THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF INDUSTRIAL AND MANUFACTURING ENGINEERING

Neonatal Abstinence Syndrome Vulnerability Index for Counties in Pennsylvania

NIKAELA HAAS
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Industrial Engineering
with honors in Industrial Engineering

Reviewed and approved* by the following:

Soundar Kumara
Allen E. Pearce and Allen M. Pearce Professor of Industrial and Manufacturing Engineering
Thesis Co-Supervisor

Kamesh Madduri
Associate Professor of Computer Science and Engineering
Thesis Co-Supervisor

Andris Freivalds
Lucas Professor of Industrial and Manufacturing Engineering
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Neonatal abstinence syndrome (NAS) has been a growing concern in Pennsylvania. NAS occurs when a mother uses opioids during pregnancy, resulting in the infant having a dependence on the drug [1]. NAS stays are costly compared to the costs of newborn infants without NAS [2] and understanding the features of a county that contribute to NAS will allow for improved early intervention rather than treating the symptoms after birth. The objective of this study is to understand which features contribute to high NAS rates in Pennsylvania by looking at data from previous years. The features will then be used to create a NAS Vulnerability Index (NVI) for each county in Pennsylvania. This NVI can be used by researchers and other policy makers to better allocate limited resources to counties most affected by NAS.

By collecting data on a variety of factors including demographics, socioeconomic and health related factors, an XGBoost Regressor based Machine Learning model was fit and the impact of each factor on the output was calculated using SHAP values to reveal the three most impactful features on NAS rates were 1) the percentage of women who smoked during pregnancy, 2) the rate of opioid use disorders among pregnant women, and 3) the estimated opioid dispensing rate to pregnant women. These three features were then used to calculate NVI for each county in Pennsylvania.

The top five counties with the highest NVI were Venango, Elk, Fayette, Clearfield, and Lawrence. Their NVIs were 0.973, 0.964, 0.944, 0.918, and 0.899, respectively. Counties in the southeast region of Pennsylvania on an average had lower NVIs than other regions of the state. The NVI provides an easily explainable metric that can be used by government officials, researchers, or other policymakers. By examining the specific features that increase NAS rates in a county, efforts can be better targeted towards mitigating these risks. Future work should expand to consider other data and apply the methodologies developed in this thesis to the states outside of Pennsylvania to see how NVI trends in other areas of the US.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Soundar Kumara for ultimately having such a large impact on me during my time at the Pennsylvania State University. His mentorship to not only my research but as well as my professional and personal development will be something that I will carry with me through my lifetime. The opportunities Dr. Kumara has given me have undoubtedly changed the course of my career. Thank you, Dr. Kumara. This work could also not be completed without the help from Dr. Kamesh Madduri. His knowledge and expertise data analytics were essential in ensuring this work is value. Lastly, I would like to thank my colleague Vishnu Kumar for his invaluable help throughout the completion for this project. Our weekly meetings, discussion, and work allowed for this research project to come to fruition. I wish you luck with your future endeavors as you have most definitely shaped mine.

Chapter 1

Introduction

1.1 Overview

NAS occurs when newborn infants experience withdrawal symptoms from opioid exposure prior to birth [1]. Drugs consumed by the mother such as oxycodone, methadone, codeine, heroin, and buprenorphine can pass through the placenta during pregnancy and result in a drug dependency for the infant at birth [2]. Symptoms of withdrawal include vomiting or diarrhea, poor feeding, seizures, excessive crying, muscle tightness, trembling, and fevers [3].

Because of the nature of the healthcare industry, large amounts of data are recorded for many people both affected and unaffected by disease. Using big data analytics may give insight to trends and patterns among populations that may be found by a computer that may not be as noticeable on a case-by-case diagnosis as seen by doctors [4]. In this thesis we develop a vulnerability index based upon Pennsylvania data, which is generalizable to all the other counties in the USA.

1.2 Background and Motivation

Opioids are prescribed to help manage pain. Back and pelvic pain is one of the most common pregnancy complications, and although there are risks, opioids may still be prescribed to mitigate the pain [2]. Opioids such as Buprenorphine and Methadone may also be used to treat drug addictions of mothers during pregnancy in medication-assisted treatment (MAT) [5]. However, treatments may still result in NAS, but to a lesser degree of severity of the symptoms in the infant.

NAS results in extended and costly hospital stays for newborns. In 2012, it is estimated that the annual cost of NAS admission (nationally) was $315,665,913 [6]. The average total hospital cost of an infant affected by NAS is approximately $16,893 whereas an unaffected infant hospital stay is approximately $5,610.

NAS has been a growing concern in the United States according to data reported by the Agency for Healthcare Research and Quality (AHRQ) and Healthcare Cost and Utilization Project (HCUP) [7]. They also report that Pennsylvania has seen similar growth but with a higher rate per 1000 newborn hospitalizations than the average for the United States. Figure 1 shows the NAS rate per 1,000 deliveries in both the US and Pennsylvania.

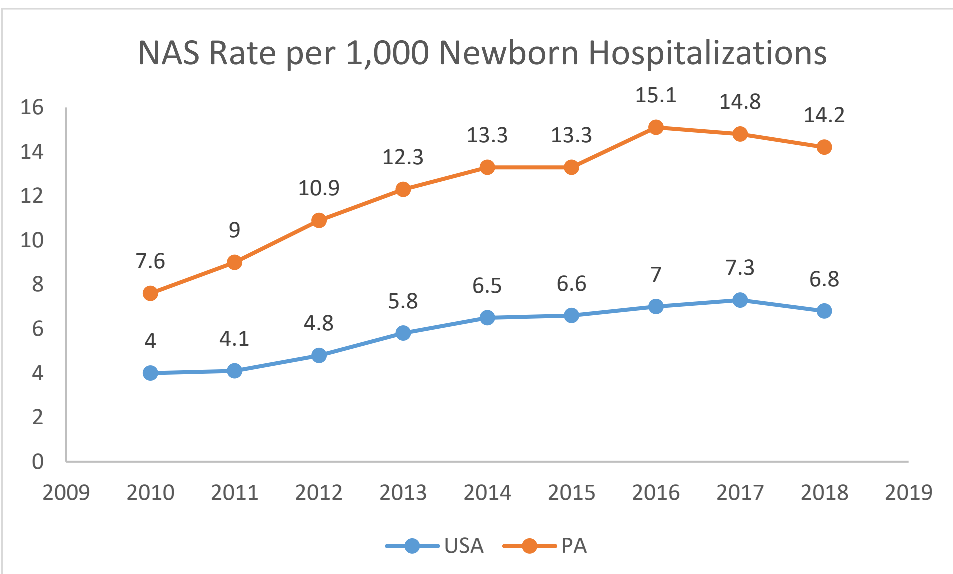

Figure 1: NAS Counts among 1000 Newborn Hospitalizations. Source: Healthcare Cost and Utilization Project (HCUP)

The Pennsylvania Department of Health labels the opioid epidemic as “the worst public health crisis in Pennsylvania” [8]. This is not only reflected in the rate of opioid overdoses and hospitalizations, but also in the growing NAS rates. From 2000 to 2017, the rate of NAS cases per 1,000 newborn stays has increased by 1,096%, according to estimates by the Pennsylvania Health Care Cost Containment Council (PHC4) [9]. In the fiscal year of 2017, PHC4 estimates $14.1 million was spent in NAS-related stays.

1.3 Problem Discussion

Although data is available relating to NAS for Pennsylvania, there lacks a comprehensive model that analyzes a range of characteristics for the mother and the county she resides in. Understanding what factors increase the rate of NAS specifically within Pennsylvania will allow for government and healthcare officials to not only better target the counties most affected, but also better address what factors are causing the increased risk. Data analytics can be used to understand and evaluate factors within Pennsylvania's population.

In specific this thesis addresses the design and development of a county-wide vulnerability index for NAS with PA as a use case.

1.4 Research Objectives

The objective of this study is to leverage data to find trends in NAS rates in Pennsylvania to create a NAS Vulnerability Index (NVI) for every county. A vulnerability index calculated will allow for better allocation of limited resources and to formulate policies to help the counties most susceptible to NAS cases. This study focuses on the characteristics of the mothers to understand the risk of an infant developing NAS. Therefore, characteristics of an infant that are measured after birth will not be included. Only information that is available prior to birth about the mother and other contributing factors will be used to create the vulnerability index.

1.5 Thesis Structure

This thesis is split into six chapters. The second chapter details the background for this paper including previous predictive models for NAS cases that were used to aid in selecting initial variables for analysis and description of the model used. The third chapter describes the methods used in this study. This includes the data descriptions, parameters used for the model, and explanation of feature scores. The fourth chapter reports the results from the analysis. The fifth chapter is a discussion of the results and validation data as well. The sixth chapter deals with the conclusions and final thoughts and future work.

Chapter 2

Previous Work

Extensive data analyses have been completed across the nation data to analyze NAS and its contributing factors [10-12]. This section will examine several NAS models used to aid in selecting important factors for initial analysis for developing NAS Vulnerability Index and apply it to Pennsylvania county level data. An opioid mortality vulnerability index for the state of Indiana will also be briefly discussed.

2.1 Tennessee Medicaid NAS Predictive Model

A model has been created using Tennessee Medicaid data to assess the risk of an infant developing NAS based on a variety of factors measured at birth [10]. Mother and infant dyads were examined to create a model to predict the likelihood of an infant developing NAS. 80% of infants diagnosed with NAS are covered by Medicaid, so examining the Medicaid data captured the majority of the NAS population. The maternal age at delivery was between ages 15 and 44 years. Similar to the method used in this study, all independent variables were chosen a priori based on existing literature concerning NAS and other information known about the mother. The outcome measured for the study was a NAS or no NAS diagnosis using the ICD-9 code 779.5.

A multivariable logistic regression was used to develop the model for predicting a NAS diagnosis for a given mother-infant dyad. The study resulted in two models: a general population model and a high-risk model. The general population model is calibrated well for infants with a risk calculated to be less than 40%. However, for a higher risk infant, another model exists with factors holding different levels of contribution to a NAS diagnosis to avoid underestimating the risk for the infant.

Since this model is created only using data from Tennessee, the factors that contribute the most to NAS cannot be expected to be the same for Pennsylvania. However, they offer a basis for which factors to look for in Pennsylvania. The three factors that contributed the most in the general population model were use of maintenance opioids, the smoking status of the mothers, and the mother having Hepatitis-C. For the high-risk model, the most significant factors were use of maintenance opioids, immediate release opioids, and/or extended-release opioids. The data used in this model came from the Tennessee Medicaid Program, TennCare. Therefore, it cannot be applied accurately to privately insured dyads.

Due to the limitations of this study using publicly available data, not all of this data could be found that was used in the study completed in the TennCare study. However, the use of opioids, smoking status, and median maternal age at birth are used for initial analysis in this study.

2.2 Additional NAS Risk Factors

Another study examined Medicaid from 46 states to estimate the risk of NAS based on the length and timing with respect to delivery of an opioid prescription during pregnancy when other NAS risk factors were present [11]. This study examined a cohort of pregnant women who were prescribed at least one opioid during pregnancy. Prescriptions were characterized as (a) short term (less than 30 days) vs. long term (more than 30 days), (b) early use (use in the first two trimesters of pregnancy) vs. later use (use in the last trimester of pregnancy), and (c) the cumulative dose in morphine milligram equivalent. Additional risk factors were used to estimate the risk of the neonate developing NAS during the pregnancy by splitting the cohort into four groups. The additional risks were (1) history of opioid misuse, (2) history of any other drug or alcohol misuse, (3) non-opioid psychotropic drug use, and (4) smoking during pregnancy.

Several logistic regression models were created for each study group with a NAS diagnosis as the outcome variable. One model had the cumulative dose of the prescription as the predictor variable, and the other had the total days of opioid use as the predictor variable. Besides prescription characteristics and additional NAS risk factors, other features studied were the maternal age, race, region of residence of the mother, presence of pain conditions, and use of benzodiazepines of the mother.

The study found that using prescribed opioids without having any of the additional risk factors resulted in only a small risk of NAS. Once pregnant women had any of these additional risk factors, NAS risk increased dramatically. Variables relating to these additional risk factors will be as closely replicated as possible with data that is publicly available for the calculation of the NVI for counties in Pennsylvania.

2.3 Opioid vulnerability index of Indiana

A study to create an opioid vulnerability index in Indiana was completed using 2017 opioid related data and other socioeconomic factors [12]. Similar to this study, all features for analysis were publicly available. The vulnerability index was calculated using mean standardized variables with the opioid-related mortality rate as the outcome variable. The index was then used to evaluate accessibility to OUD treatment, harm reduction services, and other opioid response programs.

The study found that opioid-related emergency room visits, opioid-related arrest rates, chronic hepatitis-C infection rates, opioid prescription rates, unemployment rates, and the percentage of female-lead households were positively associated with opioid mortality. Counties with higher population densities were at more risk than counties with low population densities.

Since the analysis in this paper is related to NAS rather than opioid mortality, this analysis cannot make use of all the independent variables used in the opioid vulnerability index from Indiana. However, the opioid-mortality vulnerability index illustrates the range of applications the index may be used for.

Chapter 3

Methods

In this chapter a discussion on data and the methods used in this thesis are discussed in detail. The reasons for using particular data sets, methodology and inferencing are addressed. The quantitative aspects of NVI and importance of SHAP values are discussed.

3.1 Data Description

This section will give a detailed description of the data that was used to create the model that will later be discussed. All the data used in this study is publicly available data and was chosen prior to the analysis based on existing literature. Table 1 summarizes the data and their sources. A total of 17 independent variables from 2016 to 2019 were used in the model with the response variable as the rate of NAS per 1000 livebirths. Governor Tom Wolf came into office in 2015 bringing attention to the opioid epidemic in Pennsylvania [13]. 2016 marks a year where more data is collected and published related to opioids in Pennsylvania. Data in more recent years has yet to be published for many of the variables.

Table 1: Features and response variable data used in the model

Description	Source
Birth rate per 100,000 total population	PA DOH
The median age of mothers at delivery	PA DOH
Percentage of livebirths where mother is black	PA DOH
Percentage of livebirths where mother is white	PA DOH
Percentage of livebirths where mother is other than black or white	PA DOH
Percentage of mothers who smoked during pregnancy	PA DOH
Percentage of mothers insured at delivery	PA DOH
Rate of Pregnant Women with Average Daily MME > 90 per 10,000 Population	PA DOH
Count of drug and alcohol treatment facilities	DDAP
Rate of Pregnant Women with OUD Diagnosis at Delivery per 1000 deliveries	PHC4
Percentage of births to mothers with less than nationHS education	KCDC

Percentage of pregnant women in poverty	CB, SAIPE
Estimated rate of opioids dispensed to pregnant women out of 10,000 total population	CDC
Percent of unemployed pregnant women out of female population 15-39 years	BLS
Percentage of pregnant women with early & adequate prenatal care	PA DOH
Percentage of mothers with healthy weight before birth	PA DOH
Percentage of pregnant women who drank excessively during pregnancy	CHRR
Rate of NAS per 1000 livebirths	PA DOH

Legend: PA DOH = Pennsylvania Department of Health, DDAP = Department of Drug and Alcohol Programs, PHC4 = Pennsylvania Health Care Cost Containment Council, KCDC = Kids Count Data Center, CB = Census Bureau, SAIPE = Small Area Income and Poverty Estimates, CDC = Centers for Disease Control and Prevention, BLS = Bureau of Labor Statistics, CHRR = County Health Rankings & Roadmaps

The Pennsylvania Department of Health (PA DOH) reports birth statistics[1] for every year at the county level. The statistics provided by the PA DOH were used to calculate the birth rate per 100,000 residents of a county, the median age of mothers at the time of delivery, the race of the mother, smoking status during pregnancy, and insurance status during pregnancy. The birth rate was calculated by using the total population estimates from the US Census Bureau estimates. The total population estimates from the Census Bureau were used for a number of different factors used in the model that will be discussed later.

The risky prescribing measures is the estimated rate of pregnant women who have an average daily morphine milligram equivalent (MME) over 90 mg per 10,000 population. The PA DOH provides the data for the rate of individuals within a county receiving a daily MME over 90 mg, and this rate was used to estimate the rate of pregnant women receiving over 90 mg daily. According to the Opioid Dispensing Guidelines published by the Commonwealth and the Pennsylvania Pharmacists Association, the Pennsylvania Medical society recommends a prescription of 100 mg as the maximum daily dosage [14]. Examining the rate of pregnant

[1] "These data were provided by the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions."

women receiving over 90 mg will reflect pregnant women who are near or exceeding the recommended maximum dose.

The count of drug and alcohol treatment facilities is provided by the PA Department of Drug and Alcohol Programs (DDAP). The counts are from 2018, but they were applied to every year in the study under the assumption that the number of facilities is not changing dramatically with other population parameters.

The rate of pregnant women diagnosed with an opioid use disorder (OUD) at the time of delivery out of 1,000 deliveries is provided by PHC4. The percentage of women giving birth with less than a high school education was estimated from the Annie E. Casey Foundation Kids Count Data Center. The data for the percentage of mothers with less than a high school education is available from 2000 to 2017. Forecasting methods were used to estimate the percentages for 2018 and 2019.

Estimates for the percentage of pregnant women in poverty was based on the Census Bureau data reported for percentage of women in poverty. The Census Bureau only reported 40 out of 67 counties, so in order to estimate the remaining 27 counties, the percentage of total population in poverty from the Small Area Income and Poverty Estimates (SAIPE) Program was used. A linear regression model with an R^2 value of 0.93 was used to estimate the percentage of females in poverty.

The opioid dispensing rate from the CDC was used to estimate the rate of opioids dispensed to pregnant women with OUDs. The dispensing rate is a measure of opioid prescriptions dispensed to pregnant women per 10,000 county residents. The unemployment rate for pregnant women was based on the US Bureau of Labor Statistics. The calculations make the critical assumption that all pregnant women are working or seeking a job. The percent of pregnant women with a healthy weight and percentage of women with early and adequate prenatal care is provided by the PA DOH. The percentage of pregnant women drinking alcohol excessively was estimated using the percentage of adults who drink excessively reported by the County Health Rankings & Roadmaps data published by the University of Wisconsin.

The response variable used to create the model is the NAS rate per 1,000 live births. The numbers are reported by the PA Bureau of Epidemiology. There are only NAS counts related to opioid drug use, not any other drug that may cause withdrawal symptoms.

3.2 Correlations

A correlation matrix was used to check for multicollinearity between the features described above. Variables with a correlation of greater than 0.8 were eliminated. After examining the correlation matrix, the percentage of mothers who are black and the percentage of mothers who are white were removed from the rest of the analysis. The analysis of feature importance consists of 15 features in total. The correlation matrix of all the independent variables can be found in Table 8 in the Appendix.

3.3 NAS Rate Model Fitting

The eXtreme Gradient Boosting model (XGBoost) is used in the study to create the model for predicting the county-wide rate of NAS. The XGBoost model is based on gradient boosting decision trees and were designed to optimize the speed and performance of the model [15]. Gradient boosting is a technique that uses new models to predict the residuals of prior models. The final prediction model will be the sum of the old models and the new model. A gradient descent algorithm is used to minimize loss when adding new models. In addition, regularized boosting is used to minimize overfitting of the data. Studies have shown that the XGBoost model has consistently performed at a faster computation speed than other implementations of gradient boosting [16].

The following equations can be used to describe the XGBoost model [17]. The XGBoost objective model can be split into two parts:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \qquad (1)$$

$L(\theta)$ denotes the training loss function, which is used to measure the model performance of the training data. The model aims to minimize $L(\theta)$. $\Omega(\theta)$ denotes the regularization term. This is used to minimize the effects of overfitting the training data. The next equation will describe the complexity of the tree, $\Omega(f)$.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \qquad (2)$$

T is used to denote the number of leave of each tree, and ω is the vector score of each leaf. ω_j are all independent of each other. The objective function of the XGBoost model can be rewritten as:

$$Obj = \sum_{j=1}^{T} \left[G_j\omega_j + \frac{1}{2}\left(H_j + \lambda\right)\omega_j^2\right] + \gamma T \qquad (3)$$

The XGBoost Model was completed using Python.

3.4 Explainability using feature importance

Feature importance is often used to give features a numerical score to better understand how valuable a feature used in the creation of a model is using boosted decision trees [18]. A higher importance score indicates the variable was used more in a decision tree. Features importance scores offer a number of different uses. They can be used to better understand the data by examining the relationship between variables and the dependent variable. It also aids in understanding which variables are irrelevant and may be left out of the model to improve performance. The feature importance scores will be used in this study to choose the most important variables for the NVI.

This study uses the SHapley Additive exPlanation (SHAP) values to calculate the importance score of the 17 features. The approach used in this study to calculate SHAP values was proposed by Lundberg and Lee in a paper published in 2017 [19]. SHAP values interpret the impact of each feature at a certain value and compare it to a prediction made if the feature took some baseline value [20]. The equation below can be used to describe the SHAP values [17]:

$$\phi_i = \sum_{S \in N} |S|!\,(n - |S| - 1)!\,\frac{1}{n!}\,[v(S \cup \{i\} - v(S)] \qquad (4)$$

Where,

ϕ_i = contribution of feature *i*

N = group of *n* features

v(N) = output based on marginal contribution of group N

3.5 Development of NVI

The results from the feature importance analysis were used to calculate the NVI. By computing the SHAP values for all the variables in the initial studies, this study proposes to use the top three most impactful features to calculate the NVI for all the counties in Pennsylvania. Isolating the top three most impactful variables ensures that these features are the only ones reflected in the vulnerability index, rather than giving less important variables influence on the index. Other vulnerability indexes exist such as the CDC/ATSDR Social Vulnerability Index [21]. This vulnerability index takes 15 variables into account to help determine which counties are at most risk to disease outbreaks & human or natural disasters. The development of a vulnerability index allows for a flexible range of use by other researchers and policy makers. The NVI will give others the ability to target and prioritize the counties that are most at risk. The vulnerability index also allows for easy comprehension and comparison between counties.

The calculation of the NVI is a weighted function of the top three most important features. The equation and notation are explained below [22]:

$$\text{NVI}_j = \sum_{i=1}^{n} W_i C_{ij} \quad \text{for each j (5)}$$

Where,

i = 1, 2, 3…n feature under consideration (n will depend upon the number of important features derived from SHAP values, in the limiting case $n = p$ dimensionality of the features)

j = 1, 2, … 67 county under consideration

W_i = Weighted contribution of the i^{th} most impactful feature

C_{ij} = percentile score of feature i in county j

NVI_j = NAS vulnerability index of county j

The weighted contribution, Wi, was calculated using the following equation:

$$\text{Wi} = \frac{s_i}{\Sigma s_i} \qquad (6)$$

Where,

s_i = mean SHAP value for feature i

$\Sigma_{i=1}^{n} W_i = 1$, where n refers to the number of features considered as important

Chapter 4

Results

4.1 Trends and Distribution

The average NAS rate between 2016 and 2019 was calculated to view how NAS rates vary throughout the state of Pennsylvania on an aggregate scale. The average NAS rates per 1000 deliveries are shown in Figure 2. On average, the southeast region of Pennsylvania tended to have lower average NAS cases than other regions of the state. The five counties with the highest NAS rate per 1000 deliveries were Greene County, Venango County, Fayette County, Elk County, and Lawrence County. Their average NAS rates were 61.58, 54.88, 54.83, 40.78, and 34.81, respectively. Results for the NVI will later conclude that these counties are high-risk based on the top 3 risk factors used to quantify the index.

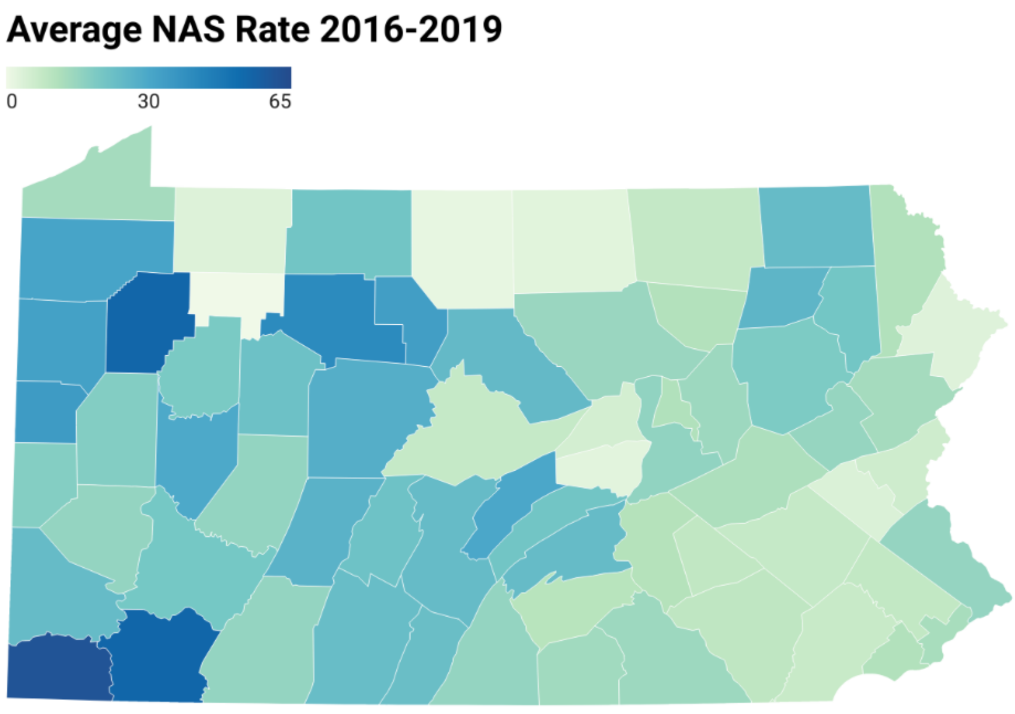

Figure 2: Average NAS rate per 1000 deliveries from 2016 to 2019

4.2 XGBoost NAS Model

The model produced by XGBoost was used to understand the relationship between the variables. Changing the parameters such as the number of trees, the learning rate, and the split between training and test data allowed for finding the best model. However, this study is not used for predictive purposes. Instead, it is to better understand how the variables affect each other and the dependent variable. The model allows for the feature importance scores to be calculated, and these results are discussed in the next section.

4.3 Feature Importance

This section will describe how the feature importance using the SHAP values is interpreted.

4.3.1 Mean SHAP Values

Figure 3 shows the mean SHAP values calculated from the XGBoost based NAS prediction model. Based on this graph, the top three features will be used to calculate the NVI using the method previously described. The top three factors are (a) the percentage of women who smoked during pregnancy, (b) the rate of women diagnosed with OUDs at delivery out of 1,000 deliveries, and (c) the dispense rate to pregnant women. Running the XGBoost model with different parameters consistently resulted in these top three most contributing factors. The mean SHAP values were approximately 4.0, 2.3, and 1.6, respectively for the top three features. These values were used to find the weighted contributions of the features for the NVI. After these top three features, the remaining 11 independent features had less of an effect on NAS.

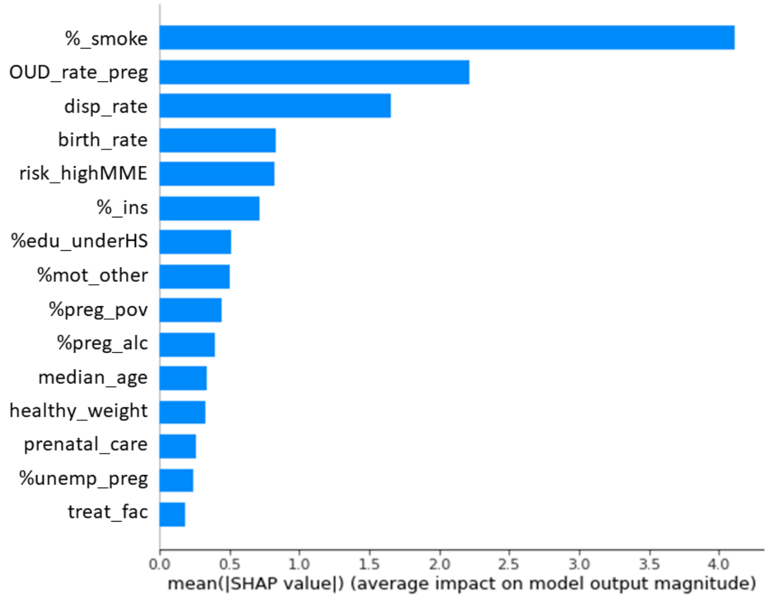

Figure 3: Mean SHAP values calculated from XGBoost Model

Figure 4 shows the SHAP summary plot created from the XGBoost based NAS prediction model. The SHAP summary plot will be interpreted for the two most contributing factors to aid in SHAP value clarification.

Interpretation 1: Smoking status during pregnancy

When examining Figure 4, for *"%_smoke*", red points represent high percentages of women smoking during prengancy in a county, and blue points represent low percentages of women smoking during pregnancy. As the percentage of women smoking during pregnancy increases in a county (points become more red), the SHAP values become more positive, meaning they have a "positive" effect on the NAS rate by increasing its value. As the percentage of women smoking during pregnancy decreases (points become more blue), they have a more negative SHAP value, having a "negative" the NAS rate. In other words, as percentage of women smoking during pregnancy increases, they have an increasingly positive impact on the NAS rate.

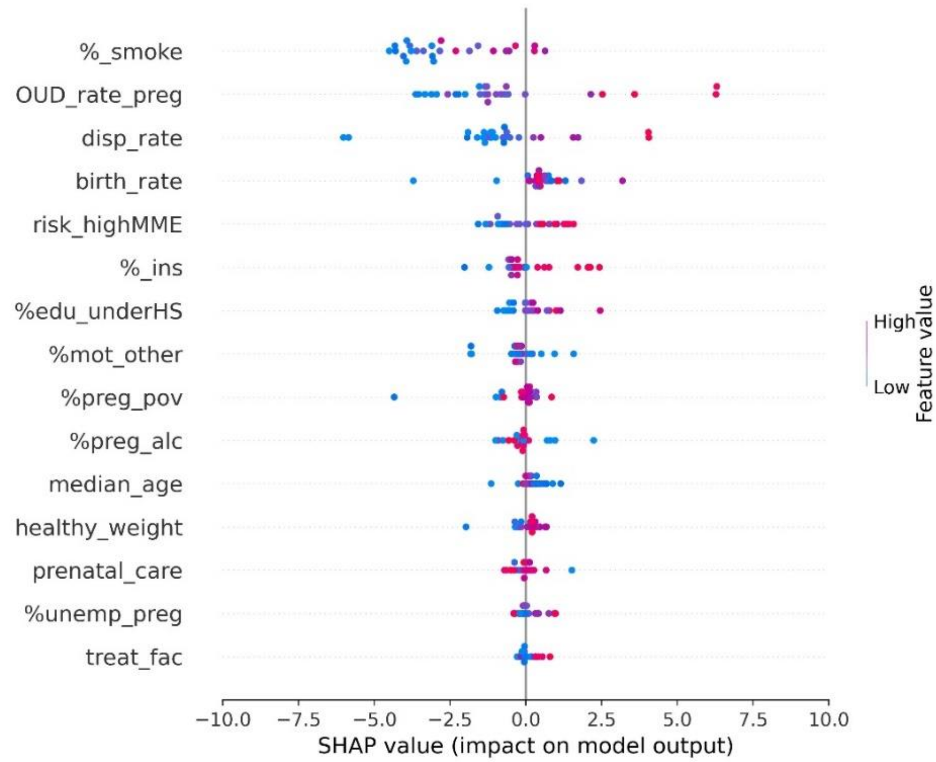

Figure 4: SHAP summary plot for the XGBoost based NAS prediction model

Interpretation 2: OUD rate out of 1000 deliveries

Looking at "*OUD_rate_preg*" in Figure 4, as the OUD rate gets increasingly higher (more red points), OUD rates have a more positive effect on the NAS rates. This means that higher OUD rates tend to push the NAS rates in a positive direction. On the otherside of the argument, the lower OUD rates (more blue points) have a more negative SHAP value. This means that lower OUD rates have a tendency to decrease the NAS rate. In other words, as OUD rates increase, they have an increasingly positive impact on the the NAS rate.

4.4 NAS Vulnerability Index (NVI)

The NVI is a measure of the counties vulnerability to NAS based on the top three most important measures explained in the previous section. The NVI values are between 0 and 1.0. To interpret the NVI, high values of the NVI mean that the county is at a higher risk of having more NAS cases, whereas a low NVI reflects a lower risk of NAS in a county. Table 2 below shows the top five highest and lowest counties with the highest NVI and their percentile scores for the three impactful features.

Table 2: Top five counties with the highest NVI

County	Smoke Pct	OUD Pct	Dispense Rate Pct	NVI
High NVI				
Venango	28.8	46.88	1.12	0.973
Elk	26.72	64.62	1.64	0.964
Fayette	27.28	31.41	1.73	0.944
Clearfield	22.78	52.8	1.28	0.918
Lawrence	23.58	31.28	1.09	0.899
Low NVI				
Delaware	6.25	12.96	0.34	0.121
Berks	8.05	8.81	0.17	0.081
Lancaster	6.58	9.97	0.17	0.076
Chester	4.32	9.8	0.16	0.037
Montgomery	4.65	8.39	0.18	0.031

Note: Smoke Pct = Percentile Score of percentage of women who smoke during pregnancy, OUD Pct = Percentile Score of OUD rate among 1,000 pregnant women, Dispense Rate Pct = Percentile Score of rate of opioids dispensed to pregnant women

In Figure 5, counties with a higher NVI are shaded in dark blue, while counties with a low NVI are of lighter shade. The map reveals that the southeast region of Pennsylvania has consistently low NVI values, whereas counties in the west have higher NVI values. High values are between 0.8 to 1.0 NVI, and low values are less than or equal to 0.2 NVI.

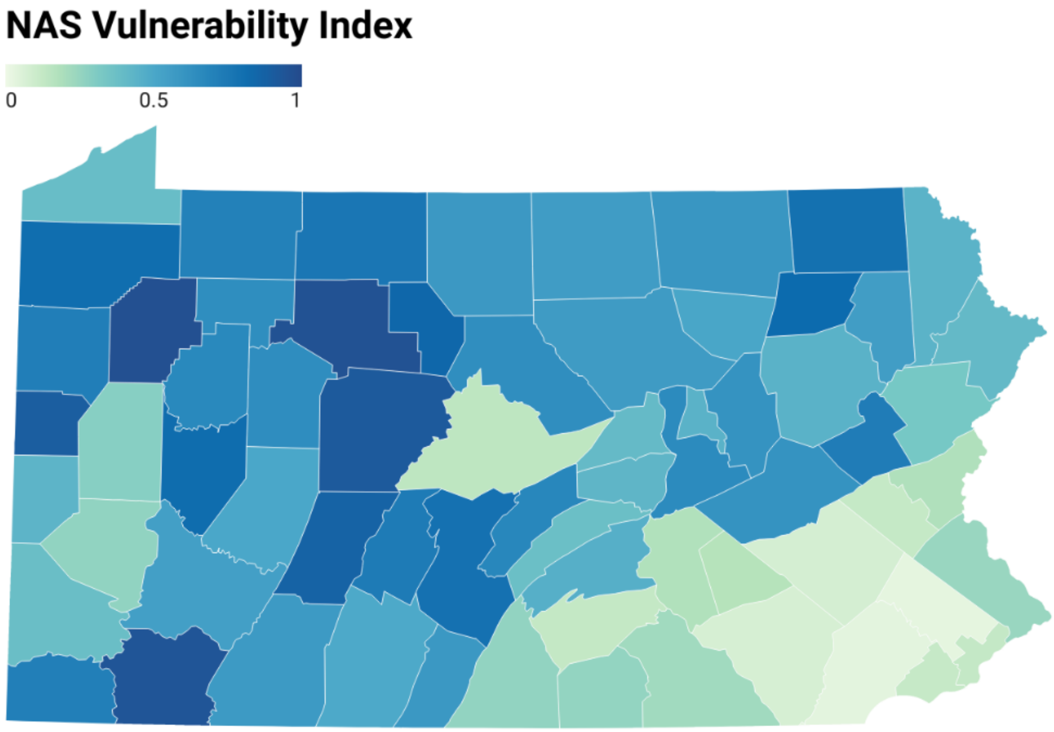

Figure 5: NAS Vulnerability Index for PA Counties

Chapter 5

Discussion

5.1 Validity of Feature Effects

Dependency plots for SHAP values show the value of the feature on the x-axis plotted against the SHAP value for the measure on the y-axis [23]. These plots allow for the SHAP summary plot to be validated by examining the range of SHAP values for all the feature values contained in the dataset [24]. Figure 6 shows the top three most important features used in the NVI. The colors of the points are used to show the value of the feature. For example, in Figure 6.a, low values of percentages of pregnant women who smoked correlates with a blue point, and a high percentage correlates with a red point. For the plots generated below, these colors correspond with the x-axis value. The SHAP values vary with the value of the examined feature.

In the case of these top 3 features shown below, the SHAP values tend to increase with higher values of a feature. In other words, when the value of a given feature is high, it has more of an impact on NAS rates than if it were a low value. Note that the dependency plots all have different x- and y- axes when comparing them. However, examining each feature separately will show how the feature value effects its SHAP value, thus effecting the overall prediction power of the model.

For the sake of clarity Figure 6.a is used as an illustration.. The line of best fit has an increasing slope. As the percentage of women who smoked during pregnancy increased, the SHAP value increased as well. This positive slope shows that higher percentages had an increasingly larger impact on the prediction model, whereas lower value had lesser effect on the model. This shows that high percentages of women who smoked during pregnancy tends to result in higher NAS rates. For this reason, the top three features— percentage of women who smoked, the OUD rate among pregnant women, and the dispensing rate of opioid to pregnant women— are used to create the NVI.

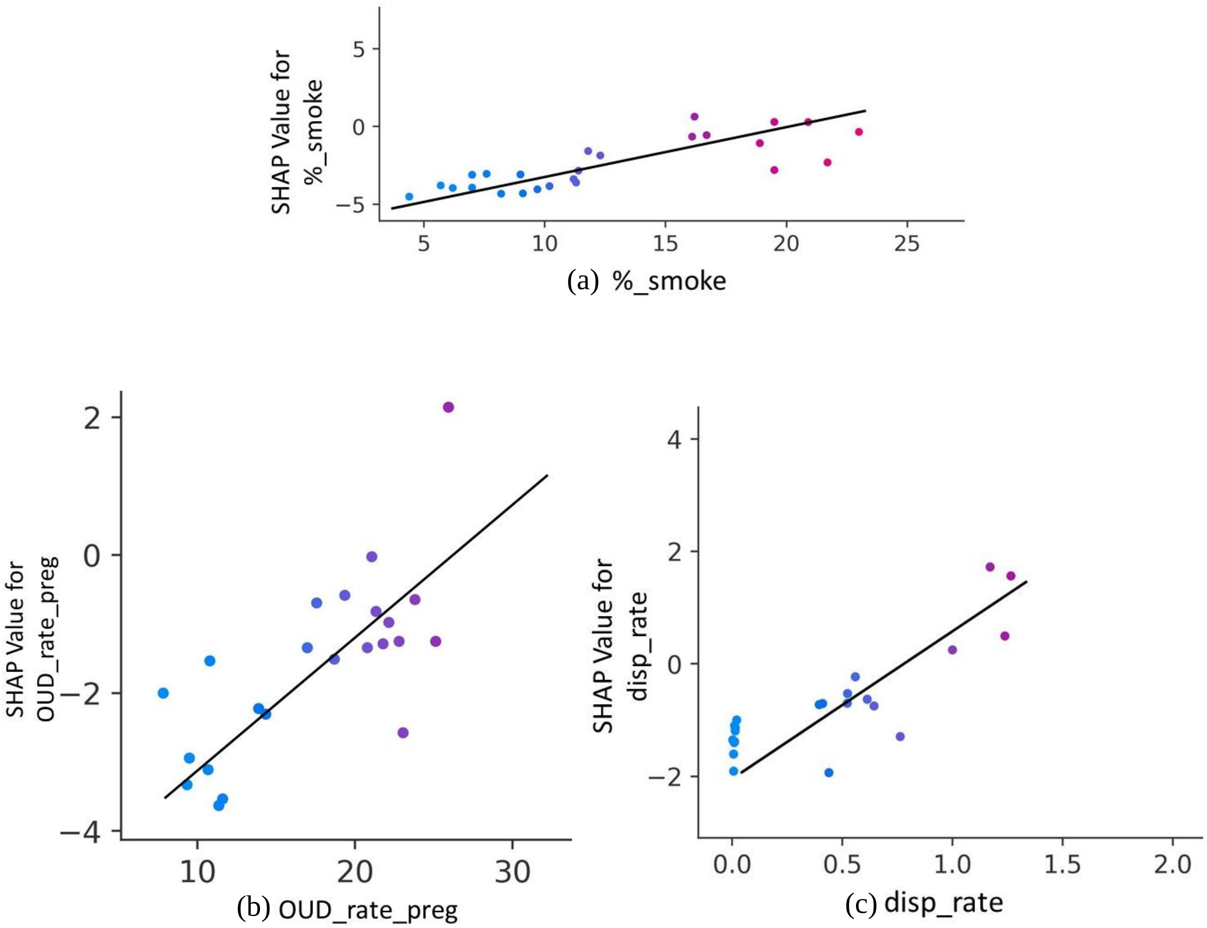

Figure 6: Dependency plots for top three most important county-level features.
(a) Smoked during pregnancy. (b) Diagnosed with an OUD during pregnancy, (c) Opioid dispense rate to pregnant women

5.2 Validity of NVI

The NVI can be validated with a number of different characteristics of a county. This section will examine the rate of maternal stays with opioid use, individuals receiving medication-assisted treatment (MAT), and economic conditions. In order to analyze the data, the NVI has been split into five categories: Lowest Risk (NVI ≤ 0.2), Low Risk (0.2 < NVI ≤ 0.4), Medium Risk (0.4 < NVI ≤ .06), High Risk (0.6 < NVI ≤ 0.8), and Highest Risk (NVI > 0.8). A map with the counties categorized by the risk-level is shown in Figure 7. It must be noted that these ranges are arbitrary, albeit realistic.

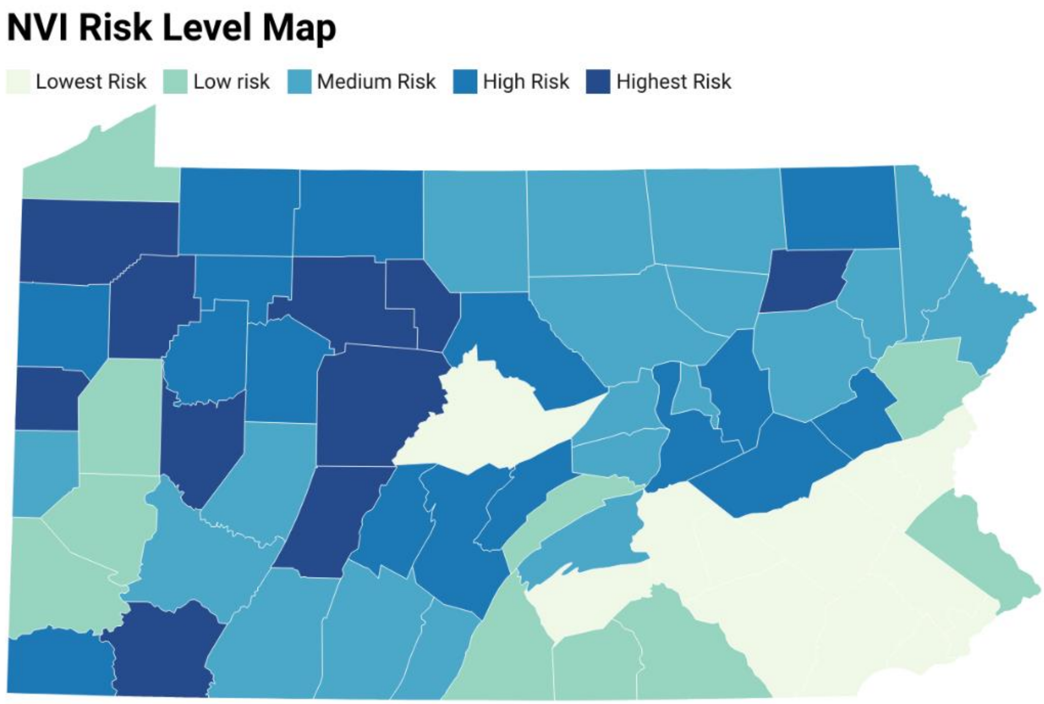

Figure 7: Map showing the counties categorized by risk-level

5.2.1 Validation with the Rate of Maternal Stays with Opioid Use

The rate of maternal stays with opioid use per 1,000 maternal stays can be used to validate the NVI. A maternal stay is defined as any hospital stay involving a delivery or other pregnancy-related stays [25]. Opioid use includes the abuse of opioids, use of prescribed opioids for pain, or medication-assisted treatment for drug addiction. The data published by PHC4 reveal the counties with the highest maternal stays - which are Elk County, Clearfield County, Venango County, Armstrong County, and Cambria County. These counties and their rates and NVI are summarized in Table 3. These five counties with the highest rate of maternal stays with opioid use all have an NVI that is in the range considered to be Highest Risk (0.8-1.0).

Table 3: Top five counties with the highest rate of maternal stays with opioid use

County	Rate of Maternal Stays with Opioid Use	NVI	Risk-Level
Elk	75.8	0.964	Highest
Clearfield	54	0.918	Highest
Venango	51.8	0.973	Highest
Armstrong	51.6	0.819	Highest
Cambria	36.3	0.872	Highest

5.2.2 Validation with the Rate of Individuals on Medical Assistance Receiving MAT

Individuals may receive medication-assisted treatment to treat OUDs. Opioids such as buprenorphine may be prescribed to help relieve symptoms of addiction in a controlled manner while decreasing dependence on opioids [26]. Data published by PHC4 shows that the five counties with the highest rate of individuals on medical assistance receiving MAT in 2019 were Fayette County, Lawrence County, Blair County, Cambria County, and Mifflin County. Higher rates of individuals receiving MAT corresponds with a higher NVI in a county. Table 4 summarizes the counties and their corresponding rates, NVI, and risk-level.

Table 4: Top five counties with highest rate of individuals on MA receiving MAT

County	Individuals on MA receiving MAT per 1000 residents	NVI	Risk-Level
Fayette	16.32	0.944	Highest
Lawrence	15.45	0.899	Highest
Blair	14.28	0.741	High
Cambria	12.52	0.972	Highest
Mifflin	12.22	0.676	High

5.2.3 Validation with economic conditions

Using the data reported from SAIPE in 2019, the counties with the lowest percentage of all ages in poverty are Bucks County, Chester County, Montgomery County, Cumberland County, Adams County. All five of these counties are in the Low or Lowest Risk level for NAS. Table 5 below summarizes the lowest five counties and their percentage of all ages in poverty, NVI, and their risk-level.

Table 5: Bottom 5 counties with lowest percentage of all ages in poverty

County	% of All Ages in Poverty	NVI	Risk-Level
Bucks	5.7	0.246	Low
Chester	5.9	0.037	Lowest
Montgomery	6.0	0.031	Lowest
Cumberland	7.2	0.128	Lowest
Adams	7.6	0.26	Low

A study found that unemployment rates had a high impact on NAS cases across the US, although they were unable to find an explanation for this relationship [27]. The NVI values calculated in this study show similar results. The five counties with the lowest rates of unemployment in 2019 are Chester County, Adams County, Centre County, Cumberland County, and Lancaster County. All five of these counties are in the Lowest or Low risk-levels based on their NVI. Adams County has the second lowest unemployment rate at 3.4%, and although it is in the Low-risk category, it is at the lower limit of the range. The counties and their corresponding rates, NVI, and risk-level are summarized in Table 6.

Table 6: Lowest 5 counties with lowest unemployment rates in 2019

County	Unemployment Rate	NVI	Risk-Level
Chester	3.2	0.037	Lowest
Adams	3.4	0.260	Low
Centre	3.4	0.145	Lowest
Cumberland	3.4	0.128	Lowest
Lancaster	3.5	0.076	Lowest

5.3 Access to Healthcare

Access to healthcare for pregnant women or any individual struggling with substance abuse disorders can vary greatly between counties. For example, rural adults have significantly less access to resources for treatment, prevention, and recovery [28]. MAT, which has shown evidence of reducing illicit drug use and the rate of accidental overdoses, is only allowed to be provided by those with a DEA DATA waiver [29]. However, only 16% of psychiatrists received the waiver across the US, and most were concentrated in urban areas. Only 3% of primary care physicians, who are the largest group of caregivers in rural America, received the waiver [30]. Because of this, out of the 2.5 million people suffering from substance abuse disorders in the US, less than 40% have access to MAT [31].

Figure 8 shows the locations of all hospitals in Pennsylvania along with the shade of the county showing the NVI. Each red dot represents a hospital. Clusters show the number of overlapping hospitals in an area. For example, Erie County, the most northwestern county in Pennsylvania, has 6 hospitals in close proximity to each other. This is indicated by the red circle with the “+6”. The urban areas with lower NVIs such as Philadelphia and Allegheny Counties and the surrounding region have more hospitals than the rural areas. The counties with the highest NVIs- Venango, Elk, Fayette, Clearfield, and Lawrence- all have only one or two hospitals to serve the region.

Programs such as the Free2BMom provided by Geisinger hospitals have been implemented to help mothers or soon-to-be mothers struggling with addiction [32]. The program began in Luzerne County in 2019, and it has since expanded to Northumberland, Montour, and Columbia Counties. The multidisciplinary team provide women with social support, access to MAT, and counseling for women who are pregnant and to mothers for up to two years after birth free of charge. As of January 4, 2021 (most recent data), the program had helped 141 women in the program.

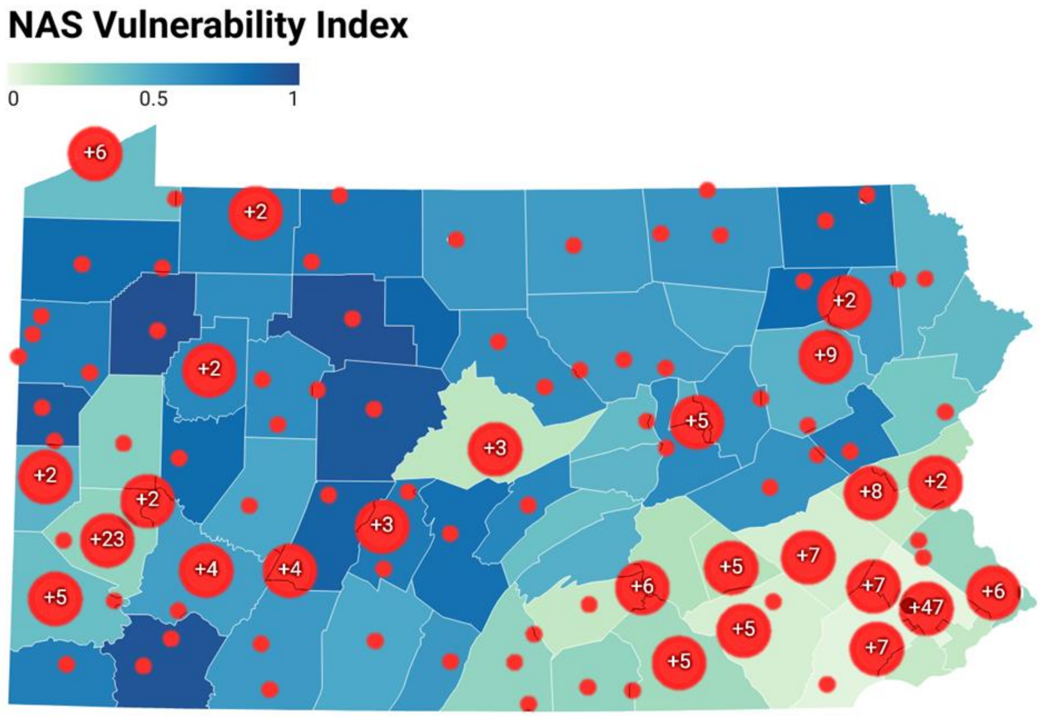

Figure 8: NVI with locations of hospitals

5.4 Study Limitations

The study is limited to only data that is publicly available. Because of this, the data is coming from a number of different sources. Discrepancies in how healthcare data is reported must be considered. The NAS case counts are often based on administrative data, and study had shown that using administrative data had a high positive prediction value for identifying clinically diagnosed cases, but there is always the chance for a wrong or missed diagnosis [33].

The NAS ICD-10 code is P96.1 [34]. The ICD-10 was implemented in 2015, so there is no conflict with this study since it is focused on 2016 to 2019. However, expanding the time range for future studies may result in inaccuracies between counts from ICD-9 codes prior to implementation of ICD-10 codes.

NVI was computed in this thesis using the 3 most contributing features. However computing NVI with 4, 5, and other number of features will be needed to study the sensitivity of NVI with respect to the feature selection.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

NAS is a growing health concern in the US. With NAS rates increasing rapidly every year, it is necessary to implement targeted NAS control and prevention efforts. This thesis examined NAS trends in all the Pennsylvania counties during the 4-year period between 2016-2019 to gain an understanding of the NAS distribution across the state. By collecting data on a variety of factors including demographics, socioeconomic and health related factors, an XGBoost Regressor model was fit and the impact of each feature on the output was observed using SHAP values. Using the top 3 features that influenced the output, a NAS Vulnerability Index (NVI) was developed for Pennsylvania counties. NVI revealed the counties prone to high NAS rates and the results were visualized and documented. It is hoped that gaining insights regarding the factors that affect the rate of NAS within Pennsylvania counties and tools such as the NVI will help government and healthcare officials in devising measure to control the factors causing increased NAS risk, effectively allocating resources and boosting NAS control and prevention efforts.

6.2 Future work

Future work should focus on expanding this method of finding the NVI for counties in states surrounding Pennsylvania. Doing so will give insight into what features affect populations the most in regard to NAS rates and how the risk of NAS is distributed across the mid-Atlantic region of the United States, as well as other neighboring states such as Ohio. Examining the NVI for every county in the USA may also reveal more trends that can be targeted on a country-wide scale. Further examination of the effect of adding more features to the calculation of the NVI should also be examined to understand the sensitivity of the NVI to the features included.

Appendix A

Table 7: NVI of all PA Counties

County	NVI
Adams	0.260
Allegheny	0.267
Armstrong	0.819
Beaver	0.434
Bedford	0.493
Berks	0.081
Blair	0.741
Bradford	0.596
Bucks	0.246
Butler	0.292
Cambria	0.872
Cameron	0.847
Carbon	0.733
Centre	0.145
Chester	0.037
Clarion	0.665
Clearfield	0.918
Clinton	0.640
Columbia	0.615
Crawford	0.811
Cumberland	0.128
Dauphin	0.180
Delaware	0.121
Elk	0.964
Erie	0.390
Fayette	0.944
Forest	0.640
Franklin	0.264
Fulton	0.587
Greene	0.721
Huntingdon	0.793
Indiana	0.504
Jefferson	0.644
Juniata	0.377

County	NVI
Lackawanna	0.558
Lancaster	0.076
Lawrence	0.899
Lebanon	0.168
Lehigh	0.124
Luzerne	0.445
Lycoming	0.567
McKean	0.765
Mercer	0.724
Mifflin	0.676
Monroe	0.332
Montgomery	0.031
Montour	0.448
Northampton	0.183
Northumberland	0.654
Perry	0.461
Philadelphia	0.127
Pike	0.409
Potter	0.583
Schuylkill	0.614
Snyder	0.429
Somerset	0.575
Sullivan	0.509
Susquehanna	0.797
Tioga	0.563
Union	0.403
Venango	0.973
Warren	0.709
Washington	0.384
Wayne	0.442
Westmoreland	0.547
Wyoming	0.828
York	0.224

Table 8: Correlation matrix for all variables in initial analysis

	birth_rate	median_age	%mot_black	%mot_white	%mot_other	%_smoke	%_ins	risk_highMME	treat_fac	OUD_rate_preg	%edu_underHS	%preg_pov	disp_rate	%unemp_preg	prenatal_care	healthy_weig	%preg_alc
birth_rate	1	0.025	0.386	-0.407	0.358	-0.281	-0.229	0.254	0.38	-0.304	0.377	0.357	-0.012	0.103	-0.375	0.241	-0.093
median_age	0.025	1	0.418	-0.52	0.529	-0.703	0.085	-0.162	0.456	-0.369	-0.401	-0.445	-0.019	-0.632	-0.122	0.413	0.141
%mot_black	0.386	0.418	1	-0.908	0.666	-0.436	0.211	0.154	0.832	-0.352	-0.201	0.333	-0.065	-0.136	-0.217	0.082	0.028
%mot_white	-0.407	-0.52	-0.908	1	-0.917	0.579	-0.154	-0.082	-0.774	0.504	0.206	-0.196	0.104	0.261	0.204	-0.251	-0.028
%mot_other	0.358	0.529	0.666	-0.917	1	-0.617	0.073	-0.001	0.586	-0.562	-0.175	0.032	-0.125	-0.335	-0.156	0.369	0.023
%_smoke	-0.281	-0.703	-0.436	0.579	-0.617	1	0.112	0.273	-0.443	0.522	0.026	0.217	0.106	0.596	0.284	-0.38	-0.136
%_ins	-0.229	0.085	0.211	-0.154	0.073	0.112	1	-0.033	0.18	0.047	-0.654	-0.124	0.015	0.044	0.376	-0.133	-0.005
risk_highMME	0.254	-0.162	0.154	-0.082	-0.001	0.273	-0.033	1	0.115	0.069	0.031	0.216	-0.066	0.233	-0.06	0.048	-0.233
treat_fac	0.38	0.456	0.832	-0.774	0.586	-0.443	0.18	0.115	1	-0.322	-0.19	0.328	-0.085	-0.167	-0.069	0.088	0.025
OUD_rate_preg	-0.304	-0.369	-0.352	0.504	-0.562	0.522	0.047	0.069	-0.322	1	0.037	0.025	0.385	0.247	0.186	-0.349	0.099
%edu_underHS	0.377	-0.401	-0.201	0.206	-0.175	0.026	-0.654	0.031	-0.19	0.037	1	0.387	0.056	0.172	-0.498	-0.025	0.02
%preg_pov	0.357	-0.445	0.333	-0.196	0.032	0.217	-0.124	0.216	0.328	0.025	0.387	1	-0.004	0.453	-0.098	-0.213	-0.124
disp_rate	-0.012	-0.019	-0.065	0.104	-0.125	0.106	0.015	-0.066	-0.085	0.385	0.056	-0.004	1	-0.156	0.073	-0.149	0.617
%unemp_preg	0.103	-0.632	-0.136	0.261	-0.335	0.596	0.044	0.233	-0.167	0.247	0.172	0.453	-0.156	1	0.061	-0.355	-0.426
prenatal_care	-0.375	-0.122	-0.217	0.204	-0.156	0.284	0.376	-0.06	-0.069	0.186	-0.498	-0.098	0.073	0.061	1	-0.26	0.024
healthy_weight	0.241	0.413	0.082	-0.251	0.369	-0.38	-0.133	0.048	0.088	-0.349	-0.025	-0.213	-0.149	-0.355	-0.26	1	-0.156
%preg_alc	-0.093	0.141	0.028	-0.028	0.023	-0.136	-0.005	-0.233	0.025	0.099	0.02	-0.124	0.617	-0.426	0.024	-0.156	1

BIBLIOGRAPHY

1. Kocherlakota, P. (2014). Neonatal abstinence syndrome. *Pediatrics*, *134*(2), e547-e561.

2. U.S. National Library of Medicine. (2019, September 9). *Neonatal abstinence syndrome: Medlineplus medical encyclopedia*. MedlinePlus. Retrieved August 5, 2021, from https://medlineplus.gov/ency/article/007313.htm

3. *Neonatal Abstinence Syndrome*. Stanford Children's Health . (n.d.). Retrieved August 5, 2021, from https://www.stanfordchildrens.org/en/topic/default?id=neonatal-abstinence-syndrome-90-P02387

4. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*(1), 1-10.

5. Jones, H. E., Kaltenbach, K., Heil, S. H., Stine, S. M., Coyle, M. G., Arria, A. M., ... & Fischer, G. (2010). Neonatal abstinence syndrome after methadone or buprenorphine exposure. *New England Journal of Medicine*, *363*(24), 2320-2331.

6. Corr, T. E., & Hollenbeak, C. S. (2017). The economic burden of neonatal abstinence syndrome in the United States. *Addiction*, *112*(9), 1590-1599.

7. HCUP Fast Stats. Healthcare Cost and Utilization Project (HCUP). September 2021. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/faststats/nas/nasquery.jsp.

8. PA DOH. (n.d.). *Opioid epidemic*. Department of Health. Retrieved March 4, 2022, from https://www.health.pa.gov/topics/disease/Opioids/pages/opioids.aspx

9. PHC4 (2018, March). *Hospitalizations for Newborns with Neonatal Abstinence Syndrome.* PHC4.org. https://www.phc4.org/reports/researchbriefs/neonatal/17/docs/researchbrief_neonatal2017.pdf

10. Patrick, S. W., Slaughter, J. C., Harrell Jr, F. E., Martin, P. R., Hartmann, K., Dudley, J., ... & Cooper, W. O. (2021). Development and validation of a model to predict neonatal abstinence syndrome. *The Journal of Pediatrics*, *229*, 154-160.

11. Desai, R. J., Huybrechts, K. F., Hernandez-Diaz, S., Mogun, H., Patorno, E., Kaltenbach, K., ... & Bateman, B. T. (2015). Exposure to prescription opioid analgesics in utero and risk of neonatal abstinence syndrome: population based cohort study. *bmj*, *350*.

12. Sawyer, J. L., Shrestha, S., Pustz, J. C., Gottlieb, R., Nichols, D., Van Handel, M., ... & Stopka, T. J. (2021). Characterizing opioid-involved overdose risk in local communities: An opioid overdose vulnerability assessment across Indiana, 2017. *Preventive medicine reports*, *24*, 101538.

13. *Gov. Wolf on passage of bills to fight opioid epidemic*. Governor Tom Wolf. (2016, October 27). Retrieved April 2, 2022, from https://www.governor.pa.gov/newsroom/governor-wolf-statement-on-passage-of-legislation-to-combat-heroin-crisis/

14. Commonwealth of Pennsylvania & The Pennsylvania Pharmacists Association. (2016, January). Opioid Dispensing Guidelines. https://www.health.pa.gov/topics/Documents/Opioids/PA%20Guidelines%20on%20the%20Dispensing%20of%20Opioids.pdf

15. Brownlee, J. (2021, February 17). *A Gentle Introduction to XGBoost for Applied Machine Learning*. Machine Learning Mastery. https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/

16. Sundog Education with Frank Kane (2020, Feb. 11) XGBoost: How it works, with an example [Video]. YouTube. https://www.youtube.com/watch?v=OQKQHNCVf5k

17. Yang, C., Chen, M., & Yuan, Q. (2021). The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accident Analysis & Prevention*, *158*, 106153.

18. Brownlee, J. (2020, August 27). *Feature Importance and Feature Selection With XGBoost in Python*. Machine Learning Mastery. https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/

19. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

20. B., D., & Cook, A. (n.d.). *SHAP Values*. Kaggle. https://www.kaggle.com/code/dansbecker/shap-values/tutorial

21. CDC. (2022, March 15). *CDC/ATSDR Social Vulnerability Index (SVI)*. CDC.Gov. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

22. Kumar, V., Sznajder, K., Kumara, S., Machine Learning based Suicide Prediction and Development of Suicide Vulnerability Index for US Counties, NPJ Mental Health (Accepted for Publication), March 2022. (2022).

23. *shap.dependence_plot — SHAP latest documentation*. (n.d.). SHAP. https://shap-lrjball.readthedocs.io/en/latest/generated/shap.dependence_plot.html

24. Becker, D., & Cook, A. (n.d.). *Advanced Uses of SHAP Values*. Kaggle. https://www.kaggle.com/code/dansbecker/advanced-uses-of-shap-values/tutorial

25. *Maternal Opioid Use Hospital Stays 2016–2017 County Health Care Cost Containment Council (PHC4) | PA Open Data Portal*. (2019, March 14). Open Data PA. https://data.pa.gov/Opioid-Related/Maternal-Opioid-Use-Hospital-Stays-2016-2017-Count/r2n4-n2i4

26. Substance Abuse and Mental Health Services Administration. (22–03-23). *Medication-Assisted Treatment*. SAHMSA.Gov. https://www.samhsa.gov/medication-assisted-treatment

27. National Institute on Drug Abuse (NIDA). (2020, June 4). *Higher rates of NAS linked with economic conditions*. National Institute on Drug Abuse. https://nida.nih.gov/news-events/news-releases/2019/01/higher-rates-of-nas-linked-with-economic-conditions

28. The Rural Health Information Hub. (n.d.). *Substance Use and Misuse in Rural Areas Overview - Rural Health Information Hub*. RHIhub. https://www.ruralhealthinfo.org/topics/substance-use

29. Diversion Control Division. (n.d.). Informational Documents. USDOJ.Org. https://www.deadiversion.usdoj.gov/pubs/docs/index.html

30. Rosenblatt, R. A., Andrilla, C. H. A., Catlin, M., & Larson, E. H. (2015). Geographic and specialty distribution of US physicians trained to treat opioid use disorder. *The Annals of Family Medicine*, *13*(1), 23-26.

31. Volkow, N. D., Frieden, T. R., Hyde, P. S., & Cha, S. S. (2014). Medication-assisted therapies—tackling the opioid-overdose epidemic. *New England Journal of Medicine*, *370*(22), 2063-2066.

32. *Geisinger's Free2BMom program expands to Columbia County*. (2020, December 1). News Articles | Geisinger. https://www.geisinger.org/about-geisinger/news-and-media/news-releases/2020/12/10/12/47/geisingers-free2bmom-program-expands-to-columbia-county

33. Maalouf, F. I., Cooper, W. O., Stratton, S. M., Dudley, J. A., Ko, J., Banerji, A., & Patrick, S. W. (2019). Positive predictive value of administrative data for neonatal abstinence syndrome. *Pediatrics*, *143*(1).

34. *Identifying, Diagnosing and Coding for Intrauterine Substance Exposure and Neonatal Abstinence Syndrome*. (2020, August). Oklahoma Mothers and Newborns Affect by Opioids. https://opqic.org/wp-content/uploads/2019/09/OPQIC-Guidance-on-NAS-and-Intrauterine-Exposure-Coding.pdf

ACADEMIC VITA

EDUCATION

Bachelor of Science in Industrial Engineering – Minor in Spanish
The Pennsylvania State University, University Park, PA Graduation: Spring 2022
Schreyer Honors College

RELEVANT COURSEWORK AND EXPERIENCE

Operations Manager Intern
Target, Suffolk, VA *Summer 2021*

- Learned and performed function of operations manager at Target Upstream Distribution Center in a packing solutions department over the course of 8-week internship program
- Led, motivated, and engaged group of 70 to 85 team members in order to ensure department reached daily goals in production, quality, and safety
- Used root cause analysis and other problem-solving techniques such as the 5-whys to manage issues within department and find opportunities for continuous improvement

Covid-19 Hotspot Analytics
Engineering Analytics Course *Fall 2020*

- Student-led honors project for Engineering Analytics class completed remotely to develop skills in online communication
- Using publicly available data to analyze the vulnerability of every county in the USA concerning COVID-19 hotspots
- Performing linear and logistics regressions using R and Excel to find trends and patterns in data

TECHNICAL SKILLS

- Microsoft Office
- R Programming
- MATLAB Programming
- SolidWorks
- Simio
- CNC Machining
- Technical Writing
- LINGO
- Time Study (MTM & MOST)
- Process Flow Analysis
- Currently enrolled in Six Sigma Green Belt Training

OTHER EXPERIENCE

Shift Lead
Jersey Mike's, State College, PA *July 2020- December 2021*

- Demonstrate strong interpersonal skills through customer service and improve multi-tasking skills in fast-paced environment
- Prioritize classwork while maintaining 20-hour work week

Shift Lead
Duck Donuts, State College, PA *March 2018- March 2020*

- Communicate with other shift leads and manager to assure that all store operations and customer orders are carried out smoothly and trained employees
- Maintained record of monetary transactions both in the registers and the store safe

Study Group Facilitator
Women in Engineering Program at Penn State *August 2019- December 2019*

- Held weekly group tutoring sessions and one-on-one tutoring sessions for Ordinary Differential Equations class
- Prepared material to assist other women in engineering with passing and performing well in class necessary for graduation