Bayesian Data Analysis Project

Sina Mokhtar smokhtar@unm.edu, Dan Kowalczyk danrk@unm.edu, Imuseoluwa Obembe iobembe@unm

1. Abstract

2. Introduction

a) Background

Beta-carotene is a natural pigment found in plants, giving them their orange colors. It is a provitamin A, meaning the body uses it to make vitamin A, which is essential for maintaining healthy vision, skin, and immune function. However, high doses over a long time can lead to a condition called carotenemia, resulting in a yellowish-orange tint to the skin.

Pharmacokinetic studies, also known as PK studies, are a type of research in pharmacology to investigate the effect of the body on a drug. One common topic for PK studies is how a drug builds up in bloodstream over time. These type of studies collecting blood, to measure the concentration of the drug in the blood plasma. In this study, researchers investigated the concentration of Beta-carotene in 46 patients who received different doses of the treatment. This data helps researchers to understand how Beta-carotene behaves in the body over time, providing valuable insights for clinical practice.

b) Questions of Interest

The specific aim of this study is to (1) determine how different dose levels of beta-carotene affected the serum beta-carotene levels in blood over time. In addition to measuring the plasma concentrations of beta-carotene by dose, we are also interested in examining (2) whether there is any effect of beta-carotene supplementation on vitamin E levels in the plasma. Since both beta carotene and vitamin E are lipid soluble (they are dissolved in fats rather than water), it might be possible that serum vitamin E levels are correlated with serum beta-carotene levels over time. We are also interested in (3) whether there are interactions between treatment and age, gender, BMI, or cholesterol.

3. Materials and Methods:

(a) Source of Data

In this dataset, there are 46 volunteers who were randomly assigned to receive one of five doses of beta-carotene (0, 15, 30, 45, or 60 mg/day) for up to 15 months in a double-blind manner. Each volunteer's progress was monitored monthly, resulting in a total of 699 observations. The dataset contains 11 variables for each observation: *ptid* (patient ID): a unique identification number assigned to each of the 46 patients, *month*: Indicates the month of the study, with values ranging from 0 to 15. Months 0 to 3 serve as a baseline, and the beta-carotene treatment begins at month 4. Note that some patients have fewer than 15 months of data. *bcarot* (Plasma beta-carotene levels): the concentration of beta-carotene in the patient's blood, measured in micrograms per milliliter. *vite* (Plasma vitamin E levels): The concentration of vitamin E in the patient's blood, measured in micrograms per milliliter. *dose* (Dose of beta-carotene): The amount of beta-carotene administered to the patient daily as part of the treatment. *age*: The age of the patient. *male*: An indicator variable that denotes the patient's gender. *bmi*: (Body Mass Index): A measure of the patient's body weight in relation to their height. *chol*: (Serum cholesterol level): The level of cholesterol in the patient's blood, measured in milligrams per deciliter *cauc* (Area under curve for serum beta-carotene): The average level of

serum beta-carotene over the months 4 and onwards. *vauc* (Area under curve for serum vitamin E): The average level of serum vitamin E over the months 4 and onwards.

Patient number 46 showed an unusual increase in beta-carotene levels at month 4, reaching 2452 ($\mu g/ml$), which is significantly higher than the average baseline level of 297.37 ($\mu g/ml$). Although the levels returned to normal after month 10, this behavior led us to remove this patient from the study. Patient number 40 was excluded from the study as he only participated in the baseline period, without taking any treatment. Patient number 24, initially categorized in the placebo group with a dose of 0, unexpectedly received a dose of 30 only at month 0. Considering the preceding patient (ptid 23) was assigned to the dose 30 group, this inconsistency seems to be a typographical error therefore, we corrected the dose to 0. For some patients, we have two sets of data at the same month regarding their beta-carotene. We have decided to retain both values for our analysis. Patient 26 which is in placebo group (zero dose) shows an unusual behavior in their beta-carotene level during the 13th month, recording a value of 1014. This value stood out as it was higher than both the preceding month's value of 114 and the next month's value of 218. We have decided to retain this data point in our analysis. Additionally, we identified and removed six missing values from the dataset which seems randomly distributed through the data.

It is important to note that we tried to estimate the missing value by using the Area Under the Curve (AUC) calculations, for the *cauc* and *vauc* variable, available in the dataset. However, due to the existence of multiple methods and formulas for AUC calculation, we could not find any approach that resulted in an AUC value consistent with our dataset. Therefore, further investigation about the origin of data (AUC formula) is necessary for this analysis.

The distribution of vitamin E levels appears to be right-skewed, with a mean of 7.6 and a range between (2.03, 12.82). This observation aligns with expectations, considering that the normal range for vitamin E in adults is between 5.5 and 17 ($\mu g/ml$).

(b) Statistical Methods:

We attempt to build a normal linear model of the data utilizing the Bayesian methodology. Each serum beta carotene observation is modeled as a normal random variable with a mean that depends on a linear combination of the covariates of age, sex, BMI, and cholesterol. The parameters of interest are the shared precision, the beta coefficients for the covariates, and the patient specific mean adjustment.

We specify priors on each of these parameters according to the plausible range of values they may take on. In all cases, the parameters may take on both negative and nonnegative values, so the normal distribution is a reasonable choice of prior.

The complexity of the data led us to simulate the posterior distribution since an analytic derivation did not seem to be reasonable. For these purposes, we chose to simulate the model using the JAGS framework through the `rjags` package in `R`.

4. Results

(a) Descriptive Statistics:

To explore the primary aim of the study to examine how treatment affects serum beta carotene levels, we plotted a line graph of serum beta carotene levels against months for each patient. This information is also presented in a time series of box blots faceted by the dosage level. For the control of zero dosage level, there is no change in beta carotene level with time. For the other positive dosage levels, serum beta carotene is close to zero, and then during the treatment period the serum values increase. It appears that the change from baseline serum levels is greater for larger dosage values on average.

Across the positive dosages treatment levels and while patients are on treatment after month 3, there appears to be constant serum levels until month 10, after which a spike occurs during months 11 and 12, which is then followed b a sharp decrease in months 13, 14, and 15.

To examine the secondary aim of the study, we also plotted a line graph of vitamin E levels over time, grouped

by treatment. In this graph, vitamin E levels appear to be flat during the placebo period during months 0-3, then there appears to be a slight increase until month 8 when the levels decrease (corresponding to just before the serum beta carotene levels spike in month 11). During months 12-15, levels appear to be constant, however they are lower than the baseline serum levels during the placebo period.

Because the beta-carotene levels shows an increase during the first month of treatment (month 4 in original data), followed by a decrease after the ninth month. Therefore, using a quadratic linear regression model seems to be a suitable approach for fitting this data.

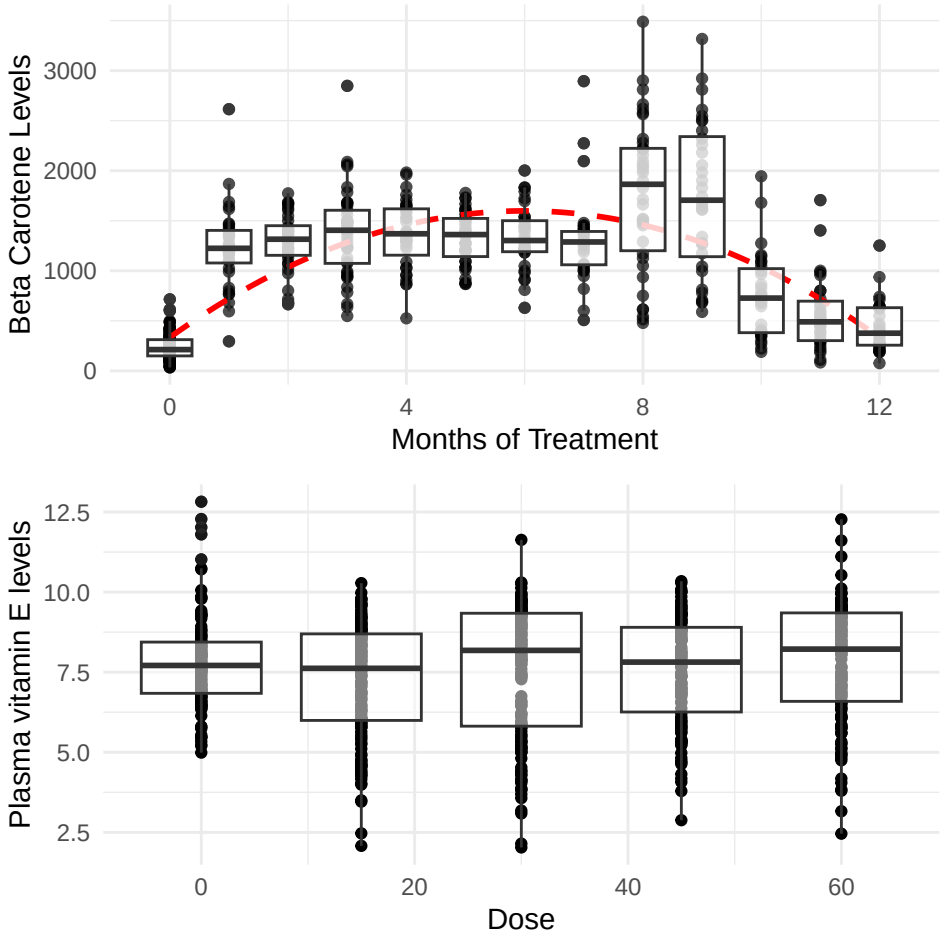


Figure 1: Fitted quadratic regression for Beta Carotene Levels over time following treatment initiation (top). Boxplot displays the spread of Plasma Vitamin E levels across different doses of beta-carotene. (bottom)

(b) Models / Model Building

The specific aim of this study is to determine how different dose levels of beta-carotene affected the serum beta-carotene levels in blood over time. for this part we fit the following model:

$$\begin{aligned}\text{BetaCarotine}^{(i)} = \beta_0 + \beta_1(\text{dose}^{(i)}) + \beta_2(\text{month}^{(i)}) + \beta_3(\text{month}^{2(i)}) + \beta_4(\text{age}^{(i)}) \\ +\beta_5 I(\text{male}^{(i)}) + \beta_6(\text{bmi}^{(i)}) + \beta_7(\text{cholestrol}^{(i)}) + \epsilon^{(i)}\end{aligned} \tag{1}$$

for the above model we have run the Gibbs sampling and we got the following distribution for each parameters.

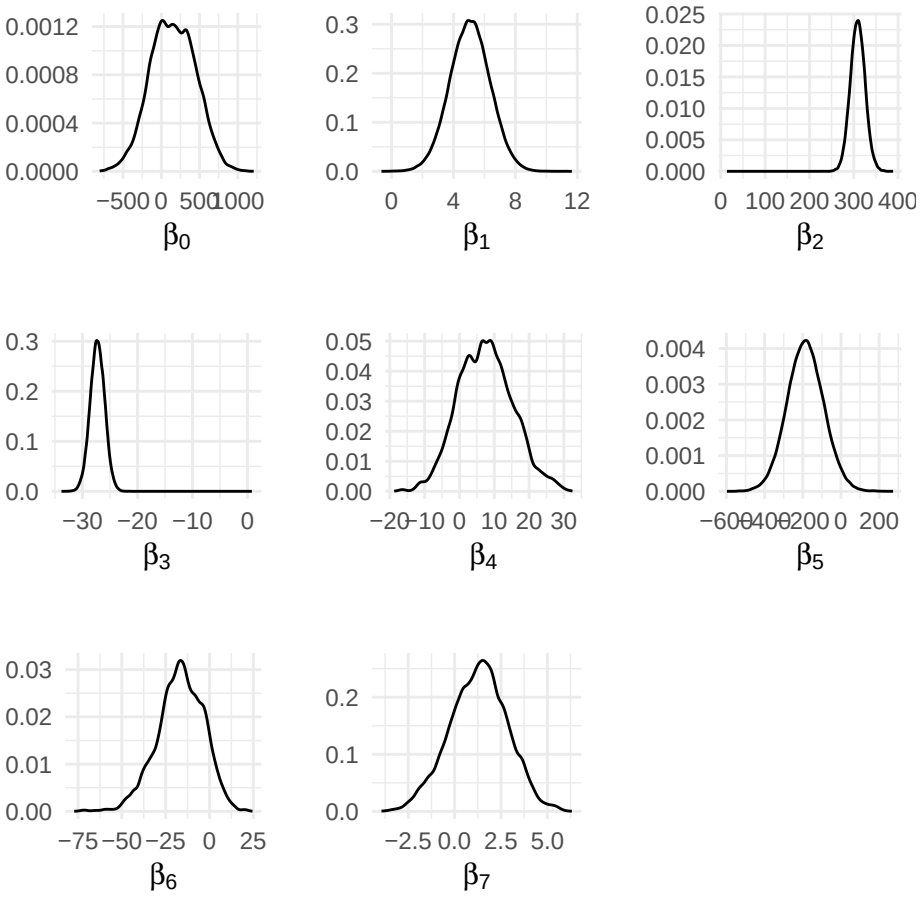


Figure 2: Posterior distribution for each parameters for model 1

Table 1: Posterior estimates and each parameter for model 1

	Mean	SD	2.5%	97.5%
beta[0]	154.144034	292.580172	-412.463994	711.921307
beta[1]	5.065613	1.296921	2.521285	7.606627
beta[2]	308.873734	17.010104	276.442569	342.203036
beta[3]	-27.272690	1.360936	-29.914803	-24.674903
beta[4]	7.502088	7.811046	-7.028346	23.681183
beta[5]	-185.994825	96.072486	-372.765604	4.779030
beta[6]	-16.611470	13.250111	-44.338085	7.038155
beta[7]	1.284489	1.553547	-1.873160	4.217780

In conclusion, the positive mean of β_2 for the month variable and the negative mean ofβ_3 for the month^2 variable suggest that, on average, we expect an initial increase in beta-carotene levels during the early months of treatment, followed by a decrease towards the end of the treatment period. Additionally, since zero is within the 95 percent confidence interval for $\beta_4, \beta_5, \beta_6$ and β_7 it indicates that the treatment effect on serum beta-carotene does not significantly differ based on age, gender, BMI, or cholesterol levels (Question 3).

to answer if there is any effect of beta-carotene supplementation on vitamin E levels in the plasma we use simple linear regresion model:

$$\text{vitaminE}^{(i)} = \beta_0 + \beta_1(\text{dose}^{(i)}) + \epsilon^{(i)} \tag{2}$$

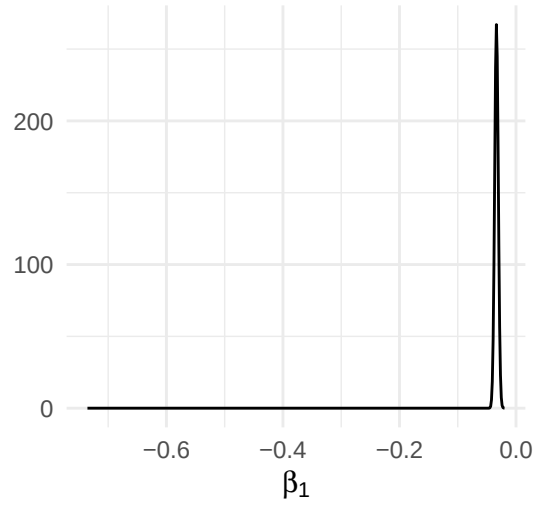


Figure 3: Posterior distribution for each parameters for model 2

Table 2: Posterior estimates and each parameter for model 2

	Mean	SD	2.5%	97.5%
beta[0]	7.6631451	3.2758204	0.9038189	8.5066924
beta[1]	-0.0335535	0.0039957	-0.0393702	-0.0276941

In conclusion, the negative slope of β_1 suggests that with an increase in dosage, we anticipate lower vitamin E levels, which is consistent with our initial expectations.

(c) Model-Checking:

Due to the time constraints of the project, we did not develop multiple models. We therefore did not go through model selection and fit the model that we could make work.

Additionally, we did not decide on criteria for model fit ahead of time, so we did not compute numerical measures of lack of fit. We do see from the graph that the general trend is captured by the quadratic model, however this functional form is not able to represent the step and spike features from the discussion of descriptive statistics.

Lastly, we did consider simulating data from the model to visually compare if our model produced data similar to the sample, but we prioritized exploration of the model.

5. Discussion:

We have found that the analysis supports that an increase in serum beta carotene levels occurs with treatment and that increasing dosage does lead to increasing serum levels. Vitamin E levels tend to decrease with treatment and this is consistent with our expectations as well.

6. Bibliography:

7. Appendix:

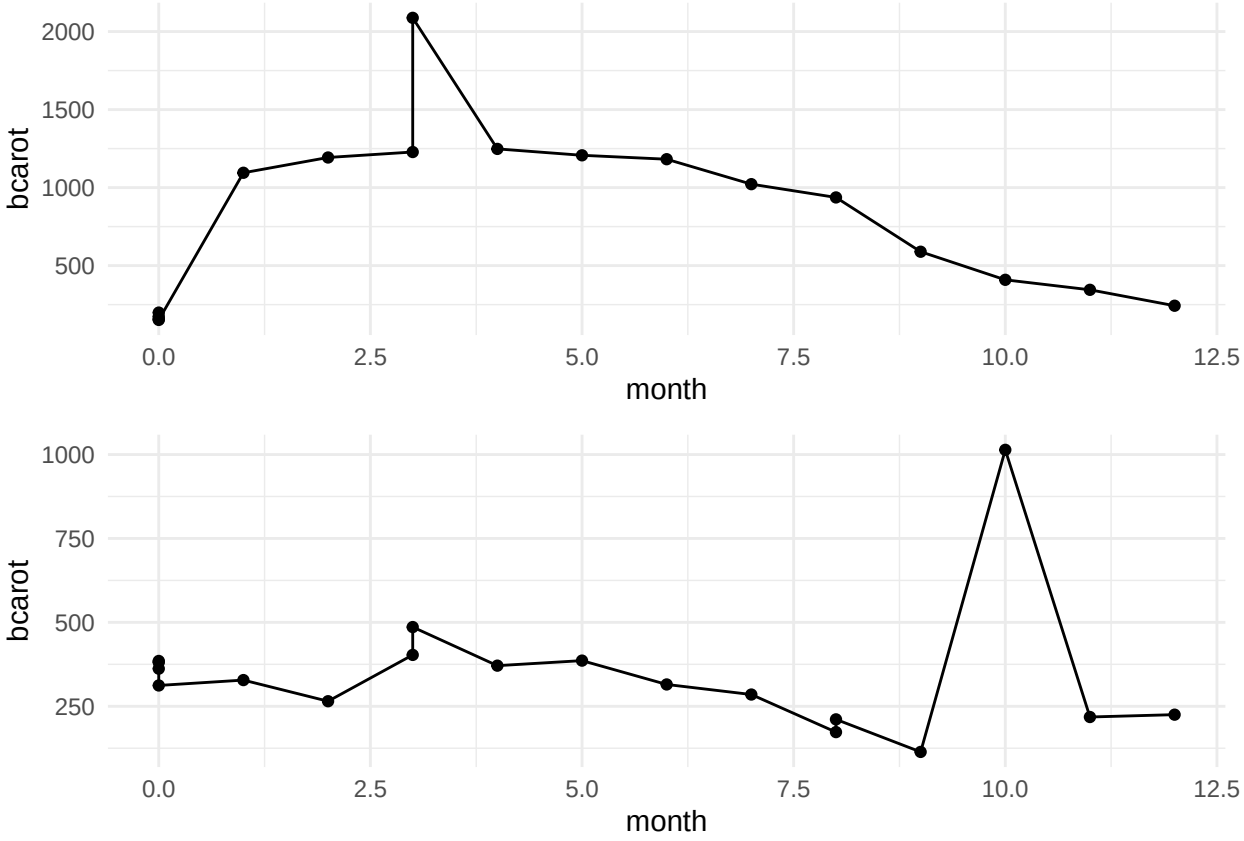
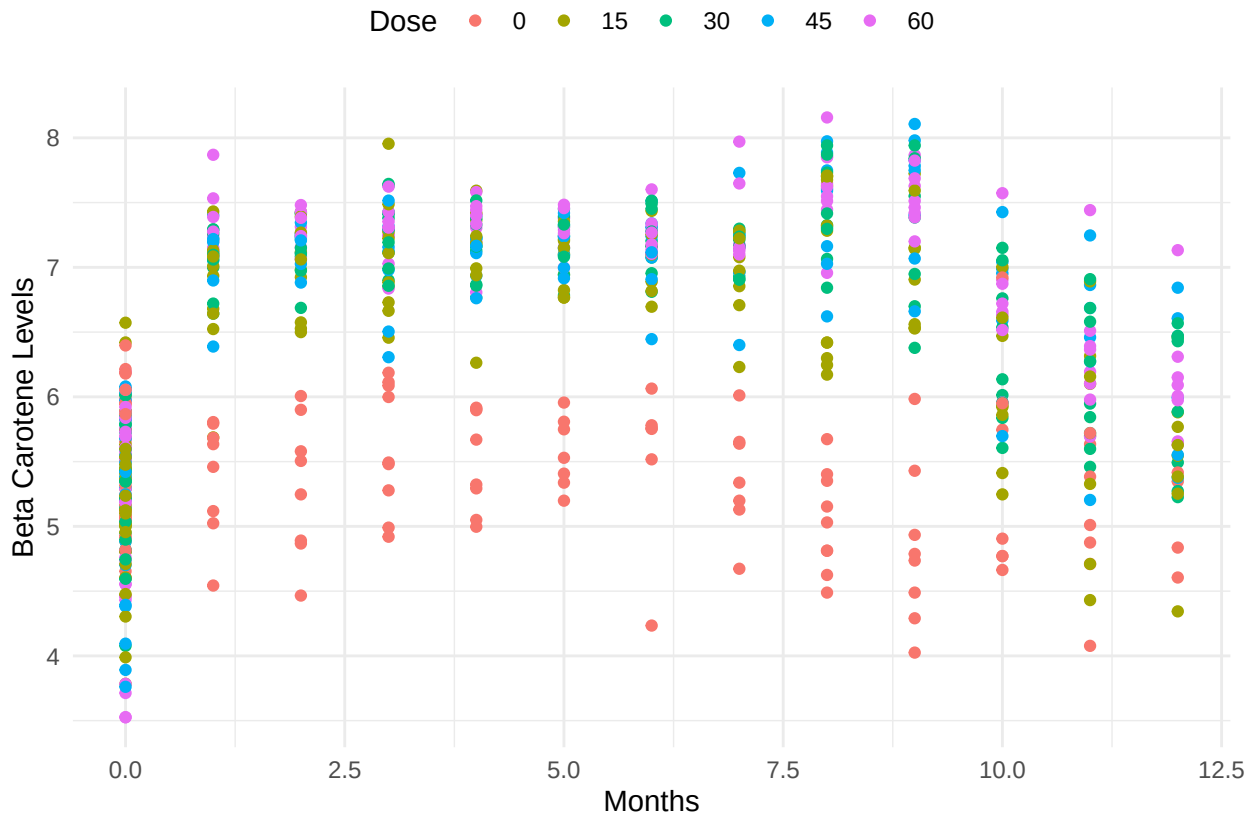
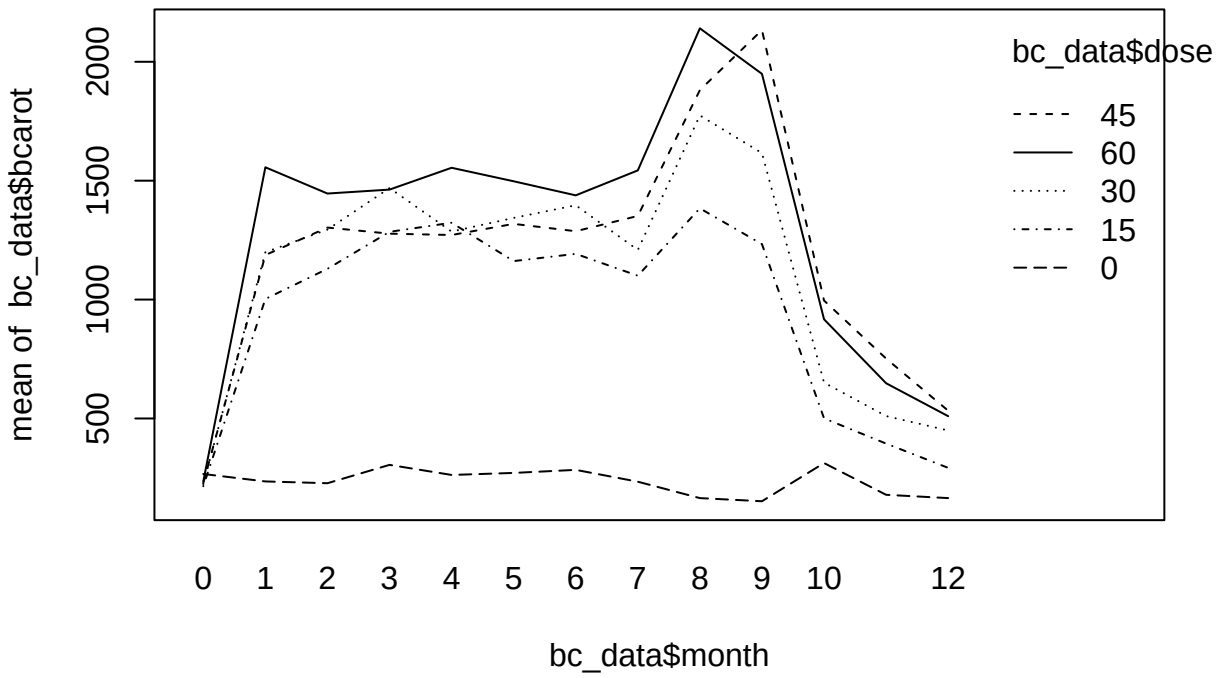


Figure 4: Beta Carotin behavier for patient 1 (top) and patient 26 (bottom) during treatment period

```r
ggplot(data = bc_data,
       aes(x = month,
           y = log(bcarot),
           group = ptid,
           col = as.factor(dose))) +
    geom_point() +
    theme_minimal()+
    labs(x = "Months",
         y = "Beta Carotene Levels",
         col = "Dose")+
    theme(legend.position = "top")
```




```r
interaction.plot(response = bc_data$bcarot , bc_data$month,bc_data$dose)
```



In the field of pharmacokinetics, the area under the curve (AUC) is the definite integral of the concentration of a drug in blood plasma as a function of time (this can be done using liquid chromatography–mass spectrometry[1]). In practice, the drug concentration is measured at certain discrete points in time and the trapezoidal rule is used to estimate AUC. In pharmacology, the area under the plot of plasma concentration of a drug versus time after dosage (called "area under the curve" or AUC) gives insight into the extent of exposure to a drug and its clearance rate from the body. The AUC (from zero to infinity) represents the total drug exposure across time.

In this dataset, there are 46 volunteers who were randomly assigned to receive one of five doses of beta-carotene (0, 15, 30, 45, or 60 mg/day) for up to 16 months in a double-blind manner. Each volunteer's progress was monitored monthly, resulting in a total of 699 observations.

The dataset contains 11 variables for each observation: *ptid* (patient ID): a unique identification number assigned to each of the 46 patients, *month*: Indicates the month of the study, with values ranging from 0 to 15. We interpret this to be categorical where each number represents a whole calendar month so that there are a total of 16 months of observations. The first four months (0 to 3) serve as a baseline, and the beta-carotene treatment begins at the month labelled 4. Note that some patients have fewer than 16 months of data due to dropout. *bcarot* (Plasma beta-carotene levels): the concentration of beta-carotene in the patient's blood, measured in micrograms per milliliter. *vite* (Plasma vitamin E levels): The concentration of vitamin E in the patient's blood, measured in micrograms per milliliter. *dose* (Dose of beta-carotene): The amount of beta-carotene administered to the patient daily as part of the treatment. *age*: The age of the patient. *male*: An indicator variable that denotes the patient's gender. *bmi*: (Body Mass Index): A measure of the patient's body weight in relation to their height. *chol*: (Serum cholesterol level): The level of cholesterol in the patient's blood, measured in milligrams per deciliter *cauc* (Area under curve for serum beta-carotene): The average level of serum beta-carotene over the months 4 and onwards. *vauc* (Area under curve for serum vitamin E): The average level of serum vitamin E over the months 4 and onwards.

Patient number 46 showed an unusual increase in beta-carotene levels at month 4, reaching 2452 ($\mu g/ml$),

which is significantly higher than the average baseline level of 297.37 ($\mu g/ml$). Although the levels returned to normal after month 10, this behavior led us to remove this patient from the study. *Patient number 40* was excluded from the study as he only participated in the baseline period, without taking any treatment. *Patient number 24*, initially categorized in the placebo group with a dose of 0, unexpectedly received a dose of 30 only at month 0. Considering the preceding patient (ptid 23) was assigned to the dose 30 group, this inconsistency seems to be a typographical error therefore, we corrected the dose to 0. For some patients, we have two sets of data at the same month regarding their beta-carotene. We have decided to retain both values for our analysis. *Patient 26* which is in placebo group (zero dose) shows an unusual behavior in their beta-carotene level during the 13th month, recording a value of 1014. This value stood out as it was higher than both the preceding month's value of 114 and the next month's value of 218. We have decided to retain this data point in our analysis.

We identified and removed six missing values from the dataset which seems randomly distributed through the data. We tried to impute the missing value by using the Area Under the Curve (AUC) calculations using the *cauc* and *vauc* variable, however we could not reproduce existing AUC values so did not follow through with this method.

The distribution of vitamin E levels appears to be right-skewed, with a mean of 7.6 and a range between (2.03, 12.82). This observation aligns with expectations, considering that the normal range for vitamin E in adults is between 5.5 and 17 ($\mu g/ml$).