Risk of Osteoporosis Based on Dietary Vitamin D and Calcium (NHANES 2017-2018)

A. Renganathan & Y.Baldonado

04/07/2023

Abstract

Introduction

Multivariate logistic regression models are a common statistical methodology used when trying to model a response that is binary, or dichotomous. It aims to derive the model with the best fit and the highest efficiency that explains the relationship between the predictor and response variables.

Unlike typical regression models, a logistic regression transforms the model into a probability model for the outcome which can be interpreted as the probability that the response will occur based on the predictors in the model.

Logistic regression is ideal for modeling our question because it allows us to test both continuous and categorical predictors against a binary response.

For this project, we use a multivariate logistic regression model to assess the risk of osteoporosis diagnosis based on a variety of predictors, including dietary and supplementary calcium and Vitamin D, as well as fractured wrists.

Osteoporosis is a chronic metabolic bone disease that is characterized by decreased bone mineral density and bone mass, resulting in weakened bone structure and strength. It presents in patients as wrist, hip and spine fractures, which are often caused from gentle stressors like falling from standing, lifting heavy objects or even coughing (REFERENCE).

The impact of osteoporosis should not be understated considering that is it one of the most common chronic metabolic diseases in old age. In 2010, an estimated 10.2 million people above the age of 50 had osteoporosis in the United States (Wright et al, 2014). The projected cost of bone fractures and osteoporosis in the United States is expected to reach 25.3 billion by 2025 (Burge et al, 2006).

Clinically significant risk factors for osteoporosis include sex, age, race, body and bone size, family history, diet and lifestyle factors (REFERENCE). For diet, calcium and vitamin D levels are most important as they're both involved in the biochemical pathways involved in bone health (REFERENCE). Lifestyle factors include levels of physical activity, alcohol use and smoking (REFERENCE).

Objectives

The objective of this analysis is to identify which dietary factors affect risk of osteoporosis diagnosis in older adults.

Methodology

Data Source

The data was sourced from the National Health and Nutrition Examination Survey (NHANES) run by the Center for Disease Control (CDC).The NHANES study has been run since the 1960's and provides population health data on adults and children in the United States.

The data is both publicly available and deidentified, therefore not requiring IRB approval. The survey includes demographic, socioeconomic, medical, dental, physiological, laboratory and dietary data.

NHANES is a nationally representative survey of the US civilian population that is achieved through clever use of survey weights. The survey itself involves oversampling minority populations, accounting for non-response and then a post-stratification adjustment to match Census population counts.

R Studio

Using the R package RNHANES, we extracted the pertinent information for our analysis directly from the website. We focused on the Demographic, Dietary, Laboratory, and Questionnaire data from the 2017-2018 cycle. The following variables were selected:

- Age, Gender and Race/Ethnicity from the demographic files
- Dietary Calcium, Dietary Vitamin D, Calcium Vitamin Supplementation, Vitamin D Supplementation from the Dietary files
- Fractured Wrist from the Questionnaire files.

After data management completed with t, we were left with 3,069 observations. The continuous predictors were Age, Calcium Intake, Vitamin D intake, Calcium Vitamin Supplementation, Vitamin D Supplementation and the categorical predictors were Gender, Race/Ethnicity, Dairy milk consumption, Nondairy milk consumption, Fractured Wrist, and Osteoporosis.

Data Management

Survey Weighting

Analysis

Exploratory Analysis

We first ran a simple logistic regression to test each variable against our response Osteoporosis to see which variables were individually significant. We found that Age, Gender (Female), Calcium Intake, and Calcium Supplementation were significant.

Descriptive Statistics

Model Assumptions

Independent Observations

Multicollinearity

Linearity of Continuous Variables Across Strata

Variable Selection

Multivariate Logistic Regression

Initial Model

Subset Models

Best Fit Model

Results

Interpretation of everything here.

Conclusions

Limitations

Future Research

References