Bayesian Data Analysis Project

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. Therefore, understanding how Beta-carotene moves through the body is crucial.

Pharmacokinetic studies, also known as PK studies, are a type of research in pharmacology to investigate the effect of a drug on body. 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 betacarotene 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. also we are interested (3) whether the effect of treatment on serum beta-carotene differs by 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.

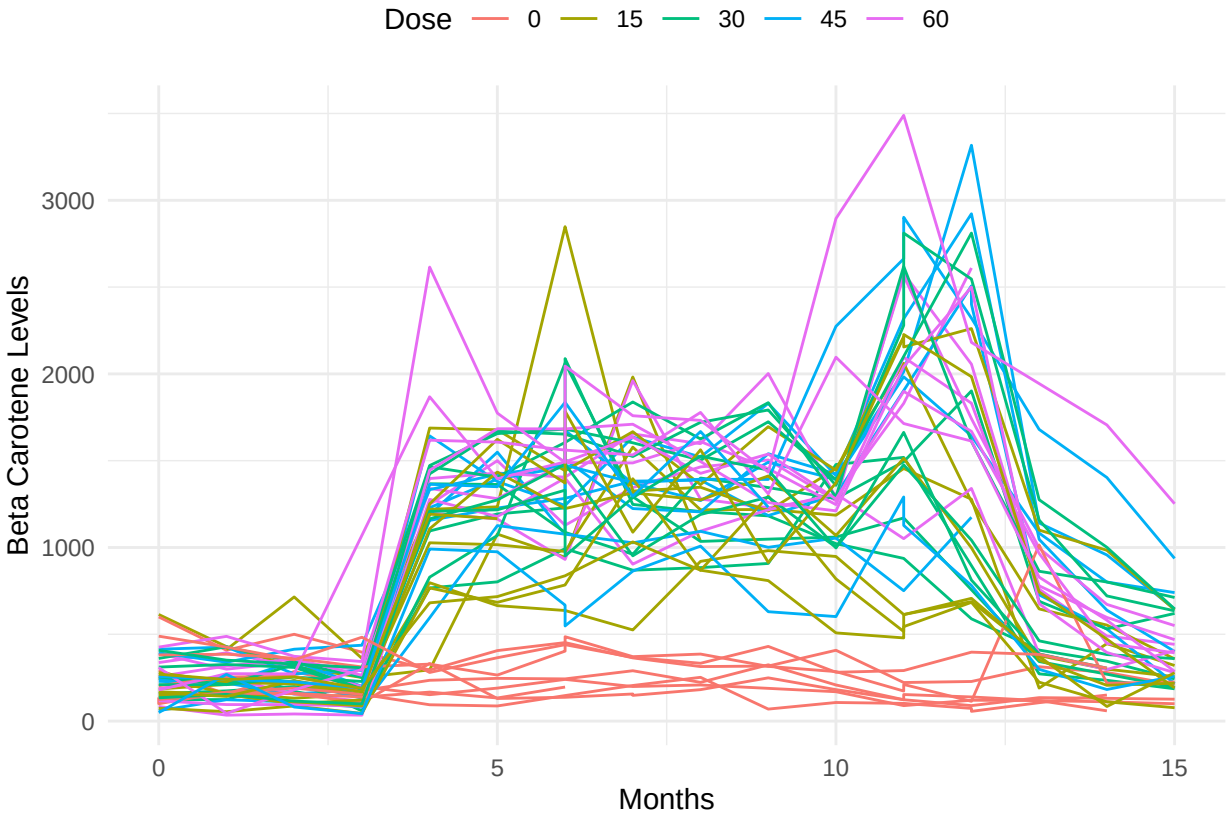
Patient number 46 showed an unusual increase in beta-carotene levels at month 4, reaching 2452 ($\mu g/ml$), which is significantly higher than the average baseline level of 297.37 ($\mu g/ml$). Although the levels returned to normal after month 10, this behavior led us to remove this patient from the 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. Additionally, we identified and removed six missing values from the dataset which seems randomly distributed through the data.

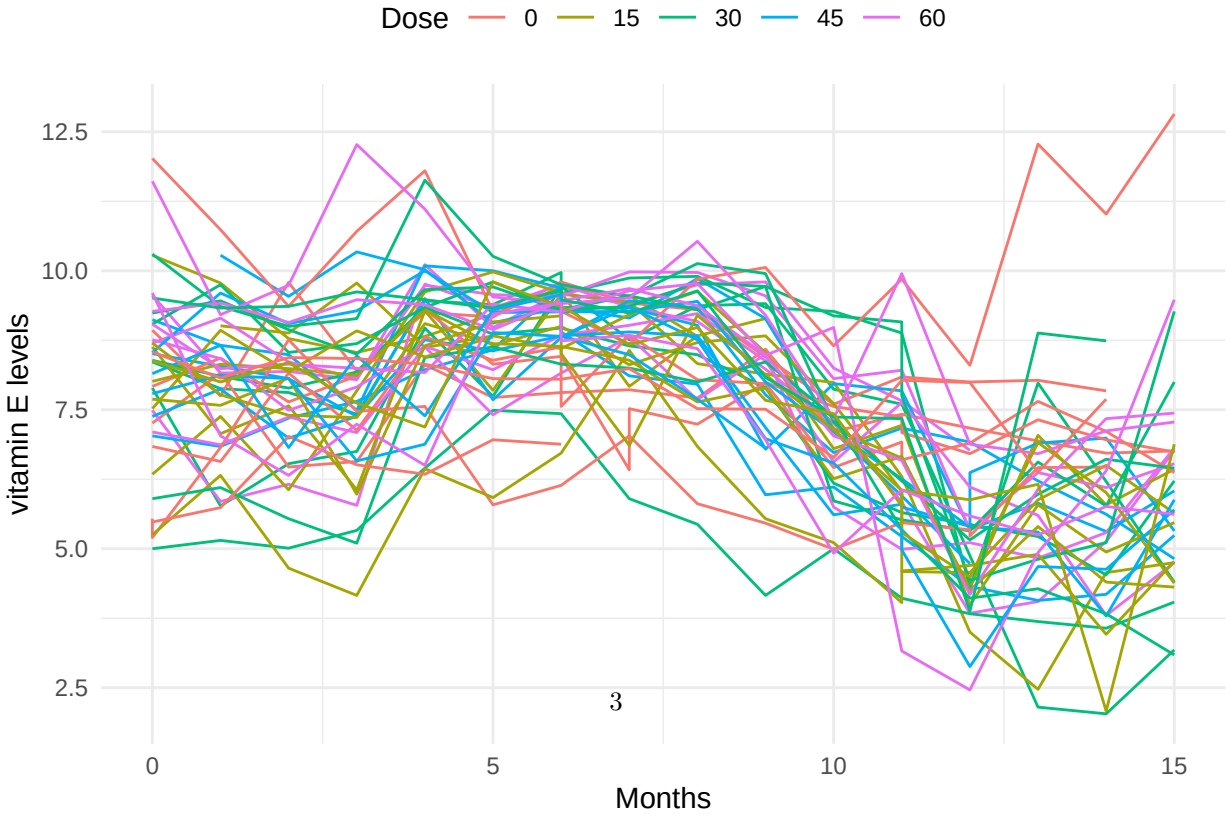
The distribution of vitamin E levels appears to be right-skewed, with a mean of 7.6 and a range between (2.03, 12.82). This observation aligns with expectations, considering that the normal range for vitamin E in adults is between 5.5 and 17 ($\mu g/ml$).

(b) Statistical Methods:

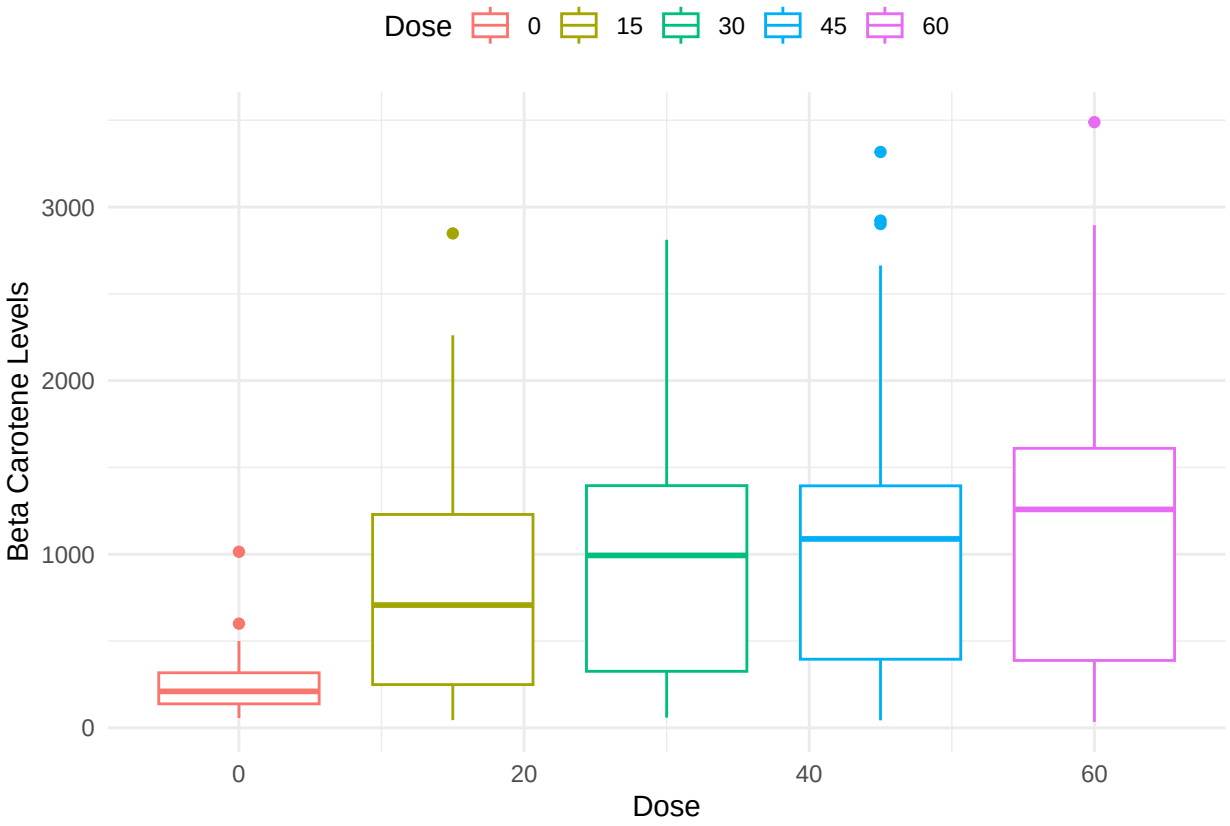
4. Results

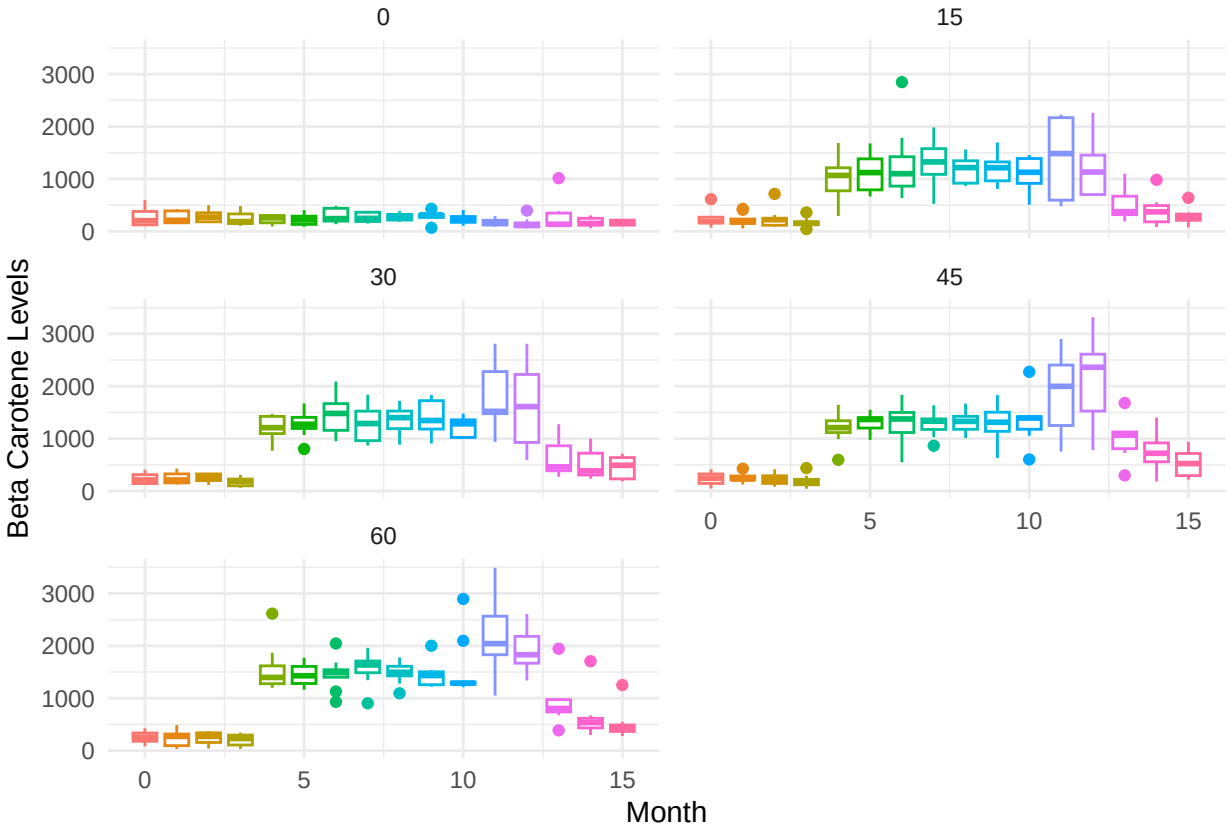
(a) Descriptive Statistics:






ptid
5
6
18
24
26
27
36
44
Beta Carotene Levels
1000
750
500
250
0
5
10
15
Months


Dose
0
15
30
45
60
Beta Carotene Levels
3000
2000
1000
0
0
20
40
60
Dose



(b) Models

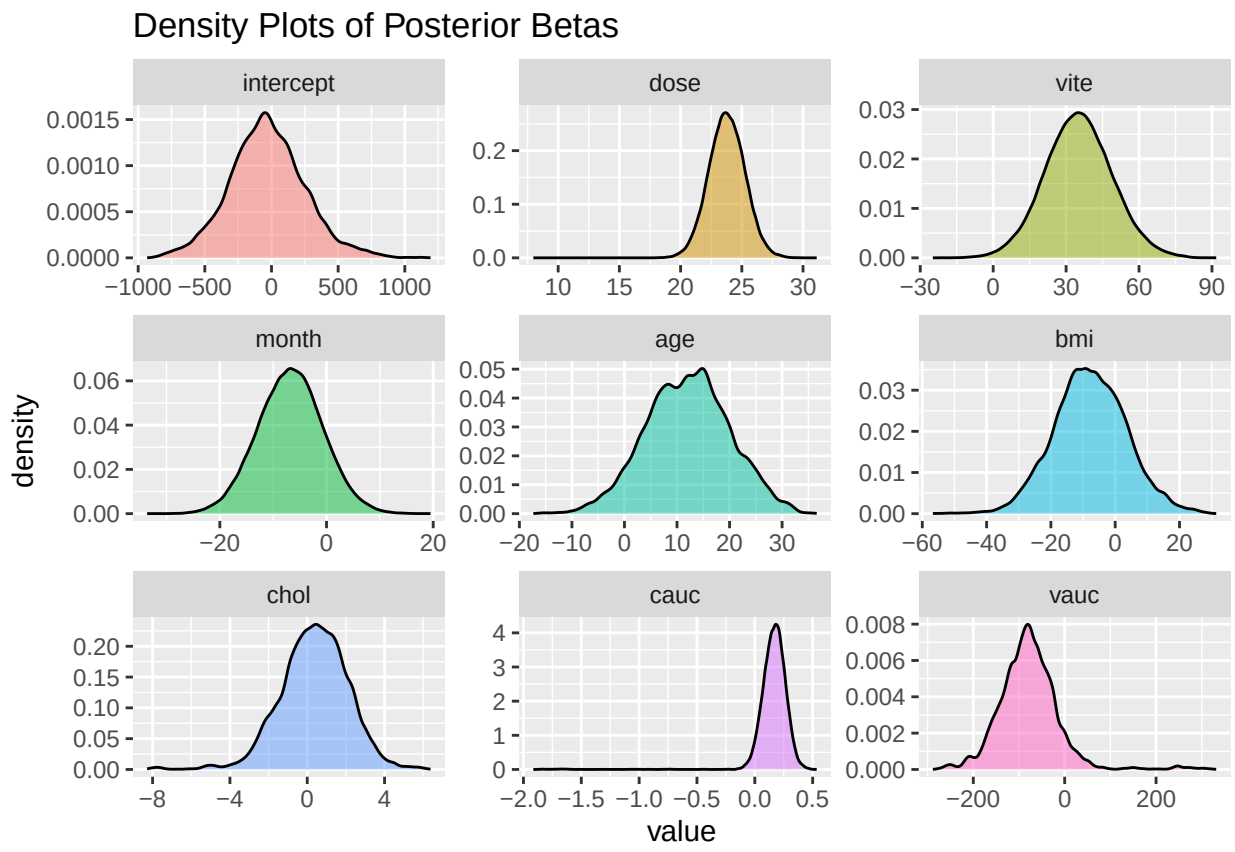
```
##  Set MCMC parameters.
chains <- 1            ##  One chain of MCMC values is obtained
burn_in <- 1000        ##  The first 1000 iterates from the MCMC chain are discarded
iterations <- 100000    ##  10,000 values are sampled from the posterior distribution
thin <- 1              ##  No thinning is performed


##  Define new variables for JAGS to use.
n <- dim(bc_data)[1]
n_patients <- length( unique(ptid) )

intercept <- rep(1,n)
tx <- intercept - (month<4)


##  Define the design matrix for the model and specify the number of covariates, for use in JAGS.
x <- cbind(intercept, tx*dose, vite, month, age, bmi, chol, cauc, vauc)
n_covariates <- dim(x)[2]
tau_b <- 0.00001


##  Establish data and parameter lists for use with OpenBUGS; define function for generating initial pa
```

```
data <- list( "bcarot"=bcarot,
              "n"=n,
              "n_patients"=n_patients,
              "n_covariates"=n_covariates,
              "ptid"=ptid,
              "x"=x,
              "tau_b"=tau_b )

inits <- function() {
  list( beta = rnorm( n_covariates, 0, 1 ),
        gamma = rnorm( n_patients, 0, 1 ),
        tau_bc = runif( 1, 0, 2 ),
        tau_g = runif( 1, 0, 2 ) )
}

parameters <- c( "beta", "gamma", "sigma_bc", "sigma_g" )



###############   CREATE MODEL   ###############

bc_modelstring<-"model
  {
    for(i in 1:n){
        bcarot[i] ~ dnorm( mu[i], tau_bc )
          mu[i] <- inprod( x[i,], beta[] ) + gamma[ ptid[i] ]
    }
      for(j in 1:n_patients){
        gamma[j] ~ dnorm( 0, tau_g )
      }
    for(k in 1:n_covariates){
          beta[k] ~ dnorm( 0, tau_b )
      }
    tau_bc ~ dgamma(0.001, 0.001)
      tau_g ~ dgamma(0.001, 0.001)
    sigma_bc <- pow(tau_bc,-0.5)
      sigma_g <- pow(tau_g,-0.5)
  }"

beta_carotene.m <- jags.model( data=data,
                               inits=inits,
                               file=textConnection(bc_modelstring),
                               n.chains=chains )
```

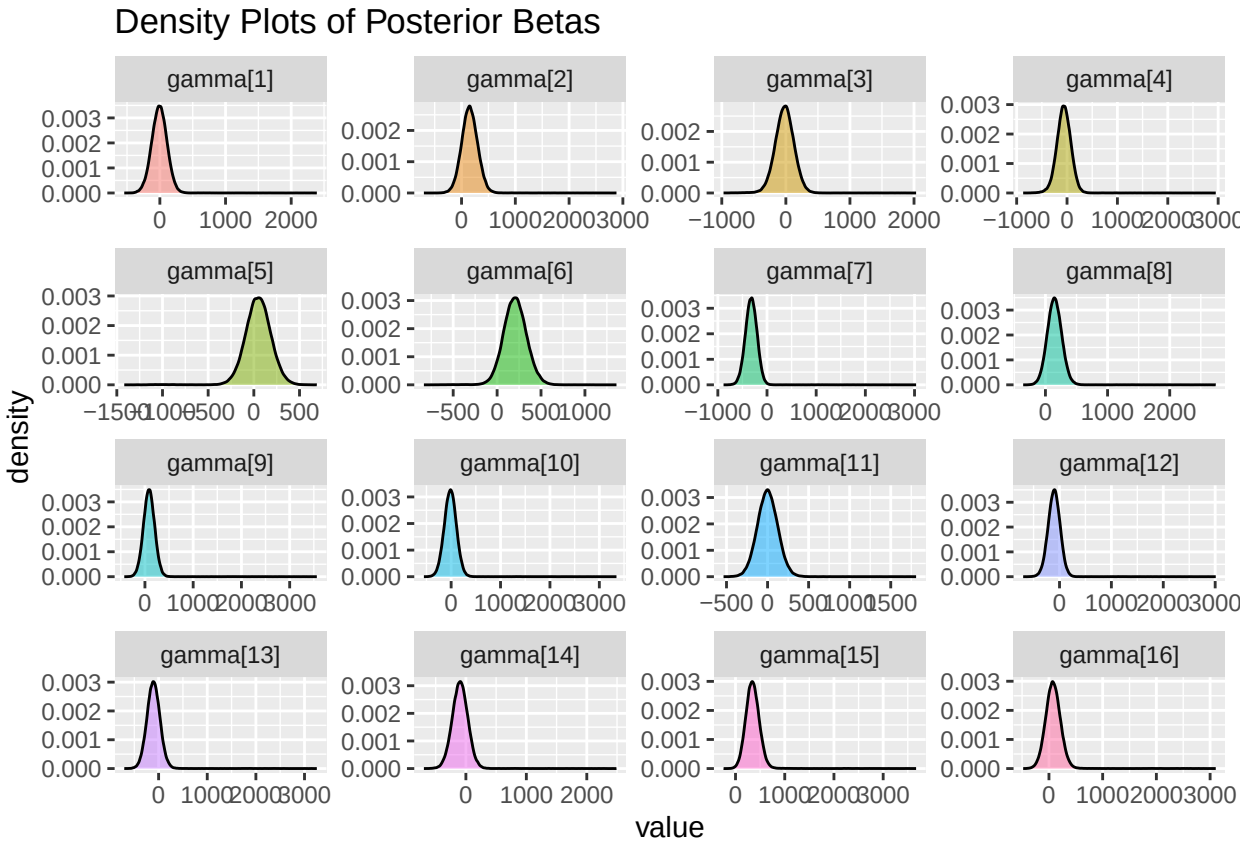
```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 674
##    Unobserved stochastic nodes: 55
##    Total graph size: 9502
##
## Initializing model
```

```r
beta_carotene.sim <- coda.samples(beta_carotene.m,
                                  parameters,
                                  n.iter=iterations,
                                  thin=thin,
                                  n.burn=burn_in)

## Assign variable names to posterior samples.
beta_carotene.iterates <- as.matrix(beta_carotene.sim)
beta_carotene.iterates <- as.data.frame(beta_carotene.iterates)

posterior_betas <- beta_carotene.iterates[,(1:n_covariates)]
names(posterior_betas) <- c("intercept", "dose", "vite", "month", "age", "bmi", "chol", "cauc", "vauc")
posterior_gammas <- beta_carotene.iterates[,((n_covariates+1):(n_covariates+n_patients))]
posterior_sigmas <- beta_carotene.iterates[,((n_covariates+n_patients+1):(n_covariates+n_patients+2))]
```

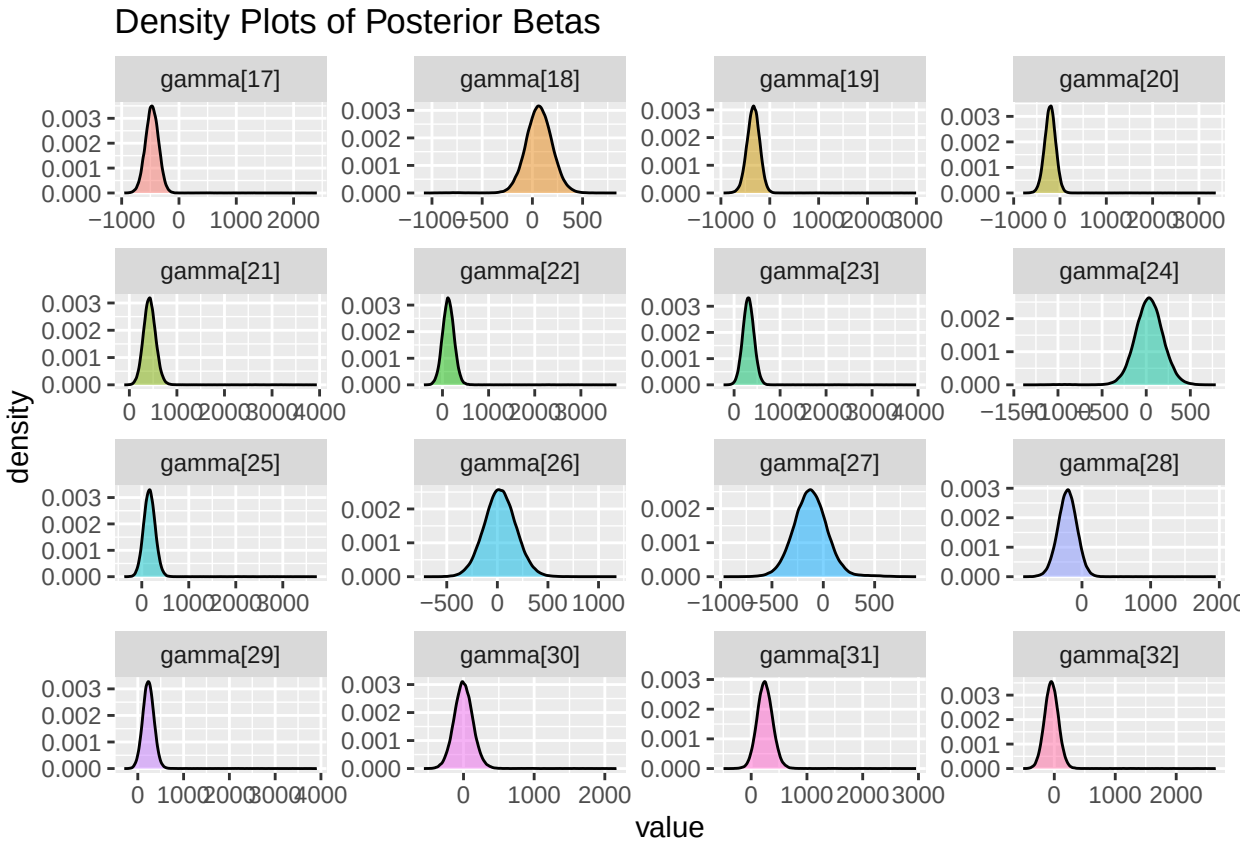
```
## No id variables; using all as measure variables
```



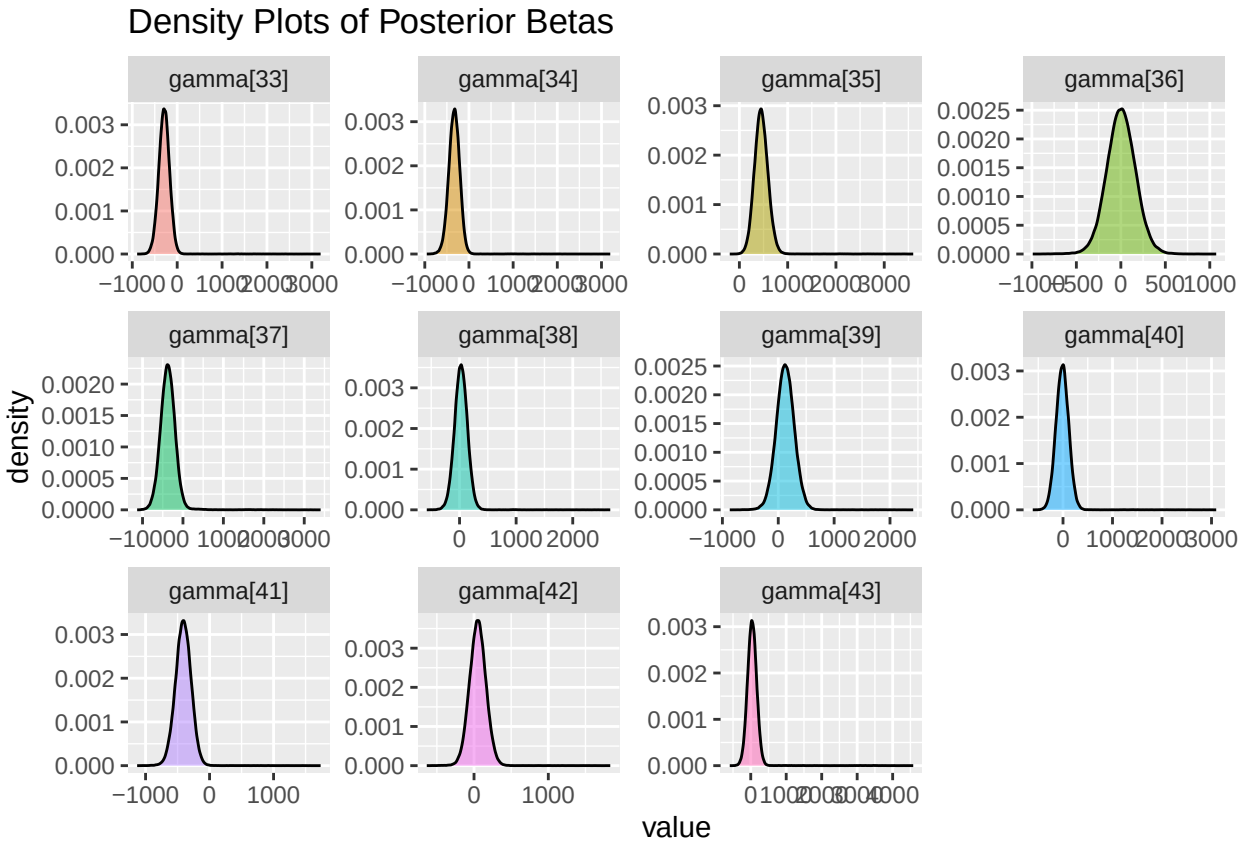

```
## No id variables; using all as measure variables
```



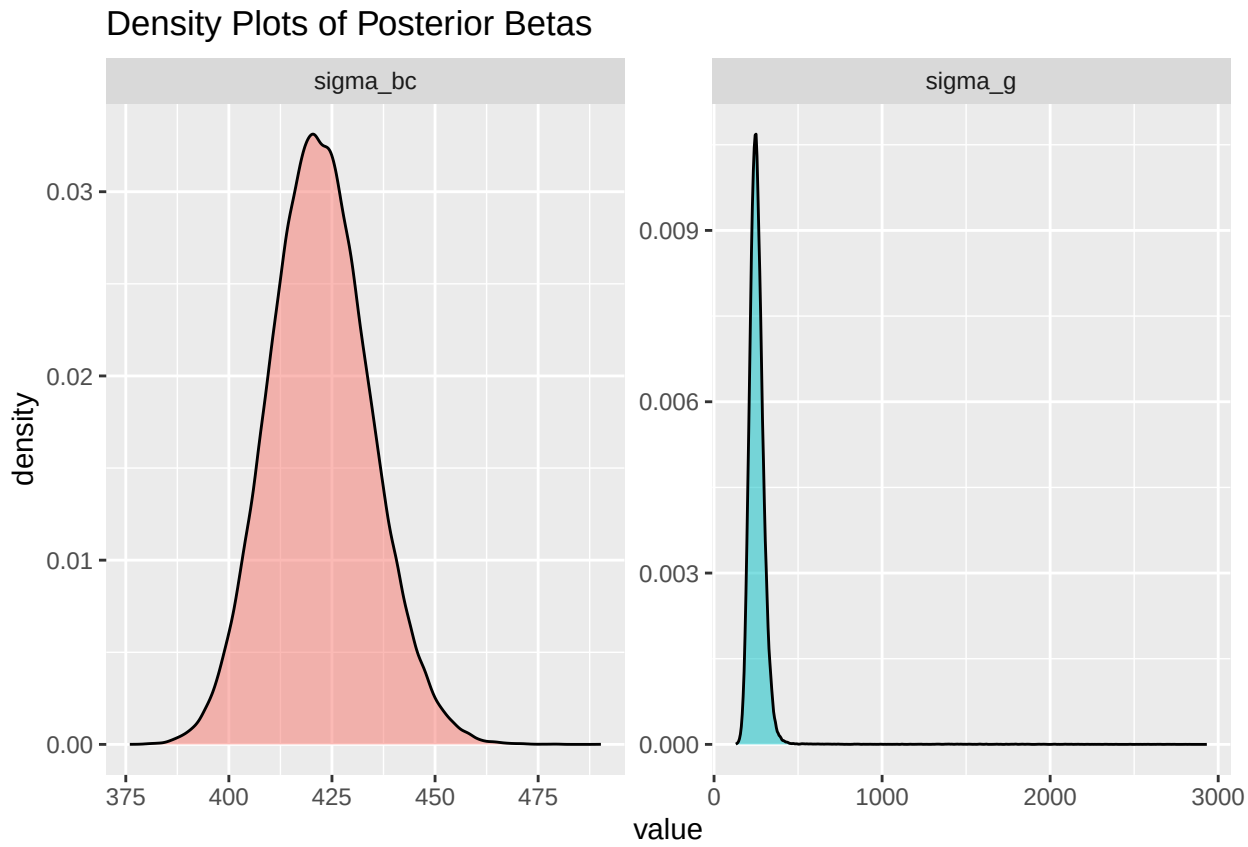

```
## No id variables; using all as measure variables
```




```
## No id variables; using all as measure variables
```




```
## No id variables; using all as measure variables
```



(c) Model-Building / Model-Checking:

5. Discussion:

6. Bibliography:

7. Appendix:

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.