P8105 Final Project Proposal

Inflammatory Indices and Chronic Disease Outcomes Using UKB Data

2024-11-08

Group Members

Zicheng Wang (zw3088), Xiaohui Kang (xk2163), Mengyuan Chen (mc5698), Shuai yuan (sy3270), Dang Lin (dl3757)

Tentative Project Title

Exploring the Role of Inflammatory Biomarkers in Predicting Mortality Risks Among UK Adults with Fatty Liver or Cirrhosis, Using Evidence from UKB Data

Motivation

Chronic diseases like non-alcoholic fatty liver disease (NAFLD), metabolic-associated fatty liver disease (MAFLD), and metabolic-associated steatohepatitis-related liver disease (MASLD) are becoming increasingly common and can progress to cirrhosis, liver failure, or hepatocellular carcinoma. These conditions are closely tied to systemic inflammation and immune dysregulation, making it essential to identify reliable, non-invasive biomarkers that can predict disease progression and provide insights into the underlying mechanisms. Such biomarkers could improve early detection, guide interventions, and support better clinical outcomes.

Immune-related indices like the Lymphocyte-to-Monocyte Ratio (LMR), Systemic Immune-Inflammation Index (SII), Naples Prognostic Score (NPS), and Neutrophil-Percentage-to-Albumin Ratio (NPAR) have gained attention for their potential utility in assessing inflammation and immune responses. These indices are derived from routine blood tests, making them accessible and cost-effective. LMR, for instance, measures the balance between lymphocytes, which are linked to immune surveillance and anti-inflammatory responses, and monocytes, which are associated with inflammation and tissue damage. SII integrates neutrophil, platelet, and lymphocyte counts to provide a comprehensive view of immune activity and thrombosis risk. NPS combines multiple prognostic factors, including inflammatory markers, albumin levels, and platelet counts, offering a broader perspective on patient health. NPAR evaluates the relationship between neutrophil percentages and albumin levels, connecting immune function with nutritional status.

The UK Biobank (UKB) dataset is well-suited for studying the relationships between these immune markers and chronic disease outcomes. With its extensive biomarker, demographic, and longitudinal health data on a representative UK population, the UKB provides an opportunity for detailed cross-sectional and longitudinal analyses. These analyses could help clarify how inflammation and immune stress contribute to chronic disease progression, ultimately advancing both clinical and public health approaches.

Exposure Variables

Lymphocyte-to-Monocyte Ratio (LMR):

$$\text{LMR} = \frac{\text{Absolute Lymphocyte Count}}{\text{Absolute Monocyte Count}}$$

Systemic Immune-Inflammation Index (SII):

$$\text{SII} = \frac{\text{Neutrophil Count} \times \text{Platelet Count}}{\text{Lymphocyte Count}}$$

Naples Prognostic Score (NPS):
This score is based on a combination of factors, typically including:

- Albumin level (<4 g/dL = 1 point).
- Neutrophil-to-Lymphocyte Ratio (NLR) (>3 = 1 point).
- Lymphocyte-to-Monocyte Ratio (LMR) (<4.44 = 1 point).
- Platelet-to-Lymphocyte Ratio (PLR) (>150 = 1 point).

The final NPS is the sum of these points (range: 0–4).

Neutrophil-Percentage-to-Albumin Ratio (NPAR):

$$\text{NPAR} = \frac{\text{Neutrophil Percentage}}{\text{Albumin Level (g/dL)}}$$

Intended Final Products

1. Exploratory Data Analysis: Descriptive statistics and visualizations of inflammatory biomarkers (e.g., LMR, SII, NPAR) across demographic groups (e.g., age, alcohol consumption, smoking status).

2. Survival Analysis: Modeling mortality outcomes in individuals with fatty liver disease and cirrhosis, assessing the predictive power of inflammatory biomarkers. We plan to include a cohort study design where survival outcomes are calculated for each individual starting from the time they enter the cohort. Survival at 1, 3, 5, and 7 years will be assessed to explore long-term outcomes. We will use a multivariable logistic regression model to evaluate the associations between inflammatory biomarkers and survival outcomes. This addition will enhance the longitudinal scope of our analysis and provide deeper insights into the predictive power of these indices for mortality risks.

3. Predictive Modeling and ROC Analysis: Use ROC curves and Net Reclassification Improvement (NRI) to evaluate the predictive effect of each biomarker index.

Anticipated Data Sources

The UK Biobank (UKB) dataset, which includes: - Biomarker Data: Relevant indices such as LMR, SII, NPAR, etc. - Demographics: Age, gender, ethnicity, socioeconomic status, etc. - Health Outcomes: Mortality data, disease diagnoses, metabolic syndrome indicators, and more.

Planned Analyses/ Visualizations

1. Exploratory Data Analysis:

- Distribution of inflammatory biomarkers by demographic factors (age, gender, socioeconomic status) using visualizations like histograms, boxplots, and heatmaps.
- Correlation analysis between biomarkers and other demographic/clinical variables.

2. Survival Analysis:

- Apply Cox proportional hazards models to explore associations between biomarkers and mortality in patients with liver disease or cirrhosis.
- Kaplan-Meier survival curves and subgroup analyses to investigate the role of each biomarker.

3. Cross-sectional Analysis:

- Logistic regression to assess associations between inflammatory biomarkers and indicators of metabolic syndrome.

4. Prediction and ROC Analysis:

- Assess predictive ability of biomarkers with ROC curves and NRI, and use weighted quantile sum regression to evaluate each factor's contribution.

Coding Challenges

Throughout the project, we anticipate addressing several coding challenges, including efficiently handling the large-scale UKB dataset, managing missing or inconsistent data, and implementing survival analysis and multivariable regression models in Python and R. Optimizing code for computational efficiency and ensuring reproducibility through clear documentation will be key priorities in overcoming these challenges.

Planned Timeline

November 1-8: Data collection, cleaning, and preliminary exploration.
November 9-15: Start survival and cross-sectional analyses on mortality and metabolic syndrome.
November 16-25: Conduct analyses related to additional outcomes (e.g., kidney function) and refine survival models.
November 26-30: Finalize all analyses and begin drafting the report.
December 1-7: Complete the draft report with visualizations and interpretation.
December 8-12: Review and finalize the project for submission.

Project Structure

1. Introduction: Background on the role of inflammatory biomarkers in chronic disease and the relevance of the UKB dataset.

2. Data Cleaning and Exploration: Initial data processing, including distributions of biomarkers and demographic overviews.

3. Association and Survival Analysis: In-depth statistical modeling and correlation analysis between biomarkers, mortality, and other health outcomes.

4. Results and Discussion: Interpretations of key findings and implications for chronic disease management and public health.

5. Recommendations: Suggestions for implementing these biomarkers in routine screenings and health policies.