MathWorks Math Modeling Challenge 2019
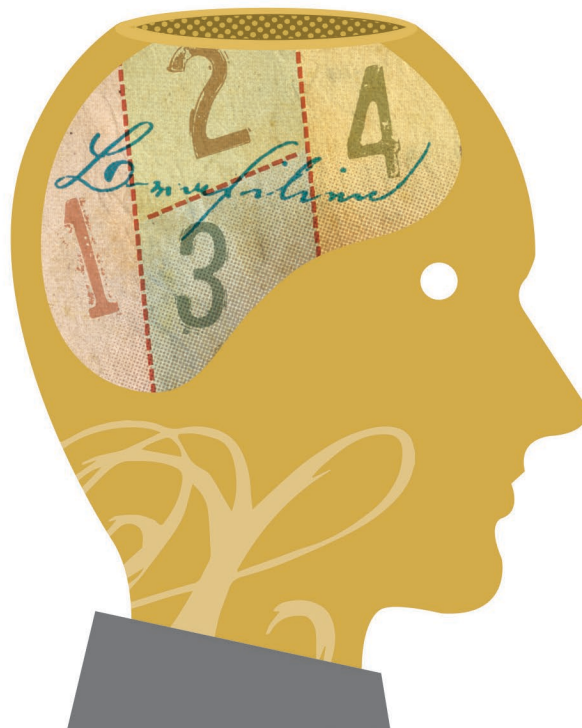
Richard Montgomery High School–

Team #12300 Rockville, Maryland
Coach: Matt Davis
Students: Matt Kolodner, Clarissa Xia, Jack Yang, Laura Yao, Lauren Zhou

MathWorks Math Modeling Challenge First Runner Up

$15,000 Team Prize

1   Executive Summary

It's no secret that the use and subsequent abuse of drugs and addictive substances has a negative impact on the mental and physical health of users. This problem is especially important among teenagers, since younger users are much more likely to become addicted to these drugs.[16] As a result, this paper explores the spread and usage of various substances, especially among high school aged students, as well as the societal implications of drug usage.

Our first model uses a Susceptible-Infected (SI) model and solves it with Euler's Method in order to predict the spread of nicotine vaping over the course of 10 years. The model is based on the idea that vaping can be modeled similarly to the spread of disease with the basis that someone who interacts with many others who vape is at a higher likelihood to begin vaping themselves.

Based on the results of the model, nicotine addiction was shown to increase by nearly 7.7% over the 10 year interval, surpassing the use of cigarettes in just 8 years. Additionally, this model also suggests that nicotine addiction will grow at an increasing rate. Although this model produces sensible results, it could be improved in the future with more information on the interactions between non-addicted and addicted people. Additionally, this model also suggests that the growth of the nicotine-vaping addicted population will exceed the population growth of the US in just 2 years.

The second model explores the likelihood a given individual uses a given substance at least once within the span of a year. While other factors were considered, the model evaluates likelihoods of substance use based on the factors of race, gender, and socioeconomic status (SES). The model makes use of a weighted average system, using weights based on the variance of characteristics. The subsequent weighted averages represent the likelihood the individual partakes in a specific drug, based on their risk factors.

The model was then applied to a population of 300 high school seniors to determine how many would use a drug within the year. We used a randomized program and the obtained probabilities to generate the results that 49 students would use nicotine, 53 would use marijuana, 112 would use alcohol, and 14 would use an unprescribed opioid.

Our third model ranks the severity of impacts from the four different drugs evaluated in the second model. It considers the factors of economic burden, deaths per year, and number of users, split into minors and adults. The model uses population values and different weights to represent the impact that these different factors have on a person's life. The values are converted into percentages, weighted, and then summed in order to best rank the drugs by impact. The results of our model are that the substances are ranked nicotine, alcohol, opioids, and marijuana, with nicotine being the most detrimental to American society.

Contents

2   Introduction

Substance and drug use is the usage of potentially addictive substances with negative impacts on physical and mental health. Teenagers are more likely to become addicted after exposure to the substances, due to their physiological susceptibility to addiction.[16] Thus, the problem is more serious in younger populations due to the potential dire future implications. Recently, several new drug problems have arisen, notably the rise of e-cigarettes, also known as vaping, as well as the recent opioid crisis with the abuse of unprescribed opioids.[29] This paper explores the problem that drug and substance abuse present to our society, especially in relation to younger populations.

3   Problem 1: Darth Vapor

3.1   Defining the Problem

- We are asked to find a model that predicts nicotine use due to vaping over the next 10 years.

- Let "nicotine use" be defined as repeated usage of nicotine through vaping, also known as nicotine addiction.

3.2   Assumptions and Justifications

- Every person has the same probability of addiction, regardless of individual factors.

    Justification: In order to apply an SI model, we needed to standardize the factors used. Although more young individuals vape nicotine than the older population, it can be argued that these adolescents will eventually move into adulthood over the 10-year period of the model, and that cigarette users may use alternatives.

- Addiction is not a spectrum. One can only be quantified as addicted or not addicted.

    Justification: This is needed in order to compartmentalize the model. Furthermore, addiction itself often is a definite term and not vague.

- The United States is representative of more economically developed countries' (MEDCs) nicotine use through vaping.

    Justification: Vaping tends to be more of a problem in MEDCs than LEDCs countries due to its cost, making the United States a good representation of spread in nicotine usage throughout the world.

3.3 Defining the Variables

Symbol	Definition	Units
S	Not addicted population	Number of people
I	Addicted population	Number of people
N	Total population	Number of people (S + I)
c	Probability of addiction given one smoke	%
p	Rate of decay in which the addicted population would increase from the susceptible population per year	%
γ	Rate of negative interactions between the not addicted and addicted population per year	%
Ω	Rate of recovery from addicted to not addicted per year	%
α	Rate of change in total population per year	%

3.4 Developing the Model

In order to model the spread of nicotine, we decided to treat nicotine use as a disease, using various models to predict how many of the nicotine addicted population would change over 10 years. We specifically used a Susceptible-Infected (SI) Compartmentalizing Model. This method classifies people into two different groups based on their nicotine addiction status and measures the change within these two groups each year by applying Euler's Method over both populations.

In the model, we mapped two primary methods by which people begin smoking nicotine. The first is a natural decay between the susceptible and infected group through people who will just start vaping on their own as a natural progression. The second method is peer pressure, occurring when the already infected group convinces the people from the susceptible group to become addicted, thus joining the infected group. This would happen more frequently given a larger S or I population. Our model also accounts for the population growth α (births - deaths + immigrants). For this model, we used the 2017 US growth rate of 0.71% per year, thus deducing the total increase due to births, deaths, and immigration as $0.0071\times$ total population N.[13] Thus far, we have

$$\frac{dS}{dt} = \alpha N$$

The next step involves deducing an algorithm for the natural decay of people from S to I. In order to calculate this value, we used the probability of becoming addicted given one smoke c and rate of decay per year p, where p only applies for one smoke. We determined both the values of c and p. The value c was determined through a PubMed paper, citing the probability of an individual becoming addicted to nicotine as 30 percent.[4] Thus, $c = 0.3$. In a study by The Denver Post, it was cited that 12.6% of adults tried vaping at least once in 2014 and 15% of adults tried vaping in 2016.[2] Thus, we can estimate the value of p through the subtraction of these two percentages and finding the increase in percentage as the rate for p. $p = \frac{0.15-0.126}{2} = 0.014$. The SI model now has equations

$$\frac{dS}{dt} = -cpS + \alpha N$$
$$\frac{dI}{dt} = cpS$$

The final part of our initial model involved accounting for peer pressure through the interactions between the two compartments S and I.[18] In order to account for this interaction, we used the standard SI model ordinary differential equation using the coefficients c and γ, where γ refers to the rate of interactions between the addicted and not addicted populations per year. Given that peer pressure can affect both teens and adults, 13.8% of the world is in adolescence, and near 90% of teens have stated they have been influenced by peer pressure, we arbitrarily estimated the upper and lower bounds for the value of γ as 0.05 and 0.35, respectively, thus assigning an average value of 0.2 for our initial model.[14,15] We will examine variations in this parameter later in the paper. Thus, we have the final group of differential equations:

$$\frac{dS}{dt} = -\frac{c\gamma SI}{N} - cpS + \alpha N$$
$$\frac{dI}{dt} = \frac{c\gamma SI}{N} + cpS$$

3.5   Applying the Model

Using this model and the previously determined coefficients, we applied Euler's Method to each differential equation, using a step size of 1 representing 1 year intervals.[4] The initial values of S and I were identified as 311,900,260 and 16,416,150, respectively, summing to the total US population N of 328,316,410.[1, 17]

Table 1: Predicted Spread of Nicotine Use over 10 Years

Years Since Start of Model	Not Addicted Population(number of people)	Not Addicted Population Change (number of people)	Addicted Population(number of people)	Addicted Population Change(number of people)
0	311,900,260	—	16,416,150	—
1	311,985,606	85,346	18,661,851	2,245,701
2	311,966,349	-19,257	21,028,704	2,366,854
3	311,838,311	-128,038	23,521,007	2,492,303
4	311,597,362	-240,949	26,143,007	2,622,000
5	311,239,446	-357,916	28,898,880	2,755,872
6	310,760,609	-478,838	31,792,700	2,893,820
7	310,157,023	-603,585	34,828,413	3,035,714
8	309,425,024	-731,999	38,009,809	3,181,396
9	308,561,137	-863,887	41,340,484	3,330,674
10	307,562,112	-999,024	44,823,810	3,483,326

Based on the data from the model, a clear increase was observed in the rate in which people in the US vape nicotine products. This can be seen in how the increase in addicted population

per year was 2.2 million after 1 year and 3.5 million after 10 years. This means that after 10 years, our model predicts that 12.7% of the US population is expected to be addicted to nicotine vaping, an increasing 7.7 percentage point from the current 5%. It was also observed that this increase in nicotine vaping users surpassed the US growth rate after 2 years, as the slope changed from 85,346 to -19,257 between years 1 and 2.

Graph 1: Predicted Spread of Nicotine Use over 10 Years
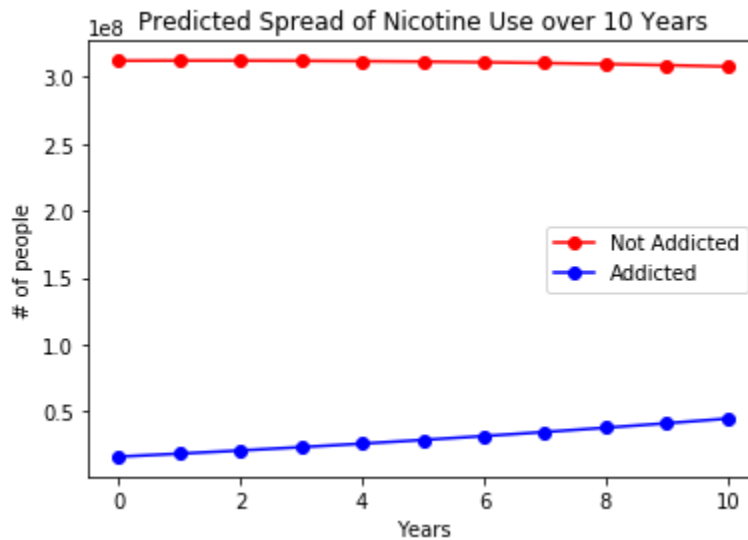
Although the use of nicotine in vaping is growing at an accelerated rate, the above graph shows that many years will need to go by before the addicted population will surpass the not addicted population. The graphs below show a closer image of the growth/decay of each compartment.

Graph 2: Predicted Spread of Nicotine Use over 10 Years for Those Addicted
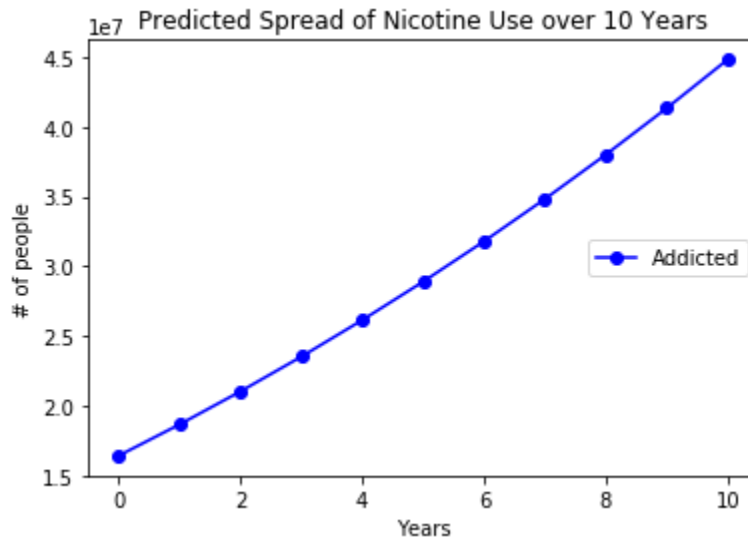
Graph 3: Predicted Spread of Nicotine Use over 10 Years for Those Not Addicted
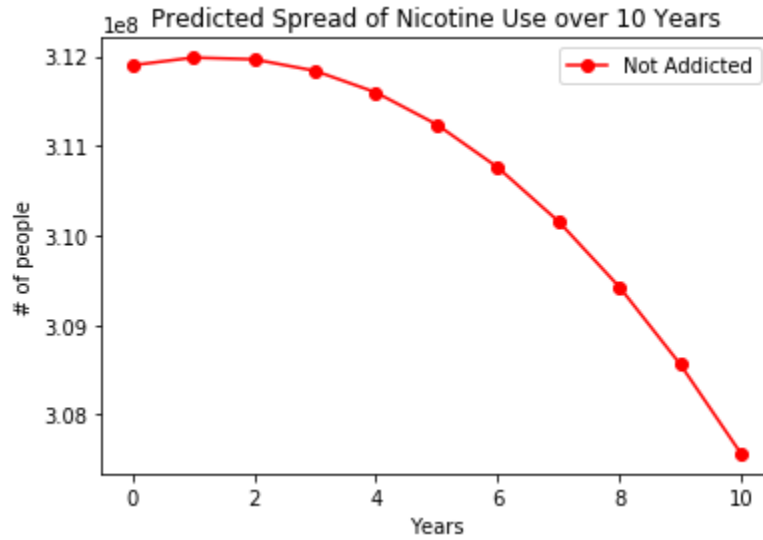

In contrast to this growth in use of nicotine for vaping, cigarettes have seen a decline in usage over the past several years. In 2018, cigarette usage reached an all-time low, as only 14.4% of adults smoked. Using the fact that approximately 42% of adults smoked in 1960, we can use a linear regression to predict when vaping with nicotine will surpass the use of traditional cigarettes.[19] Using the linear equation $y = N((0.42 - 0.14)/58) + 0.24N$, we can predict that the use of nicotine-based vaping products will surpass the usage of cigarettes in just under 8 years.

Graph 4: Nicotine Usage Versus Cigarette Usage
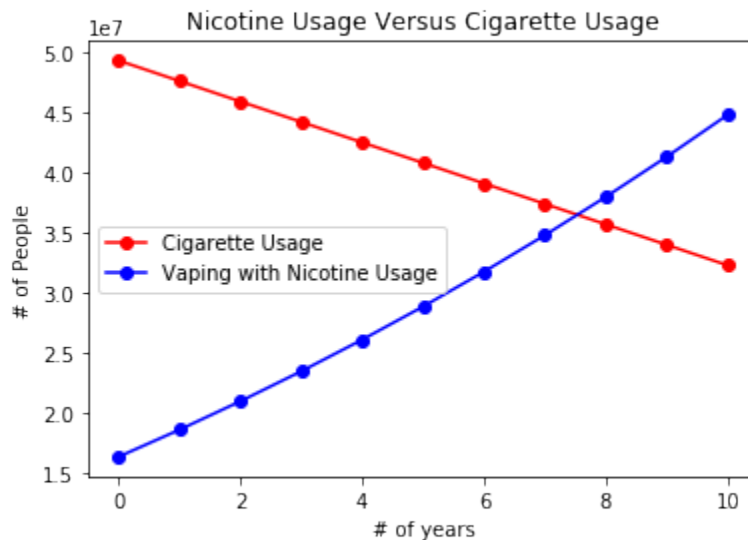
3.6   Evaluating the Model

3.6.1   Validation

The results of the model make sense in context. Nicotine has been growing in popularity in recent years, and the concave up nicotine-addicted graph reflects this trend. The magnitude of the increase in nicotine vaping is also reasonable, being validated by the the 1.5x increase in cigarette usage between 1940 and 1950.[20]

3.6.2   Sensitivity Analysis

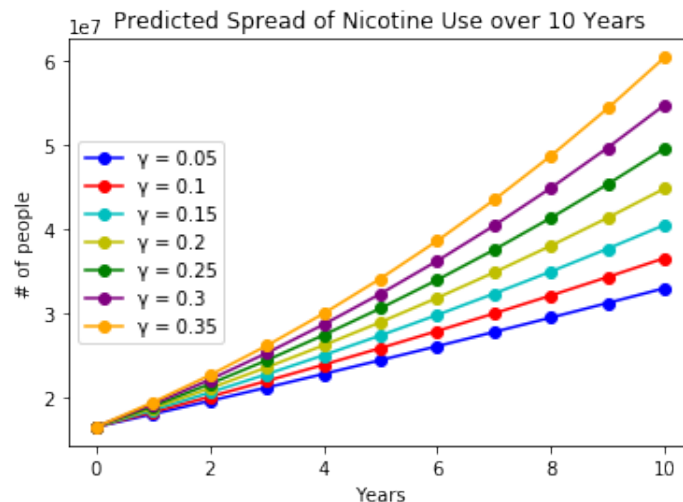One of the more uncertain variables in the model was the value of γ, the rate of interactions between I and S per year. The value used in the model, 0.2, was taken as an average of a 0.05 lower bound and a 0.35 upper bound. We can perform a sensitivity analysis of this variable.

Graph 5: Predicted Spread of Nicotine Use over 10 Years

Based on Graph 5, we can see that the graph appears to be less linear over the 10 years as the value of γ increases. This shows that increasing the value of γ makes the model less linear over the 10 years. After 10 years, the lower bound value was approximately 68% of the existing model value. Likewise, the upper bound value was approximately 33% higher than the existing model value. Overall, this means that the graph is still somewhat sensitive to fluctuations in its parameter γ, having a somewhat large range of 3×10^8.

3.6.3   Strengths and Weaknesses

The strengths of this model include its ability to create a realistic model that makes sense in context to the overall growth of nicotine vaping in the population. Additionally, this model is also flexible, being able to be applied reliably should the parameters or populations change at a future time.

However, this model also has several weaknesses. One of these weaknesses lies in the assumption that every human is equally likely to engage with nicotine vaping. This is a somewhat unrealistic assumption to make given that vaping is more common among younger individuals than older, but nevertheless it was a necessary assumption to make in order to use an SI model. In addition, this model also is unable to map nicotine-vaping trends over long periods of time, as the I compartment doesn't have a method of leaving that category. This currently is due to the lack of data available for recovering from nicotine-vaping addiction. Finally, this model is somewhat weakened by the vagueness of the γ parameter, again due to the limited amount of data available.

3.6.4   Extensions of the Model

One potential expansion of this model is including an incubation rate from the I compartment to the S compartment. Currently, there is not significant data regarding rates at which people are able to recover from nicotine-vaping addiction. As a result, certain values were assigned for an incubation rate variable, Ω, that would return infected individuals to the omega group. This would change the differential equations to

$$\frac{dS}{dt} = -\frac{c\gamma SI}{N} - cpS + \alpha N + \Omega I$$
$$\frac{dI}{dt} = \frac{c\gamma SI}{N} + cpS - \Omega I$$

Graph 6: Predicted Spread of Nicotine Use over 10 Years
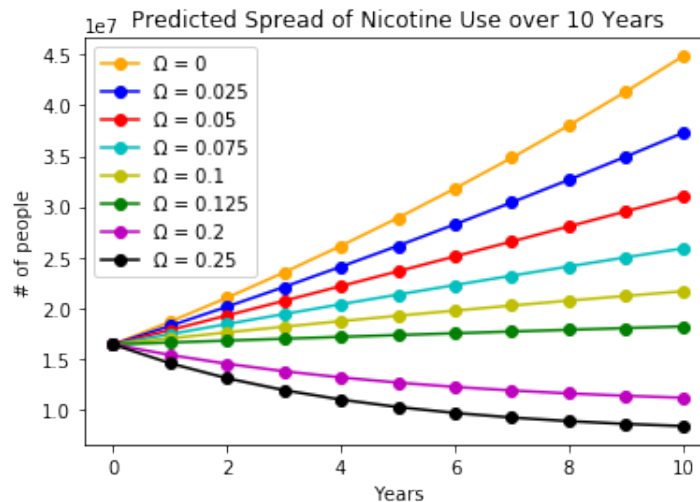
Graph 6 demonstrates how more information on this variable would influence the model, with the value we used for our current model being $\Omega = 0$. As shown by the graph, in order to reduce the number of nicotine-addicted people, more than 12.5% of the infected population need to be moved to the S compartment each year.

4 Problem 2: Above or Under the Influence?

4.1 Defining the Problem

We are asked to:

- Determine the likelihood an individual will use nicotine, marijuana, alcohol, or unprescribed opioids at least once within a given year, based on characteristics such as socioeconomic status, race, and gender.

- Determine the characteristics of the student body at an average high school.

- Apply the model to a student body of three hundred high school seniors to determine their drug usage likelihood during that year.

4.2 Definition of Terms

Lifetime usage: The usage of a drug at least once in an individual's life.
Substance use: The usage of a drug at least one time, regardless of addiction status.
Nicotine: Drugs derived from tobacco, including cigarettes, cigars, e-cigarettes, and smokeless tobacco.
Opioids: Drugs derived from the opium poppy plant, specifically, heroin and opioid pain killers.
Opioid abuse: The usage of illegal or unprescribed opioids.

4.3 Assumptions and Justifications

- Race and gender distributions in high schools are roughly the same as their distributions in the United States as a whole.

    Justification: Based on available data, the differences between high school demographics and overall United States demographics can be considered to be negligible in order to generalize available data on adults to a high school situation. For example, approximately 5 percent of high-schoolers are Asian, while about 5.8 percent of the total United States population is Asian.[7,8]

- Academic success should not be used as a quantifiable measurement of drug usage.

    Justification: There is no consistent grading system across the United States to measure academic success. While SAT scores could be used as a standardized system, not all individuals take the SAT. Due to a lack of a consistent measurement mechanism across different regions of the United States, this factor will not be considered in this model.

- Income can be used as a representative measure of socioeconomic status.

Justification: While there are many measures of socioeconomic status such as income, wealth, and parental education, income can be used as a representative measure for simplicity. Data trends follow similar patterns, independent of the measure of socioeconomic status being used. For example, in an NIH study, 19.3 percent of those in the bottom quartile of income smoked marijuana, while 20.2 percent of individuals in the bottom quartile of wealth smoked marijuana. The differences are therefore negligible and can be ignored.[9]

- The 300 high school seniors are composed of an average distribution of traits.

  Justification: The problem does not define how the characteristics of the students are distributed. For the purposes of maximum applicability, we assume that the student body is representative of an average high school trait distribution.

- The only races considered in this model are white, African American, Asian American, and Hispanic or Latino.

  Justification: The four races considered have the highest percentages in the United States population and were therefore chosen as representative of the population.[8]

4.4 Defining the Variables

Symbol	Definition	Units
P(trait)	Probability that a person of a trait would use the given drug	%
R(trait)	Range of the percentages for a certain trait and for the given drug (highest percentage - lowest percentage)	%
W	Weight given to each of the factors when determining its effect on likelihood of drug use	%

4.5 Developing the Model

To develop the model, we considered an individual's likelihood to partake in the four substances considered: alcohol, nicotine, unprescribed opioids, and marijuana. We identified several factors which would affect the likelihood of substance use, mainly race, gender, and socioeconomic status. Other identified factors, such as usage of other drugs, time since last usage of the substance, age when first exposed to the substance, family history of drug use, and percent usage of the substance by peers in immediate contact were also considered but are not directly evaluated in this model due to time and data constraints. The three main identified factors are used as representative ones to display the model's mechanism.

Data on an individual's likelihood to use a substance was collected based on their race, gender, and socioeconomic status (SES). Race was distinguished into white, African American, Hispanic or Latino, and Asian American. Socioeconomic status was distinguished into the categories "high" and "low," namely whether the individual was in the top half of income or the bottom half.[9] The data collected was from an age range of 12-17 year olds, given that

the context of the problem is a high school. All of the data was for substance use within the past year, since by the definition of the problem, the model is designed to determine an individual's likelihood of substance use within a given year.

The compiled data is displayed below. High school refers to the percent trait distribution of an average high school and is not related to drug use.

Table 2: Percent Likelihood of Substance Use within 12 Months Based on Individual Characteristics for 12-17 Year Olds[21,22,23]

Race	Nicotine	Marijuana	Alcohol	Opioids	High Schoolers
White	14	13	13.2	3.9	49
African American	7.9	12.6	8.6	4.1	15
Hispanic or Latino	9.3	12.6	11.2	4.1	26
Asian	3.5	4.5	4.5	1.9	5.0

Gender	Nicotine	Marijuana	Alcohol	Opioids	High Schoolers
Male	12.7	12.9	9.3	3.6	50
Female	10.4	12.3	9.9	4.2	50

Socioeconomic status	Nicotine	Marijuana	Alcohol	Opioids	High Schoolers
High	16.5	39.1	24	2.7	50
Low	26.7	19.3	16	6.3	50

Using the likelihood data from each category, an individual could be assigned a percent change of substance use in a given year. The model makes use of a weighted average system, where the variance of a particular trait, such as race, is used to determine the amount a trait should be weighted. Traits with a larger range are weighted more heavily because they indicate a greater importance. If a trait has little variance and thus a low range, it is reasonable to say that that trait does not significantly affect a person's likelihood to use a substance. If a trait has a much larger range, then that would indicate that there is more of a correlation between that factor and substance use.

The calculation can be represented through the following formula where W is the weight of the trait and R represents range. The total range is the sum of the ranges of all of the traits.

$$W = \frac{R(trait)}{R(total)}$$

As an example calculation, the amount race should be weighted for nicotine use is found as follows.

The variation of the effect of race on nicotine use is equal to the maximum nicotine use percentage, which is 14 percent for white people, minus the minimum nicotine use percentage, which is 3.5 percent for Asian Americans. There is therefore an $R(race)$ value of 10.5 percentage points.

The total variation for nicotine would be the sum of $R(race)$, $R(gender)$, and $R(SES)$, which yields a value of 23 percentage points. The W value for the effect of race on nicotine usage is the ratio of $R(race)$ and $R(total)$, or $\frac{10.5}{23}$. This yields the value 0.456.

The W value for each of the characteristics is found and used in the following equation where $P(T)$ is the overall likelihood of use of a given substance and the numbered P values are different traits, in this case, race, gender, and SES.

$$P(T) = W_1(P(1)) + W_2(P(2)) + W_3(P(3)) + \cdots + W_n(P(n))$$

Using the example of nicotine, the probability of a white male of high socioeconomic status can be calculated as a demonstration.

$$P(T) = 0.456(0.14) + (0.1)(0.127) + 0.443(0.165)$$

Solving this equation provides a value of 0.150, or a 15 percent likelihood that this specific individual uses nicotine at least once during a given year. The same process is applied across the different drugs considered with a potential variety of individual characteristics considered.

While the development of this model only takes into account the three factors, additional factors could easily be added by employing the same formula with more data values. The benefit to the weighted average system is that the weight value used will never exceed one, regardless of how many factors are added. As a result, large numbers of factors could potentially be used to characterize an individual's risk of substance use. The addition of more factors would serve to refine the accuracy of the probability of use; however, the three factors used currently are enough to provide a fairly accurate estimate.

4.6   Applying the Model

We considered the problem scenario of 300 high school seniors with varying characteristics, modeled after the average high school in the United States. From these characteristics, we used the model to predict, out of the 300 students, how many would use nicotine, marijuana, alcohol, and unprescribed opioids at least one time in their senior year. For this simulation, we only considered the three factors of race, gender, and socioeconomic status when predicting the students' drug use habits. However, other characteristics could easily be added to our simulation with the usage of more data.

To create the simulation of the 300 high school students, we researched the average distribution of race and socioeconomic status for the average high school in the United States. We assumed that the gender distribution in schools would be 50/50. Using this information, we wrote a Python program (see appendix), in which we created 300 "individuals" with three traits each (race, gender, and socioeconomic status), and each trait would be randomly assigned using a number generator and the distributions. For an example, there would be a 49% chance for a student to be white, a 50% chance for a student to be male, and a 50% chance for a student to have a low socioeconomic status. Therefore, there would be about a 12.25% chance for a student produced by the simulation to be a white male of low socioeconomic status.

In order to find how many of the 300 students would have used each of the drugs, we incorporated the model from above into the Python program. The model took into account the characteristics of each individual, as well as the likelihoods of a person of each of these characteristics to be using the given drug. The program used the characteristics of each individual and the model to determine the likelihood the individual would use each of the four drugs, and then used a random number generator with those likelihoods to determine whether or not the individual would use each of the four drugs. Finally, the program outputted the number of students, out of the 300, that used each of the drugs. We compared these percentages to the percentages given by reputable sources about the demographics of drug use to see how our results matched up.
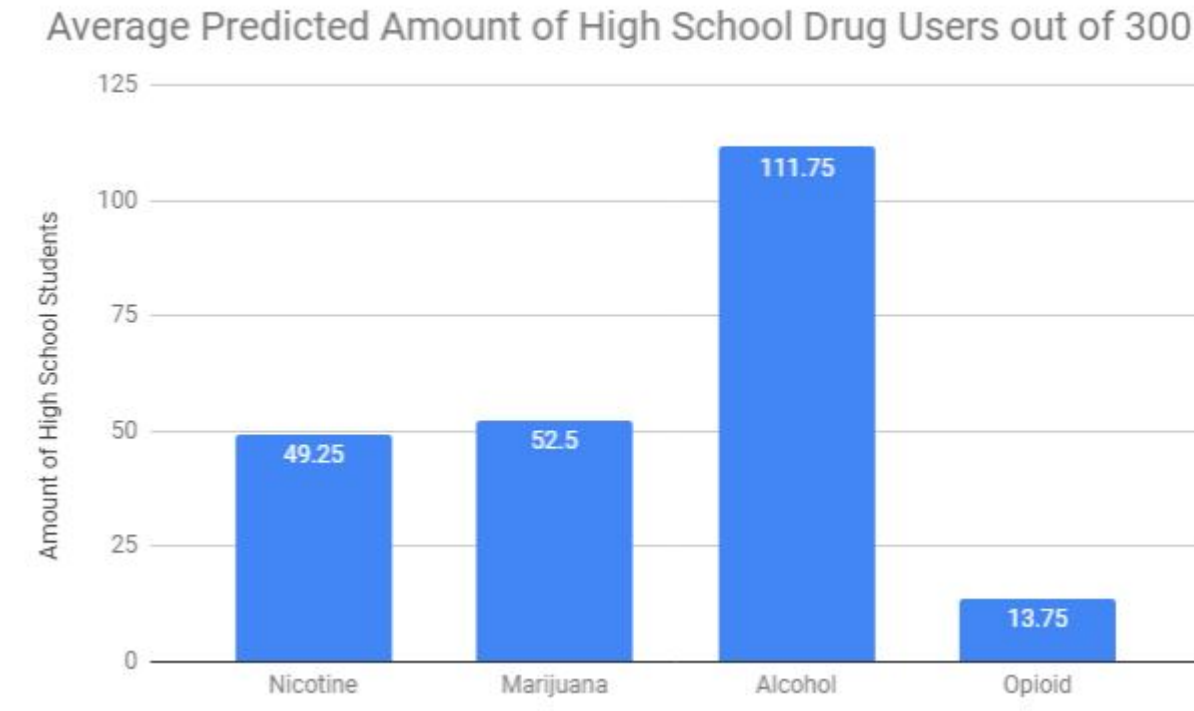
4.6.1 Conclusions of the Model

Upon running the simulation for five trials, the following averages were collected.

Table 3: Average Number and Percent of Simulated High School Drug Users (out of 300 Students)

Number who will use	Percent who will use
49.25	16.42
52.5	17.5
111.75	37.25
13.75	4.58

Graph 7: Average Predicted Amount of High School Drug Users out of 300 Students

The model indicates that, of an average high school senior class of 300 students, 49 would use nicotine, 53 would use marijuana, 112 would drink alcohol, and 14 would use unprescribed opioids. These conclusions are similar to accepted values, where approximately 22.7 percent of seniors smoked marijuana in the past month, which is comparable to the 17.5 percent of seniors smoking marijuana obtained in our simulation.[12] In actuality, approximately 16.2 percent of high school students smoked a cigarette in the past month, while our simulation yielded an average value of 16.4 percent, which indicates a high degree of accuracy.[12] There was not enough data on opioid use, due to the recency of the problem. However, there was a larger discrepancy between the accepted value of 26 percent of high school students who drank in the last month and the 37 percent obtained through our simulation.[12] This could be due to the fact that our simulation accounts for the potential for alcohol consumption within the time frame of a year. Since alcohol is more likely to be consumed than the other substances, more students could have consumed alcohol in the additional time frame.

While our estimates are not completely accurate, they are largely reasonable. Additionally, these values would likely be more accurate given the usage of more factors to refine our probability values. Given the three factors used and the data available, the estimates from the model are reasonable.

4.7   Evaluating the Model

4.7.1   Sensitivity Analysis

Our model was based on data from 12-17 year olds. However, the age range used in the model has an effect on the percentage of individuals who are partaking in the drug. Therefore, here we evaluate the sensitivity of our model to the age range used. We will use marijuana with race and gender as a simplified representative model.

Table 4: Percent Probability of Marijuana use by Race and Gender, Compared between 12-17 Year Olds and Total Population[21]

	12-17 Years	Total Population
Males	12.9	16.1
Females	12.3	11
White	13	13.6
African American	12.6	16.8
Asian American	4.5	5.7
Hispanic or Latino	12.6	11.9

Using these values to calculate the percent likelihood a white male uses marijuana, using the values of the 12-17 year old yields a probability of marijuana use of 13.0 percent. Using the values for the general population yields a value of 14.4 percent. While there is a discrepancy between the values, the difference is relatively small. Therefore our model is only slightly sensitive to our age assumption, indicating a more robust model and an ability to generalize the model to a wider population, although a more accurate result could be obtained using data specific to a given situation.

Additionally, our model is not very sensitive to the assumption that income is used as a measure of socioeconomic status. The percentage variation between drug usage in relation to income and other markers of SES is minuscule and thus does not greatly affect the range of drug usage probability (9). As a result, using different markers of SES would not affect the weighted average or the resulting probability value.

4.7.2   Strengths and Weaknesses

One of the strengths of this model is that it used data that applied to people aged 12-17, which means that it fits the situation of high school seniors relatively well. The model was therefore highly specific to the situation and can be considered to be relatively accurate for the situation. However, this can also be considered a weakness, since the model would not be able to be applied to the general population, due to the substantial impact of age on drug use. New data would have to be added, making age a factor in drug use, to make the model broader and more applicable.

Another strength of the model was that it took into account three of the most important factors for drug use, which are race, gender, and socioeconomic status. However, it also left

out many factors that could have a substantial impact on drug use, such as genetics, social circles, and mental health issues, due to time constraints and data availability. Despite this, our model is flexible and robust, and new factors could easily be considered if the general data is given. Our model thus has the strength that it can easily be adapted to include additional data, since the weighted average system serves to accommodate any number of potential factors.

One of the weaknesses of the model and simulation was a lack of a comprehensive data source. The data was taken from many different sources, all of which had different methodologies and slightly different units of data. For example, some data sources gave slightly different age ranges than others, and some data sources defined opioids slightly differently than others. Additionally, due to the lack of consistency, there were varying values for the same information, such as the total number of individuals who use a specific drug. While we had a limited amount of time and no method of obtaining more reliable and coherent data, we attempted to work around this by using one data source when possible and finding similar units between data sources.

Another weakness of the model was that not every factor was independent of each other. For example, it's commonly known that there's a correlation between race and socioeconomic status, while our model would have treated them as completely unrelated characteristics. Furthermore, if more factors were to be added that also relate to other factors, the model wouldn't be able to account for the relationships between different factors, and therefore wouldn't be as accurate. In order to take into account the effects of certain factors upon other factors, we would have had to use factor analysis. Unfortunately, the time constraint did not allow us to do so.

We also only took into account the risk factors involved with drug use—not the protective factors, such as anti-drug programs and prevalence of health education. These factors are often tied to socioeconomic status and the affluence of the community in which the school is located, so to prevent overcounting, we did not take them into account in this model. However, we acknowledge that risk factors and protective factors don't have a direct inverse relationship with each other, and that the model must be adapted in order to consider both types of factors.

4.7.3 Extensions of the Model

One way the model could be extended is to predict the likelihood for a senior in high school to use other drugs, such as cocaine or LSD. This could not only give us a better idea of the frequency of their usages, but also help us better verify the accuracy of our model. Similarly, the model could also be improved by including more factors, both risk and protective factors, to make the model more accurate. It's important to find a way that the effects of the risk factors and the effects of the protective factors do not simply cancel each other out but rather accurately reflect how these factors interact with each other. When adding in the

factors, the model could be extended to include factor analysis in order to acknowledge the relationships between the factors that may affect the results. A factor analysis would give us a better idea of how to weigh the different factors and see which would affect the likelihood that a given individual would use a certain drug the most.

Another way the model could be improved would be to obtain the data all from studies that formatted the data in the same way. This would help ensure that all the definitions of keywords and the population the sample was taken from are standardized, so that the data can be compared more accurately. Furthermore, our model only took into account data from people between the ages of 12 and 17. Using a broader range of ages would allow us not only to generalize our model but also to compare the effects of age on the likelihood of using a certain drug.

5   Problem 3: Ripples

5.1   Defining the Problem

- Determine what factors to consider in the impact of a drug.

- Use these factors to rank the overall impact of that drug as compared to other drugs.

5.2   Assumptions and Justifications

- Economic burdens have similar impacts.

    Justification: This allows us to evaluate and standardize the economic burdens within our model.

- We can evaluate all of the factors together.

    Justification: Our model evaluates all the factors together in order to be able to rank the different drugs.

- The United States is a sufficient representation of the world.

    Justification: Based on available data, we were able to find a solid set of data for the United States, making it the best option to consider as representation within our model. Due to a lack of other data, we decided to utilize data from the United States population.

- The different factors have weights determined by impact on life.

    Justification: These weights make our model more consistent with reality as these different factors do not impact life in the same way and thus need to be weighted.

5.3   Developing the Model

- Economic Burden on the US: The economic burden is a broad factor that takes into account the cost of law enforcement, health care, and loss in productivity.

- Scope of Issue (measured by number of users): A greater number of users means that in general, more people will be impacted by users. A distinction was made between users 18 and older and those under 18 because minors will have their futures impacted as well and thus were more significant.

- Deaths per year: This reflects the various different types of death caused per year including sickness, overdose, and accidents.

Table 5: Impact of Factors on the US Population per Year[30,31,32,33]

	Nicotine	Marijuana	Alcohol	Opioids	Total
Economic Burden (in billions of dollars)	500	16.4	249	78.5	843.9
Deaths per Year (number of people)	480,000	0	88,000	47,600	615,600
Number of Users 18+	49,500,000	24,000,000	134,500,000	10,909,000	218,909,000
Number of Users 12-17	855,000	1,600,000	2,300,000	891,000	5,646,000

To combine this data, we compared the values against each other and then weighted them. We compared the cost to the total cost; for example the economic burden of nicotine is $\frac{500}{843.9}$. After that, we multiplied by a weighted importance of a factor. We assigned an importance 0.5 to economic burden, 0.5 to deaths, 0.2 to 18+ users and 0.3 to 12-17 users. Our formula overall is

$$\text{score} = .5\left(\frac{\text{EB}}{T(\text{EB})}\right) + .5\left(\frac{\text{DPY}}{T(\text{DPY})}\right) + .2\left(\frac{\text{MA}}{T(\text{MA})}\right) + .5\left(\frac{\text{MI}}{T(\text{MI})}\right)$$

where:
EB = economic burden in billions of dollars
DPY = deaths per year (number of people)
MA = number of users 18 or above
MI = number of users ages 12-17
T(x) = total

5.4   Applying the Model

We inserted our values into our equation in order to get the following values:

Table 6: Impact Scores of Nicotine, Marijuana, Alcohol, and Opioids

Substance	Score
Nicotine	0.776761843
Marijuana	0.116659651
Alcohol	0.4640968372
Opioids	0.1424816689

5.5   Conclusion: Ranking

Nicotine was the substance with the largest negative impact on the US, with a score of 0.776. Although it was less widely used than alcohol, the larger number of economic burden and deaths meant that it had a greater negative effect per user. Alcohol was the next worst, with a score of 0.464, being the most widespread and also having the second highest number of deaths attributed and economic burden. Opioids rank at having the third most harmful impact, with a score of 0.142. Finally, marijuana seems relatively harmless as, although it has a large number of users, the impact per user is low. Marijuana had a score of only 0.117.

5.6   Evaluating the Model

5.6.1   Sensitivity Analysis

In our model, we used different weights to address the different factors considered. These weights have an impact on the ranking of the drugs, making it sensitive to our model. However, the values utilized in our model are sound, which makes it less sensitive to the difference in weights. Furthermore, these weights reflect impact on life but could be made more accurate and robust to the situation.

5.6.2   Strengths and Weaknesses

One of the strengths of this model is that its variables are an aggregate of different factors that can be important in ranking the significance levels of different drugs based on their impact. Another strength of the model is that we take into consideration how the factors would magnify the result of others.

However, this model also has several weaknesses. One of the weaknesses of this model is that it only accounts for a few factors. Another weakness of the model is that there is some arbitrariness in deciding how the factors influence each other. This occurred due to the difficulty of quantifying the relationship between different factors.

5.6.3   Extensions of the Model

One of the ways we could improve this model would be adding more factors. Although this could potentially introduce more arbitrariness to the model due to vague relations between

more factors, this could also expand the application of this model. Finding more data on the relationship between the different factors could help reduce the arbitrariness as well.

6   Works Cited

1. http://worldpopulationreview.com/countries/united-states-population/

2. https://www.denverpost.com/2018/05/15/adults-try-vaping-study/

3. https://www.atr.org/new-cdc-data-more-9-million-adults-vape-regularly-united-states

4. http://faculty.washington.edu/heathml/2018spring307/euler_method.pdf

5. https://www.ncbi.nlm.nih.gov/pubmed/22231818

6. http://math.colgate.edu/~wweckesser/math312Spring06/handouts/IMM$_S I_M$odel.pdf

7. https://nces.ed.gov/programs/raceindicators/indicator_rbb.asp

8. https://www.census.gov/quickfacts/fact/table/US/PST045217

9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3410945/

10. https://www.nih.gov/news-events/news-releases/vaping-popular-among-teens-opioid-misuse-historic-lows

11. https://www.statista.com/statistics/639344/rate-of-opioid-abuse-us-area-income-level/

12. https://www.drugabuse.gov/sites/default/files/drugfactsmtf.pdf

13. http://www.worldometers.info/world-population/us-population/

14. https://teens.lovetoknow.com/Statistics_on_Peer_Pressure

15. https://www.hhs.gov/ash/oah/facts-and-stats/changing-face-of-americas-adolescents/index.html

16. https://www.drugabuse.gov/related-topics/treatment/what-to-do-if-you-have-problem-drugs-teens-young-adults

17. https://www.reuters.com/article/us-health-ecigs-us-adults/almost-one-in-20-u-s-adults-now-use-e-cigarettes-idUSKCN1LC2DN

18. https://institutefordiseasemodeling.github.io/Documentation/general/model-si.html

19. https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm

20. https://www.chegg.com/homework-help/graph-shows-cigarette-consumption-billions-united-states-yea-chapter-1.
    2-problem-54e-solution-9780077356651-exc

21. https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2015/NSDUH-DetTabs-2015/NSDUH-DetTabs-2015.pdf

22. https://www.hazeldenbettyford.org/education/bcr/addiction-research/socioeconomic-advantage-edt-818

23. https://www.statista.com/statistics/639344/rate-of-opioid-abuse-us-area-income-level/

24. https://www.drugwarfacts.org/table/mj_prevalence

25. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3600369/

26. https://pubs.niaaa.nih.gov/publications/arh40/152-160.htm

27. https://pubs.niaaa.nih.gov/publications/arh27-1/79-86.htm

28. https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis

29. https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm

30. `https://drugabuse.com/marijuana-vs-alcohol/`

31. `https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates`

32. `https://health.mo.gov/data/opioids/pdf/opioid-dashboard-slide-29.pdf`

33. `https://www.drugscience.org/Archive/bcr4/1Fed_costs.html`

7   Appendix

7.1   SI Compartmentalizing Addiction Model using Euler's Method

```python
import numpy as np
from matplotlib import pyplot as plt

#years
xi = 0
xf = 10

y0 = 16416150 #addicted population
y1 = 311900260 #not addicted population
yt = 328316410 #total population

n = 11 #number of points
var = 0.2 #rate of negative interactions per year between addicted and not
    addicted

deltax = (xf-xi)/(n-1) #step size

x = np.linspace(xi, xf, n) #array of years
y = np.zeros([n]) #array of not addicted people
y2 = np.zeros([n]) # array of addicted people
y[0] = y1
y2[0] = y0

#euler's method
for i in range (1, n):
    #not addicted
    y[i] = y[i-1]-(0.3 * var * (yt - y[i-1])*y[i-1]/(yt)) - 0.3 * 0.014 * y[i-1]
        + 0.0071*yt
    #addicted
    y2[i] = y2[i-1] + ((0.3 * var * (yt - y2[i-1]) * y2[i-1])/(yt)) + 0.3 * 0.014
        * (yt-y2[i-1])
    #population change
    yt = 1.0071 * yt
#print values
for i in range (n):
  print ("Not addicted:", x[i], y[i])
  print ("Addicted:", x[i], y2[i])
plt.plot (x, y, 'o-', color = 'r', label = 'Not Addicted') #not addicted
    population
plt.plot(x, y2, 'o-', color = 'b', label = 'Addicted') #addicted population
plt.legend(loc= 'center right')
```

```python
plt.xlabel("Years")
plt.ylabel("# of people")
plt.title("Predicted Spread of Nicotine Use over 10 Years")
plt.show()
```

7.2   Simulation of Drug Usage in 300 High School Students

```python
from numpy.random import choice
from collections import Counter

#probability of usage of each type of drug based on characteristics
white = [14,13,13.2,3.9]
aa =[7.9, 12.6, 8.6, 4.1]
his = [9.3,12.6,11.2,4.1]
asi = [3.5,4.5,4.5,1.9]
male = [12.7,12.9,9.3,3.6]
fem = [10.4,12.3,9.9,4.2]
high =[16.5,39.1,75.8,2.7]
low = [26.7,19.3,51,6.3]

#different characteristics
category = [white,aa,his,asi,male,fem,high,low]

#ranges of each drug based on characteristic
nic = [10.5,2.3,10.2]
mar = [8.5,0.6,19.9]
alc = [8.7,0.6, 8]
opi = [2.2,0.6,3.6]
probcate = [nic,mar,alc,opi]

#total ranges of each drug
ranges= [23,29,17,6.4]

#defines characteristics
def organ(x):
    stats = []
    if(x[0]=="high"):
        stats.append(6)
    else:
        stats.append(7)
    if(x[1]=="white"):
        stats.append(0)
    elif (x[1]=="aa"):
        stats.append(1)
```

```python
        elif (x[1]=="hispanic"):
            stats.append(2)
        else:
            stats.append(3)
        if(x[2]=="male"):
            stats.append(4)
        else:
            stats.append(5)
        return stats


#utilizes probability equation
def equation(cate, drug):

    #probability of drug based on characteristic
    socioprob = category[cate[0]][drug]/100
    raceprob = category[cate[1]][drug]/100
    gendprob = category[cate[2]][drug]/100

    #equation
    addictprob = socioprob*(probcate[drug][0]/ranges[drug])
        +raceprob*(probcate[drug][1]/ranges[drug])
        +gendprob*(probcate[drug][2]/ranges[drug])
    return addictprob


#socioeconomic status weighted percentages
socioecon = ["high","low"]
sprobs = [0.5,0.5]

#race status weighted percentages
race = ["white","aa", "hispanic", "asian"]
rprobs = [0.52,0.16,0.27,0.05]

#gender status weighted percentages
gender = ["male","female"]
gprobs = [0.5,0.5]

#counters
scount = Counter()
rcount = Counter()
gcount = Counter()
ncount = Counter()
mcount = Counter()
acount = Counter()
ocount = Counter()
```

```python
#random simulation
for i in range(300):
    #random assignment of characteristics
    c1=choice(socioecon,p=sprobs)
    scount[c1] += 1
    c2=choice(race,p=rprobs)
    rcount[c2] += 1
    c3=choice(gender,p=gprobs)
    gcount[c3] += 1

    person = [c1,c2,c3]
    person = organ(person)

    #choosing of usage based on probabilities
    nicotine = equation(person, 0)
    nicaddict = ["use", "not"]
    naddprob = [nicotine,1-nicotine]
    marijuana = equation(person,1)
    maddict = ["use", "not"]
    maddprob = [marijuana,1-marijuana]
    alcohol = equation(person, 2)
    aaddict = ["use", "not"]
    aaddprob = [alcohol,1-alcohol]
    opiod = equation(person, 3)
    oaddict = ["use", "not"]
    oaddprob = [opiod,1-opiod]

    n =choice(nicaddict,p=naddprob)
    m =choice(maddict,p=maddprob)
    a =choice(aaddict,p=aaddprob)
    o =choice(oaddict,p=oaddprob)

    ncount[n] +=1
    mcount[m] +=1
    acount[a] +=1
    ocount[o] +=1

    #print for each person
    print(i,c1,c2,c3,n,m,a,o)

#print counts
print()
print(scount)
print(rcount)
print(gcount)
```

```python
#print usage counts
print()
print("Nicotine:",ncount)
print("Marijuana:",mcount)
print("Alcohol:",acount)
print("Opioid:",ocount)
```