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

Sacha Robbins

June 113, 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 confirmed 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 is 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 blantantly 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 ith 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. $month_{ij} = 1, 2, .., n_i$ keeps track of the measurements during treatment phase, where n_i is the number of treatment measurements for i^{th} patient.

To build the predictive model, we used Bayesian Inference for linear mixed effects (LME) model. Here, we were able to place non-informative priors and through Gibbs sampling technique 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. Implementation of the model was conducted using 'rjags' package version 4-6.

$$E[Y_{ij}|\vec{b_i}] = \beta_0 + \beta_{C1} Tx_i + \beta_{C2} Male_i + \beta_{C3} Chol_i + \beta_{C4} base_i + \beta_{L1} month_{ij} + \beta_{L2} Tx_i \cdot month_{ij}$$

Results

Appendix A

Appendix B