

Initial Problem Description

Problem 1: Preventive health care for disease deaths linked to modifiable risk factors

Introduction:

1. Of the leading causes of death in the United States (not including unintentional accidents and suicide), diseases and deaths related to lifestyle choices caused over a million deaths in 2019. These lifestyle choices are called modifiable risk factors like obesity, smoking, alcohol consumption and an unhealthy diet. Modifiable risk factors are linked to diseases like heart disease, cancer, cerebrovascular diseases, type 2 diabetes, and nephrosis. In fact, 45% of cancer deaths are linked to modifiable risk factors.
2. Disease burden is the impact of a health problem on a given population and can be measured by indicators like quality of life, financial cost, mortality and morbidity. By measuring disease burden governments can plan for preventive action, identify high-risk populations and plan for future needs.

Motivation:

I'm inspired by a project called [Searching for Health](#). It's a collaboration between the Google News Lab, Alberto Cairo and Schema Design. The project is a visualization of the "top searches for common health issues" in the US and compares that data to the "location of occurrences for those same health conditions" to understand how search data is related to people's live in America.

The problem:

Is there a way to predict when a public health intervention is necessary to prevent the rise of the modifiable factors that inevitably lead to the diseases listed above? By public health intervention, I mean programming done by government, nonprofits and even hospitals to raise awareness before the diseases become major issues in a community/area. We know that health disparities exist by race and gender but is there a way to prevent deaths due to a disease before it becomes an issue. Disease burden, too, is assessed after communities are already high-risk and have had to suffer the consequences of disease. Can we predict that a disease will be an issue in a community before it becomes a burden?

Data: There will be many different datasets that will be required for this. Some of them are:

1. [Google trends data](#) on searches related to smoking, alcohol consumption and diseases.
2. [Data](#) on tobacco purchase.
3. [Nutrition, physical activity and obesity data](#).

Bibliography:

Leading Causes of Death in the United States:

<https://www.cdc.gov/nchs/fastats/leading-causes-of-death.html>

More than 4 in 10 cancer deaths are linked to modifiable risk factors:

<https://www.cancer.org/latest-news/more-than-4-in-10-cancers-and-cancer-deaths-linked-to-modifiable-risk-factors.html>

Kidney diseases linked to modifiable risk factors:

[https://www.medscape.com/viewarticle/742119#:~:text=May%205%2C%202011%20\(Las%20Vegas,Kidney%20Foundation%202011%20Clinical%20Meetings.](https://www.medscape.com/viewarticle/742119#:~:text=May%205%2C%202011%20(Las%20Vegas,Kidney%20Foundation%202011%20Clinical%20Meetings.)

The State of US Health, 1996- 2016: <https://pubmed.ncbi.nlm.nih.gov/29634829/>

Measure of disease burden:

<https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/measures-disease-burden>

CDC National Health Report: Leading Causes of Morbidity and Mortality and Associated Behavioral Risk and Protective Factors—United States, 2005–2013:

<https://www.cdc.gov/mmwr/preview/mmwrhtml/su6304a2.htm>

Problem 2: Vaccine development for COVID-19

Introduction: A key area of research in vaccine development is predicting epitopes. A molecule found on the outside of a pathogen is called an antigen. Antigens are what trigger an immune response in our body. Part of the immune response is production of antibodies that can attach to the antigen and this is handled by a type of cell known as a B-cell. They induce antigen-specific immune responses to produce large amounts of antigen-specific antibodies by recognizing the epitope regions of antigens. The epitope is the part of the antigen to which an antibody attaches to. Being able to predict the molecular structure of the epitope means that we can understand the type of antibodies required to defeat a pathogen like COVID-19 and whether or not the body is able to produce that particular antibody.

Motivation: I am currently taking a class about pathogens and vaccines. I'm extremely fascinated by the vaccine development process and seeing if particular antibodies are effective or not is an important part of that process.

The problem: Predicting epitopes on antigen of SARS-CoV-2 for antigen-specific antibodies.

Data:

1. [Immune Epitope Database and Analysis Resource](#)
2. [UniProt](#)

Bibliography:

Epitope Prediction of Antigen Protein using Attention-Based LSTM Network:

<https://www.biorxiv.org/content/10.1101/2020.07.27.224121v1>