Bioinformatics Summer School Report

Name: Sulaiman Alalwani
Internship Organization: Duke University
Supervisor's Name: Jay Lusk, Brian Mac Grory, Fan Li
Project title: Data Science to Optimize Cardiovascular Disease Prevention

1. Introduction

Diabetes and obesity are some of the most important risk factors for cardiovascular disease. In recent years, a new class of medications, Glucagon-Like Peptide-1 (GLP-1) agonists, and other drugs with similar mechanisms of action have proven highly effective in protecting patients with diabetes or obesity from cardiovascular events. Our goal in this project is to use a large source of routinely collected healthcare data, provided by Truveta, which contains electronic health record data for millions of patients across the United States.

Using U.S. national electronic health record (eHR) data provided by Truveta, we compare the safety and efficacy of GLP-1 drugs with traditional medications. We hope to find the best ways to distribute GLP-1 drugs, targeting patients who would benefit the most.

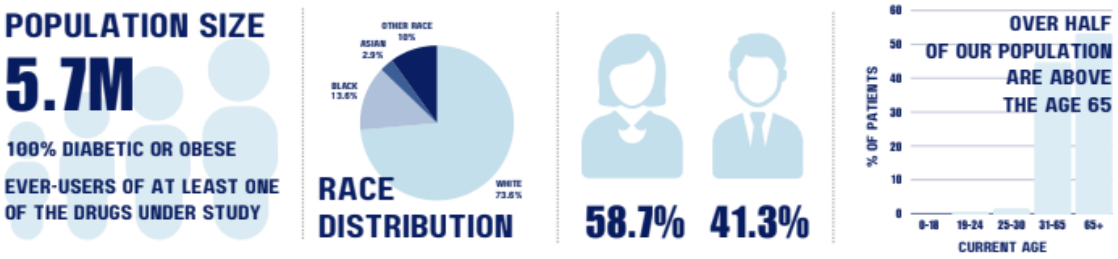
Using Truveta's eHR data, we aimed to address the following research questions:

How do GLP-1 receptor agonists affect—alleviate or worsen—obesity, type II diabetes, and cardiovascular and brain health compared to other available medications?

If there are effects, whether positive or negative, how do they vary across populations based on age, race, sex, and baseline health conditions?

2. Materials and Methodology

We used electronic health records (EHR) data from Truveta, which provides comprehensive health records on over 100 million patients across the United States. Our method is Target trial emulation: is a framework for designing and analyzing observational studies that aim to estimate the causal effect of interventions. We create a Statistical Analysis Plan (SAP) for obesity, type II diabetes, cardiovascular disease prevention, and neurodegenerative diseases. From this population, we selected a subset for our study.
Our final study population has the following key features:



3. Results and Discussion

We don’t have results yet because our project is a Bass Connections project that will last until next summer. But for our future work, we aim to do the following:

- Data Processing: Extract each patient's covariates, outcomes, and baseline features. Handle missing data through imputation or removal to ensure dataset completeness.

- Descriptive Analysis: Generate tables to visualize data, including summary statistics and pre-adjustment covariate balance. Check for multicollinearity before modeling.

- Statistical Analysis: Perform statistical analysis according to our analysis plan.

4. Skills and Knowledge Gained

The most important skills that I gained are:

- Familiarity with routinely collected healthcare data
- Communication
- Teamwork
- Presentation

5. Challenges and Solutions

- Initially, I was not assigned to bioinformatics-related projects, but I managed to adapt and found the experience very informative. Starting from scratch, I was unfamiliar with the research and methodology, but it was fascinating to see how quickly I was able to grasp the concepts.

- The data access was approved quite late, and I only received access in the final two weeks.

6. Conclusion

The experience enhanced my understanding of some diseases and how doctors prescribe different medications in various doses for patients individually according to their conditions. As I prepare my application for medical school and decide if I want to pursue research in the future, the simple answer is yes—I want to be both a clinician and a researcher.

7. Acknowledgments

I would like to extend my heartfelt gratitude to KAUST Academy for their incredible support. It was truly amazing to be cared for by such an outstanding organization. Thank you once again for enabling us to achieve things we never imagined possible.

8. References (if applicable)

[1] Truveta, https://www.truveta.com/.