.

Wing, R. R., & Phelan, S. (2005). Long-term weight loss maintenance. *The American journal of clinical nutrition, 82*(1), 222S-225S.

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 121-130.

Zheng, Y., Klem, M. L., Sereika, S. M., Danford, C. A., Ewing, L. J., & Burke, L. E. (2015). Self‐weighing in weight management: A systematic literature review. *Obesity, 23*(2), 256-265.