

## **“What Are Your Odds?”:**

### **An Interactive Web Application to Visualize Health Outcomes**

#### **Abstract**

Spreading health knowledge and promoting healthy behavior can impact the lives of many people. Our project aims to create an interactive visualization system to spread knowledge of the effects of health habits and encourage healthy lifestyles. We used survey data about people's health behaviors and analyzed how these different behaviors affect an individual's likelihood of developing certain diseases, such as high blood pressure, high cholesterol, diabetes, and heart attack. Then we created a web application for users to see their chances of getting these diseases and what they can do to reduce these risks. From this project, we find that the aforementioned diseases are, although common, the most avoidable by living healthier. In addition, smoking, gender, eating at least some fruits and vegetables, and exercise intensity have the strongest effects on reducing the risks of these health problems, while drinking and eating more vegetables do not matter much. We also find that to significantly reduce the risks, changing one aspect of lifestyles (quitting smoking or starting exercising) is not enough. To have sufficient effects, one has to develop healthy habits in every aspect: smoking, diet, weight loss, and exercising.

## Introduction

This project analyzes how people's health behavior (diet, smoking, drinking, and exercising habits) can predict their chances of having high blood pressure, high cholesterol, diabetes, and heart attack. The predictive model is then made into a visualization application, available at <https://angiedinh.shinyapps.io/WhatAreYourOdds/>.

## Data and Methods

The data comes from the 2015 Behavioral Risk Factor Surveillance System. This survey data is built from landline and cell phone data from all 50 states, the District of Columbia Guam, and Puerto Rico and there are 441,456 records for 2015. The survey data includes answers to questions relating to people's health behaviors (smoking, drinking, exercising habits, etc) as well as their most recent diagnoses for conditions such as **high blood pressure, high cholesterol, diabetes, and myocardial infarctions (heart attacks)**.

Most of our data cleaning involved converting categorical into the right format (0 for no and 1 for yes) and selecting important variables out of hundreds of survey questions. We then used the data to build a logistic regression equation that predicts one's probability of developing the aforementioned diseases based off of a user's inputted variables for:

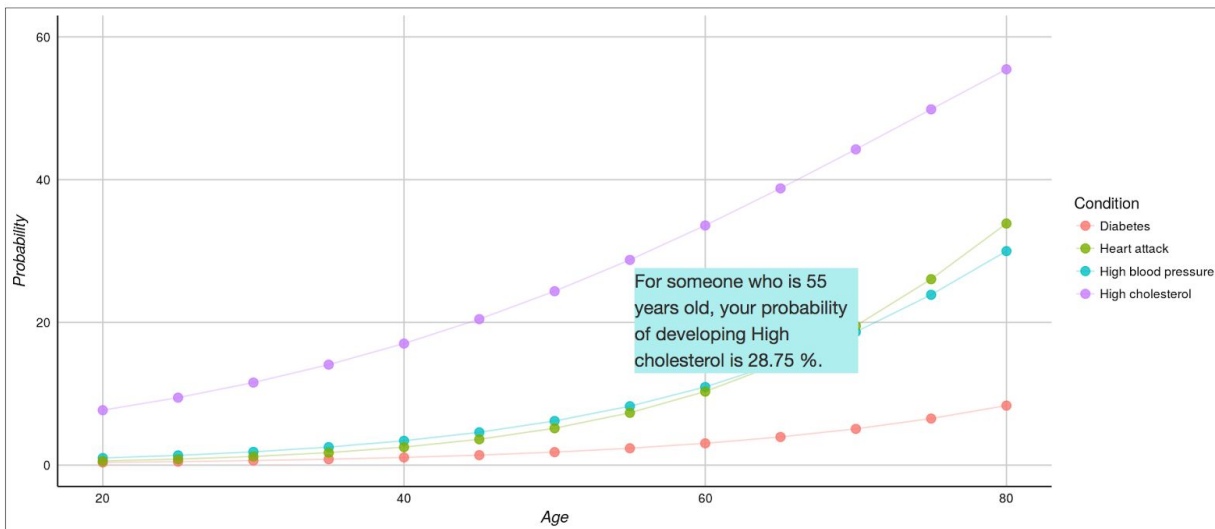
- **Individual characteristics:** Height, weight, sex (male or female)
- **Diet choices:** eating at least one serving of fruit everyday, eating at least one serving of vegetables everyday, number of alcoholic drinks per week
- **Health behaviors:** Smoking habits (everyday, some days, quit, never smoked), exercise intensity (moderate/no exercise or vigorous), exercise duration

We chose the logistic regression model because the goal of this project is to allow users to see their chances of health problems and what they can do to reduce the risks. The logistic model gives us that interpretable probability output. We chose our independent variables based on statistical significance (if the p-value of that variable is less than 0.05), practical significance (if the effects are high enough to matter and be visualized), and importance (if the variable is interesting and is something that the user can change). Then we extracted the coefficients of this regression and use that model to predict new inputs from the user. The output are the probabilities of getting the aforementioned diseases, which are visualized in a line graph, as discussed in the next section.

## Results

We find several interesting results. First of all, not all diseases are results of lifestyles: cancer does not correlate much with health habits. Of all diseases in the survey, high blood pressure, heart attack, high cholesterol, and diabetes have the strongest relationships with lifestyles. They are diseases that can be avoided by living healthier.

The plot below shows the probabilities of an example men to have these diseases over his lifetime. This men is 5'7 tall, weighs 170 lbs, smokes everyday, drinks 13 drinks per week, does not eat fruit or vegetables, and exercise moderately for 6 minute each time.



*Figure 1: Probabilities of having four diseases of an example men*

We notice that some diseases are more common than others: most people have a fairly high risk of high cholesterol but a low risk of diabetes. Also, health problems are most likely to occur at age 60-80. Therefore, changing lifestyles may not matter at young age but can impact one’s life down the road.

*Table 2: Odd ratios of the four aforementioned logistic regressions. All variables are statistically significant. Most practically significant variables are in bold.*

Coefficient	Odd ratio			
	Blood Pressure	Diabetes	High Cholesterol	Heart Attack
Intercept	0.7387	0.2188	1.8239	0.0140
Age	1.0644	1.0545	1.0461	1.0775
<b>Male</b>	<b>1.4245</b>	<b>1.5972</b>	<b>1.4423</b>	<b>2.5871</b>
<b>Height</b>	<b>0.9121</b>	<b>0.9104</b>	<b>0.9344</b>	<b>0.9436</b>
Weight	1.0003	1.0003	1.0001	1.0001
Smoke everyday	Reference			
Smoke some day	1.0450	0.9534	0.8572	1.0418
<b>Quitted smoking</b>	<b>0.8157</b>	<b>0.8107</b>	<b>0.8761</b>	<b>0.6711</b>
<b>Never smoked</b>	<b>0.7024</b>	<b>0.6573</b>	<b>0.7208</b>	<b>0.4155</b>
Number of drinks per week	1.0000	0.9993	1.0000	0.9998
<b>Eat at least one piece of veggies</b>	<b>0.8862</b>	<b>0.8327</b>	<b>0.8452</b>	<b>0.9170</b>
<b>Eat at least one piece of fruit</b>	<b>0.8518</b>	<b>0.8559</b>	<b>0.9629</b>	<b>0.8744</b>
<b>Moderate or no exercise</b>	<b>1.2807</b>	<b>1.9937</b>	<b>1.2035</b>	<b>1.7578</b>
Vigorous exercise	Reference			
Duration of workout	0.9994	0.9993	0.9995	0.9989

Table 2 tells us that drinking alcohol does not have much effect on the risks of having these four diseases, but smoking does. Quitting smoking decreases the odds of having blood pressure by 18%, diabetes by 19%, high cholesterol by 12%, and heart attack by 33%. People who never smoked have an even lower risk of having these health problems. However, reducing smoking frequency from everyday to some days only lowers the risk of high cholesterol, which means that smoking is harmful no matter how often you do it. Most people associate drinking with liver diseases and smoking with lung diseases, but this research shows that smoking is also correlated with other seemingly irrelevant health problems.

Also, people who eat at least one serving of vegetables everyday have 10-15% lower odds of having these diseases. The same is true for people eating fruits. However, once you already eat fruits and vegetables, increasing the amount of green consumption does not significantly matter. Additionally, people who exercise vigorously have 20-75% lower odds of having these diseases. Increasing exercise duration does not affect much unless the increase is large enough. For this example men, raising workout duration from 0 to 100 minutes reduces the probability of getting high cholesterol at the age of 80 by almost 2%. Gender and body type also play a role in these health risks. Taller people (with the same weight) have lower odds. Losing weight reduces one's odds, but only significantly does so if the weight loss is large enough. Men have much higher odds (42-258%) of having these diseases than women.

To sum up, to reduce the risks of having these four diseases requires a complete change of lifestyle; changing only one habit does not have much effect. However, if the aforementioned example men does everything: quitting smoking, starting to eat fruits and vegetables, and working out vigorously for 90 minutes each time, he reduces his probability of having high cholesterol by 15%, heart attack by 17%, blood pressure by 13%, and diabetes by 5%, at the age of 80.

### **Discussion of Visualization Techniques**

We use Shiny for our visualization technique. In our app, the sliders include questions about individual's height, weight, gender, number of servings of fruit, vegetables and alcoholic drinks per week, smoking habit, exercise intensity and exercise duration. The plot's x-axis specifies user's age (0-80), and the y-axis, with the hovering feature, shows the user's probability of developing diseases, calculated using their answers to the above questions. As a user mouses over the plot's points, a sentence appears summarizing the plot's output. We felt that a line-graph with dots most easily encapsulated the information we wanted to share with the user, both the probabilities at each dot and the overall trend. The app interface is shown in the Appendix section. With our visualization, users can gain insights and see for themselves how they should change their health behaviors to achieve a certain level of overall health.

### **Evaluation**

While we are pleased with our final product, there still remain several aspects that we would like to change in our visualization. We received feedback suggesting that our visualization could be improved by having a more interactive line graph. For instance, including a feature so that users could drag points to a certain probability and receive some recommended steps the user could take to reach that targeted probability. In addition, we think that that survey respondents can change their health habits after knowing that they have diseases, which underestimates our coefficients. Therefore, we agree that using a longitudinal dataset - a dataset that includes survey responses from the same individuals over time - would give us a more accurate depiction of how a lifetime of health behaviors can impact our health prognosis across our lives. However, such a dataset proved difficult to find.

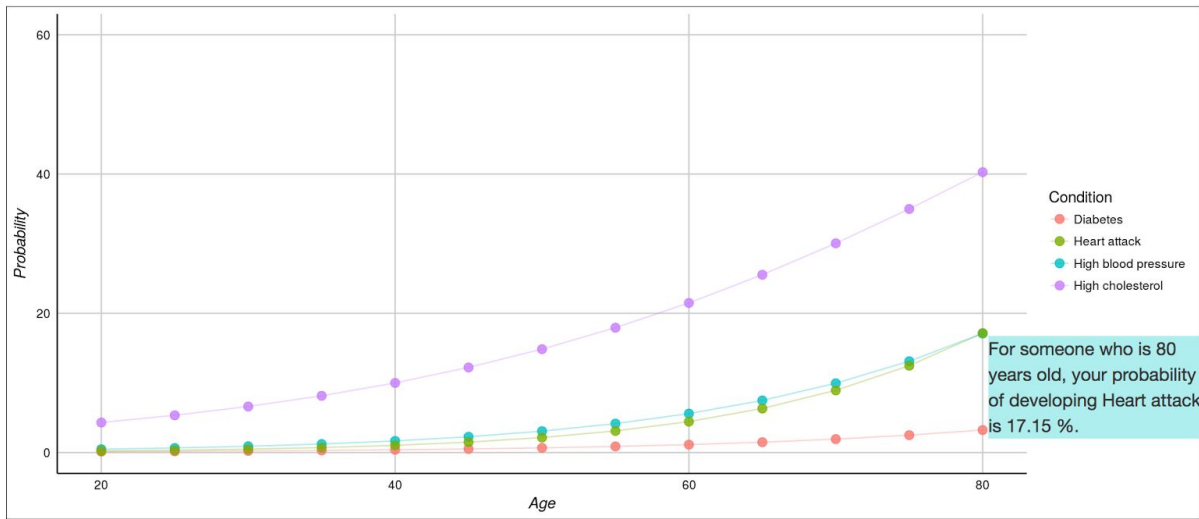
### **Future Directions**

There is a bright future for easy-to-use, interactive, and informative health applications. As individuals continue to focus on healthy lifestyle choices and their impact, the importance of applications like ours becomes increasingly evident. Going forward, modifications such as integrating additional variables and longitudinal data, would increase the interpretability and impact of this application.

### Appendix:

Our app interface, showing health outcomes of the aforementioned example men once he changes his lifestyles: quitting smoking, eating fruits and vegetables, and working out vigorously.

## What Are Your Odds?



[How It Works](#)

How do you compare?

#### How tall are you?

Ft

5

In

7

#### How much do you weigh? (lbs)

170

#### Are you:

Male

#### Think back over the last week:

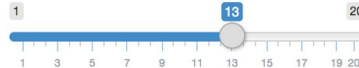
Did you eat at least one serving of fruit every day?

Yes

Did you eat at least one serving of vegetables every day?

Yes

How many alcoholic drinks did you have?



#### Select the smoking habit that most resembles yours

Quit

#### In the last week,

When you exercised, what was the intensity?

Vigorous

#### How long was your workout?

