

# Understanding the Longitudinal Impact of Beta-Carotene (BC) Supplementation on Serum BC Levels

*Sacha Robbins*

*June 13, 2018*

## Abstract

Since previous studies have shown that the highly antioxidant beta-carotene (BC) is associated with a decreased risk of cancer, its pharmacokinetics needs to be better understood. In this study, we used generalized estimating equations (GEE) to confirm these associations: BC supplementation impacts serum BC levels over time and it varies depending on dosage level, and, after BC supplementation is ceased, the rate at which patients return to baseline does not differ by dosage. It is estimated that the relative difference in mean serum BC comparing patients in the treatment group with patients in the placebo group throughout the 6 month treatment phase is 823.09 ug/mL (95% CI: [598.18, 1048.00]). This estimate increases to 1055.52 ug/mL (95% CI: [991.29, 1119.74]) after the potential outlier observations of patient 57 are removed. As secondary analyses, we explored the potential differences in effect of BC supplementation on serum BC levels over time by other covariates: gender, cholesterol, age, and body mass index (BMI). We also explored the effect of BC supplementation on serum vitamin E over time. Lastly, we created a predictive linear mixed effects model using Bayesian methods and provided some predictions for randomly chosen patients if they were to extend their duration of the BC supplementation. The biggest challenge with this study was that the data set was greatly unbalanced between those receiving treatment to those receiving placebo (4:1 ratio). These challenges may have affected our models and inferences made.

## Introduction

Statistics in the 2012 American Cancer Society journal predicted that by year 2030, there will be 21.7 million new cancer cases and 13 million cancer deaths worldwide. When you adjust for other environmental or lifestyle factors like smoking, diabetes, pollution, or fewer childbirths, the numbers can seem very daunting. Studies like this one help us move toward a better understanding of cancer and what we can do to prevent it. Many antioxidants like BC have been used to prevent and battle against cancer. It has been shown that BC is associated with a decreased risk of cancer, but that some physiological factors can affect the absorption, storage, and utilization of the drug, which could then affect the impact it has against cancer. Our analysis seeks to confirm that BC supplementation does significantly impact the serum BC levels over time and that that impact is magnified by higher dosages. We explore the effect of other covariates on the impact of the supplementation, and we also explore the effect that BC supplementation has on serum vitamin E levels over time. We were able to confirm these associations, but, we were unable to create a good predictive model. However, through our association analyses, we made some interesting discoveries that are worth further investigation.

## Methods

These data were collected over the course of 15 months (double-blind) that was split into three phases, which we will call *pre-treatment*, *treatment*, and *post-treatment*. At the start of the study, these baseline measurements were collected from each patient: age, gender, BMI, and cholesterol (time-invariant covariates).

On each monthly visit, plasma/serum BC (ug/mL) and plasma/serum vitamin E (ug/mL) were also collected from patients. From months 0 to 3, all patients were on placebo. 46 volunteers were randomized to receive one of five doses of BC (0, 15, 30, 45, or 60 mg/day) for a duration of 6 months (months 4 to 9). Then, after the ninth visit, all patients ceased supplementation and their levels were tracked for another 6 months. However, it is important to note that none of the 46 patients had data on months 11 and 12, and some patients had missing monthly observations. Another interesting note is that there seems to be some patients missing in the study as the patient id numbers extend to 57 even though there is only data for 46 patients.

Our primary subset of data that we worked with was that during the treatment phase. We did some exploratory analysis in the pre-treatment phase just to get an understanding of our sample in terms of the distributions of serum BC, serum vitamin E, age, gender, BMI, and cholesterol. We also grabbed a baseline serum BC and serum vitamin E measurements by taking the average of those in the pre-treatment phase. This provided a better association model with more efficient estimates and confidence intervals. When investigating the rate at which patients returned to baseline after supplementation, we of course opened up our analysis to the data in post-treatment phase.

Patient 31 had missing serum BC and serum vitamin E levels for month 6. In order to utilize patient 31 in the analysis, that observation was filled in by the fitted value of a simple linear regression model based on the subject-specific trajectory during treatment phase. This seemed to be a reasonable decision instead of throwing away data. Patient 57 did not have missing information, but did seem like an abnormality. This patient's trajectory was clearly showing trends of someone taking BC supplement rather than a placebo. The deviation is blatantly noticeable. It is possible that Patient 57's dose was wrongly noted. Analysis with Patient 57 did not change the majority of our conclusions, but it did have a strong influence as briefly stated in the Abstract. For simplicity, our final models and inference will not include Patient 57.

The randomization of this study eliminates the need to adjust for other covariates besides dose, an indicator for treatment or placebo (predictor of interest), and serum BC (continuous response variable) for our primary scientific questions of interest. We chose to take a semi-parametric approach by fitting generalized estimating equations (GEE). Here we do not need to assume a full distribution on our continuous response variables, serum BC and serum vitamin E, and we could implement a covariance structure since we have inherently correlated data. Huber-White tells us that our regression estimates are consistent even if we misspecify the dependence model. GEE allowed us to gain efficiency in our estimators by assuming an exchangeable covariance structure based on our exploratory data analysis (EDA) in Appendix A. The appropriateness of the models were assessed using residual plots, and we used empirical correlation/covariance matrices and variograms for the EDA on our  $\Sigma_i = Cov[Y_i]$ , which can also be found in Appendix A. To safeguard our inference, we used the robust variance estimates. Implementation of the models was conducted using 'gee' package version 4.13-19 and RStudio version 1.1.414.

$Y_{ij} = \beta_0 + \beta_1 Tx_i + \beta_2 mc.base_i + \beta_3 month_{ij} + \beta_4 Tx_i * month_{ij} + \epsilon_{ij}$  is the GEE model we used for the  $i^{th}$  patient and  $j^{th}$  measurement where  $Y_{ij}$  is the plasma BC levels (ug/mL),  $Tx_i$  is an indicator variable for whether or not a patient is on any dose treatment of BC supplementation besides the placebo.  $mc.base_i$  is the mean-centered averages of plasma BC measurements from the pre-treatment phase. We mean-centered the covariate for better interpretability.  $month_{ij} = 1, 2, \dots, n_i$  keeps track of the measurements during treatment phase, where  $n_i$  is the number of treatment measurements for  $i^{th}$  patient. Most patients have all six observations during the Treatment phase, but some have as low as 2 observations. Patients with less than three observations in any of the three phases were not used. Error terms are assumed independent and centered at zero with an independent covariance matrix.

To investigate whether the impact of BC supplementation on serum BC levels over time was dose-dependent, we focused on Treatment Phase and only those who received supplementation. Our mean model here was  $E[Y_{ij}] = \beta_0 + \beta_1(dose_i/15) + \beta_2 mc.base_i + \beta_3 t_{ij} + \beta_4 t_{ij}(dose_i/15)$ . Instead of an indicator variable for treatment, we tranformed dose by dividing by 15 for better interpretability.

Our final secondary aim was to build a predictive model for serum BC level if patients were to extend their supplementation another three months. We used Bayesian Inference for linear mixed effects (LME) model in order to create a predictive model. Here, we were able to place non-informative priors, and through Gibbs sampling technique we were able to draw from the posterior distribution of the parameters.

From there, we were able to obtain marginal estimates as well as subject-specific random intercepts and slopes. Our covariance matrix for our random effects was assumed to follow an inverse Wishart distribution. Here, we ignored vitamin E levels, BMI, and age based on our exploration of these covariates and their effect on the impact that supplementation had on serum BC levels. Here is our mean model:  $E[Y_{ij}|\vec{b}_i] = \beta_0 + \beta_{C1}Tx_i + \beta_{C2}Male_i + \beta_{C3}Chol_i + \beta_{C4}base_i + \beta_{L1}month_{ij} + \beta_{L2}Tx_i \cdot month_{ij}$ . Implementation of the model was conducted using ‘rjags’ package version 4-6.

## Results

Recall our mean model fitted in order to investigate the impact of BC supplementation on serum BC levels over time:  $E[Y_{ij}] = \beta_0 + \beta_1 Tx_i + \beta_2 mc.base_i + \beta_3 month_{ij} + \beta_4 Tx_i * month_{ij}$ . If  $\beta_1 = 0$  and  $\beta_4 = 0$ , the longitudinal model would be the same regardless of what type of treatment a patient receives. Figure 1 visually shows some support against these null hypotheses as well as the potential influence of Patient 57.

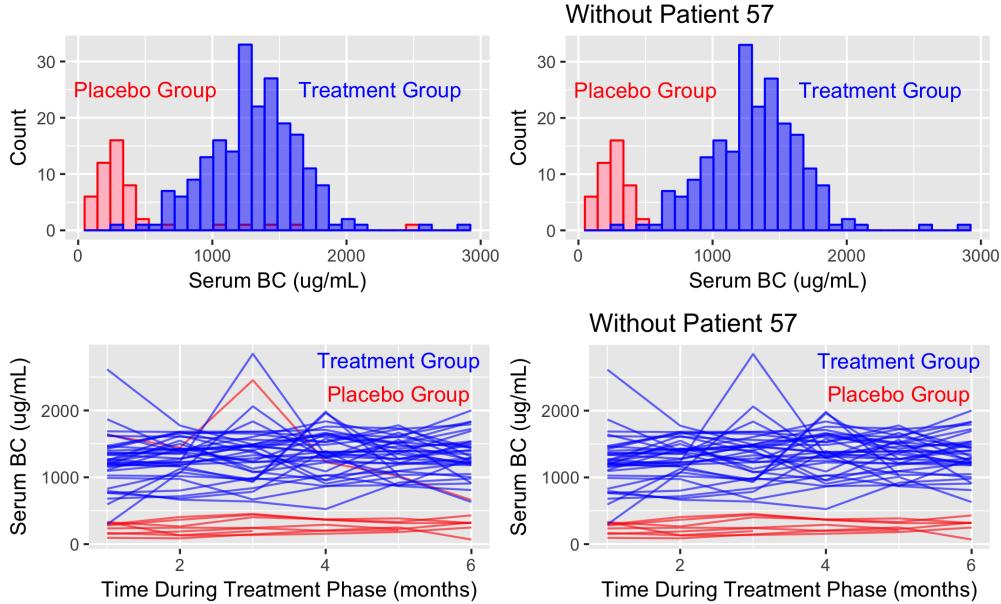


Figure 1: Histograms and Spaghetti Plots Comparing Distributions of Serum BC Levels with and without Patient 57

Table 1: Output from GEE Model Comparing with and without Patient 57

	Estimates	Est (w/o 57)	Robust SE	Robust SE (w/o 57)
Intercept	432.8930070	201.5480753	215.5282979	20.9030602
Tx	823.0911857	1055.5153804	224.9112015	64.2235247
Mc.Base	0.6997623	0.8913019	0.3226287	0.2634526
Month	-20.1219211	6.1636016	24.8136216	3.4466712
Tx:Month	38.0069463	11.8422225	26.5544136	10.0969343

Table 1 shows the estimates for these regression parameters after fitting models with and without Patient 57. Let’s first consider the model with Patient 57.  $\widehat{\beta}_1 = 823.09$  is the estimated relative difference in mean serum BC comparing treated patients with placebo-treated patients with same average baseline BC measures after one month into the Treatment Phase (95% CI: [598.18, 1048.00] ug/mL). There is strong evidence that BC supplementation does marginally impact serum BC levels. But, our estimate for  $\beta_4 + \beta_3$  is insignificant for this model; we don’t have strong evidence that suggests different rates of serum BC between treated and placebo-treated patients.

However, when we eliminate the observations from Patient 57, our inferences change. Our estimate for  $\beta_1$  increases to 1055.52 and the robust standard error decreases so that our new 95% CI is [991.29, 1119.74] ug/mL. This gives us the same conclusion as the previous model, except with more precision and magnitude. The big difference is in our assessment of the trajectories.  $\widehat{\beta}_4 + \widehat{\beta}_3$  is now significant. We estimate that 18.00 ug/mL is the relative change in serum BC for every 1 month increase in time among treated patients who start the trial with similar baseline serum BC (95% CI: [4.46, 31.55] ug/mL). In other words, the estimated trajectory of serum BC among treated patients is 11.84 ug/mL higher than that of placebo-treated patients (95% CI: [1.75, 21.94] ug/mL). Relatively speaking, this is only a slight difference in slope within a six-month period. It's also very important to remember that our sample for placebo-treated patients ( $N_P = 9$ ) is very small compared with the treated group ( $N_T = 35$ ). So, we must be cautious with our inferences here in general.

The next question of interest is if the impact of BC supplementation on serum BC levels over time is dose-dependent, and if so, does the rate at which patients return to baseline after ceasing supplementation also differ by dosage levels. Recall our mean model fitted:  $E[Y_{ij}] = \beta_0 + \beta_1(dose_i/15) + \beta_2mc.base_i + \beta_3t_{ij} + \beta_4t_{ij}(dose_i/15)$ . We will answer both questions with the same model; however, using data from different study phases. For the first part of the question, we focused on the Treatment Phase; for the second part of the question, we focused on the Post-Treatment Phase ( $mc.base_i$  will be the mean-centered average of the serum BC levels in the Treatment Phase for the  $i^{th}$  patient). In the post-treatment phase, the EDA eluded to an autoregressive-1 working correlation structure. So, unlike the previous models, we fitted the GEE post-treatment model using an AR-1 covariance structure. Our residuals plot shows slight funneling as fitted values increase, so we again used robust variance estimates (ref: Appendix A).

If  $\beta_1 = 0$  and  $\beta_4 = 0$ , then all treated patients ( $N_T = 35$ ) would have the same model regardless of their dosage level. Again, here we are only focusing on the treated patients, so Patient 57 is of no concern. Our mean plots in Figure 2 shows that dosage level may be influential during the treatment, but the dosage level may not be influential in the decrease to baseline of serum BC levels post-treatment. The post-treatment trajectories seem to stay parallel with one another with highest dose above and lowest dose below.

Table 2: Output from GEE Dose-Dependent Model

	Estimate	Robust S.E.
Intercept	916.2169386	162.196535
Dose/15	137.1023010	57.954428
Mc.Base	0.9501305	0.290889
Month	54.6013635	26.451417
(Dose/15):Month	-14.6417539	10.271601

Table 2 shows the output from our model, which will answer our first question about dose-dependency.  $\widehat{\beta}_1 + \widehat{\beta}_4 = 122.46$  ug/mL is the estimated relative difference in the trajectory of serum BC comparing two subpopulations of patients who differ in dosage by 15 ug/mL with similar baseline serum BC (95% CI: [54.23, 190.69] ug/mL). The higher the dosage that a patient receives, the larger their increase of serum BC over time. We have strong evidence to reject our null hypothesis. Our results support our observations from the EDA in Figure 2.

Table 3: Output from GEE Dose-Dependent Model Post-Treatment

	Estimate	Robust S.E.
Intercept	1003.886830	143.5083531
Dose/15	137.071552	63.8799162
Mc.Base	1.879735	0.5323811
Month	-153.371669	28.4277506
(Dose/15):Month	-11.377829	11.9275631

Table 3 shows the output from our model, which will answer our second question about dose-dependency. Our suspicions were supported by the results.  $\widehat{\beta}_4 = 11.38$  is statistically insignificant, which tells us that all

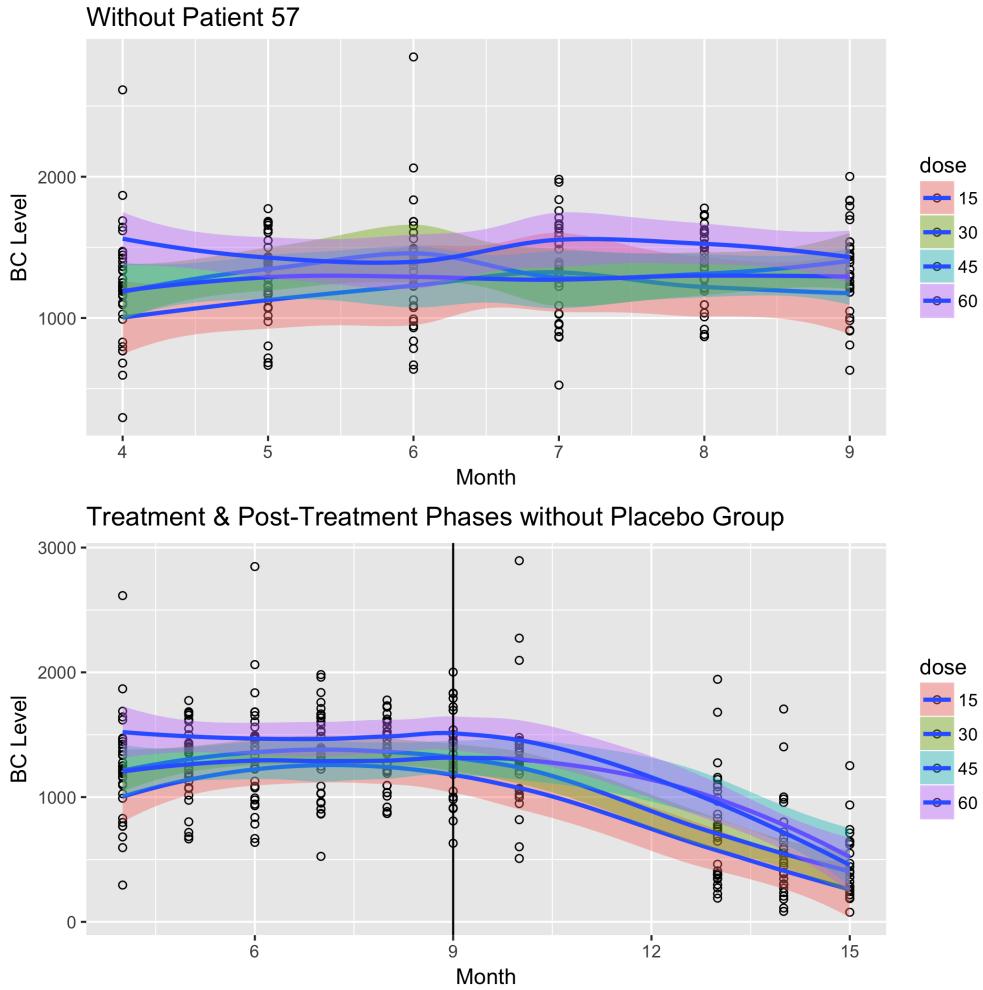


Figure 2: Mean Plots Comparing Various Doseage Levels

patients have the same estimated rate of decrease in serum BC levels post-treatment.

## Secondary Analysis

First, we did extensive EDA (Appendix A: scatterplots, mean plots, and spaghetti plots) on the other time-invariant covariates (age, gender, body mass index (BMI), and cholesterol). Here are some of the things we noticed. BMI seems to have a negative association with average serum BC levels. Female patients seemed to have a mean serum BC level consistent above the male patients in all phases of the study. Then, we did EDA on age, BMI, and cholesterol as categorical and continuous covariates. Looking at them as categories, we saw slight mean serum BC level differences. Lower cholesterol and older ages seemed associated with lower serum BC levels, whereas the opposite stayed true with BMI.

After fitting one model for gender and two models per all other covariates (age, BMI, and cholesterol), one assuming continuous type of variable and another for categorical, we obtained the results laid out in Tables 4 and 5.

Our results support males and females having different intercepts, but similar trajectories of BC levels over time during treatment. We estimate that males have 148.53 ug/mL lower mean difference in serum BC levels over time during the treatment than females with similar baseline BC measures (95% CI [-288.94,-8.11]).

There was not strong enough evidence to reject the null for age and BMI; age and BMI do not seem to effect the impact of BC supplementation on serum BC levels over time. It is interesting to note that when BMI is treated as a categorical variable parsed out using quantiles, individuals with BMI between 25.35 and 27.63  $kg/m^2$  have a statistically significant lower mean serum BC levels than individuals with BMI less than 23.06  $kg/m^2$ . More exploration with the association between BMI and serum BC levels should be investigated with larger sample sizes.

For cholesterol, our conclusions for both models, treating cholesterol as a categorical variable and then as a continuous variable, are similar. Both support a difference in intercepts, but not a difference in trajectories. There is strong evidence for a positive correlation between cholesterol and serum BC levels. We estimate that the mean difference in BC levels over time during the treatment comparing two subpopulations that differ in cholesterol by 50 mg/dL with similar baseline BC measures is 133.52 ug/mL (95% CI [68.61, 198.43]).

Table 4: Model Outputs of the Effects of Various Covariates on Impact of BC Supplementation on Serum BC Levels over Time

	Male Est.	Male Rob. SE	Chol. Est.	Chol. Rob. SE	BMI Est.	BMI Rob. SE	Age Est.	Age Rob. SE
Intercept	1319.057	72.171	949.708	368.961	1721.073	581.254	1617.619	721.577
Mc.Base	0.656	0.453	0.981	0.335	0.668	0.374	0.943	0.369
Month	34.103	28.118	-68.925	141.259	75.651	193.150	-237.161	253.280
X	-92.801	151.912	73.827	86.790	-17.556	22.065	-6.042	12.968
X:Month	-34.566	43.426	19.681	32.197	-2.266	7.303	4.523	4.541

Table 5: Model Outputs of the Effects of Various Categorical Covariates on Impact of BC Supplementation on Serum BC Levels over Time

	Chol. Est.	Chol. Robust SE	BMI Est.	BMI Robust SE	Age Est.	Age Robust SE
Intercept	1052.88	55.88	1406.41	157.92	1378.39	69.95
Mc.Base	1.01	0.36	0.52	0.36	1.25	0.40
Month	39.47	32.75	18.00	63.54	-0.63	28.38
X1	389.31	124.13	-102.94	184.26	-121.80	187.39
X2	452.79	118.35	-275.42	199.56	-305.30	144.62
X3	45.73	142.69	-175.37	187.44	52.95	146.36
X1:Month	-66.04	56.26	-0.02	66.62	-17.22	62.58
X2:Month	-58.19	42.86	35.97	77.11	68.58	52.17
X3:Month	49.42	53.68	-26.98	73.29	17.33	54.45

Another big secondary question of interest is whether BC supplementation has an effect on serum vitamin E levels during and after treatment. Figure 3 shows us that serum vitamin E Levels appear to be similar to serum BC levels in trends during the Treatment Phase. However, there seems to be a substantial decrease below the mean levels of the placebo group. This is definitely different than how serum BC levels behave post-treatment. Figure 3 also contains a spaghetti plot that shows a slight increase in vitamin E levels in most subjects after month 14. There is not enough data here to explore than phenomena further. If we were to continue follow up visits with the treated patients after month 15, maybe there trends move back towards the normal levels of vitamin E demonstrated by the placebo group.

Table 6: Model Outputs of the Effects BC Supplementation on Serum Vitamin E Levels over Time

	Tx Phase Estimates	Tx Phase Robust SE	Post-Tx Phase Estimates	Post-Tx Phase Robust SE
Intercept	9.0472056	0.2679017	5.2559163	1.5541936
Tx	0.5280594	0.3272453	4.5254350	1.6937215
Mc.Base	0.5325729	0.0825343	0.6770584	0.1031741
Month	-0.1313394	0.0488460	0.1794820	0.1371150
Tx:Month	0.0171029	0.0562328	-0.4730793	0.1464734

The variogram (Appendix A) for serum vitamin E levels eluded to an exchangeable covariance structure. Here we fit two models: one for Treatment phase and one for Post-Treatment phase. Table 6 contains our

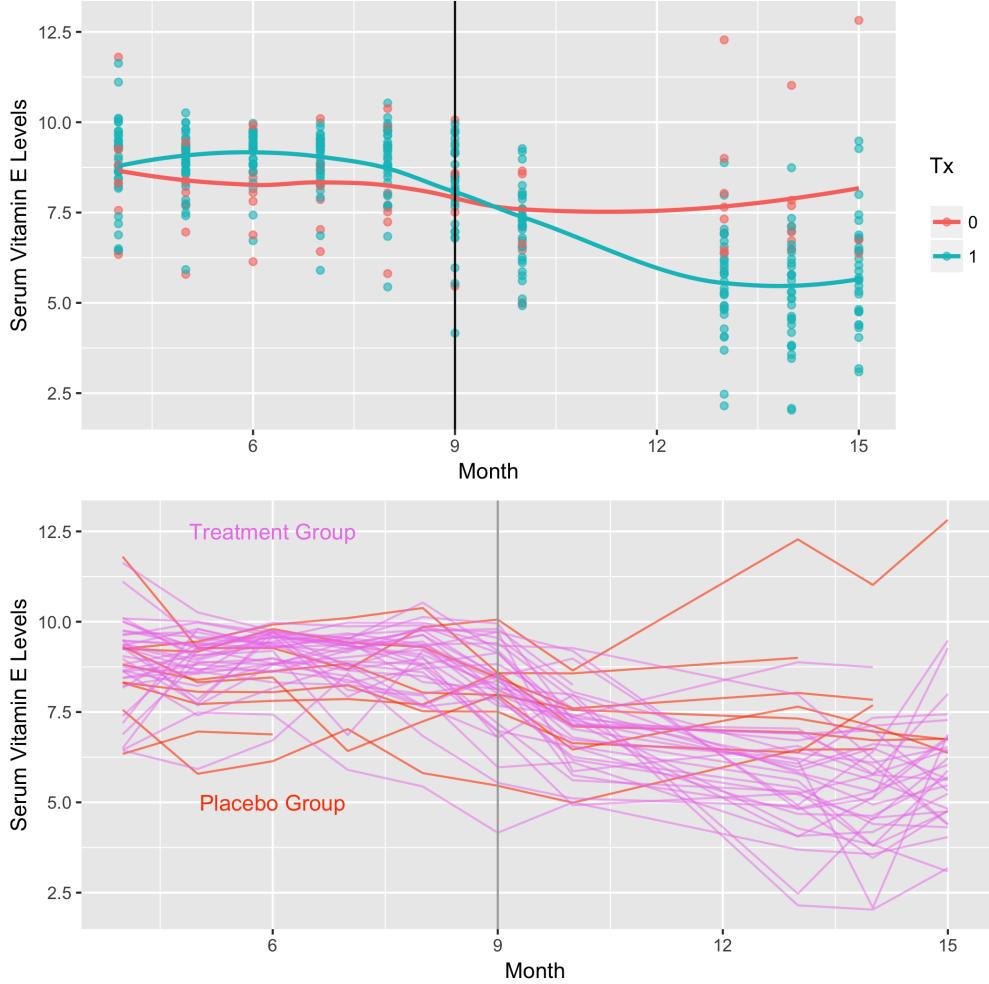


Figure 3: Mean and Spaghetti Plots of Serum Vitamin E Levels Adjusted for Treatment

results from both models. Our results agree with what we saw in the EDA. We estimate the mean difference in serum vitamin E levels between treated patients and placebo-treated patients is 0.528 ug/mL (95% CI: [0.201, 0.855]) during the Treatment phase and 4.53 ug/mL (95% CI: [2.83, 6.22]) during the Post-Treatment phase. There is no statistically significant difference in trajectory during the Treatment Phase. However, in the Post-Treatment phase, we estimate that difference in rate of serum vitamin E levels decreases by 0.473 ug/mL for every 1 month increase among treated patients compared to placebo-treated patients with similar baseline serum vitamin E level pre-treatment (95% CI: [-0.620, -0.327] ug/mL).

Lastly, we created a predictive model. Recall our mean linear mixed effects model:  $E[Y_{ij} | \vec{b}_i] = \beta_0 + \beta_{C1}Tx_i + \beta_{C2}Male_i + \beta_{C3}Chol_i + \beta_{C4}base_i + \beta_{L1}month_{ij} + \beta_{L2}Tx_i \cdot month_{ij}$ . Trace, density, and autocorrelation plots were used to confirm convergence. Table 7 gives us means and standard deviations from the posterior distributions of each regression parameter and those for the standard deviations of the random effects. Predictions are shown in Figure 4; two patients from each dosage group was randomly selected and their serum BC levels were predicted if they were to continue supplementation for another three months past month 9.

Table 7: Jags Output from Predictive Model

	Mean	SD
Tx	77.334573	33.5128863
Male	-10.979870	60.2720893
Cholesterol	4.586649	0.6889218
Mc.Base	0.954882	0.3840056
Tx:Month	5.867379	4.9191405
Intercept	88.118153	106.4042588
sd(b0)	228.374352	43.1015306
sd(b1)	15.207155	3.7458627

## Discussion

It is important to note some limitations with our data. With all the models we fit, we never once used all 46 patients. The age distribution of these 46 volunteers ranged from 50 to 65 years. All our models and inference can only be applied to that type of population. Other potential confounding factors like diet and environmental exposures may also contribute to our results. There was some missing data for Patient 31 and outliers associated with Patient 57.

We propose that there is still more investigation that needs to be done. Why were there gaps in the list of patient id's? What happened to those other patients/volunteers? These serum BC level trends definitely differed by gender; females had consistently higher mean serum BC levels than males. Other confounding variables that might explain this discrepancy needs to be investigated further. How can we explain the drastic dip above marginal average in serum vitamin E levels immediately after ceasing BC supplementation and the sharp increase around month 14. Lastly, BMI is highly correlated with serum BC levels and seems to effect these levels over time during Treatment phase. In general, even though model results confirmed most of our suspicions drawn from extensive EDA, there is clearly still much to explore beyond our limited data set and small time frames.

## Appendix A

Table 8: Patient 31 Missing Value Replaced with Average (968.4 ug/mL)

Obs.	Patient ID	Month	BC	Dosage	Tx	Baseline BC
372	31	1	595	45	1	93.25
373	31	2	1126	45	1	93.25
374	31	3	NA	45	1	93.25
375	31	4	1027	45	1	93.25
376	31	5	1093	45	1	93.25
377	31	6	1001	45	1	93.25

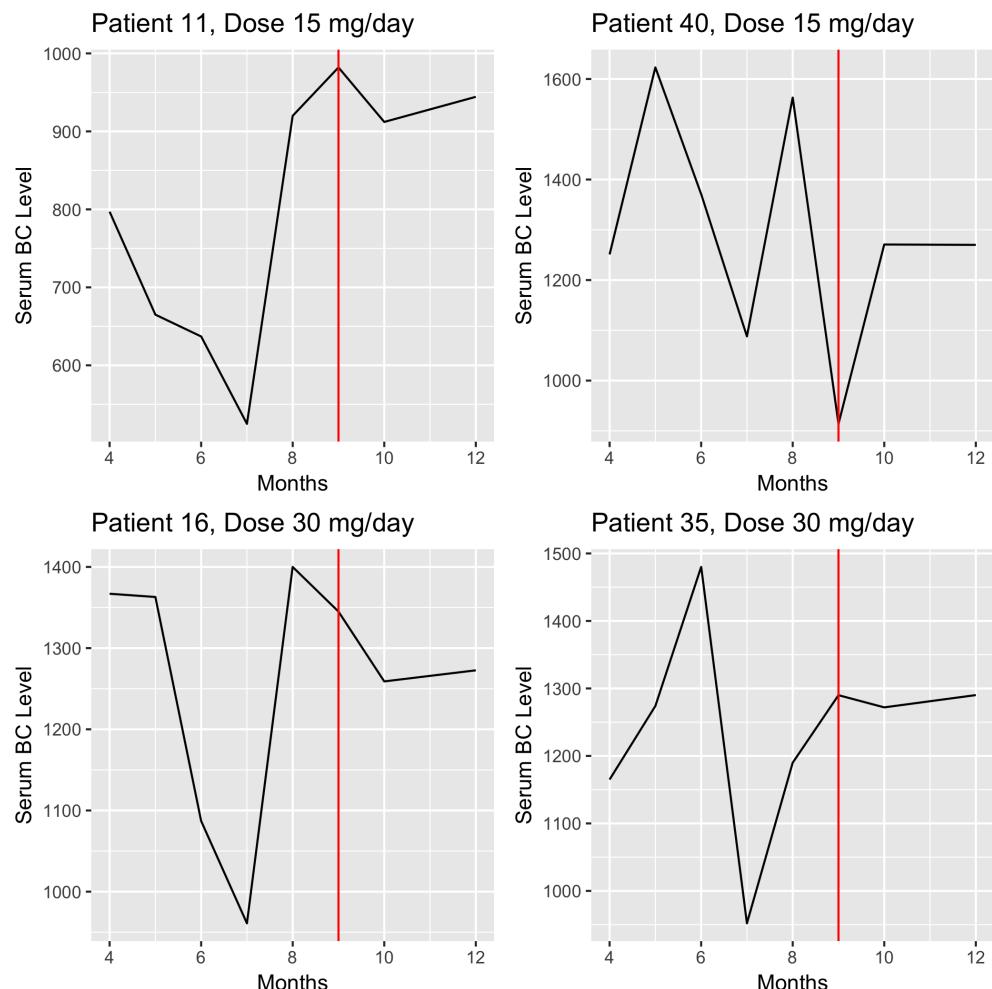


Figure 4: Prediction Plots for Two Randomly Chosen Patients in Each Dose Level

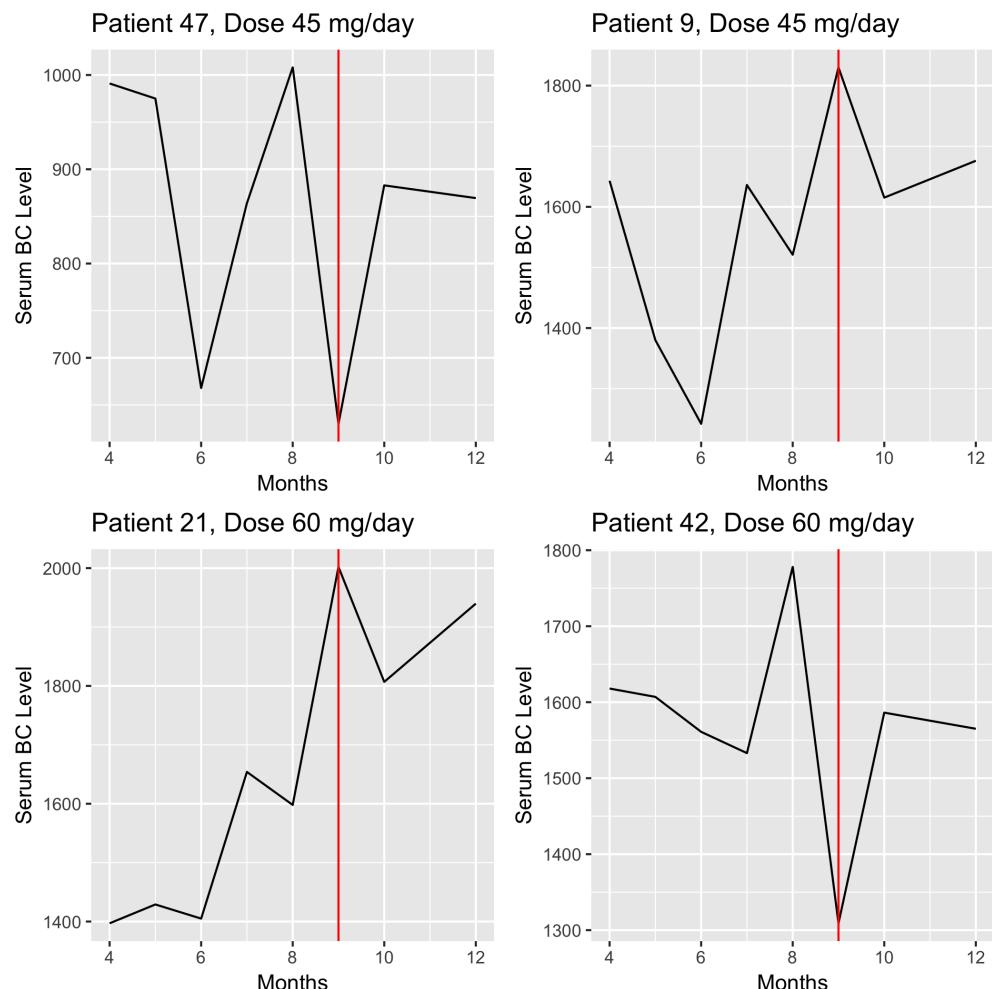


Figure 5: Prediction Plots for Two Randomly Chosen Patients in Each Dose Level

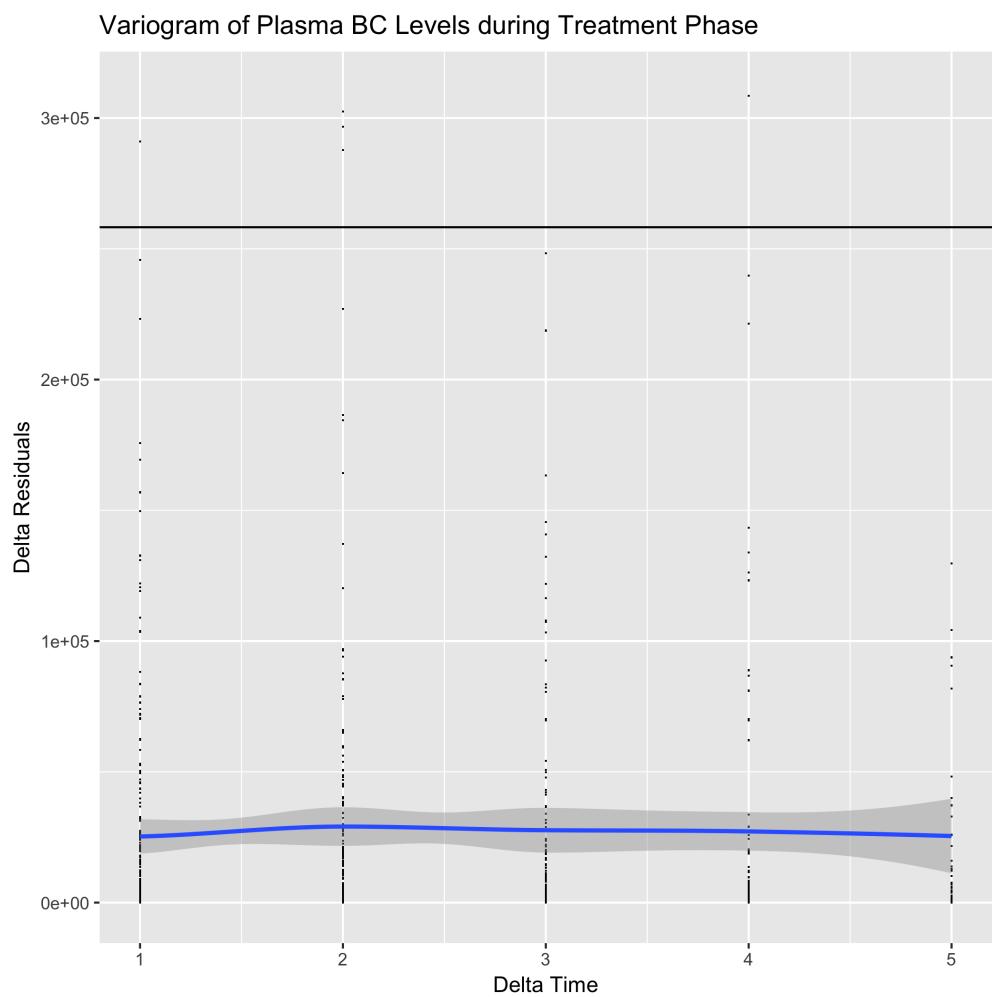


Figure 6: Variograms

Variogram of Plasma BC Levels during Post-Treatment Phase

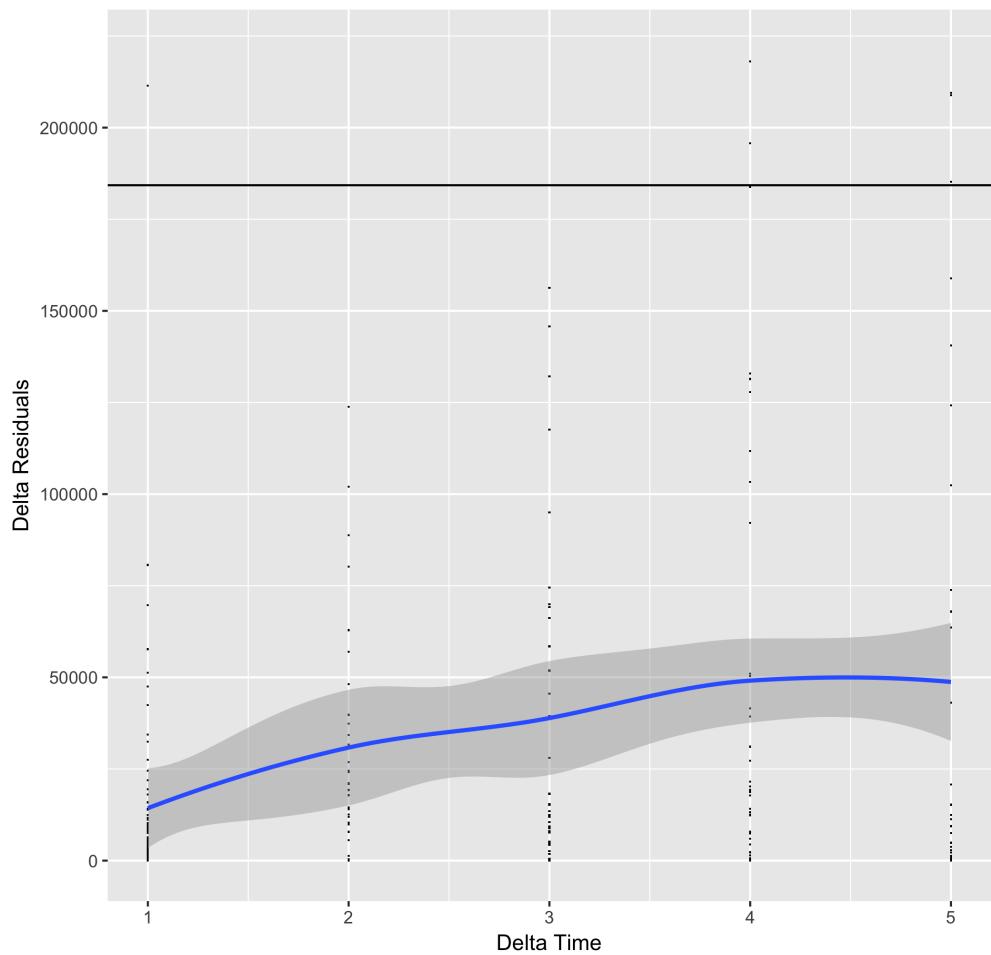


Figure 7: Variograms

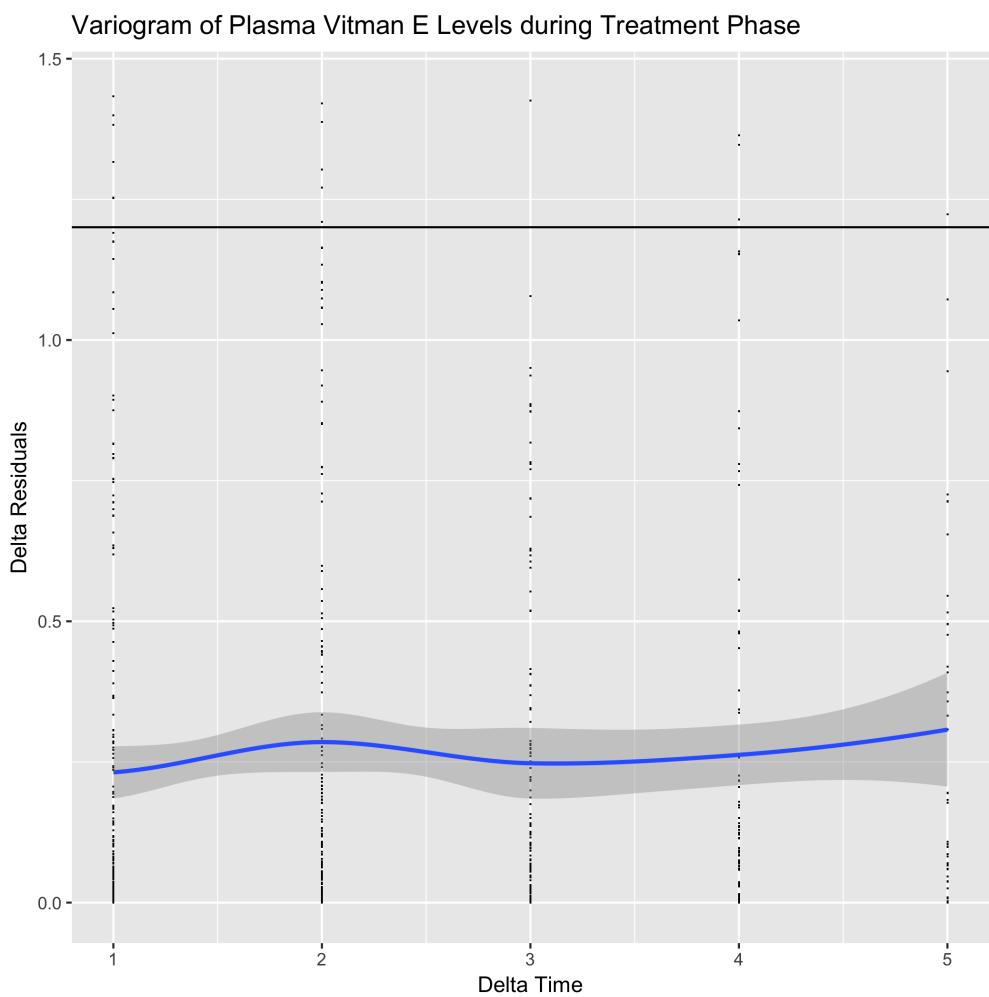


Figure 8: Variograms

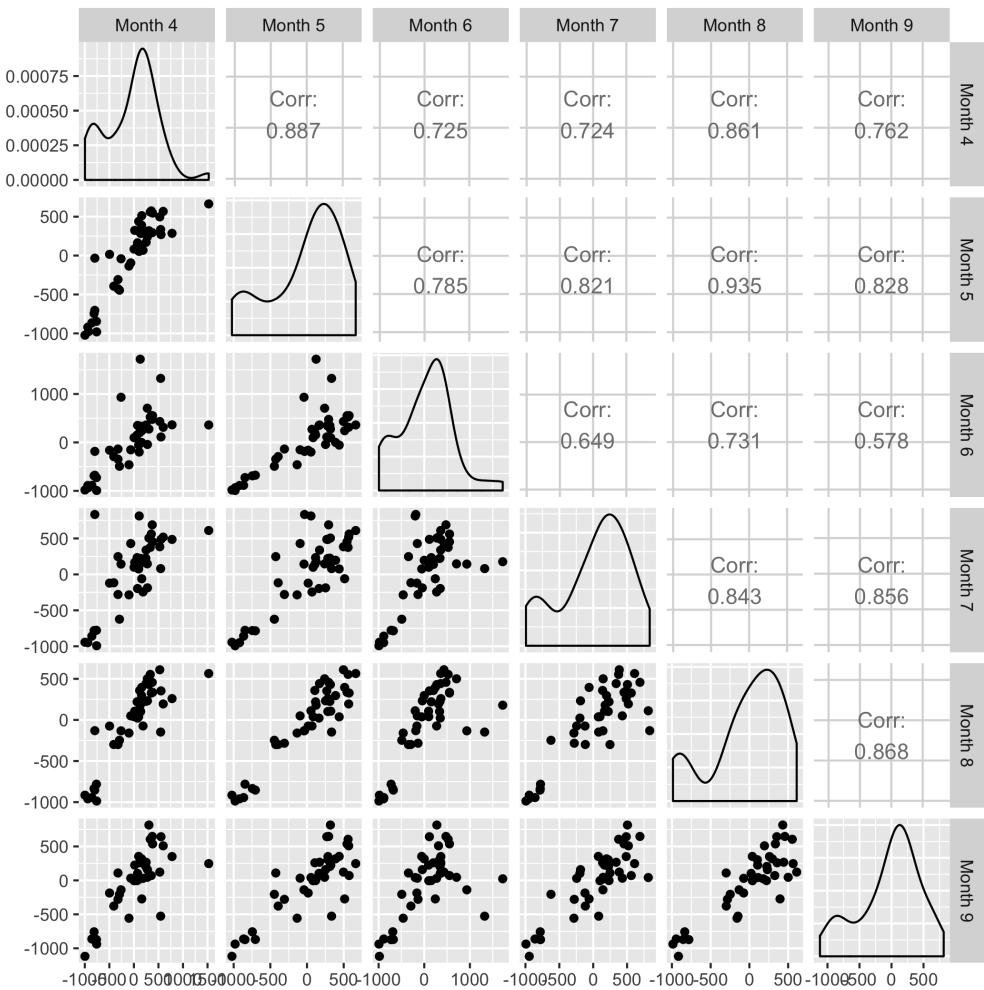


Figure 9: Empirical Correlation Matrix during Treatment Phase

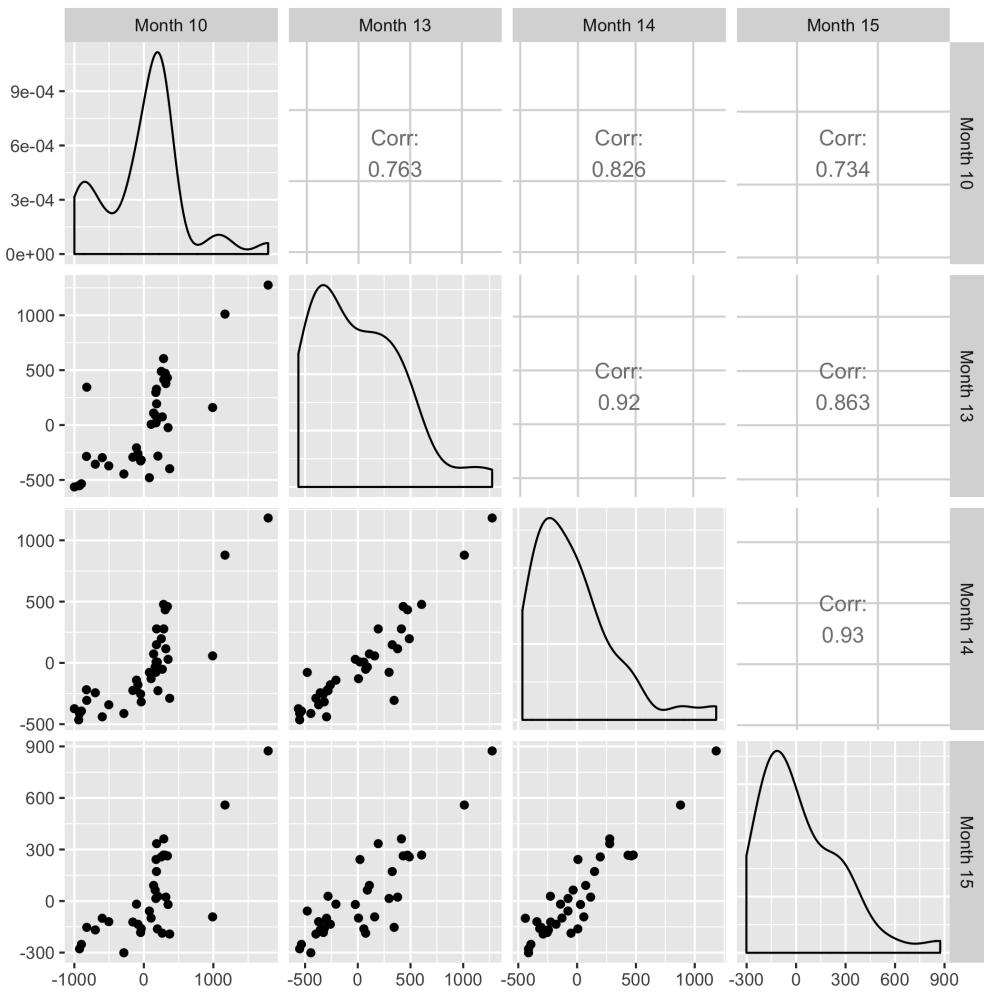


Figure 10: Empirical Correlation Matrix during Post-Treatment Phase for Patients Having Received Treatment

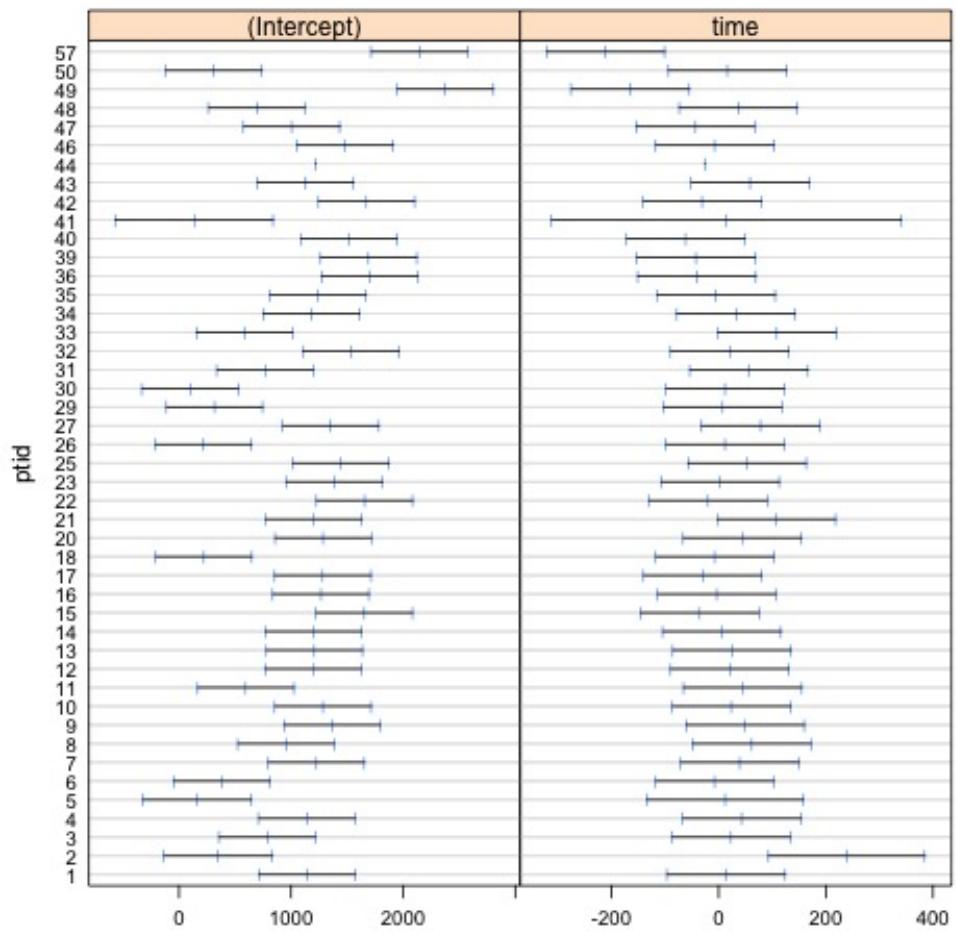


Figure 11: Subject-Specific Random Effects During Treatment Phase

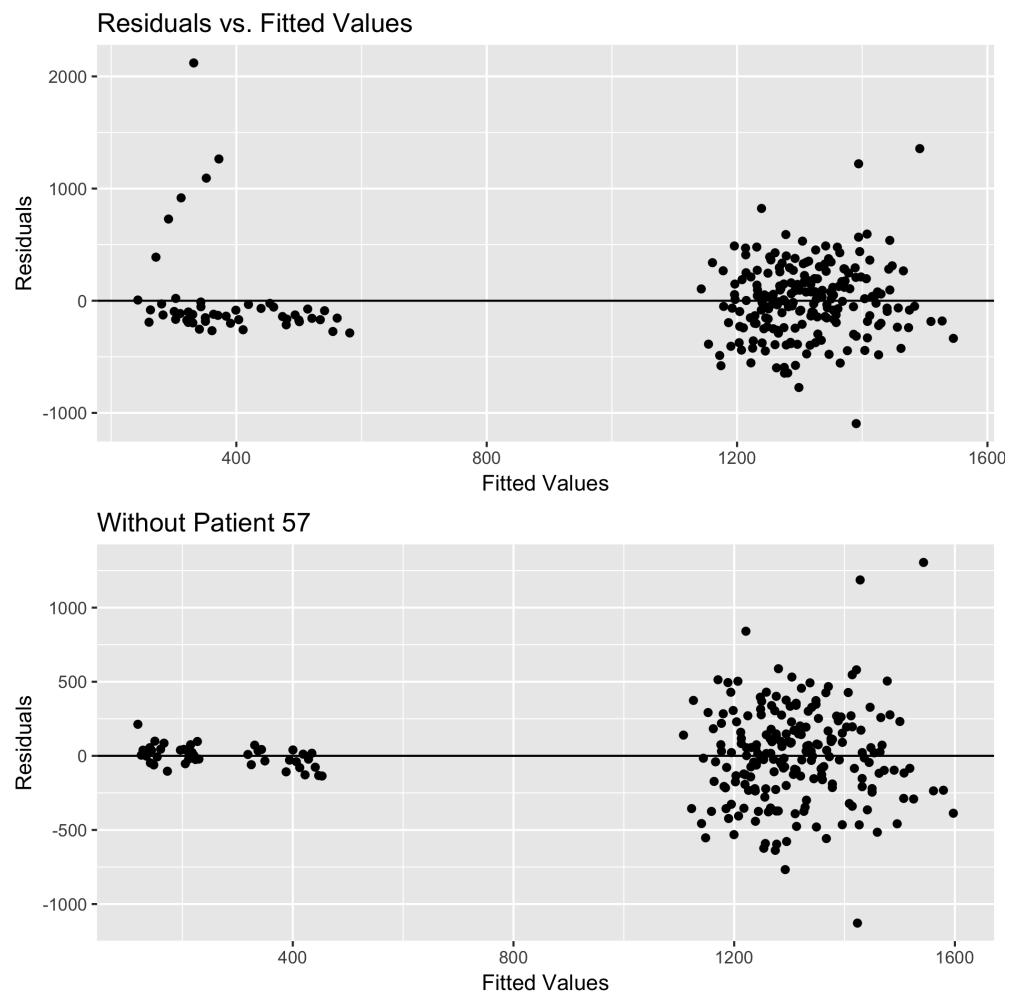


Figure 12: Residual Plots

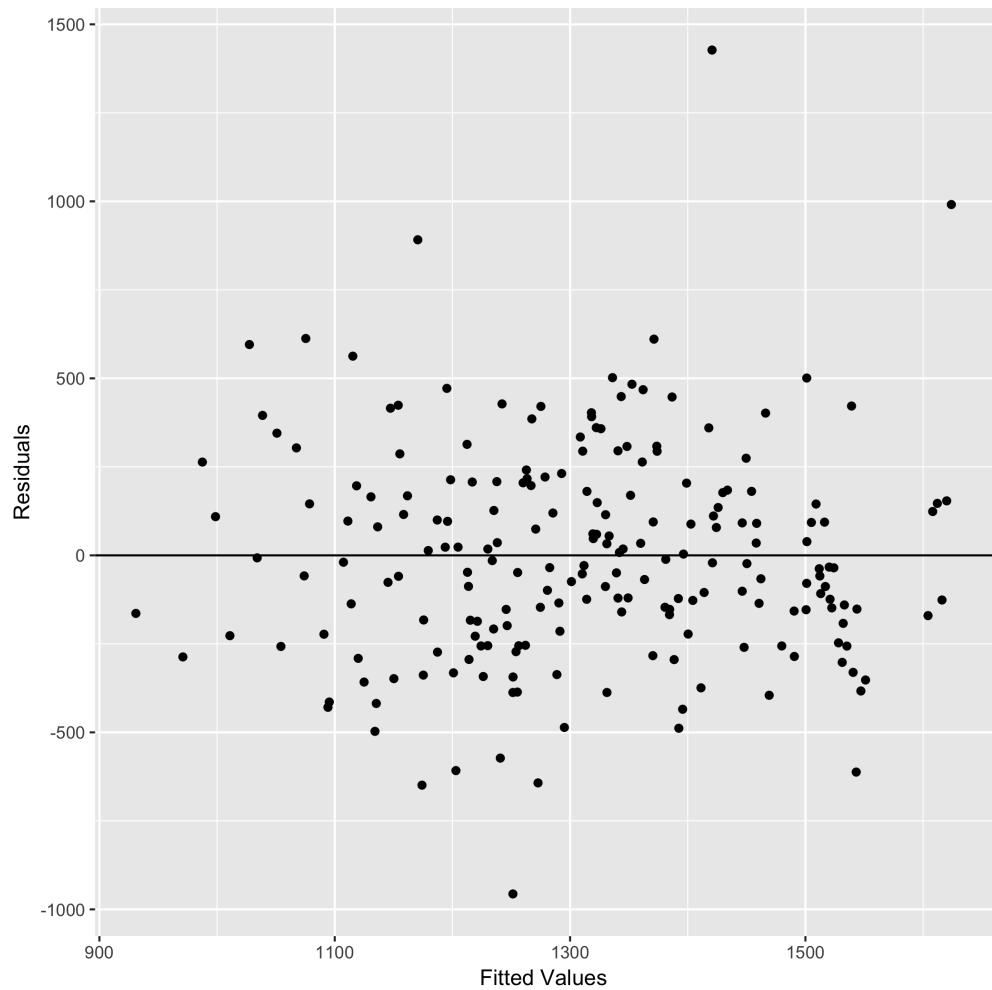


Figure 13: Residual Plots

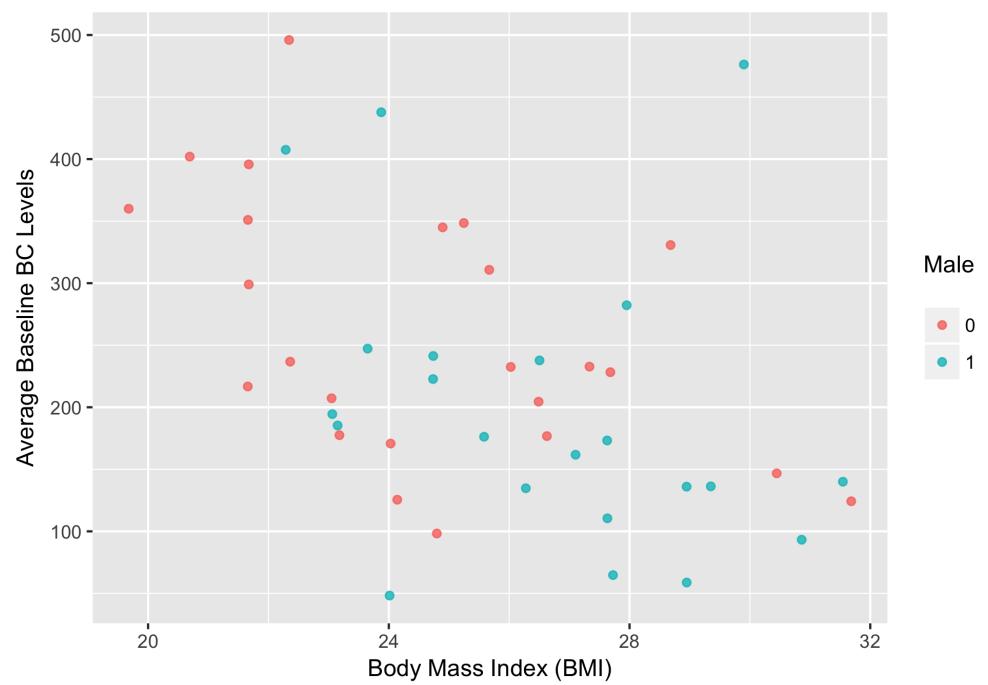


Figure 14: Association Between BMI and Average Serum BC Levels

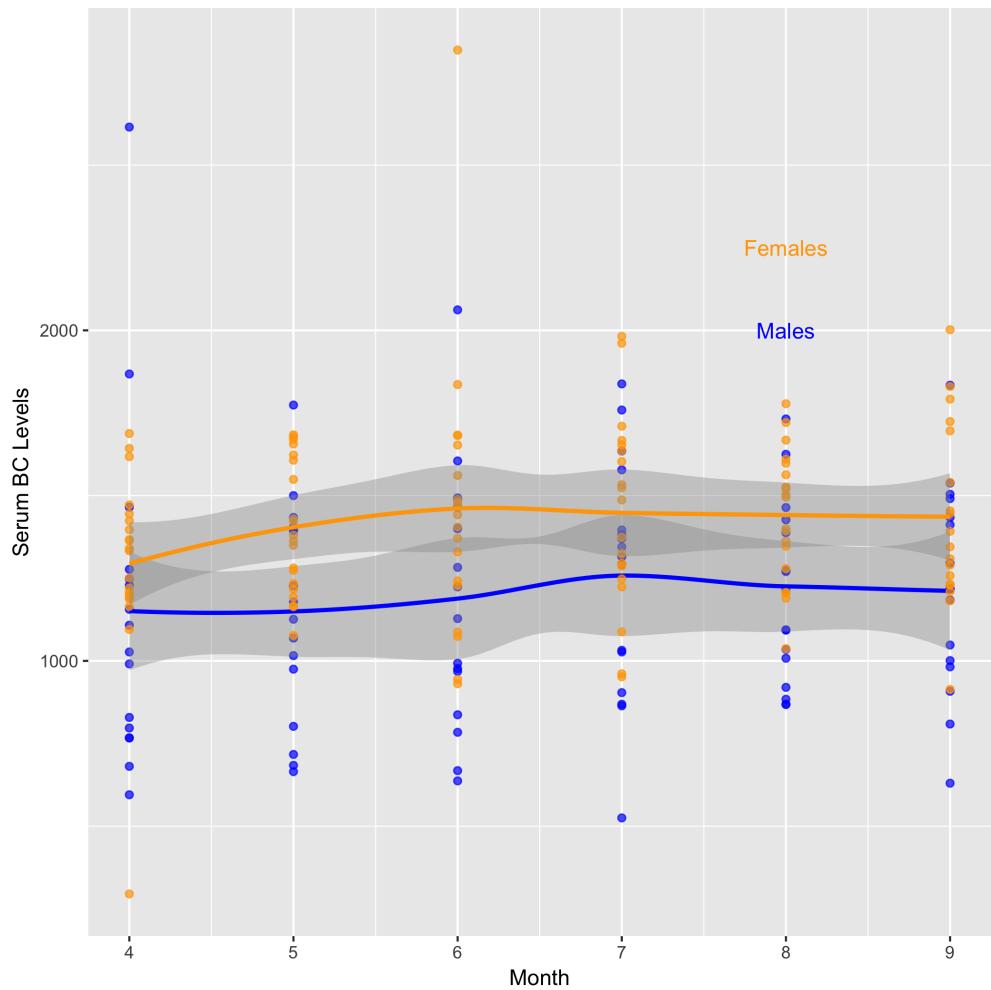


Figure 15: Mean Plots by Gender, Cholesterol, BMI, and Age, Respectively

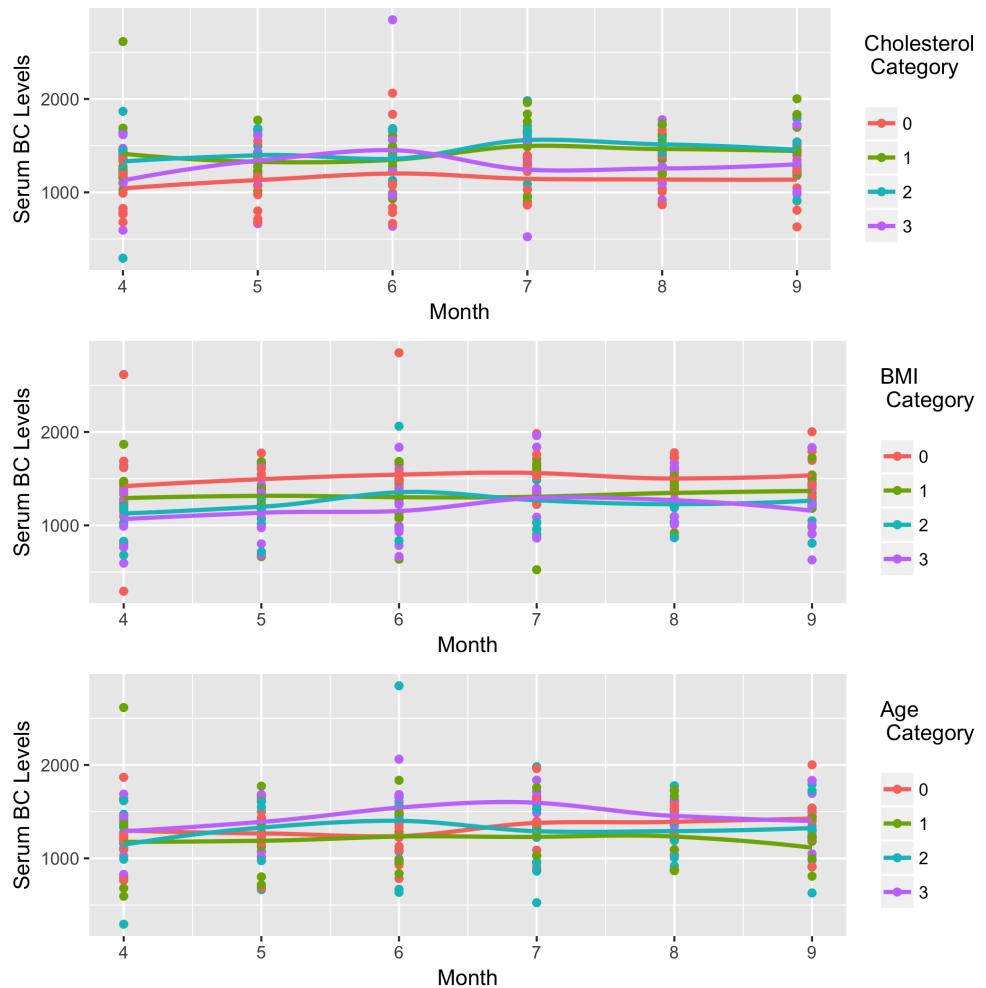


Figure 16: Mean Plots by Gender, Cholesterol, BMI, and Age, Respectively

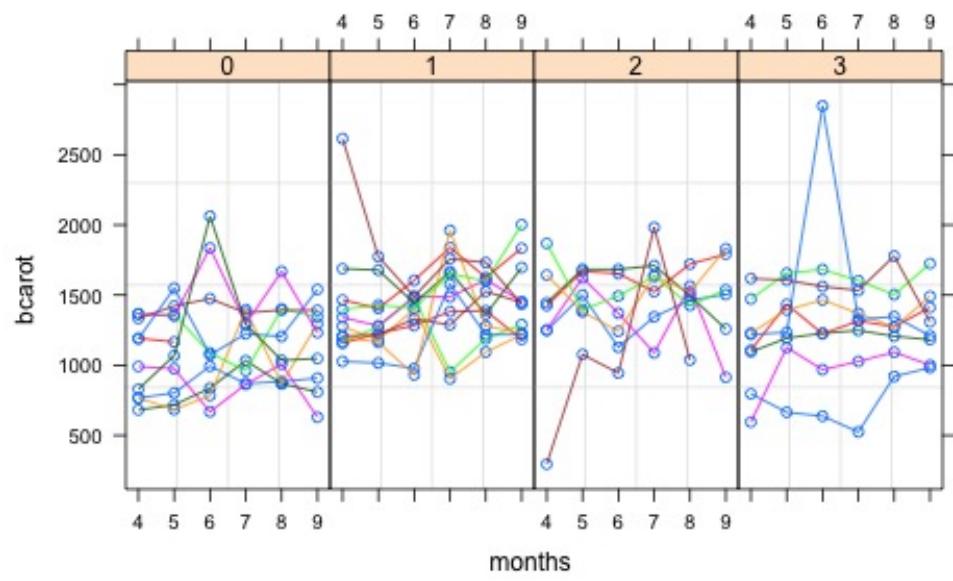


Figure 17: Spaghetti Plots by Cholesterol, BMI, and Age Categories, Respectively

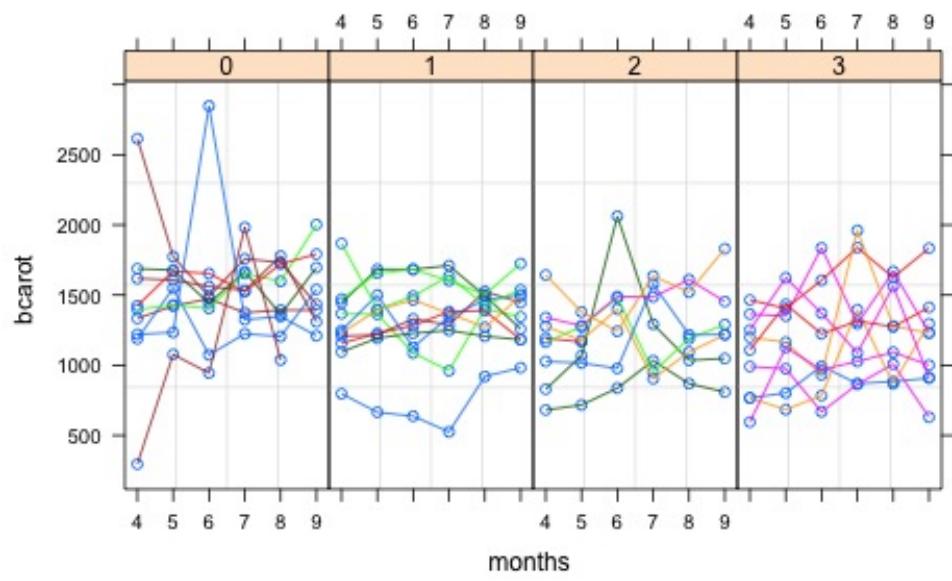


Figure 18: Spaghetti Plots by Cholesterol, BMI, and Age Categories, Respectively

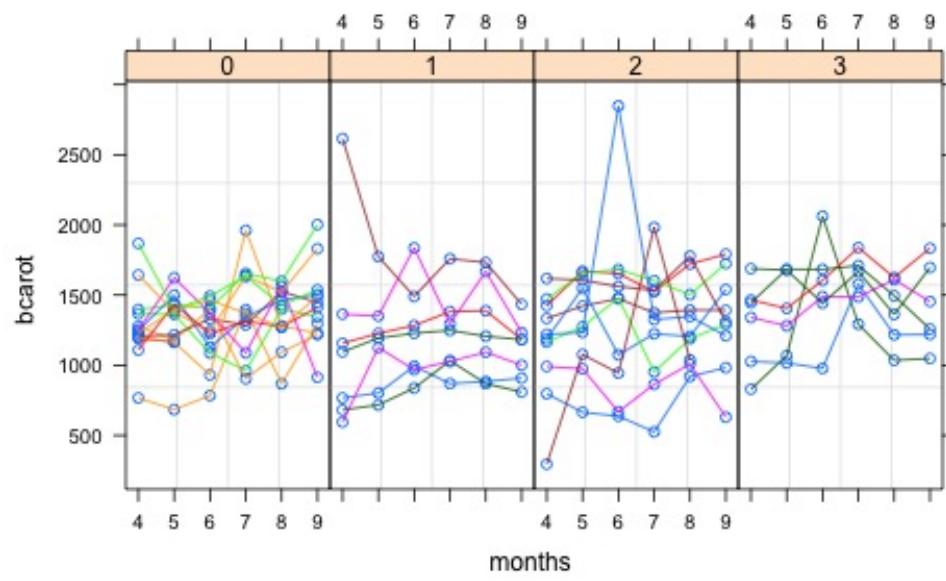


Figure 19: Spaghetti Plots by Cholesterol, BMI, and Age Categories, Respectively