

Predicting Systemic Inflammation: A Data-Driven Analysis of CRP Levels Using NHANES

Takreem Virk

University of Virginia, Department of Statistics

Introduction

- C-reactive protein (CRP) is a biomarker of systemic inflammation commonly used in clinical and public health research. Elevated CRP levels are associated with chronic diseases such as cardiovascular disease, diabetes, and obesity.
- Understanding what factors contribute to higher inflammation levels can help identify at-risk individuals, shape preventative strategies, and improve health outcomes.
- This project uses a nationally representative dataset (NHANES 2021–2023) to explore the relationship between CRP and a range of demographic, lifestyle, and nutritional factors.

Research Question

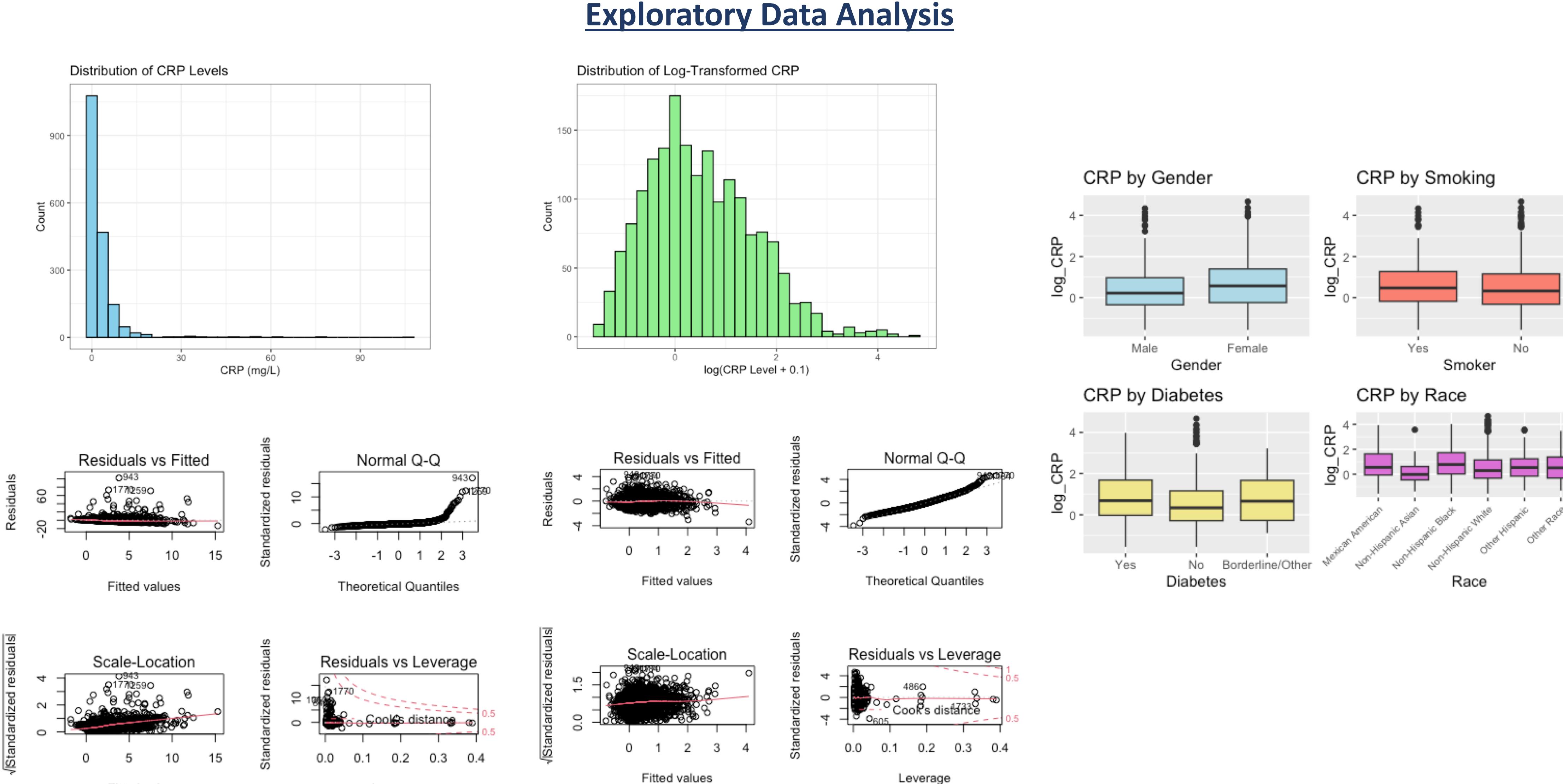
- Which demographic, behavioral, and nutritional factors best predict systemic inflammation, as measured by C-reactive protein (CRP), in U.S. adults?

Methods

- Data were obtained from the 2021–2023 cycle of the National Health and Nutrition Examination Survey (NHANES), a nationally representative study of the U.S. population. After filtering for complete cases, 1,918 adult participants were included.
- Predictor variables included age, gender, race, income-to-poverty ratio, height, weight, dietary protein and calorie intake, physical activity, diabetes status, and smoking status. The outcome variable was C-reactive protein (CRP), a biomarker of systemic inflammation.
- CRP values were log-transformed to meet linear regression assumptions. Analyses included multiple linear regression and a Random Forest model to identify key predictors.

Data

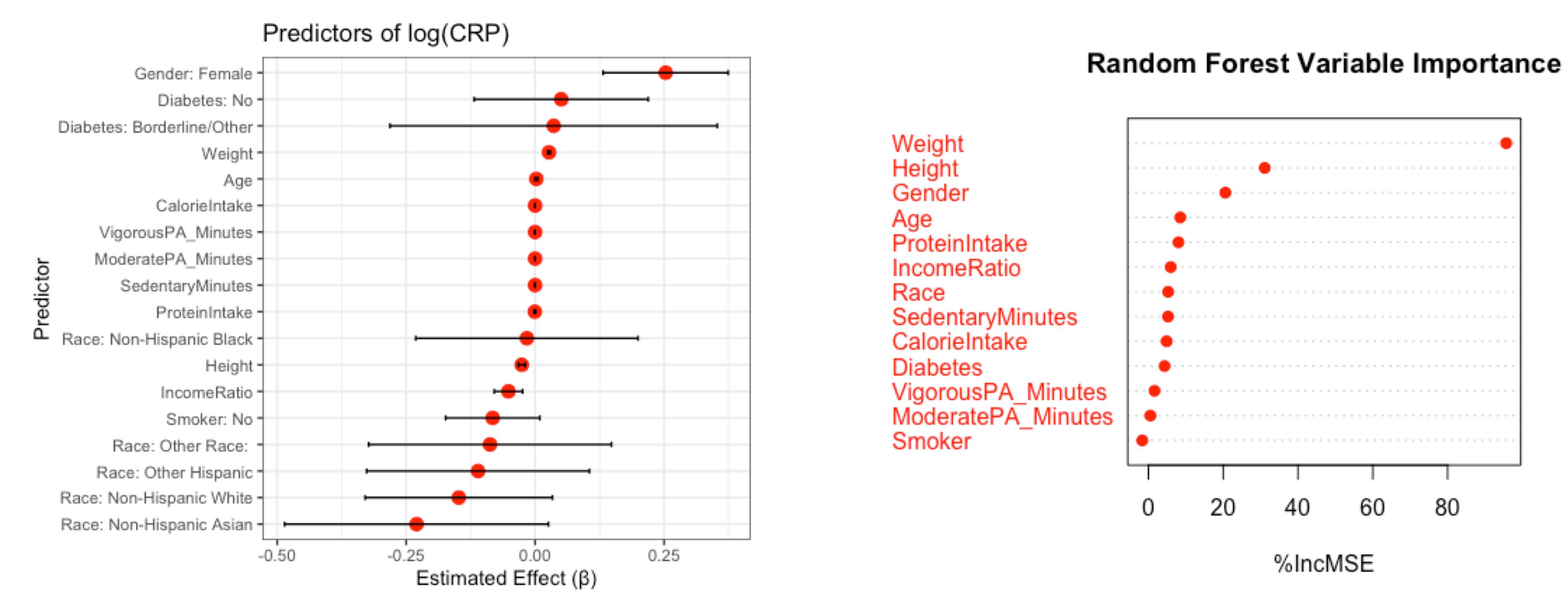
Variable	Type	Description
1 CRP	Continuous	CRP level
2 Age	Continuous	Participant's age
3 Gender	Categorical	Biological sex
4 Race	Categorical	Race/ethnicity
5 IncomeRatio	Continuous	Income-to-poverty ratio
6 Weight	Continuous	Body weight
7 Height	Continuous	Height
8 ProteinIntake	Continuous	Protein intake (daily)
9 CalorieIntake	Continuous	Calorie intake (daily)
10 ModeratePA_Minutes	Continuous	Moderate physical activity (weekly)
11 VigorousPA_Minutes	Continuous	Vigorous physical activity (weekly)
12 SedentaryMinutes	Continuous	Sedentary time (daily)
13 Smoker	Categorical	Smoker status
14 Diabetes	Categorical	Diabetes status



Model

$$\log(CRP) = \beta_0 + \beta_1(Age) + \beta_2(Gender) + \beta_3(Race) + \beta_4(IncomeRatio) + \beta_5(Weight) + \beta_6(Height) + \beta_7(ProteinIntake) + \beta_8(CalorieIntake) + \beta_9(ModeratePA) + \beta_{10}(VigorousPA) + \beta_{11}(SedentaryMinutes) + \beta_{12}(Smoker) + \beta_{13}(Diabetes) + \epsilon$$

Results



- The multiple linear regression model revealed several significant predictors of log-transformed CRP levels. Gender, weight, height, and income-to-poverty ratio showed the strongest associations. Specifically, females and individuals with lower income tended to have higher CRP levels, while greater body size was also linked to increased inflammation.
- To validate these findings and account for possible nonlinear effects, a Random Forest model was used to assess variable importance. Weight, height, gender, and age emerged as the top predictors, aligning closely with the regression results.
- Together, these results suggest that demographic and anthropometric factors play a key role in predicting systemic inflammation.

Conclusions

This study identified several meaningful predictors of systemic inflammation, as measured by C-reactive protein (CRP) levels, in a large, nationally representative sample of U.S. adults.

1. Body size is the strongest predictor of CRP.

Both weight and height were consistently important across linear and machine learning models. This reinforces the established link between greater body mass and increased inflammation.

2. Sociodemographic factors matter.

Gender and income-to-poverty ratio were significant predictors. Females had higher CRP levels than males, and individuals with lower income showed elevated inflammation — highlighting the influence of socioeconomic disparities on health.

3. Lifestyle behaviors showed weaker effects.

Physical activity and dietary intake had relatively modest predictive power compared to body size and demographics, suggesting that inflammation is driven by complex, multifactorial influences.

Together, these findings emphasize the importance of addressing both biological and social determinants in understanding and reducing systemic inflammation risk.

Limitations and Further Study

Limitations:

- The study is based on cross-sectional data, which limits the ability to make causal conclusions.
- Some variables, including diet and physical activity, were based on self-reported recall, which may introduce bias.
- Other relevant inflammation biomarkers (e.g., IL-6, TNF- α) were not included in this analysis.

Future Study:

- Explore longitudinal datasets to examine how inflammation changes over time.
- Incorporate additional biological markers and genetic data for deeper insight.
- Investigate interactions between lifestyle factors and social determinants to uncover more nuanced risk patterns.

References

<https://www.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?Cycle=2021-2023>