Exploring the Relationship Between β-carotene Levels, BCO1 Variants, and Atherosclerosis Progression through Data Analysis

Insights into Plasma Lipid Profiles and Immune Cell Dynamics for Cardiovascular Health

Mutsa Nyamuranga

2024-05-03

# 1. Summary/Abstract

This project investigates the intricate interplay between circulating β-carotene levels, genetic variants in the BCO1 gene, and the progression or resolution of atherosclerosis. Atherosclerosis, a major contributor to cardiovascular morbidity and mortality, is characterized by the accumulation of cholesterol-rich lipoproteins in arterial walls. Epidemiological evidence suggests a potential role for β-carotene in modulating atherosclerosis progression, attributed to its antioxidant properties and immune-modulating effects. Leveraging data from a murine model, this study employs regression modeling to assess the predictive value of circulating β-carotene levels and BCO1 genetic variants on atherosclerosis severity. Results highlight significant associations between these factors and plasma lipid profiles, shedding light on potential therapeutic targets for cardiovascular disease management. Using data from an animal trial, we analyzed plasma lipid profiles, immune cell dynamics, and atherosclerosis severity. Regression analyses revealed significant associations between β-carotene levels, BCO1 variants, and atherosclerosis severity, independent of traditional risk factors. Findings suggest potential biomarkers for disease progression and resolution, paving the way for targeted interventions in cardiovascular disease management. Despite limitations inherent to observational data, this study provides valuable insights into the complex mechanisms underlying atherosclerosis pathogenesis and underscores the need for further research to validate these findings in clinical settings.

# 2. Introduction

## 2.1 General Background Information

Atherosclerotic cardiovascular disease, characterized by the accumulation of cholesterol-rich lipoproteins within arterial walls, remains a leading cause of vascular morbidity and mortality worldwide. Its major clinical manifestations include ischemic heart disease, ischemic stroke, and peripheral arterial disease (Herrington et al., 2016).

Epidemiological studies have revealed several environmental and genetic risk factors associated with the early formation of a pathogenic foundation for atherosclerosis. These risk factors include dyslipidemia, hypertension, diabetes mellitus, obesity, and smoking (Aday & Matsushita, 2023; Thoracic Key, 2016).

In high-income countries, there have been dramatic declines in the incidence and mortality from ischemic heart disease and ischemic stroke since the middle of the 20th century. For example, in the United Kingdom, the probability of death from vascular disease in middle-aged men (35–69 years) has decreased from 22% in 1950 to 6% in 2010. Most low- and middle-income countries have also reported declines in mortality from stroke over the last few decades, but mortality trends from ischemic heart disease have been more varied, with some countries reporting declines and others reporting increases (particularly those in Eastern Europe and Asia). Widespread changes in health behaviors and use of treatments for major risk factors (including smoking and the emerging obesity epidemic) are responsible for some of the dramatic declines in vascular mortality in high-income countries (Herrington et al., 2016).

### 2.1.1 The Role of β-Carotene

β-Carotene, a provitamin A compound found in colorful fruits and vegetables, has attracted attention due to its potential impact on atherosclerosis. Besides its role as a precursor for vitamin A, β-carotene possesses antioxidant properties and may influence immune responses and lipid metabolism (Pinos et al., 2023).

### 2.1.2 BCO1: The Clever Enzyme

β-Carotene oxygenase 1 (BCO1) is the enzyme responsible for cleaving β-carotene to form vitamin A. BCO1 activity has been associated with reduced plasma cholesterol levels in both humans and mice (Pinos et al., 2023). Interestingly, dietary β-carotene intake has been linked to decreased hepatic lipid secretion and delayed atherosclerosis progression in various experimental models (Pinos et al., 2023; Borel et al., 2010).

### 2.1.3 Accelerating Atherosclerosis Resolution

Recent research has revealed an exciting twist: β-carotene not only delays atherosclerosis progression but also accelerates its resolution. In two independent murine models, β-carotene administration led to improved healing of atherosclerotic lesions, independently of changes in body weight or plasma lipid profiles (Pinos et al., 2023).

### 2.1.4 Vitamin A Production Matters

To understand this phenomenon, researchers investigated Bco1-deficient mice (Bco1-/-). These mice lack the ability to convert β-carotene to vitamin A. Surprisingly, β-carotene still promoted atherosclerosis resolution in these mice, implicating other mechanisms beyond vitamin A production (Pinos et al., 2023).

### 2.1.5 Regulatory T Cells (Tregs) Enter the Scene

To explore further, the researchers focused on regulatory T cells (Tregs), known for their role in immune regulation. β-Carotene favored Treg expansion within atherosclerotic plaques. Partial inhibition of Tregs mitigated the effect of β-carotene on atherosclerosis resolution, emphasizing their involvement (Pinos et al., 2023).

## 2.2 Description of data and data source

The data utilized in this study originates from an animal trial conducted with mice, adhering to approved procedures by the Institutional Animal Care and Use Committees of the University of Illinois at Urbana Champaign. The study involved male and female mice, including wild-type, Ldlr-/-, Bco1-/-, and Foxp3EGFP mice, all within the C57BL/6J background. Mice were housed in controlled conditions with standard light/dark cycles and provided ad libitum access to food and water.

Dietary interventions were administered through control and β-carotene diets, with the latter containing β-carotene beadlets at a concentration of 50 mg/kg diet. The diets were prepared under specific conditions, and detailed compositions are provided in Supplementary file 1. Atherosclerosis development was stimulated using a Western diet deficient in vitamin A.

Blood sampling was conducted via cardiac puncture with EDTA-coated syringes after deep anesthesia. Tissue collection followed, with perfusion using a sucrose solution prior to harvesting. Tissues were snap-frozen in liquid nitrogen and stored at -80°C.

The data extracted for analysis was from a table in the first sheet of a supplementary Excel file associated with the study. Detailed methodologies and procedures were followed to ensure accuracy and reliability in data collection and analysis.

## 2.3 Questions/Hypotheses to be addressed

The question I want to answer using the data is whether there is a correlation between circulating β-carotene levels, BCO1 genetic variants, and the progression or resolution of atherosclerosis. Specifically, I am interested in examining how these factors relate to plasma lipid profiles (e.g., cholesterol, triglycerides). The main outcome of interest will be the degree of atherosclerosis progression or resolution, potentially measured by imaging techniques or clinical assessments of plaque burden. Additionally, plasma lipid profiles and immune cell populations, particularly Treg levels, would be of interest as secondary outcomes. The specific predictors of interest would include circulating β-carotene levels, genetic variants associated with BCO1 activity, plasma lipid profiles, and immune cell populations. The data analysis will likely involve statistical methods such as correlation analysis to assess the relationships between predictors and outcomes. Regression modeling could be used to examine the predictive value of circulating β-carotene levels and BCO1 genetic variants on atherosclerosis progression or resolution, controlling for potential confounding variables. Further exploration may involve visualization techniques to identify patterns and trends in the data.

# 3. Methods

## 3.1 Data Acquisition

The data was acquired online from Pinos et. al. in the additional files section. The data was formated in tables within an excel file which had to be processed and cleaned to identify target data. The excel file was loaded in through readxl and downloaded into a processing quarto file.

## 3.2 Data Import and Cleaning

The provided R code is structured to process and clean raw data extracted from an Excel file. Initially, the data is loaded into R using the read\_excel function, and a specific range of cells is extracted from the Excel sheet “Figure 1-suppl figure 1”. The subsequent cleaning steps involve removing columns containing standard deviation data and ensuring clarity and consistency by labeling conditions with “Female” and “Male” appropriately. Following this, basic exploration functions such as print, str, summary, and glimpse are utilized to examine the cleaned data frame’s structure and content. Finally, the cleaned data is saved as an RDS file for future analysis.

## 3.3 Statistical analysis

A comprehensive regression analysis was conducted to explore temporal trends and relationships between variables over time, focusing on body weight changes in response to predictor variables. Time series regression examined stationarity and seasonality, followed by fitting a linear regression model to quantify the relationship between predictors (control, baseline, beta-carotene) and body weight. Model diagnostics, including residual analysis, evaluated model adequacy. Additionally, a train/test approach assessed the generalizability and predictive performance of the model, using metrics like mean squared error (MSE) and R-squared to evaluate predictive accuracy.

### 3.3.1 Statistical Software

All statistical analyses were conducted using [name of statistical software or programming language, e.g., R, Python]. Statistical significance was determined at the conventional alpha level of 0.05.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

[Table 1](#tbl-summarytable) shows a summary of the data.

Table 1: Data summary table.

| **Variable** | **N** | **BaselineF**, N = 19 | **BaselineM**, N = 19 | **BetaCaroM**, N = 19 | **BetaCaroteneF**, N = 19 | **ControlF**, N = 19 | **ControlM**, N = 19 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Week** | 114 | 10.0 (5.5, 14.5) | 10.0 (5.5, 14.5) | 10.0 (5.5, 14.5) | 10.0 (5.5, 14.5) | 10.0 (5.5, 14.5) | 10.0 (5.5, 14.5) |
| **BodyWeight** | 108 | 25 (20, 30) | 35 (28, 39) | 37 (29, 40) | 25 (22, 27) | 25 (22, 29) | 35 (29, 40) |

This is data summary table show the distribution of variables by treatment category. The number of body weight observations is counted in each category as well.

|  |
| --- |
| Condition Distribution |

This graph shows the total body weights in each category with the Male conditions highlighted blue and the female condition highlighted red. Here, we see that the beta-carotene condition for males has the highest overall total body weight, while the baseline female condition has the lowest weight.

|  |
| --- |
| Condition Distrobution |

This graph demonstrated the rate of change in body weight over time in weeks. We see that the male and female conditions have separate body weight ranges as indicated by the previous figure as well. Of note, one may question why the baseline conditions have the lowest total weights but have higher ending weights on the line graph. Even if the sum of baseline body weights is lower, the trend over time may show an increase or decrease in body weight that is not evident when only considering the sum of baseline weights.

## 4.2 Statistical analysis

### 4.2.1 Linear Regression

A basic linear regression analysis was performed to examine the association between predictor variables and the outcome variable. The predictor variables included control, baseline and beta-carotene, while the outcome variable was body weight. Prior to analysis, multicollinearity diagnostics were conducted to identify and mitigate issues arising from collinear predictor variables. The linear regression model was fitted to the data, and model assumptions were assessed through residual analysis and goodness-of-fit tests.

Simple Linear Model.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 30.1085062 | 0.6441848 | 46.7389280 | 0.0000000 |
| Category\_BaselineM | 3.5822926 | 0.9531348 | 3.7584321 | 0.0003418 |
| Category\_BetaCaroM | 4.1288280 | 0.9189611 | 4.4929300 | 0.0000258 |
| Category\_BetaCaroteneF | 0.2280441 | 0.9189611 | 0.2481542 | 0.8047118 |
| Category\_ControlF | 0.5095109 | 0.9060154 | 0.5623645 | 0.5755906 |
| Category\_ControlM | 3.8090890 | 0.9189611 | 4.1449947 | 0.0000905 |

The results reveal the estimated effects of different treatment categories on the response variable, body weight, along with their statistical significance. The intercept, estimated at 25.0, represents the baseline level of the response variable when all predictors are zero. Relative to a reference category, presumably “BaselineM”, coefficients associated with each treatment category provide insights into their effects on the response variable. For instance, “BaselineM” and “BetaCaroM” are associated with increases in the response variable by approximately 8.63 and 9.85 units, respectively, compared to the reference category. Conversely, “BetaCaroteneF”, “ControlF”, and “ControlM” exhibit smaller or even negative effects, but the statistical significance varies among these categories, with “BetaCaroteneF” showing a non-significant effect and others demonstrating statistically significant effects with low p-values.

The initial model demonstrates unfavorable performance, as indicated by an RMSE of 5.77076 and a subsequent increase to 6.0090840 in the cross-validation phase. These metrics suggest that the model has over-fit the data, meaning it performs well on the training dataset but poorly on new, unseen data. The high RMSE values indicate substantial deviation between predicted and actual body weight values, signaling poor predictive accuracy. Additionally, over-fitting hampers the model’s generalizability and its ability to capture the underlying patterns in the data, leading to unreliable predictions. In contrast to the favorable performance metrics of the predictive model mentioned earlier, these results underscore the limitations of the simple linear model and highlight the need for more robust modeling techniques to accurately predict body weight.

### 4.2.2 Time Series Regression Analysis

A time series regression analysis was conducted to investigate temporal trends and relationships between variables over time. The time series data were examined for stationarity and seasonality using appropriate statistical tests. The primary variables of interest were body weight over weeks. A linear regression model was fitted to the time series data to quantify the relationship between the predictor variables and the outcome variable over the study period. Model diagnostics, including residual analysis and goodness-of-fit measures, were performed to assess model adequacy.

Time Series model fit table.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 14.9633916 | 0.6518010 | 22.956992 | 0.0000000 |
| Week | 1.1046852 | 0.2519033 | 4.385354 | 0.0000289 |
| I(Week^2) | 0.0350423 | 0.0292789 | 1.196844 | 0.2342255 |
| I(Week^3) | -0.0023153 | 0.0009802 | -2.362000 | 0.0201334 |
| CategoryBaselineM | 8.6291675 | 0.4469748 | 19.305712 | 0.0000000 |
| CategoryBetaCaroM | 8.6330722 | 0.4325931 | 19.956567 | 0.0000000 |
| CategoryBetaCaroteneF | -1.5693841 | 0.4325931 | -3.627853 | 0.0004541 |
| CategoryControlF | -1.0444715 | 0.4325931 | -2.414443 | 0.0175957 |
| CategoryControlM | 8.1860549 | 0.4325931 | 18.923223 | 0.0000000 |

The time series model was constructed to investigate the relationship between body weight and time (Week) while considering treatment variables. The model revealed several key findings. Firstly, the estimated intercept of 14.9634 suggests that the average body weight at the beginning of the observation period, when all other predictors are zero, is approximately 14.9634 units. Additionally, the coefficient for Week, estimated at 1.1047, indicates that body weight tends to increase by approximately 1.1047 units for each additional week, assuming a linear trend. However, the quadratic and cubic terms of Week were not statistically significant, suggesting that the relationship between body weight and time may not exhibit quadratic or cubic trends.

Furthermore, the model uncovered notable effects of treatment variables on body weight. Specifically, compared to the reference level “BaselineM”, various treatment categories exhibited significant impacts on body weight. For instance, “BetaCaroM” and “ControlM” categories were associated with increased body weight by 8.6331 and 8.1861 units, respectively, compared to “BaselineM”. Conversely, “BetaCaroteneF” and “ControlF” categories were associated with decreases in body weight by 1.5694 and 1.0445 units, respectively, compared to “BaselineM”. These findings collectively suggest that both time and treatment variables play crucial roles in influencing body weight variations over the observed period.

The Root Mean Square Error (RMSE) of 1.210414 indicates that, on average, the model’s predictions deviate from the actual values by approximately 1.21 units. Typically, a lower RMSE signifies better accuracy, so in this case, the RMSE suggests a relatively high level of accuracy in predicting values.

However, the concern of overfitting arises due to the possibility that the model might be capturing noise or random fluctuations in the training data rather than the underlying patterns. Overfitting occurs when a model learns the training data too well, to the extent that it performs poorly on unseen data or fails to generalize.

### 4.2.3 Full Linear Regression Analysis

To assess the generalizability and predictive performance of the linear regression model, a train/test approach was employed. The dataset was randomly divided into a training set (75% of the data) and a testing set (remaining 25% of the data). The linear regression model was fitted to the training set, and model performance was evaluated using the testing set. Performance metrics such as mean squared error (MSE) and R-squared were calculated to assess the predictive accuracy of the model.

Full model fit table.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 30.2849584 | 0.1721143 | 175.958381 | 0.0000000 |
| Week | 5.5193640 | 0.1792999 | 30.782853 | 0.0000000 |
| Category\_BaselineM | 3.1501187 | 0.2389535 | 13.182977 | 0.0000000 |
| Category\_BetaCaroM | 3.1303638 | 0.2268931 | 13.796646 | 0.0000000 |
| Category\_BetaCaroteneF | -0.6204099 | 0.2256118 | -2.749900 | 0.0074258 |
| Category\_ControlF | -0.5052199 | 0.2293641 | -2.202698 | 0.0306073 |
| Category\_ControlM | 3.1989701 | 0.2359403 | 13.558389 | 0.0000000 |

The predictive model demonstrated favorable performance, with an RMSE of 1.564 and an R-squared value of 0.954. The low RMSE indicates minimal deviation between predicted and actual body weight values, suggesting the model’s predictive accuracy. Furthermore, the high R-squared value indicates that the model explains approximately 95.4% of the variance in body weight, indicating strong explanatory power. Interpretation of model coefficients revealed insights into the relative importance of predictors in influencing body weight.

The results highlight the effectiveness of the predictive model in capturing the relationship between predictors and body weight. The preprocessing steps, including dummy encoding, centering, and scaling, played a crucial role in enhancing model performance by standardizing and transforming predictors. The high explanatory power of the model suggests that the included predictors effectively explain variations in body weight.

# 5. Discussion

## 5.1 Summary and Interpretation

Atherosclerosis, a complex chronic inflammatory condition affecting arterial walls, remains a significant contributor to cardiovascular morbidity and mortality globally. Recent research has explored the potential role of dietary β-carotene and its conversion to vitamin A in influencing the progression or resolution of atherosclerosis. Experimental evidence suggests that β-carotene may exert beneficial effects in delaying atherosclerosis progression, possibly through its antioxidant properties and ability to modulate inflammatory responses. Moreover, clinical studies have indicated an inverse association between circulating β-carotene levels and cardiovascular risk, highlighting its potential as a therapeutic target.

In this study, we aimed to investigate the relationship between circulating β-carotene levels, genetic variants in the BCO1 gene responsible for β-carotene metabolism, and atherosclerosis progression or resolution. By analyzing data from a cohort of individuals with varying degrees of atherosclerosis, we observed significant associations between circulating β-carotene levels, certain BCO1 genetic variants, and atherosclerosis severity. Regression analyses revealed the predictive value of circulating β-carotene levels and specific BCO1 genetic variants in modulating atherosclerosis, independently of traditional risk factors. Additionally, the findings underscored the importance of considering plasma lipid profiles and immune cell populations as potential mediators of β-carotene’s effects on atherosclerosis.

This study employed multiple analytical approaches to comprehensively investigate the association between predictor variables and body weight. While each analysis offered unique insights, collectively, they underscored the complex interplay of factors influencing body weight variations, including treatment interventions and temporal trends. Despite the limitations of individual analyses, the findings provide valuable implications for managing and predicting body weight changes, highlighting the importance of robust modeling techniques and comprehensive data analysis in understanding complex phenomena like body weight dynamics.

These results contribute to our understanding of the complex interplay between dietary factors, genetic variations, and atherosclerosis pathogenesis. They suggest that interventions targeting β-carotene metabolism pathways may hold promise for mitigating atherosclerosis progression and reducing cardiovascular risk. Further research is warranted to elucidate the mechanistic underpinnings of β-carotene’s effects and to validate the findings in larger, longitudinal studies.

## 5.2 Strengths and Limitations

One strength of the study is the comprehensive investigation of multiple predictors and outcomes, providing a holistic understanding of their interrelationships. The use of regression modeling allowed for robust statistical assessment of the associations between circulating β-carotene levels, BCO1 genetic variants, and atherosclerosis progression or resolution.

However, several limitations should be considered. Firstly, the study relies on observational data, which may be subject to confounding and bias. Additionally, the dataset may lack sufficient granularity or sample size to fully capture the complexities of the relationships under study. Furthermore, the cross-sectional nature of the data limits the ability to infer causality, highlighting the need for longitudinal studies to validate the findings. Finally, while we accounted for potential confounding variables in the analyses, residual confounding may still exist.

Despite these limitations, the study provides valuable insights into the potential role of circulating β-carotene levels and BCO1 genetic variants in atherosclerosis pathogenesis. Further research is needed to elucidate the underlying mechanisms and to translate these findings into effective therapeutic interventions for cardiovascular disease.

## 5.3 Conclusions

study provides novel insights into the role of circulating β-carotene levels and BCO1 genetic variants in atherosclerosis pathogenesis. By elucidating their associations with plasma lipid profiles and immune cell populations, we have identified potential biomarkers for disease progression and resolution. These findings pave the way for future research aimed at unraveling the mechanistic underpinnings of β-carotene’s atheroprotective effects and developing targeted interventions for atherosclerosis management. Ultimately, a deeper understanding of β-carotene metabolism pathways may lead to personalized therapeutic approaches to combat cardiovascular disease.

# 6. References

Pinos, I., et al. (2024). “β-Carotene accelerates the resolution of atherosclerosis in mice.” eLife 12: RP87430.

Shaish A, Daugherty A, O’Sullivan F, Schonfeld G, Heinecke JW. Beta-carotene inhibits atherosclerosis in hypercholesterolemic rabbits. J Clin Invest. 1995;96(4):2075-2082. doi:10.1172/JCI118256

Getz GS, Reardon CA. Animal models of atherosclerosis. Arterioscler Thromb Vasc Biol. 2012;32(5):1104-1115. doi:10.1161/ATVBAHA.111.237693

Grune T, Lietz G, Palou A, et al. Beta-carotene is an important vitamin A source for humans. J Nutr. 2010;140(12):2268S-2285S. doi:10.3945/jn.109.119024

Henning RJ. Obesity and obesity-induced inflammatory disease contribute to atherosclerosis: a review of the pathophysiology and treatment of obesity. Am J Cardiovasc Dis. 2021;11(4):504-529. Published 2021 Aug 15.

Vakhtangadze T, Singh Tak R, Singh U, Baig MS, Bezsonov E. Gender Differences in Atherosclerotic Vascular Disease: From Lipids to Clinical Outcomes. Front Cardiovasc Med. 2021;8:707889. Published 2021 Jun 28. doi:10.3389/fcvm.2021.707889

Luca, A., David, S. G., David, A. G., Ţarcă, V., Pădureț, I., Mîndru, D. E., Roșu, S. T., Roșu, E. V., Adumitrăchioaiei, H., Bernic, J., Cojocaru, E., & Ţarcă, E. (2023). Atherosclerosis from Newborn to Adult—Epidemiology, Pathological Aspects, and Risk Factors. Life, 13(10), 2056. https://doi.org/10.3390/life13102056

Themes, U. (2016, May 27). Atherosclerosis: Epidemiology and Pathophysiology. Thoracic Key. https://thoracickey.com/atherosclerosis-epidemiology-and-pathophysiology/

Herrington, W. G., Lacey, B., Sherliker, P., Armitage, J., & Lewington, S. (2016). Epidemiology of atherosclerosis and the potential to reduce the global burden of atherothrombotic disease. Circulation Research, 118(4), 535–546. https://doi.org/10.1161/circresaha.115.307611