

## Executive Summary

As a result of excessive prescriptions and a history of substance abuse, the United States has today found itself in the midst of an opioid addiction crisis. The U.S. government has throughout the past decades recognized the ramifications of drugs on health and society at large and instituted regulations on the use of such substances. However, these substances have continued to be abused by many. Over the course of analyzing the drug crisis, we created models to predict the scale, causes, and costs of drug addiction in our society.

In recent years, the consumption of e-cigarette and vape products has skyrocketed, especially among adolescent Americans. To better understand the scale of this consumption and its effect on nicotine consumption (assuming no policy changes to combat usage), we created a logistic growth model to predict the proportion of teenagers who used nicotine-based vape products over the next decade. Based on the model we fit, we concluded that 36.18% of teenagers in the year 2029 would use vape products, or approximately 6.1 million teens. Additionally, we modeled the growth of the adult vaping population by adding the portion of teens coming of age who vaped to the existing number of vaping adults, finding that 23.6 million adults would be using vape products in the year 2029, amounting to a total of 29.7 million total vape users in that year.

While it is impossible to definitively predict substance abuse by any person, there exist metrics documented by government and nonprofit organizations such as the Substance Abuse and Mental Health Services Administration (SAMHSA), which detail percentages of demographics that engage in drug-related activities. This information, within which exist striking trends, served a central role in our solution. We calculated weight values, ratios between a groups measure and the average measure of all groups. Combining these weight values of all the measures into a product gives a measure of how likely a person is to become addicted. We used this approach to test ten people with randomly assigned characteristics (e.g. income, residence population density) for their tendency to become addicted to the following: marijuana, nicotine, alcohol, and opioids. We found that at our hypothetical high school of 300 students, there would be 6 opiate users, 26 marijuana users, 162 nicotine users, and 34 alcohol users.

To determine a metric to rank nicotine, opioid, marijuana, and alcohol based on their ramifications to financial and non-financial matters, we quantified factors of medical cost, employment, drug costs, and life expectancies into U.S. dollars. We made per capita values in dollars for the cost of being addicted to each drug and determined that marijuana is the least harmful drug, followed by nicotine, alcohol, and opioid at the worst.

# Contents

|  |           |
|--|-----------|
| <b>Executive Summary</b>   | <b>1</b>  |
| <b>1 Introduction</b>  | <b>2</b>  |
| 1.1 Background . . . . .   | 2         |
| 1.2 Restatement of Problem . . . . .                             | 3         |
| <b>2 DARTH Vapor - The Spread of Vaping</b>                      | <b>3</b>  |
| 2.1 Defining Variables . . . . .                                 | 3         |
| 2.2 Assumptions . . . . .  | 4         |
| 2.3 Development of the Model . . . . .                           | 6         |
| 2.3.1 Teenage Proportion Growth . . . . .                        | 6         |
| 2.3.2 Adult Vaping Population Growth . . . . .                   | 7         |
| 2.3.3 Combining Teen and Adult Models . . . . .                  | 8         |
| 2.4 Results of the Model . . . . .                               | 8         |
| 2.5 Model Assessment and Analysis . . . . .                      | 9         |
| 2.5.1 Strengths . . . . .  | 9         |
| 2.5.2 Weaknesses . . . . .                                       | 9         |
| 2.5.3 Sensitivity Analysis . . . . .                             | 9         |
| 2.5.4 Comparison to Cigarettes . . . . .                         | 10        |
| <b>3 Above or Under the Influence? - Probability of Drug Use</b> | <b>10</b> |
| 3.1 Defining Variables . . . . .                                 | 10        |
| 3.2 Assumptions . . . . .  | 12        |
| 3.3 Development of the Model . . . . .                           | 12        |
| 3.3.1 M3 “High” School . . . . .                                 | 13        |
| 3.4 Results of the Model . . . . .                               | 13        |
| 3.5 Model Assessment and Analysis . . . . .                      | 15        |
| <b>4 Ripples - The Cost of Substance Abuse</b>                   | <b>15</b> |
| 4.1 Defining Variables . . . . .                                 | 15        |
| 4.2 Assumptions . . . . .  | 16        |
| 4.3 Development of the Model . . . . .                           | 16        |
| 4.4 Results of the Model . . . . .                               | 17        |
| 4.5 Model Analysis . . . . .                                     | 17        |

## 1 Introduction

### 1.1 Background

Drug abuse has contributed to growing issues in the United States such as automobile accidents, violence, child abuse, crime, unemployment, and homelessness. Drugs alter brain function and in cases such as driving, this alteration can impair the drivers ability to make

the appropriately rapid decisions necessary to be a safe driver. In 2010, a study revealed that 11% of crashes that resulted in a death involved a driver that was under the influence of a drug [3]. Addictions to substances such as alcohol, nicotine, marijuana, and opioids also impact the quality of life of an individual through health and financial factors. Substance abuse impacts over 23.6 million Americans [1] and nationally costs more than \$200 billion in healthcare [2] and \$540 billion in other expenses due to crimes and lost productivity [2]. With the recent opioid crisis, 2.1 million people have been reported to have an opioid use disorder in 2017 with overdosing accounting for 47,600 deaths [4].

The impact of substance abuse has ramifications that extend to society and therefore the race to curb substance abuse has escalated. Regulations from the FDA have increased on drugs as well as products that act as vessels, like vape pens; however, as regulations increase so does opposition. As more research is done about the impacts of drugs, it becomes increasingly imperative to determine methods of regulation based on personal health risks presented as well as societal impacts.

## 1.2 Restatement of Problem

In order to address the growing addiction to substances such as tobacco, alcohol, and narcotics, we were tasked to create solutions for the following problems:

- Generate a model to predict the impact of vaping on new usage of nicotine over the next 10 years and compare this trend to past cigarette trends.
- Generate a model that considers internal and external factors and use the model to simulate the likelihood that a given individual will use a given substance. Then use the model produced to predict how many students of 300 seniors are likely to use the following substances:
  - Nicotine
  - Marijuana
  - Alcohol
  - Unprescribed Opioids
- Develop a comprehensive metric for the impact of substance use taking into consideration financial and non-financial factors and rank the substances listed above in terms of the severity of their impact.

## 2 Darth Vapor - The Spread of Vaping

### 2.1 Defining Variables

- **Input:** Years

**Explanation:** We track the percentage of teen vape usage over time in years.

- **Output:** Proportion of teens who vape

**Explanation:** Through our logistic model, we mathematically predict the proportion of American teens who vape in a specific year,  $t$ .

- **Output:** Number of teens who vape

**Explanation:** We multiply our models prediction of the proportion of teen vape usage for a specific year,  $t$ , by the population of total teens in the United States to calculate the population of teens who vape.

- **Output:** Number of adults who vape

**Explanation:** We add the number of existing adults who use vape products to the sum of teens using vape products that transition into adulthood starting from a specific year,  $t_0$ , to calculate the total population of adults that vape.

- **Output:** Total number of Americans who vape

**Explanation:** We find the sum of the number of teens who vape and the number of adults who vape to determine the total number of Americans that vape.

## 2.2 Assumptions

In order to simplify the model and calculations, the following assumptions were made:

1. People who switch to vaping from cigarettes do not contribute to additional nicotine consumption.

**Justification:** Predictions are to be made in relation to the spread of nicotine usage due to vaping. For those who smoked in the past and later transitioned to vaping, the nicotine usage for these individuals remains the same simply their method of intaking the substance has altered.

2. Teens<sup>1</sup> who vape did not smoke before.

**Justification:** In order to simplify the calculation of how vaping impacts nicotine usage, we assumed that all teens were not previously exposed to cigarette usage and teens who vape are considered newly introduced to nicotine use.

3. Teen vape usage is equivalent to the average of vape usage among 10th and 12th graders.

**Justification:** The assumption made is that the average of data from 10th graders and 12th graders provides a fair average for high school usage. Data provided includes the statistics for 8th, 10th, and 12th graders, thus giving enough of a trend for high

---

<sup>1</sup>We use “teens” herein to refer only to those aged 14-18 (i.e. high school aged adolescents).

school vaping. Also considering the missing 9th and 11th grade data, it simplifies the process to use the 10th and 12th grade data provided to draw necessary trends.

4. E-cigarette statistics are analogous to vaping statistics.

**Justification:** Starting in the years 2015-2018, we used Monitoring the Future data for vaping tobacco because in both data sets included 2015 in which the usage of e-cigarettes and vaping were very close which implies that the title E-cigarettes versus vaping are interchangeable.

5. Teen vape trends will level off at a carrying capacity similar to teen cigarettes trends from the the past, 36.4%.

**Justification:** In order to create a model of trends in vaping among teens, it is necessary to determine a trend line that would potentially fit the data. Using the given data on teen cigarettes and realizing that the usage reaches a carrying capacity, it can be stipulated that teen vape usage will follow a similar trend.

6. The only growth of the proportion of adults who vape will come via teen vape users becoming

**Justification:** Though there is presumably some amount of existing adults who take up vaping nicotine, historical data of adult e-cigarette/vaping usage was not forthcoming beyond the most recent year of data. As a result, we could not model this kind of growth accurately. Since vaping is predominantly a youth phenomenon anyway, we opted to assume that only adults who took up vaping as teenagers would add to the existing population of adult vape users.

7. The 14-18 year old population will remain relatively constant at 16.77 million teens in the US. 4.21 million Americans enter adulthood every year.

**Justification:** In order to model predictions of nicotine spread among teens, a trend in teen population must be established. According to the Department of Health and Human Services, the teen population will remain relatively constant over the next several decades [6]. We then calculated that their approximately 16.77 million high school age children currently in the United States [7].

8. No policy initiatives will be undertaken to reduce the proportion of teens who vape.

**Justification:** Though this is not true as many organizations, including several levels of government, are planning to undertake efforts to reduce teenage vaping, we found that predicting the effects or even existence of such efforts would be arbitrary in nature. We thus chose to predict the proportion assuming the status quo holds. We feel that this assumption makes our model less arbitrary and in fact helps it to serve as a useful guide to policy makers in determining the success of certain initiatives compared to taking no action at all.

## 2.3 Development of the Model

The below steps summarize the approach we took to modeling this problem:

1. We first model the growth of the proportion of teens who use nicotine through vaping over time
2. We then multiply the proportion of teens who vape by the total number of teens in the United States (which we assumed to be constant at 16.77 million) to calculate the total number of teens who vaped nicotine/tobacco.
3. Next, we calculate the existing number of adults who vape in the United States based on the proportions provided to us.
4. To calculate the growth of the adult vaping population, we multiply the number of teens entering adulthood by the proportion of teens that year who use vape products. We add this figure to the existing vaping adult population to calculate the new total adult vaping population.

### 2.3.1 Teenage Proportion Growth

As the population of greatest concern, we first modeled the growth of teenage nicotine consumption through vaping. We chose to model the proportion of teens who vape as a logistic growth model<sup>2</sup>.

We chose logistic growth because it effectively fits the characteristics of the real-life situation: it begins with a slow period of growth, followed by a period of very quick growth, and finally, a slowing of growth as the proportion reaches some carrying capacity. The growth of e-cigarettes and vape products similarly began slowly as the products gained traction and then underwent a sustained period of quick growth among teenagers. Of course, as there is a finite proportion of teens willing to vape under any circumstances, the growth would eventually slow drastically as that upper limit was approached.

We fit our logistic model [11] with a carrying capacity of the proportion  $p_{v,T} = 0.364$  or 36.4%, where  $p_{v,T}$  is the proportion of teens at some given time who vape regularly. We chose this (see Assumption 5) as this was the proportion of teens who consumed cigarettes at its peak and therefore served as a non-arbitrary upper limit to the proportion of teens who would vape regularly.

Figure 1 shows the logistic growth model created to predict the proportion of teens who will use vape products in some given year. The logistic growth function can be defined as:

$$p_{v,T}(t) = \frac{36.4}{1 + e^{-0.42179(t-2016.933)}} \quad (1)$$

---

<sup>2</sup>The data that the model was fit to was obtained from two combined sources: the 2016 Surgeon General's report regarding e-cigarette use among young adults [10] and the annual Monitoring the Future survey [8]. We did so as they covered different periods of time and showed similar statistics in their overlapping proportions, giving us confidence that we could safely combine them.

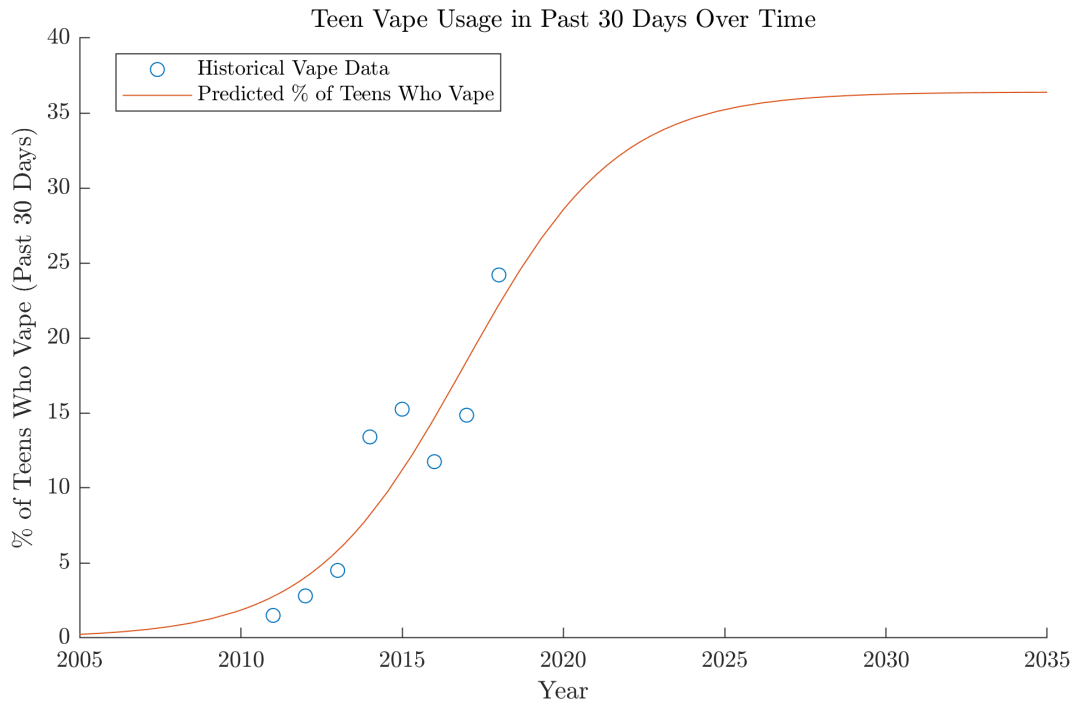


Figure 1: A graph showing the predicted proportion of teens who use vape products over time. The red line denotes the logistic growth model created for prediction purposes.

where  $p_{v,T}$  is the proportion of teens who vape in some year  $t$ .

### 2.3.2 Adult Vaping Population Growth

In order to accurately capture the use of vape products of the whole population, we must also consider adult usage of the products. We were not able to find historical data of e-cigarette or vape product usage for adults other than for the most recent year, so we chose to assume that the only addition to the adult vaping population would come from teens entering adulthood (see Assumption 6).

As such, the total number of adults using vape products can be calculated by adding the existing number of adult users to the sum of teens using vape products that have entered adulthood since the starting year. In 2017, the CDC reported that the proportion of adults in the United States who use e-cigarettes was 2.8%, or 6.93 million people [12], and, as established in Assumption 7, 4.11 million teens enter adulthood annually. Thus, the total number of adults who use vape products can be modeled as

$$P_{v,A}(t) = 6.93 + \sum_{t_0=2017}^t 4.11 \times \frac{36.4}{1 + e^{-0.42179(t-2016.933)}} \quad (2)$$

where  $P_{v,A}(t)$  is the predicted number of adults (in millions) who vape in some given year  $t$ .

| Years | Predicted % | Actual % | Years | Predicted | Actual   |
|-------|-------------|----------|-------|-----------|----------|
| 2013  | 5.821       | 4.5      | 2013  | 0.9761    | 0.75465  |
| 2017  | 18.46       | 14.85    | 2017  | 3.095     | 2.490345 |
| 2021  | 30.85       | -        | 2021  | 5.174     | -        |
| 2025  | 35.23       | -        | 2025  | 5.908     | -        |
| 2029  | 36.18       | -        | 2029  | 6.067     | -        |

Table 1: The first table shows predicted proportion vs actual proportion of teens using vaping products for the given years. The second table gives the predicted vs actual number of teens using vaping products in millions of people.

| Years | Adults | All   |
|-------|--------|-------|
| 2017  | 7.689  | 10.78 |
| 2021  | 12.1   | 17.27 |
| 2025  | 17.7   | 23.61 |
| 2029  | 23.61  | 29.67 |

Table 2: The predicted number of adults and predicted number of all Americans who use vape products in millions of people for the given years. Calculated using Equation 2 and 3.

### 2.3.3 Combining Teen and Adult Models

To obtain the total number of Americans who use nicotine through vape products, we can combine Equations 1 and 2 as such:

$$P_{v,USA}(t) = 16.77 \times \frac{36.4}{1 + e^{-0.42179(t-2016.933)}} + 6.93 + \sum_{t_0=2017}^t 4.11 \times \frac{36.4}{1 + e^{-0.42179(t-2016.933)}} \quad (3)$$

where  $P_{v,USA}(t)$  is the total number of Americans who use vape products regularly in millions.

## 2.4 Results of the Model

Table 1 shows the predicted and (when present) the actual proportion and number of teens who use vape products in various years based on Equation 1. Table 2 shows the predicted number of adults and Americans as a whole who use vaping products for the given years.

We then conclude based on our models that 10 years from now (2029), 36.18% of American teenagers, or 6.067 million teenagers, would be using vaping products (assuming no intervention in the current trends, of course). Additionally, we conclude that 23.61 million American adults would be using vape products, amounting to a total of 29.67 million Americans vaping in the year 2029.



## 2.5 Model Assessment and Analysis

### 2.5.1 Strengths

- Our logistic model accurately reflects the pattern of growth that would take place in the proportion of teens who vape.

**Explanation:** Because there is a limited population and because there will always be a portion of that population that will not intake a substance, it logically follows that a trend line for a substance like nicotine will eventually level out at an effective maximum after a period of growth.

- The model does not make any arbitrary assumptions about turning points/extrema or the undertaking of attempts to reduce nicotine consumption in the trend line.

**Explanation:** The model does not depict any attempt made by outside forces to reduce the use of vape products in the general population. This is a strength because it avoids arbitrary conclusions and serves as a useful guide for policy makers to analyze a scenario in which no action is taken and from there conclude what actions are necessary and effective.

### 2.5.2 Weaknesses

- We assume a potentially incorrect "carrying capacity" upper limit for our logistic growth matter.

**Explanation:** Because the trend line predicts future values, it follows that there is no way to know the carrying capacity of vaping until it is potentially reached. Thus, we used the maximum proportion of teen cigarette use as a guideline of what the carrying capacity for vaping might be in the future.

- Possible growth for adult nicotine usage due to existing adults beginning to vape is not factored in to the prediction models.

**Explanation:** Any increase in vaping in adult populations is attributed to teens who vape and become adults once they turn 18. The current number of adults who vape is included in our model, but no additional growth from solely adult populations is included as we could not find enough data to predict it accurately. With this data, we could easily implement this into our model, however.

### 2.5.3 Sensitivity Analysis

We created three related models in this part. The first one was the logistic growth model for the proportion of teens who would use vape products in a given year. Given the small amount of data that the model was trained on, it would be fairly variable if the inputs

were changed slightly. However, given the defined carrying capacity value and the relatively strong fit ( $RMSE = 3.01$ ), the model fits the data fairly strongly.

The second two models were the ones predicting the number of adults and Americans as a whole using vape products. Because these models are both dependent on a summation of the first model, they are largely more sensitive and thus minor variations in the data that the initial logistic model was trained on could have a great impact on the number of people that these models compute.

#### 2.5.4 Comparison to Cigarettes

Both the trends in the percentage of teens who vape and the adult per capita cigarette consumption have trends of rapid growth in usage when first introduced. Both vape and cigarettes contain nicotine and thus consumption of both products would logically have similar trends in the increase of usage at their introduction. In the graph of adult per capita cigarette consumption (see Figure 2), the rapid growth is followed by a trend of decreasing consumption. It can be presumed that this decrease is a result of research and publications about the consequences to health due to smoking and increasing regulations on cigarettes.

In our model of the percentage of teens who vape, the trend shows that consumption increases until it reaches a carrying capacity, but never decreases. The model does not show any decrease because it does not attempt to predict the effect of future research on health concerns or possible regulations on vaping. Instead, our model serves as a prediction for the trend that vaping will follow if no such regulation is put into place and therefore acts as a precaution for regulation agencies on the future of vaping should it go unchecked.

### 3 Above or Under the Influence? - Probability of Drug Use

#### 3.1 Defining Variables

- **Input:** Income

**Explanation:** The income bracket of the individual based off of the poverty line (20k). The brackets we use are as follows:

Less than \$20,000

\$20,000 - \$40,000

More than \$40,000

Because income correlates to substance abuse (inversely, most often), we found it to be a strong predictor of use of certain drugs.

- **Input:** Race

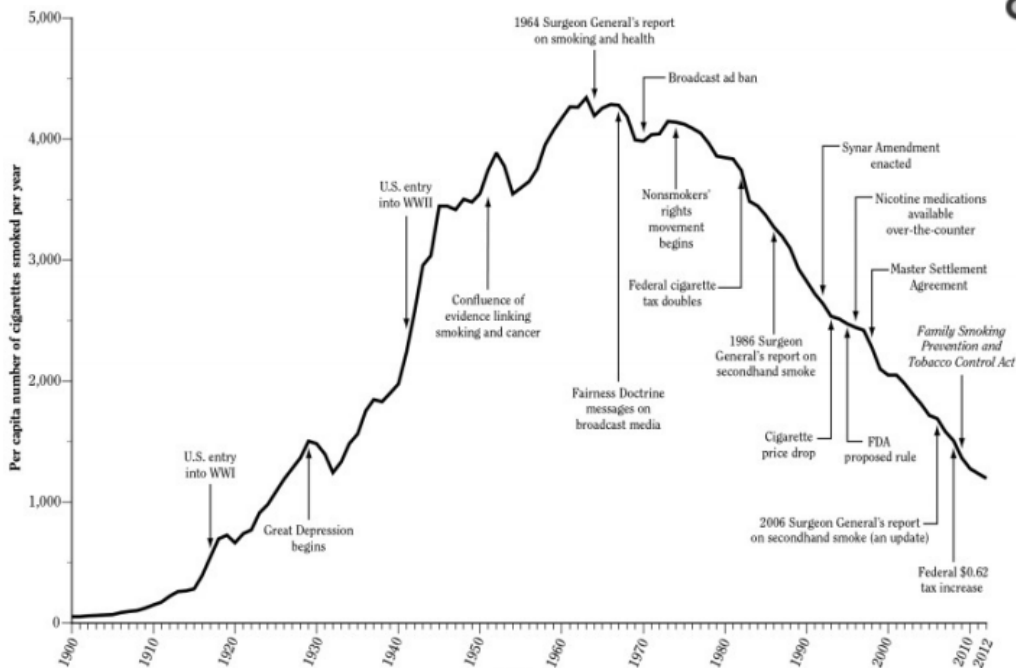


Figure 2: A graph of cigarette consumption in the United States since 1900 [10]. Note the rapid rise in cigarette consumption that is analogous to the fast growth portion of our logistic model.

**Explanation:** The race of the individual. The four categories of races considered for our model are White, Black, Hispanic, and Asian. We include race to serve as a crude indicator of genetic predispositions to substance abuse across humankind.

- **Input:** Gender

**Explanation:** The gender of the individual (male or female).

- **Input:** Urbanization/population density of location

**Explanation:** The density of population (urbanization) of the residence for the individual. We grouped this variable into the following three categories:

Urban - large metro (50,000 or more people)

Suburban (urban clusters) - small metro (2,500 - 50,000 people)

Rural - rural nonmetro (less than 2,500 people)

Definitions courtesy of the Census Bureau [5].

- **Input:** Substance

- The substance that we are determining the probability of usage for. The four drug categories we are analyzing in this section are nicotine, marijuana, alcohol, and un-prescribed opiates.

- **Output:** Probability that a given person with traits will use a specific substance.

**Explanation:** We will additionally determine the predicted number of people in a diverse high school who would use specific substances.

### 3.2 Assumptions

- Use means used the substance at least once a month.

**Justification:** For the simplicity of creating a model that can predict use of a substance there needs to be an assumption made on the definition of using a substance. For the purposes of the question we have assumed that to be considered use of a substance, the individual users it at least once a month. As a result of this assumption, we opted to use Monitoring the Future’s data of drug use at least once a month [8]

- The national average proportion of users of a certain drug can be used as a baseline probability that a given person would use it.

**Justification:** In order to predict the probability that someone would use a substance and determine the number of people within a demographic that would use a substance, a baseline must be established. From this baseline, we will make adjustments based on a person’s specific traits to determine their true probability of using a certain substance.

- The input variables are not correlated with each other.

**Justification:** Though this is not true (e.g. race *is* correlated with income) in real life, it is impossible for us to quantify the exact amount of correlation and factor it out of our predicted probabilities. We therefore assume that it does not exist, leading to over/underestimation in some cases.

### 3.3 Development of the Model

We took a weighted probability approach to this problem. First, we compiled data regarding the proportion of people with different traits (e.g. race, gender, etc.) who use the substances in question from the extensive 2016 National Survey on Drug Use and Health conducted by the Substance Abuse and Mental Health Services Administration [13].

Each of the traits of a given person were assigned a weight value  $R(i, j)$ . This value represents the proportionally greater (or lesser) likelihood that a given person with trait  $i$  is to use a certain substance  $j$ . The equation of weight  $R(i, j)$  can be given in general form as

$$R(i, j) = \frac{p_{i,j}}{p_{USA,j}} \quad (4)$$

where  $p_{i,j}$  is the proportion of people with trait  $i$  who use substance  $j$  and  $p_{USA,j}$  is the proportion of all Americans who use substance  $j$ . For example, when examining the female

| Drug Type | $P(j)$ |
|-----------|--------|
| Opioids   | 0.014  |
| Marijuana | 0.089  |
| Nicotine  | 0.107  |
| Alcohol   | 0.242  |

Table 3: The baseline probabilities for each substance type analyzed in this part. Each is derived from the proportion of all Americans who use the specified substance.

gender’s use of opiates, the weight can be calculated as

$$R(\text{female}, \text{opiate}) = \frac{p_{\text{female}, \text{opiate}}}{p_{USA, \text{opiate}}} = \frac{0.013}{0.014} = 0.929 \quad (5)$$

After calculating weights for all four of the traits we analyzed (race, gender, income, urbanization of location), we then proceeded to multiply each of these trait weights to the baseline probability (i.e. the proportion of all Americans who use a given substance  $j$ ), such that

$$P(i, j) = P(j) \times \prod_{i_c=i_1}^{i_4} R(i_c, j) \quad (6)$$

where  $i$  is the set of a person’s traits,  $P(i, j)$  is the probability that a person with given traits  $i$  will use substance  $j$ , and  $P(j)$  is the baseline probability that a person would use substance  $j$ .

### 3.3.1 M3 “High” School

In order to analyze the various probabilities of substance abuse across different types of people and predict probabilities among an entire high school of 300 students, we generated ten fictional high school students with randomly selected traits using a Matlab script. While creating 300 randomly generated students would be ideal, we were unable to write a program to calculate each person’s  $P(i, j)$  automatically due to time constraints, so we limited ourselves to ten people who we assumed to be a representative sample of the 300 student high school, whereby each person represents thirty people at the school. Table 4 shows the traits of these students.

## 3.4 Results of the Model

Table 5 shows the probability that each generated student will use each drug. The variance is quite high among various students; Student 2, for example, has a 91.64% of using marijuana based on his traits versus just a 12.93% chance for Student 9. We can see that, regardless of each student’s predispositions, certain drugs such as marijuana have higher probabilities by virtue of their higher baseline probabilities.

| Person | Geography | Gender | Income     | Race     |
|--------|-----------|--------|------------|----------|
| 1      | Suburban  | Female | <20k       | White    |
| 2      | Suburban  | Male   | <20k       | White    |
| 3      | Suburban  | Female | $\geq$ 40k | Hispanic |
| 4      | Suburban  | Female | $\geq$ 40k | Black    |
| 5      | Suburban  | Female | <20k       | White    |
| 6      | Rural     | Male   | 20k-40k    | Black    |
| 7      | Urban     | Male   | 20k-40k    | White    |
| 8      | Urban     | Male   | <20k       | Hispanic |
| 9      | Suburban  | Female | $\geq$ 40k | Asian    |
| 10     | Suburban  | Female | <20k       | Asian    |

Table 4: The ten randomly generated students created to represent the 300 students of M3 High School. M3 High is quite the magnet apparently, given that it has students from rural, urban, and suburban areas.

| Person | $P(i_P, \text{Opium})$ | $P(i_P, \text{Marijuana})$ | $P(i_P, \text{Nicotine})$ | $P(i_P, \text{Alcohol})$ |
|--------|------------------------|----------------------------|---------------------------|--------------------------|
| 1      | 0.0224                 | 0.089                      | 0.7862                    | 0.105                    |
| 2      | 0.0702                 | 0.1475                     | 0.9164                    | 0.0553                   |
| 3      | 0.0117                 | 0.0497                     | 0.214                     | 0.0892                   |
| 4      | 0.0102                 | 0.0717                     | 0.4323                    | 0.157                    |
| 5      | 0.0224                 | 0.089                      | 0.7862                    | 0.105                    |
| 6      | 0.0102                 | 0.0949                     | 0.9302                    | 0.1325                   |
| 7      | 0.0169                 | 0.1329                     | 0.6875                    | 0.1883                   |
| 8      | 0.0237                 | 0.1419                     | 0.3298                    | 0.1577                   |
| 9      | 0.0008                 | 0.0213                     | 0.1293                    | 0.0892                   |
| 10     | 0.0015                 | 0.0327                     | 0.1996                    | 0.0533                   |

Table 5: The probability that each of the ten people will use each of the four drug types.  $i_P$  denotes the traits of a person in some row  $P$ .

| Drug Type  | Opium | Marijuana | Nicotine | Alcohol |
|------------|-------|-----------|----------|---------|
| # of Users | 5.7   | 26.12     | 162.34   | 33.98   |

Table 6: The estimated number of users for each type of drug type.

Table 6 shows the predicted number of users of each substance. Of course, as we cannot have part of a person, we round and thus conclude that there will be: **6** users of opiates, **26** users of marijuana, **162** users of nicotine, and **34** users of alcohol.

### 3.5 Model Assessment and Analysis

Drug abuse is a complicated problem within countless contributing factors and effects, many of which our problem considered to an exceptional extent. Income, race, residence population density, and gender, based on statistical data, were considered the four most influential variables in substance abuse. We documented the relative percentages of the demographics and assigned them into separate bins, such as differing races, ranges of income, and genders, dividing the percentage of each demographic which abused each of the drugs. This gave us four ratios between the percentage of the group which abused a substance and the percentage of the entire population which abused that substance; we call these measures weights. To assess the likelihood of substance abuse based on all four metrics, we multiplied all the ratios gathered as described above into a single term. The product would be high if the ratios were generally high, implying a higher probability of addiction, and lower if the ratio were generally low.

This model has various strengths, such as versatility and similarity to real-life circumstance. Had another factor been identified and documented, a corresponding factor could be derived and multiplied to the product of the other four weights. The model is thus highly scalable and can be implemented into different populations with many different metrics, rendering it extremely adaptable yet illustrative. Our model is also true to the geometric nature of probability, using ratio comparisons between groups and the average, or general population, and multiplying relative probabilities of different instances to derive a representative measure of addiction probability.

A weakness for our model is the assumption that the factors did not overlap into others. This assumption is not entirely true as there are trends between factors such as urbanization and income which cause two variables to be true for the same case, but for the sake of simplicity of the problem and in order to be able to devise a prediction, this assumption had to be made.

## 4 Ripples - The Cost of Substance Abuse

### 4.1 Defining Variables

- **Input:** Per capita medical expenses for rehabilitation and addictions in USD.
- **Input:** The employment impact, specifically salary lost due to addiction in USD.
- **Input:** The cost of buying the drugs in USD.

- **Input:** Impact on life expectancy quantified as USD per year for a year of life for a person.
- **Output:** The total cost in USD of using a specific substance

## 4.2 Assumptions

1. Quality of life and happiness are not included in the metric.

**Justification:** Because quality of life and happiness are not numerically quantifiable values, we chose to exclude these factors in the metric to rank the substances.

2. Auto accidents are not factored into metric.

**Justification:** Auto accident statistics are mostly general substance abuse cases and does not often break down specific drugs involved in accidents; therefore, we assumed all drugs to have about the same impact on auto accidents and did not factor it into the metric.

3. Crime and violence are equally impacted by the substances.

**Justification:** Statistics on crime and violence in relation to drugs and substance also does not give break downs to different drugs; therefore, we also assumed that all crime and violence are equally impacted by the substances and thus do not factor into the ranking.

4. Part time is treated at half full time for salary.

**Justification:** For the sake of simplifying the calculations, part-time jobs are treated as earning exactly half the income of full-time workers.

5. Addiction costs one for the duration of their career from age 20 to age 65 (45 years).

**Justification:** This was done so that we could effectively quantify the loss in wages one experiences over the course of a career from being a drug addict.

6. A human year of life is worth \$129,000 per year.

**Justification:** This assumption is based off of the cost of dialysis, keeping a body alive [15].

## 4.3 Development of the Model

Our general approach in this section was to quantify all factors, even those not monetary in nature, into U.S. dollars as this was the most universal metric applicable to all types of factors. By utilizing statistics from various sources, we were able to compare and contrast the costs of abusing the different substances.

For medical expenses we found how much addicts to each substance costs the US per year then divided that by number of addicts for each substance in the US to find per capita



| Drug Type | Money Lost (\$) |
|-----------|-----------------|
| Marijuana | \$1,215,017     |
| Nicotine  | \$2,004,912     |
| Alcohol   | \$3,009,097     |
| Opioids   | \$5,417,072     |

Table 7: The estimated dollar value lost due to drug use over the course of the lifetime.

cost for medical care of addicts of each drug. For marijuana the costs were already given in per capita values. Money spent on drugs was researched and found as per capital per year values.

To calculate the effects of drugs on life expectancy, a year of human life was quantified into US dollars (researched value) and then the years lost when a substance is abused was multiplied by this value to determine how much money is loss to substance abuse in relation to costs. It was also assumed that addictions start at the age of 20 for the purpose of determining wages lost.

To calculate the effects of drug abuse on employment issues, we found the percentage of full-time and part-time workers that use each of the four different drugs out of the total population of Americans aged 26+ that use them, assuming that the population of Americans in the workforce under 26 years that abuse the drugs is negligible. We then found the sum of the full-time work proportion and half the part-time work proportion, and multiplied that by the annual median per capita income of an American citizen (\$31,099) [14] to determine an average annual income for full-time workers using a substance. By subtracting this number from our calculated expected value of annual income based on the total proportions of full-time and part-time workers in the United States (\$27,278) and multiplying that difference by 45 years to determine the lifetime cost to ones income when abusing each specific drug.

#### 4.4 Results of the Model

We conclude based on Table 7 that the most harmful substance is opioids, followed by alcohol, nicotine, and finally marijuana. This is in line with conventional thought, in that opioids tend to be the most addictive and life-altering while marijuana is the least harmful to one's life for the most part in both financial and personal aspects of life..

#### 4.5 Model Analysis

The strength of this metric is that it quantifies the cost of substance use in order to rank the substances. This is beneficial to give a tangible value to the cost of continuing to use a substance and thus develops a sense of real consequences for substance abuse.

Weaknesses in the model include the fact that crime and violence, auto accidents, quality of life, and happiness are not included. Violence and auto accidents are not considered

because it is assumed that substances all impact these factors at the same rates. Quality of life and happiness are not included because they are not quantifiable values due to the number of impacting factors and the psychological factors.

## References

- [1] <https://drugabuse.com/statistics-data/>
- [2] <https://www.drugabuse.gov/related-topics/trends-statistics>
- [3] <https://www.drugabuse.gov/publications/drugfacts/drugged-driving>
- [4] <https://www.hhs.gov/opioids/about-the-epidemic/index.html>
- [5] Branch, Geographic Products. 2010 Census Urban Area FAQs. Census Bureau QuickFacts, United States Census Bureau, 1 Sept. 2012, [www.census.gov/geo/reference/ua/uafaq.html](http://www.census.gov/geo/reference/ua/uafaq.html).
- [6] Office of Adolescent Health. The Changing Face of America's Adolescents. HHS.gov, US Department of Health and Human Services, 25 Feb. 2019, [www.hhs.gov/ash/oah/facts-and-stats/changing-face-of-americas-adolescents/index.html](http://www.hhs.gov/ash/oah/facts-and-stats/changing-face-of-americas-adolescents/index.html).
- [7] Child Population by Age Group — KIDS COUNT Data Center. KIDS COUNT Data Center: A Project of the Annie E. Casey Foundation, [datacenter.kidscount.org/data/tables/101-child-population-by-age-group#detailed/1/any/false/871,870,573,869,36,868,867,133,38,35/62,63,64,6,4693/419,420](http://datacenter.kidscount.org/data/tables/101-child-population-by-age-group#detailed/1/any/false/871,870,573,869,36,868,867,133,38,35/62,63,64,6,4693/419,420).
- [8] MTF Data Tables and Figures. Monitoring the Future, National Addiction & HIV Data Archive Program, [www.monitoringthefuture.org/data/18data.html#2018data-drugs](http://www.monitoringthefuture.org/data/18data.html#2018data-drugs).
- [9] Cigarette Smoking among U.S. High School Students at an All-Time Low, but e-Cigarette Use a Concern. CDC Newsroom, Centers for Disease Control and Prevention, [www.cdc.gov/media/releases/2016/p0609-yrbs.html](http://www.cdc.gov/media/releases/2016/p0609-yrbs.html).
- [10] 2016 Surgeon General's Report: E-Cigarette Use Among Youth and Young Adults — CDC. Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, [www.cdc.gov/tobacco/data\\_statistics/sgr/e-cigarettes/](http://www.cdc.gov/tobacco/data_statistics/sgr/e-cigarettes/).
- [11] De Silva, Varuna. Fit Logistic Curve to a Data Set. Matlab File Exchange, [www.mathworks.com/matlabcentral/fileexchange/31399-fit-logistic-curve-to-a-data-set](http://www.mathworks.com/matlabcentral/fileexchange/31399-fit-logistic-curve-to-a-data-set).
- [12] Tobacco Product Use Among Adults United States, 2017. Morbidity and Mortality Weekly Report (MMWR), Centers for Disease Control and Prevention, 8 Nov. 2018, [www.cdc.gov/mmwr/volumes/67/wr/mm6744a2.htm?s\\_cid=mm6744a2\\_w](http://www.cdc.gov/mmwr/volumes/67/wr/mm6744a2.htm?s_cid=mm6744a2_w).
- [13] RESULTS FROM THE 2016 NATIONAL SURVEY ON DRUG USE AND HEALTH: DETAILED TABLES. Substance Abuse and Mental Health Services Administration, Substance Abuse and Mental Health Services Administration, 2016, [www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf](http://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf).
- [14] [https://fred.stlouisfed.org/series/MEPAINUSA672N&sa=D&ust=1551682599956000&usg=AFQjCNHQIQYAg1W5cuEWZUKE\\_RacLoH8JQ](https://fred.stlouisfed.org/series/MEPAINUSA672N&sa=D&ust=1551682599956000&usg=AFQjCNHQIQYAg1W5cuEWZUKE_RacLoH8JQ)
- [15] <http://content.time.com/time/health/article/0,8599,1808049,00.html>