Effect of Nutrition and Exercise Incentive on Weekly Weight Loss

Samantha Bothwell | December 11th, 2020

1 Introduction

An estimated 71.6% of American adults over the age of 20 are overweight, according to the CDC. Health problems attributed to overweight/obesity account for $70 billion annually in health costs and consumers spend $33 billion additionally per year trying to lose weight or to prevent weight gain (R Cleland et al, 1998). Despite long-term weight loss being difficult to maintain, some evidence suggests that changes to both one's diet and physical activity are important for sustaining weight goals (NP Pronk and RR Wing, 1994).

Individuals who classified as overweight or obese were enrolled into the study to understand factors that contribute to weight loss, and to understand how well weight loss is maintained over time. Participants were separated into two groups. One group had a nutritionist, and the other group did not. Both groups received free access to the Wellness center at the University of Colorado Anschutz Medical Campus. Participants were asked to step on a Bluetooth scale once a day over the course of the study. Within the study, there are 3 cohorts. These cohorts indicate participants who started the study around the same time. This study is ongoing, so we do not have access to intervention groupings. Additionally, complete data has not been collected for all participants.

The research questions of interest are :

- What is the trajectory of weight over the duration of time in the study?
- Is there a significant decrease in weight over the course of the study, regardless of intervention group?

2 Methods

Data Cleaning

The data received were fairly cleaned for prior analysis. One subject in Cohort 2 was missing Age, Sex, Race and Ethnicity information and was removed. Some participants measured weight once a week, rather than daily. We summarized weight as the average weight each week in the study to account for some of the missing data. One participant had > 80% missing data, after computing weekly averages, and was removed. To analyze trajectories over a year, we truncated the data to 365 days. This allows for some consistency as cohorts had different lengths of measurement time. The outcome measure for analysis was defined as weekly weight difference from baseline. The original data contained baseline ages. To separate out the effects of aging, we added 1/365 to each participant's weight, for each day in the study. This allows us to use real age for every day, rather than simply baseline age.

Data Analysis

Data cleaning and visualization was performed in R, version 4.0.2 (The R Foundation, Vienna University). Longitudinal modeling was performed in SAS 9.4 (SAS Institute Inc., Cary, NC, USA).

To understand the trajectory of weight over the duration of time in the study we used the fpca.sc method for regular data in the R refund package, version 0.1-21 (Jeff Goldsmith, et al, 2019). We summarized the results of the mean trajectory for each cohort.

To assess if there is a significant decrease in weight over the course of the study, we fit a change-score linear mixed model in SAS with PROC MIXED. Since time points are equally spaced, we used an AR(1) covariance structure. The model includes a random intercept for participant and random effects for the linear and quadratic effect of Week. By including random effects for the linear and quadratic terms, we allow predictions for subjects with just a linear trend to not be influenced by the quadratic term.

Summaries for categorical variables are represented as N(%) and continuous variables are represented as Median [Range] and Mean (95% CI). P-values were computed using a Kruskal Wallis test for continuous variables and a Chi-Square test for categorical variables.

Table 1. Cohort Characteristics

	Cohort 1 (N=29)	Cohort 2 (N=25)	Cohort 3 (N=36)	P-Value
Age				
Mean (SD)	41.8 (9.36)	41.5 (10.5)	43.6 (8.30)	0.881
Median [Min, Max]	43.2 [26.1, 55.8]	43.1 [21.9, 55.2]	42.7 [21.8, 54.9]	
Sex				
Female	21 (72.4%)	19 (76.0%)	29 (80.6%)	0.739
Male	8 (27.6%)	6 (24.0%)	7 (19.4%)	
Race				
Asian	0 (0%)	1 (4.0%)	4 (11.1%)	**0.034**
Black	0 (0%)	3 (12.0%)	4 (11.1%)	
White	26 (89.7%)	21 (84.0%)	28 (77.8%)	
Other	3 (10.3%)	0 (0%)	0 (0%)	
Ethnicity				
Hispanic or Latino	4 (13.8%)	3 (12.0%)	10 (27.8%)	0.210
Non Hispanic or Latino	25 (86.2%)	22 (88.0%)	26 (72.2%)	
Baseline Weight				
Mean (SD)	214 (32.5)	210 (40.2)	220 (31.7)	0.281
Median [Min, Max]	210 [166, 285]	200 [158, 297]	219 [147, 304]	
Time Span				
Mean (SD)	547 (92.8)	412 (71.7)	181 (27.7)	**<0.001**
Median [Min, Max]	593 [210, 600]	431 [84.0, 433]	186 [41.0, 202]	

3 Results

Table 1 displays the cohort characteristics at baseline stratified by enrollment cohort. Of the 90 participants in the study, 69 are female and 21 are male. The average age for Cohort 3 is slightly higher than cohorts 1 and 2, mean 43.6 years compared to 41.8 and 41.5 years for Cohorts 1 and 2 respectively. In effect, the average baseline weight is highest for Cohort 3. For analysis, we are truncating data to 1 year, or 365 days. However, Cohort 3 is still collecting data, so this cohort has less than a year of data, 181 days on average.

Figure 1. Weight Trajectories over time

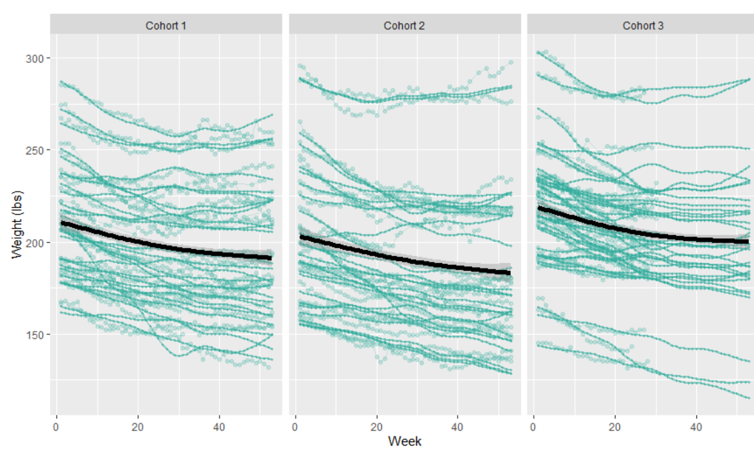


Figure 1 shows the actual weights vs the smoothed FPCA line for all individuals in their respective cohort. From the curves, there is a general decline in weights over time. The mean trajectories for Cohorts 1 and 2 are fairly comparable. The mean trajectory is higher for Cohort 3 than the other two cohorts, indicating that this group had higher weights on average. A reason for this could be that Cohort 3 was sampled during the Covid-19 pandemic, while Cohorts 1 and 2 were sampled prior to the pandemic. All three mean trajectory lines follow a similar pattern. As time increases, the slope of the line tends to flatten. The FPCA method fills in missing data from full data trends. For Cohort 3, the subjects do not have data after day 200, but FPCA fills in data for the rest of the year to show what they expect the trajectory to be for these subjects. The longitudinal model only uses the data provided and will not incorporate these predicted values given by the FPCA trajectories.

Tables 2 and 3 represent the summary output from the linear mixed model. The outcome of interest in this model is the weekly change in weight from baseline. The model incorporates a random intercept, a random effect for the linear and quadratic effects of week and repeated measures across subjects with an AR(1) covariance structure. After adjusting for cohort, sex, age and race, there is a significant relationship between week in study and change in weight ($p < 0.0001$, NDF = 1, DDF = 86, F = 305.49). On average, a participant will lose 1.1047 lbs for each week in the study, holding all other model covariates constant. There is also a significant effect for the quadratic Week term ($p < 0.0001$, NDF = 1, DDF = 88, F = 183.53). The quadratic week effect indicates that for each additional week in the study, the slope of Week increases by 0.0151, on average. This agrees with the results of Figure 1, where the trajectory curve starts to flatten over time rather than maintaining a constant decline. An overall test for fixed effects indicated that, in addition to the effects of Week, Cohort and Sex were significantly associated with weight change. From Table 2, Cohort 2 lost 2.75 more pounds, on average, compared to Cohort 3 ($p = 0.001$, NDF = 2, DDF = 3105, F = 6.95). This agrees with the effect we saw in Figure 1. Additionally, males lost 1.61 more pounds, on average, compared to females ($p = 0.025$, NDF = 1, DDF = 3105, F = 5.03). Based on the overall test for fixed effects, Age, Race and the Week by Cohort interaction, we all non-significant in relation to weight change.

Table 2. Linear Mixed Model Effect Estimates (AIC = 13130.5)

Effect	Estimate	Standard Error	DF	T Value	Pr > |t|
Intercept	-3.6197	1.6552	81	-2.19	**0.0316**
Cohort					
Cohort 1	-0.6333	0.7813	3105	-0.81	0.4177
Cohort 2	-2.7467	0.7637	3105	-3.60	**0.0003**
Sex = Female	1.6087	0.7175	3105	2.24	**0.0250**
Age	0.0612	0.0333	3105	1.84	0.0659
Race					
Asian	-0.2756	1.3865	3105	-0.20	0.8424
Black/African American	1.1706	1.2908	3105	0.91	0.3646
Other	-2.5192	1.7693	3105	-1.42	0.1546
Week	-1.0147	0.1015	86	-10.00	**<.0001**
Week*Week	0.0151	0.0011	88	13.55	**<.0001**
Week*Cohort 1	-0.1239	0.1442	3105	-0.86	0.3904
Week*Cohort 2	-0.0841	0.1497	3105	-0.56	0.5745

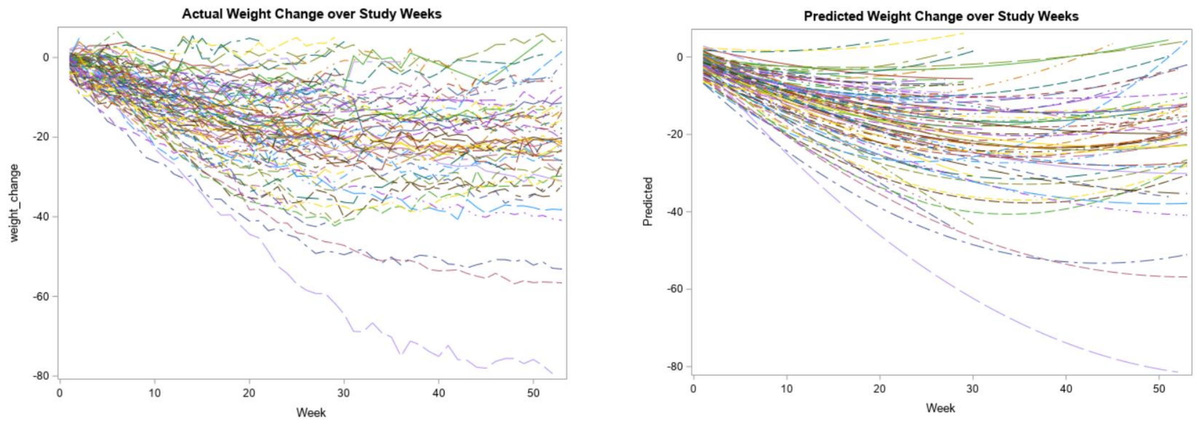
Note: Reference levels - Cohort = 3, Sex = Male, Race = White

Table 3. Type 3 Tests for Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Cohort	2	3105	6.95	**0.0010**
Sex	1	3105	5.03	**0.0250**
Age	1	3105	3.39	0.0659
Race	3	3105	0.99	0.3985
Week	1	86	305.49	**<.0001**
Week * Week	1	88	183.53	**<.0001**
Week * Cohort	2	3105	0.39	0.6798

Figure 2 depicts the actual vs predicted weight change over the course of the study. The predicted weight change lines follow a similar pattern to the actual weight change lines, indicating a good prediction from our model. For a better understanding of how well our model predicts the data, we could perform cross-validation.

Figure 2. Actual vs Predicted Weight Change over Study Weeks



4 Discussion

Our analysis found that participants, on average, lost weight during the course of the study. However, while the study progressed, our analysis found a significant positive quadratic effect for Week. This indicates that participants lost the most weight at the beginning of the study but, as the study progressed, participants started to lose less weight than the previous week and, in some cases, gained back weight.

The outcome for the final model used weekly weights. This causes some discrepancies as a participant may have 4 measurements on week and all 7 the next. To account for different numbers of measurements each week, we incorporated a weighting factor for every week. Each weighting factor took on a value from 0 – 7, indicating how many measurements were included in the week. When not adjusting for this weighting, our model AIC was 13101.4 which was lower than the model presented, which had an AIC of 13130.5. While the weight factor increased the AIC, it is a more reliable estimate.

As this is an ongoing study, we are unable to evaluate the differences in treatment groups. It may be that the found weight loss over time can be significantly attributed to treatment group, as Cohort 3 overlaps with COVID-19. However, we do not have complete data on Cohort 3 as the study is still collecting data from this cohort. Based on the data we have from Cohort 3, measurements are taken from October of 2019 to April of 2020. When we have a complete year of data from Cohort 3, nearly 60% of the data will be overlapped with COVID-19. We do not know how the treatments were adjusted for this study based on the restrictions of COVID-19. As an extension, it may be of benefit to find data of weight trajectories over the course of COVID-19 for a healthy normal population for comparison. We expect that, on average, people gained weight during COVID due to the inability to access fitness centers, lack of walking to or around school/work and a general inability to travel due to stay-at-home orders. Our results showed that participants lost weight over time, but this trend flattened out near the end of the year. A further analysis can use a similar LMM strategy to effectively evaluate if the change in weight over time is different between Cohort 3 compared to Cohorts 1 and 2, utilizing a full year of Cohort 3 data.

References

FastStats. 4 Aug. 2020, https://www.cdc.gov/nchs/fastats/obesity-overweight.htm.

Ewen Harrison, Tom Drake and Riinu Ots (2020). finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling. R package version 1.0.2. https://CRAN.R-project.org/package=finalfit

Jeff Goldsmith, et al (2019). refund: Regression with Functional Data. R package version 0.1-21. https://CRAN.R-project.org/package=refund

Jeff Goldsmith, et al. "Corrected Confidence Bands for Functional Data Using Principal Components." Biometrics, vol. 69, no. 1, 2013, pp. 41–51. Wiley Online Library, doi:https://doi.org/10.1111/j.1541-0420.2012.01808.x.

Matthew Strand and Gary Grunwald (2020). BIOS6643 Analysis of Longitudinal Data Course Notes.

NP Pronk, RR Wing (1994). Physical activity and long-term maintenance of weight loss. Obesity Research; 2:587-599.

R Cleland, DC Graybill, V Hubbard. et al (1998). Commercial Weight Loss Products and Programs: What Consumers Stand to Gain and Lose. Washington, DC: Federal Trade Commission, Bureau of Consumer Protection.

Table Appendix

Appendix Table 1. Covariance Parameter Estimates

Cov Parameter	Subject	Estimate
Intercept	Participant	4.2251
Week	Participant	0.2840
Week*Week	Participant	0.000067
AR(1)	Participant	0.8023
Residual		25.1626

Code Appendix

All code can be found on my GitHub: https://github.com/sjbothwell/BIOS6643_FinalProject

All data cleaning was performed in R. Data analysis was performed in SAS. Here is some relevant code:

```
/* Final Models */
/* 13130.5 */
PROC MIXED DATA = wt_week ;
      CLASS participant_id cohort sex race(ref = "5");
      MODEL weight_change = cohort sex age race Week Week*Week cohort*Week / 
solution outp = pred1;
      RANDOM intercept Week Week*Week / subject = participant_id;
      REPEATED / SUBJECT = participant_id type=ar(1);
      WEIGHT num_days;
RUN;

/* 13101.4 */
/* Lower AIC but does not account for week weight */
PROC MIXED DATA = wt_week ;
      CLASS participant_id cohort sex race(ref = "5");
      MODEL weight_change = cohort sex age race Week Week*Week cohort*Week / 
solution outp = pred2;
      RANDOM intercept Week Week*Week / subject = participant_id;
      REPEATED / SUBJECT = participant_id type=ar(1);
RUN;


PROC SGPLOT DATA = pred1;
   title 'Predicted Weight Change over Study Weeks';
   series x=Week y=pred / group=participant_id;
run;
PROC SGPLOT DATA = pred1;
   title 'Actual Weight Change over Study Weeks';
   series x=Week y=weight_change / group=participant_id;
run;
```

```
# Missing Data summary (in R)

explanatory = c("cohort", "sex", "race", "age")

dependent = "mean.wt"

wt.week %>% missing_compare(dependent, explanatory)
```

```
# FPCA Plots (in R)
wt_fpca = fpca.sc(wt.lb)
# Add yhat varaibles to dataframe
yhat <- as.data.frame(wt_fpca$Yhat)
wt$wt_lb_hat <- c(t(yhat))
colnames(wt.lb) = paste0("time_", 1:53)
# Actual vs predicted
wt.act <- as_tibble(wt.lb) %>%
  mutate(id = row_number(), cohort = cohort.list$cohort) %>%
  gather(time, value.act, contains("time_")) %>%
  mutate(time = str_remove(time, "time_"),
    time = as.numeric(time)) # %>%
 #  filter(id %in% (1:10))
wt.pred <- as_tibble(wt_fpca$Yhat) %>%
  mutate(id = row_number(), cohort = cohort.list$cohort) %>%
  gather(time, value.pred, contains("V")) %>%
  mutate(time = str_remove(time, "V"),
    time = as.numeric(time)) # %>%
 #  filter(id %in% (1:10))

wt.compare <- cbind(wt.act, wt.pred$value.pred)
colnames(wt.compare)[5] <- "value.pred"

wt.compare$cohort <- ifelse(wt.compare$cohort == 1, "Cohort 1",
                    ifelse(wt.compare$cohort == 2, "Cohort 2", "Cohort 3"))

ggplot(data = wt.compare[!is.na(wt.compare$value.act),], aes(time, value.act, group = id)) +
  geom_point(alpha = 0.2, color = "#32ad9d") +
```

```
  geom_path(alpha = 0.2, color = "#32ad9d") +

  geom_point(data = wt.compare[!is.na(wt.compare$value.pred),], aes(time, value.pred, group = id), alpha = 0.5, color = "#32ad9d", size = 0.75) +

  geom_path(data = wt.compare[!is.na(wt.compare$value.pred),], aes(time, value.pred, group = id), alpha = 0.5, color = "#32ad9d", size = 1) +

  facet_grid(. ~ cohort) +

  geom_smooth(data = wt.compare[!is.na(wt.compare$value.pred),], aes(time, value.pred, group = cohort), lwd = 2, color = 'black') +

  xlab("Week") + ylab("Weight (lbs)")
```