# RELATIONSHIP ANALYSIS BETWEEN ORAL HEALTH CONDITIONS AND SIX FACTORS IN THE UNITED STATES

by

**Yanan Tao**

**A Thesis**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Master of Science**

Department of Computer & Information Technology

West Lafayette, Indiana

December 2018

# THE PURDUE UNIVERSITY GRADUATE SCHOOL
# STATEMENT OF COMMITTEE APPROVAL

Dr. John A. Springer, Chair

Department of Computer and Information Technology

Dr. Chad M. Laux

Department of Computer and Information Technology

Dr. Dawn D. Laux

Department of Computer and Information Technology

Dr. Michael D. Kane

Department of Computer and Information Technology

**Approved by:**

Dr. Eric Matson

Head of the Graduate Program

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

Author: Tao, Yanan. MS
Institution: Purdue University
Degree Received: December 2018
Title: Relationship Analysis Between Oral Health Conditions and Six Factors in the United States
Major Professor: John A. Springer

Dental health is an important aspect of one's health and well-being (American Dental Association, 2015). This research analyzes six factors (income level, weather, sales tax, population density, dentist density, and water quality) to examine their relationship with oral health conditions based on 2015 state-level data in the United States. The results show that these factors indeed affect oral health conditions. The analysis results clearly show that income level, dentist density, temperature, and water quality have significant positive effects while temperature has a negative effect effects on oral health at state level.

Furthermore, this study uses a multilinear regression algorithm stepwise method to build three predictive models on different income groups, using the above factors to predict oral health. These models can be a helpful reference for further research in related areas, including but not limited to insurance companies, research institutes that work on improving public oral health, and government agencies.

Key words: Oral health, predictive model

# CHAPTER 1. INTRODUCTION

Oral health is a crucial part of general health and is critical to improving the quality of life. The health problems that poor oral hygiene can induce range from minor issues such as bad breath and tooth decay to much more severe and even life-threatening problems such as heart disease, as reported by DeStefano, Anda, Kahn, Williamson, and Russell (1993), and pancreatic cancer, as reported by Michaud, Joshipura, Giovannucci, and Fuchs (2007). Therefore, it is important to understand what determinants affect the condition of people's oral health in order to better support creation of policies and programs that can make a difference.

## 1.1 Scope

The problem of oral health is affected by multiple factors and is a complicated issue. Everyone can have his or her own oral health problems. For instance, income level is a significant factor in people's oral health (Moeller, Starkel, Quionez, and Vujicic, 2017). While regular dental examinations and good oral hygiene can prevent most dental diseases, people might not be able to visit a dentist due to limited income.

HHS (2000) show that children in low-income families suffer twice as many dental caries as their peers from families with higher incomes (HHS, 2000). Other factors that could potentially influence oral health include the distribution of dentists, population distribution, sales tax in different states, etc.

This study focuses on oral health conditions in the United States in the year 2015. The factors that will be analyzed in the study are as follows:

- Income
- Population density

- Dentist density
- Sales tax
- Average temperature
- Water quality
- Fluoride in drinking water

In chapter 1.6 first paragraph and chapter 2 literature review, we will discuss why choose those factors. In chapter 3.1, we will discuss the data resources.

## 1.2 Significance

Even though some Americans enjoy oral health and keep their natural teeth throughout their lives, oral diseases persist. In the report of National Health Expenditures Highlight from CMS (2015), the total cost for dental services has increased 4.2 percent to $117.5 billion, an acceleration from 2.4 percent growth in 2014. According to a report by HHS’ Surgeon General (2000), a considerable number of people still suffer from dental problems due to various socioeconomic factors. One of the unsolved questions before the public and policymakers is, what are the key factors with the most influence on people’s oral health condition? If this question is answered with confidence using data analysis, policymakers could find themselves in a better situation with regard to policy development.

Evaluation of the relationship between these key factors and oral health conditions becomes a very important question in public health research. Solving this problem will not only benefit government policy improvement, but also help the design of insurance companies’ public health insurance plans.

## 1.3 Research Question

1) What are the key factors with the most influence on people's oral health condition?
2) What is the correlation between oral health and various objective factors including income, dentist density, population density, and weather at the state level?

Due to the significant influence of oral health on quality of life, it is necessary to identify the key determinents factors that affect the condition of people's oral health.

Various objective factors can possibly affect the condition of people's oral health, including income, dentist density (number of dentists per 100,000 people), population density, etc. As environmental and economic conditions have changed rapidly in recent years, so has the data set for oral health. The development of a regression method that utilizes the most recent data sets to identify the top contributing factors to people's oral health on a longitudinal basis for this study. Solving this problem will not only benefit government policy improvement, but also help the design of insurance companies' public health insurance plans.

## 1.4 Assumptions

This study is designed to follow the rules for online surveys to gather data about oral health conditions and is answered by respondents to the best of their knowledge.

The data are randomly collected, and each sub-dataset can represent the oral health condition of all the people in their state.

The income levels of low, medium, and high are determined based on a state level Gini coefficient.

All factors studied in the analysis are independent from each other.

## 1.5 Limitations

Every study has its own limitations. In statistical analysis, the limitation usually involves the data collected and the method implemented by the researcher. This study is restricted by the following constraints.

For every state, the oral health online survey includes 300 people—100 each from low, middle, and high income levels. Population density is not considered in the collection; however, the study provides a weight factor for population density in the model design to take it into consideration.

This study and its results apply only to state level data in the United States, and no other countries.

## 1.6 Delimitations

The delimitations in this study are as follows.

The study focuses on factors including inequality of income, dentist density, population density, sales tax, average temperature, water quality and fluoride in drinking water. Those factors can be collected in open public resource. Other potential factors such as age and gender are not considered in this study because as these data are personal information with restricted access, the researcher cannot obtain them for this study.

All factors under study are considered to be independent from each other, as data are insufficient to study their correlation.

## 1.7 Summary

The main goal of this study is to understand the effect of various factors on people's oral health. This chapter explained the significance of the current research topic and the expected outcome of this study.  In addition, a list of assumptions, limitations, and delimitations were proposed to better describe the scope of the current study.

# CHAPTER 2. REVIEW OF RELEVANT LITERATURE

This chapter reviews related studies on public oral health and their contributing objective factors including income, dentist density, population density, sales tax, weather, water quality, and fluoride in drinking water. The statistical analysis methods employed in previous studies will be discussed as well.

## 2.1 Oral Health

Oral health as described in HHS (2000) considers not only the teeth and gums and their supporting tissues, but also branches of the nervous, immune, and vascular systems related to oral tissues and providing connections to the heart, the brain, and the rest of the body. The growth of biomedical research since World War II has brought advances in the health and well-being of the American people, particularly in the case of oral health. The National Institutes of Health (1988) took on the challenge of addressing oral disease along with many other health problems, as reported by Geiger (2008).

However, challenges remain in the area of oral health problems, such as the disparity and inequality that affect those who are not able to afford medical expenses. The American Dental Association's (ADA) (2015) Health Policy Institute has developed a new comprehensive survey to assess how Americans view their own oral health and has collected a wealth of valuable data (page 4). They provided four level of oral health condition: poor, fair, good and very good. This research pointed out that 39% of low income adults are most likely to report having problems due to the condition of their mouth and teeth while 35% of young adults are most likely to report having problems due to the condition of their mouth and teeth. This data will be analyzed thoroughly in the current study.

## 2.2 Significance of the Relationship Between Income and Oral Health Condition

After collecting data about people's perceptions of their overall oral health in the United States, a study done by Moeller, Starkel, Quionez, and Vujicic (2017) showed that income inequality is a potential factor that affects oral health in both functional and social dimensions, possibly through a psychosocial pathway. Furthermore, distribution of oral health shifts downward when income inequality emerges. Similarly, Paradise (2009) pointed out that poor and near-poor families have more oral disease than other families. Sixty percent of low-income people did not visit dentist within one year, but this number was only 40% in people who are not low-income. Meanwhile, Similarly, Paradise (2009) also found out that dental diseases cause poor children to experience 12 times more restricted activity days than children in higher-income families.

The researchers mentioned in the preceding paragraph indicated links between income level and oral health. Poor people have a greater risk of experiencing dental disease than others who are not in low-income families.

## 2.3 Dentist Density, Population Density, Taxes, Weather, Water

Besides income level, dentist density, taxes, population density, weather, and water are the other five objective factors considered in the model.

### 2.3.1 Dentist Density

Nowak and Casamassimo (2002) revealed that "dentists can provide guidance to children and parents, deliver preventive oral health services, and diagnose and treat dental disease in its earliest stages. This ongoing dental care will help both children and adults maintain optimal oral health throughout their lifetimes" (p. 1). Since 1950, the ADA (2015) has collected data on dentist distribution (number of dentists per 100,000 population) in the United States at the state level and provides this data to support researchers concerned about how dentist density influences oral health conditions at the state level and published this number every year. A report by Munson and Vujicic (2014) showed that oral health conditions have been improving with the increase in dentist density in the United States, and the supply of dentists is projected to increase through 2033.

### 2.3.2 Sales Tax

Tax is another factor investigated by this research. Sturm, Powell, Chriqui, and Chaloupka (2010) found that increasing taxes on sugar-sweetened beverages has improved children's oral health, especially in low-income families. The people's oral health would be better due to the possibility of an eight percent average reduction in the consumption of these beverages based on a ten percent increase in the price caused by the increased tax. This number will higher in low- income family and in children. Furthermore, Powell, Chriqui, Khan, Wada, and Chaloupka (2013) showed that state-level sugar-sweetened soda taxes should consider public health conditions and the potential problems caused by sugar, including oral health risks.

### 2.3.3 Weather

Weather is a factor considered in this oral health study. The impacts of weather events and climate change may potentially increase the prevalence of a wide range of health risks, including the risk of oral disease risk. McGeehin and Mirabelli (2001) studied the weather's effect and observed that weather variations indeed have an effect on oral disease and also increase the effectiveness of local and national policies.

This research used average temperature for the weather condition to analyze the impact of weather on the oral health condition predictive model.

### 2.3.4 Population Density

Population density is commonly used in health-related analysis. Cassel (1971) found that population density in the United States has a significant association with dental caries, which is one of the most common infectious dental diseases. Centers for Disease Control and Prevention (CDC, 2012) reported that about one in five Americans older than age 44 has untreated tooth decay, and more than 27.4 percent of people between ages 20 and 44 years old suffered from this disease between 2011 and 2012. This number has increased every year since 1978, especially in states with high population density, as reported by Amstutz and Rozier (1995).

Therefore, in this study population density will be investigated to decide whether or not this factor is significant with regard to dental health conditions.

### 2.3.5 Water Quality and Fluoride in Water

Water quality has been a considerable factor in dental health research since the last century. In 2003 the World Health Organization (WHO) pointed out that water quality affects human health significantly. Fluoride has a significant association with dental caries reported since 1943 by McClure (1943). His team pointed out that "dental caries are produced using natural fluoride waters, or of waters artificially fluoridated at the optimum level of 1 p.p.m. F." It is believed that adding Fluoride at this level to public water supplies does not lead to unfavorable effects.

Fluoride compounds with other elements represent not only a normal component of teeth and bones but are also present in some species of plants. Since fluoride is the most negatively charged element, it can react with the strongest oxidizing element and binds with other positive elements. The use of fluoride for improving oral health has been reported in scientific article (Medjedovic, Medjedovic, Deljo, and Sukalo,2015). Fluoride has an effect on tooth structure that cannot be ignored. For example, a 2015 article by Medjedovic, Medjedovic, Deljo, and Sukalo entitled “Impact of Fluoride on Dental Health Quality” concluded that “there is a statistically significant difference in the evaluated parameters of oral health of children in the control group compared to the studied parameters of oral health in the experimental group of children at the final dental examination.”

This study includes drinking water quality and fluoride in drinking water as parameters to represent whether water affects the condition of oral health.

## 2.4 Previous Analysis Methods

There are many studies discussing the relationship between oral health conditions and considered factors. Locker (2000) reviewed this topic and summarized that there are four types of studies in this field as follows.

1. *Simple descriptive studies* aim to identify different segments of people and measure one exact factor to study the relationship between that factor and the population subgroups about oral health condition.

2. *Comparative studies* assess whether or not measures of one indicator of status perform such as sugar tax have the same effect as another indicator of status such as household income.

3. *Explanatory studies* aim to cast light on mechanisms that link factors considered to impact the risk of oral disease. For example, according to Gratrix and Holloway (1994), comparing the characteristics of communities to the risk indicators for dental caries, the

limitations of the researches about the reasons increasing the dental caries was explained based on collecting data regularly.

4. Analytic studies introduce a third variable into analyses which examine the association between oral health condition and specific factors. When a previous study has already stated the relationship between one factor and oral health condition (that is in the range of age, sex, family income, education, and employment status), an analytic study would build a statistical model to further analyze the relationship between more than one factor and the target. For instance, based on data from 1977 to 1996, Manski, Moeller, and Maas (2001) used logistic regression to estimate the impact of socioeconomic and demographic variables on use rate and number of annual dental visits.

A similar method was also used to study the relationship between individual and contextual factors and dental care utilization by informal caregivers in the U.S (Wu, Luo, Flint, and Qin, 2015). After the chi-square test, the researchers selected the significant variables and divided the variables into three subsets: Caregiver Characteristics, Individual Caregiver Characteristics, and Contextual Characteristics. The researchers chose a multiple logistic regressions method to examine the relationship between the outcomes for dentist visit (Yes/No) and dental cleaning (Yes/No), and the variables in each part. It was observed by Wu, Luo, Flint, and Qin (2015, page 1) that caregivers with health insurance coverage and higher education are more likely to use dental care. Community characteristics were not directly related with caregivers' use of dental care.

In conclusion, the classic method followed generally included three steps:

1. Determine the relationship between the target object and each factor.
2. Analysis the impact of all factors to make sure that the chosen factors persist in the multivariate statistical model.

3. Choose the regression method to estimate the impact factors and the target object.

## 2.5 Summary

This chapter provided a review of the literature relevant to the oral health problem, related objective factors, and analysis methods.  The next chapter provides the framework and methodology to be used in the current research project.

# CHAPTER 3. FRAMEWORK AND METHODOLOGY

This chapter discusses the hypotheses in detail, which is the major aim of this research, and data collection including income, dentist density, population density, sales tax, weather, water quality, and fluoride in drinking water at the state level.

Furthermore, the statistical analysis methods employed in this research will be discussed, including both the statistics chosen and the regression methods used in the model.

## 3.1 Hypotheses

The aim of this study is to propose a new state-level oral health prediction approach based on the dataset collected in 2015. Our research hypothesis is stated below.

1. The factors under study meet the significant level of correlation with the oral health condition at the state level in the United States.
2. The polynomial regression model can reflect the relationship between oral health condition and the factors considered in this study.

## 3.2 Data Collection

Data were collected from four different sources, as explained in detail in the following subsections. Information about how the data is collected and processed will be clarified.

### 3.2.1 Oral Health Condition with Income

Oral health condition data were collected via an online survey administered by Harris Poll at the request of ADA (2015). The raw data were collected from June 23, 2015 to August 7, 2015. Harris Poll collected 14,962 responses. All the survey results were randomly selected. The research received an oversample of 300 completed surveys for every state. Three income levels were classified in each state: low, middle, and high, with 100 completed surveys for each level.

Income levels were calculated based on the federal poverty level (FPL). The low income level was between 0% and 138% FPL. Middle income was between 139% and 400% FPL, and high income was higher than 400% FPL. This data set sampling design followed the state-level analysis. Harris Poll chose the 2014 federal poverty guidelines.The low-income cutoff was 138% FPL based on Medicaid eligibility expansion guidelines outlined in the Affordable Care Act (ACA) (2010) and Huberfeld, Leonard, and Outterson (2013). The high-income cutoff of 400% FPL was based on health insurance marketplace premium tax credit eligibility as established in the ACA (2010).

Oral health condition was measured by responses describing the condition of oral health. Respondents were given five options: poor, fair, good, very good, or unknown. Ninety-nine responses of 14,926 chose unknown, which were treated as missing.

### 3.2.2 Dentist Density

The ADA Health Policy Institute's Distribution of Dentists (DOD) survey data has been updated yearly since 1978. In this report, the researchers collected the dentist density dataset from ADA. The dataset followed U.S. Census Bureau population counts and counted the total number of dentists per 100,000 population (ADA, 2016). Dentists included in this dataset are the primary occupation, including: part-time and full-time private practice, dental school faculty and staff, public health service, hospital staff dentists in hospitals, graduate students or interns in hospitals, or staff in health organizations that include dental services.

### 3.2.3 Population Density

Population density was calculated as resident population divided by total land area. Population and land area data were taken from the United States Census Bureau estimates for July 1, 2015 for the 50 states, published by US Census Bureau (2016).

### 3.2.4 Sales Tax by State

In this study, the sales tax data was collected by state from Sales Tax Clearinghouse. Sales Tax Clearinghouse published sales tax rates every year and the rates held constant from July 1, 2015 to January 1, 2016, as introduced in Drenkard and Walczak (2016). We collected data from the article published by Drenkard and Walczak (2016). State sales taxes have been stable at the state level in the 12 months since July 1, 2015.

### 3.2.5 Average Temperature

Weather data for 2015 was collected from the National Centers for Environment Information in the National Oceanic and Atmospheric Administration (NOAA National Centers for Environmental Information, 2016). NOAA provides weather data at state level every year. There are many features in the dataset, such as humility, feels-like temperatrue, wind. Since the study in McGeehin and Mirabelli (2001) article used the temperature, we also choose temperature in this study.

### 3.2.6 Surface Water Quality

Surface water quality data for 2015 was collected from the Pelican water system (2016). The dataset divides surface water quality into 4 tiers:

1: limited risk level, 90-100

2: low risk level, NSFWQI 50-90

3: median risk level, NSFWQI 50-70

4: high risk level, NSFWQI <50

The tiers are based on the National Sanitation Foundation Water Quality Index. Nine water quality parameters were selected for the index, including: dissolved oxygen (DO), fecal coliform, pH, biochemical oxygen demand (BOD) (5-day), temperature change (from 1 mile upstream), total phosphate, nitrates, turbidity, and total solids.

### 3.2.7 Fluoride in Drinking Water

In this research, the statistical data were prepared using water system data reported by states to the CDC Water Fluoridation Reporting System (2016). Data is state-level.

## 3.3 Modeling Procedure

Three steps from classic methods will be implemented in the oral health condition modeling. Each step in the study will be introduced here.

### 3.3.1 Variable Selection

As previously stated, after data collection the relationship between each factor (income level, dentist density, population density, sales tax, weather) and the target object of oral health condition will be analyzed. Under the assumption that all the factors are independent, correlation coefficients as explained by Devore (2015) will be evaluated to determine whether or not the relationship is significant.

1. Correlation coefficient:

When two random variables X and Y are not independent, the covariance between them is a very important statistic in variable selection. We use this method to state that the variables are independent in the research. The cutoff is 0.5 based on the article by Rumsey, D. J. (2015) and article Taylor, R. (1990).

2. Collinearity check and VIF:

Collinearity is a common problem in linear regression. If high correlations exist among variables the collinearity will be considered a big problem. There are several situations in which collinearity can be safely ignored. In this research, we consider the variance inflation factor (VIF). In linear regression, the VIF is just $1/(1-R^2)$. The cutoff is 4 based on the article by O'brien, R. M. (2007).

3. Linear regression:

In single linear regression, R square ($R^2$) is used to evaluate the relationship between the exact factor and oral health condition. R-squared is a statistical measure of how close the data are to the fitted regression line. In other words, the definition of R-squared is the percentage of the response variable variation, as described by Frost (2013).R-squared is between 0 and 1, where 0 indicates that the model explains none of the variability of the response data around its mean, while 1 points out that the model explains all the variability of the response data around its mean. In a sum, the R-squared value represents how the model fits the data; the higher the better.Meanwhile, we use p value to test the significance between each variable and oral health condition. The cutoff is 0.05, which is a classic cutoff in health analysis (Riggs, B. L., Wahner, H. W., Dunn, W. L., Mazess, R. B., Offord, K. P., & Melton, L. 3. ,1981).

### 3.3.2 Predictive model Multilinear Regression

Multilinear regression measures the relationship between dependent variables and one or more independent variables via a linear regression function. The residual values are assumed to be normally distributed. This assumption is checked in the chapter 4.2. In the final model, we use stepwise multilinear regression to build the predictive model between oral health condition and the final selected variable. The cutoff for the stepwise method is 0.2, which is also the classic cutoff for research in stepwise regression (Andrews, D. F., & Pregibon, D., 1978).

### 3.4 Summary

This chapter elaborated on the methodology used in this study. It provided a detailed description of data collection, statistical analysis steps including correlation coefficient analysis, simple linear regression aimed to select significant factors based on R square, and multilinear regression as the predictive model. To validate the regression predictive model, cross-validation will be implemented to test the performance of the model.

# CHAPTER 4. PREDICTIVE MODEL DESIGN

This chapter presents the study results in detail, including the statistical analysis for each factor (income, dentist density, population density, sales tax, weather, water quality, and fluoride in drinking water at the state level) involved in this research and the multilinear regression predictive model based on the dataset collected in this research.

Moreover, the potential reasons that might have led to these results will be included in this chapter as well.

## 4.1 Factors' Relationship to Oral Health Condition

Based on the methodology mentioned in the previous chapter (Chapter 3), the aforementioned factors are first tested to determine whether or not they have an effect on oral health condition at the state level. We define the oral health condition score (OHC score) for every state. The original dataset divides oral health into four levels: poor, fair, good, and very good. For example:

Alabama's average oral health condition (sample size: 300) is shown in the following table; 10% of the total are Poor, 21% are Fair, 43% are Good, and 27% are Very Good.

Table 1. Alabama's Average Oral Health Condition

| State | Poor | Fair | Good | Very Good |
|---|---|---|---|---|
| Alabama | 10% | 21% | 43% | 27% |

To calculate the OHC score we provide weight for each level: 0.1 to Poor, 0.2 to Fair, 0.3 to Good, and 0.4 to Very Good. Then the Alabama average OHC score is：

$$0.1 \times 10 + 0.2 \times 21 + 0.3 \times 43 + 0.4 \times 27 = 1 + 4.2 + 12.9 + 10.8 = 28.9$$

In the following discussion, the relationship between oral health condition and factors is based on the OHC score. This score is also at the state level.

### 4.1.1 Oral Health Condition and Income

As mentioned in section 3.2.1, income is defined using three groups: low, middle, and high. The distribution of income and oral health condition scores are plotted based on the dataset collected.

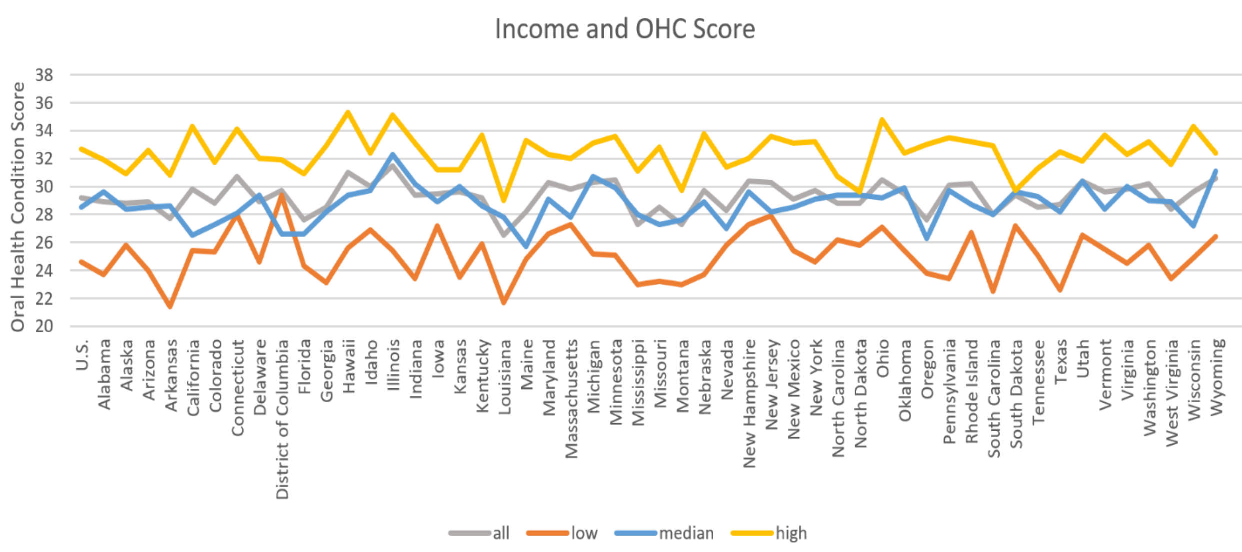


Figure 1. Income and OHC Score

The label "all" (grey) indicates all 300 records for every state. The label "low" indicates 100 low-income records. The label "median" indicates 100 mid-income records. The label "high" indicates 100 high-income records.

This plot shows that income level indeed has a significant effect on oral health condition. High-income level has an OHC score as high as 32.7, while the OHC score is 28.5 for mid-income level and 24.6 for low-income level all over the United States. The average OHC score for all states in the United States is 29.2. Oral health condition for high-income level is also considerably higher than for mid-income level, while oral health condition for mid-income level is higher than for low-income level for every state in the United States.

Therefore, we conclude that income has a large influence on oral health condition. This study will provide three models for three different income levels, and as discussed in the following model sections 4.2.3.

### 4.1.2 Oral Health Condition and Dentist Density

Chapter 2 pointed out that dentist density is a factor that is likely to affect oral health condition. Using the dataset collected from the ADA Health Policy Institute's Distribution of Dentists, we plot it and make a linear regression.

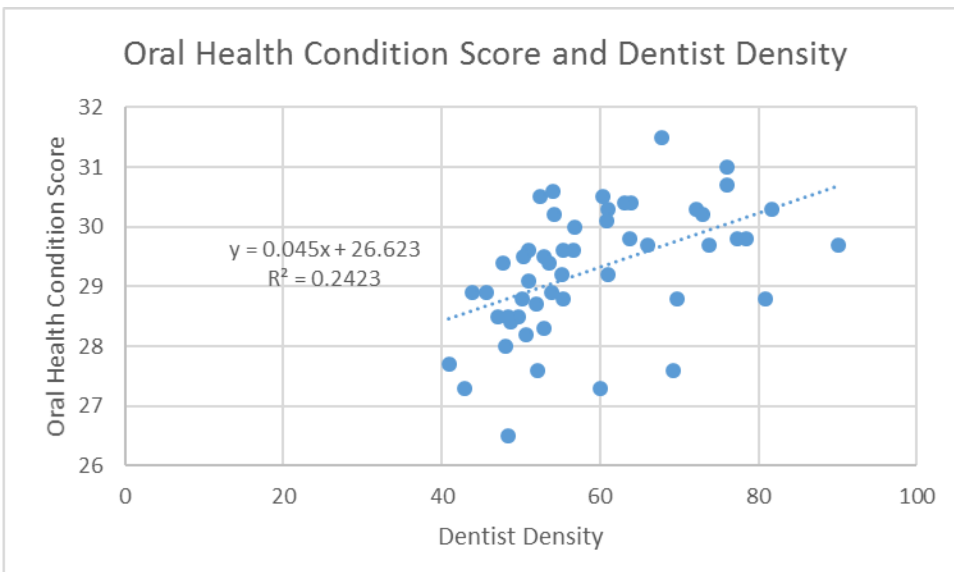


Figure 2. Oral Health Condition Score and Dentist Density

The X-axis represents dentist density counted as the sum of dentists per 100,000 population and the Y-axis represents OHC scores for all data from every state.

The linear regression result shows that when dentist density increases, the OHC score also increases. The R-Square of 0.24 is low but acceptable to represent the linear relationship between the OHC score and dentist density.

In this study, we also calculate the correlation between the OHC score and dentist density. The correlation value is 0.4927, which shows that higher dentist density leads to a higher OHC score. Consequently, dentist density is one of the most important factors to consider with regard to oral health condition.

It is worth mentioning that the population of the United States is 320,896,618 and the total number of dentists is 195,772, so dentist density for the whole United States is 60.99 per 100,000 population. This number has been increasing since 1990.

Based on the above analysis, we can conclude that dentist density affects the OHC score, and is a very important factor to consider in our model.

### 4.1.3 Oral Health Condition and Sales Tax

Sales tax is another factor identified by review of research literature. Using the dataset for sales tax and OHC score, we perform similar analyses to study the relationship between OHC score and sales tax.

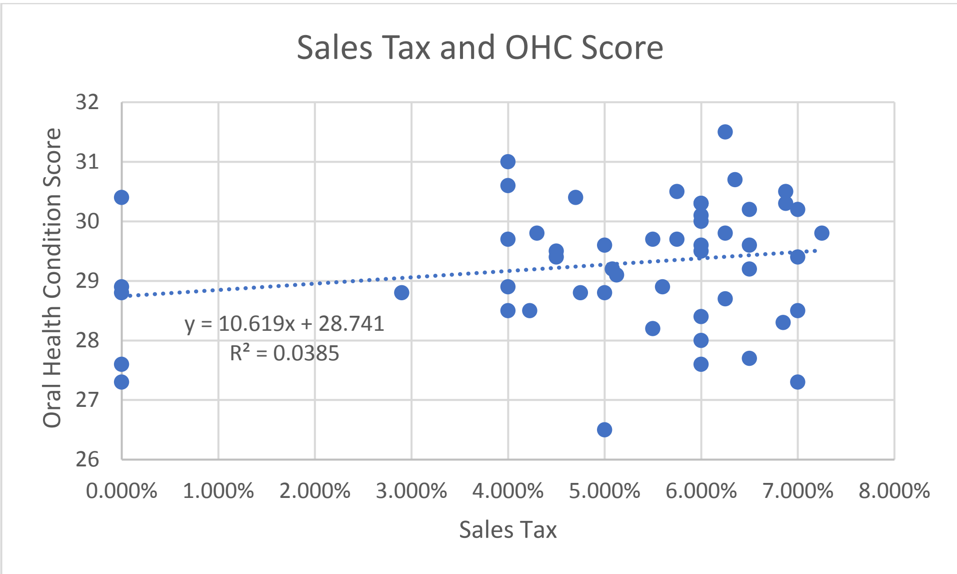


Figure 3. Sales Tax and Oral Health Condition Score

The X-axis represents sales tax and the Y-axis represents OHC score for all oral health condition data at the state level.

This chart does not directly show the effect of the relationship between sales tax and OHC score, but the regression result shows that OHC score increases slightly when sales tax is higher. We suggest that this trend is is a result of higher sales tax leading to decreased consumption of candy, and consequently a higher OHC score.

The correlation between sales tax and OHC score is 0.1963, which suggests a weak statistical relationship but also the need to be included in variable selection for predictive model development.

### 4.1.4 Oral Health Condition and Temperature

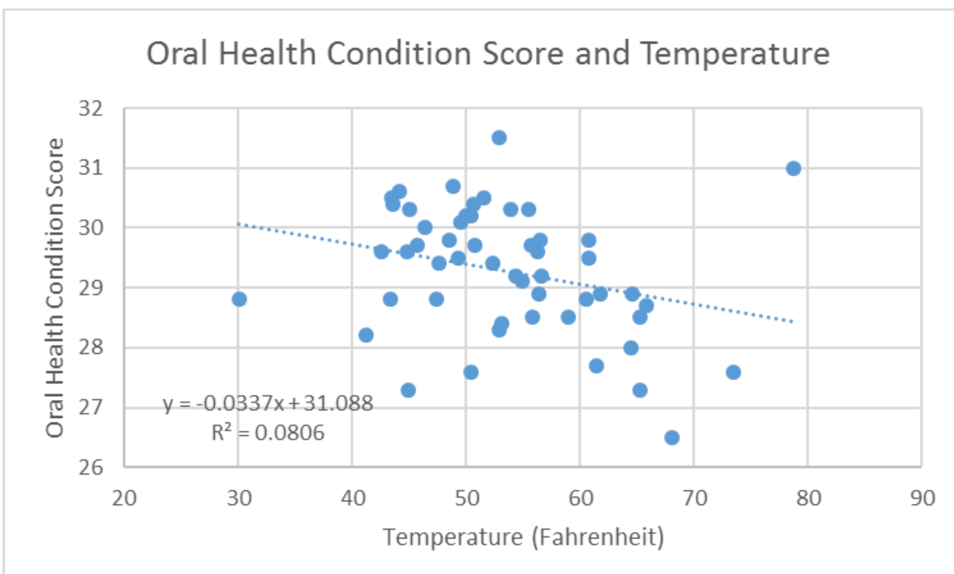


Figure 4. Oral Health Condition Score and Temperature

The X-axis represents temperature in Fahrenheit and the Y-axis represents OHC score for all oral health condition data at the state level.

The correlation for this factor is negative 0.28, which means that OHC score and average temperature have a weak relationship. This factor is tested in the predictive model and presented in greater detail in the next section.

### 4.1.5 Oral Health Condition and Surface Water Quality

Based on the four tiers mentioned in Chapter 3 (1: limited risk, 2: low risk, 3: median risk, and 4: high risk), the average OHC score for every risk level is calculated as follows:

Table 2. Risk Level OHC Scores

| Risk Level | OHC Score |
|---|---|
| 1 | 29.57143 |
| 2 | 29.32222 |
| 3 | 29.05 |
| 4 | 29.25714 |

Risk levels 1 and 2 have higher OHC scores than levels 3 and 4, which indicates that if surface water quality is better, OHC score is generally higher. However, level 4 seems to have a higher OHC score than level 3, possibly because of the limited quantity of data for level 4. There are only 7 data records for level 4, and two of them are California and Washington, which have higher average incomes than other states. If we remove the level 4 data for those two states, the OHC score decreases to as low as 29.01.

The table above and the linear regression are plotted as follows:

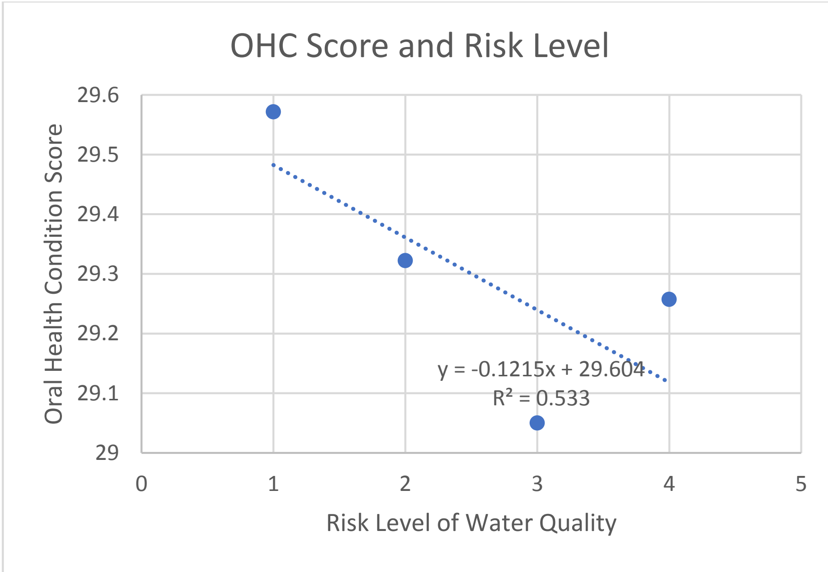


Figure 5. OHC Score and Risk Level

The X-axis represents surface water quality risk level and the Y-axis represents OHC score for all oral health condition data at the state level.

The correlation is negative 0.15, which indicates that OHC score and surface water quality risk level have a weak relationship.

Since R-square is greater than 0.5 and the regression slope is negative, the linear regression shows that water quality is one of the significant factors that affects oral health condition and needs to be considered in the predictive model.

### 4.1.6 Oral Health Condition and Fluoride in Drinking Water

Fluoride in drinking water is the last factor considered for oral health condition in this study. In the research, we collected both percentage of fluoride needed in drinking water and oral health condition score for all states based on datasets at the state level.

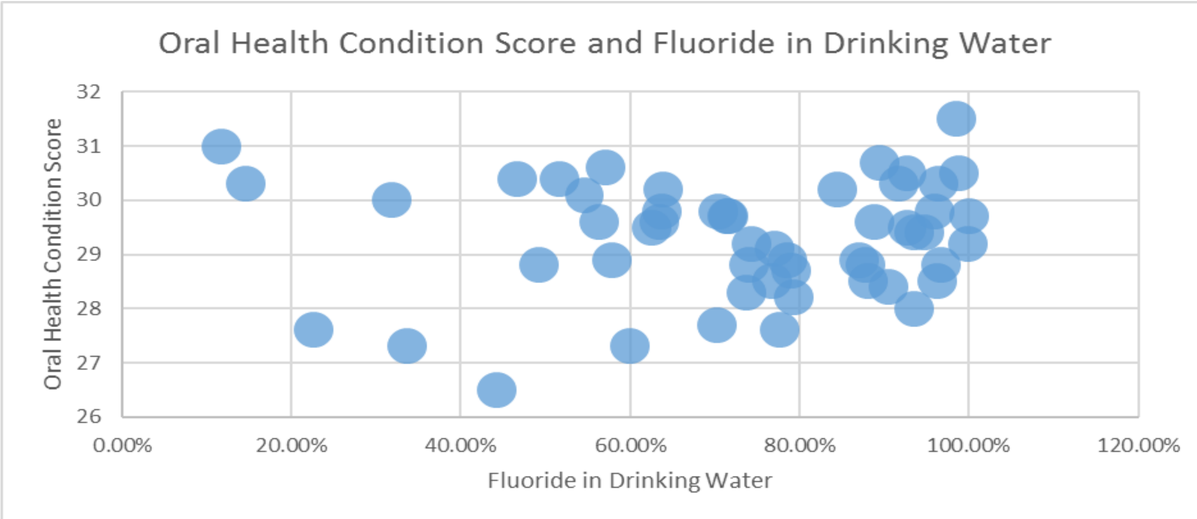


Figure 6. Oral Health Condition Score and Fluoride in Drinking Water

The X-axis represents daily percentage of fluoride required in drinking water and the Y-axis represents OHC score for all oral health condition data at the state level.

In Fig 6, the data is the accurate amount of fluoride in drinking water. The majority of data are in the range of more than 70%. To achieve higher distinction in this factor, rank was used to obtain the following result; figure 7 contains the rank of fluoride content in drinking water in all 50 states:

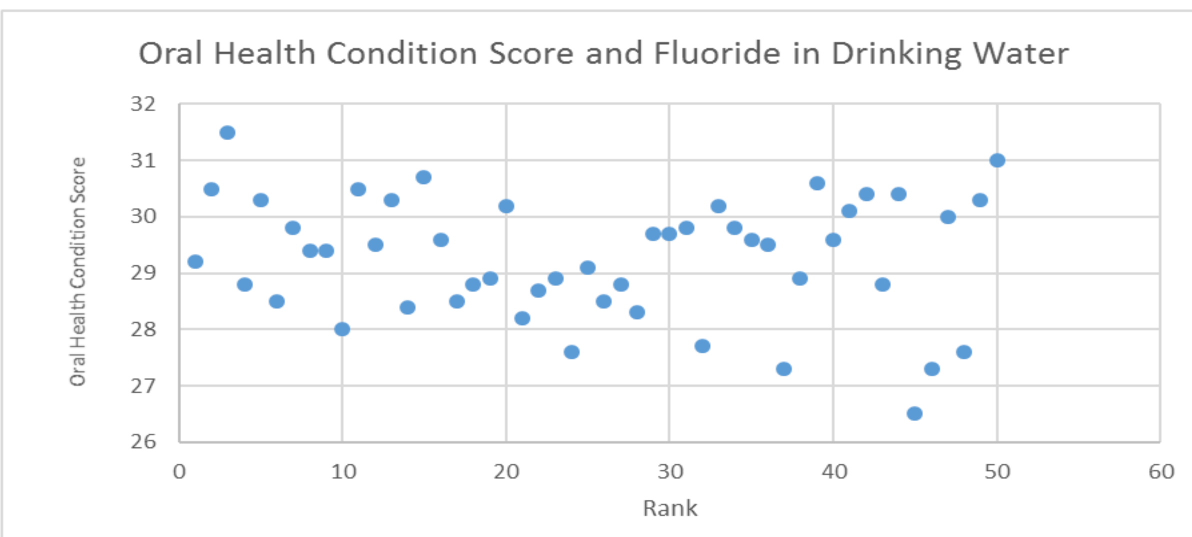


Figure 7. Oral Health Condition Score and Fluoride in Drinking Water

The X-axis represents ranked daily fluoride required in drinking water and the Y-axis represents OHC score for all oral health condition data at the state level.

Based on figure 6 and figure 7, we can see the trend from the two above: when there is more fluoride in drinking water, a higher OHC score is achieved. As a result, the fluoride in drinking water factor will also be considered in the predictive model development.

### 4.1.7 Summary

This section discussed the factors proposed for use in building the predictive model. The results of the statistical analysis demonstrate that income is a significant factor. Therefore, we can either separate the whole dataset into three subsets to discuss in the predictive model building or shift the all-data model with different intercepts. This aspect will be explored further in the next section.

Dentist density and water risk appear to relate more strongly to oral condition score than other factors such as sales tax and temperature. Fluoride in drinking water needs to be tested in the predictive model building. One method is to evaluate it using rank, and another is by using the exact number in the model. This will be discussed further in the following sections.

## 4.2 Oral Health Condition Predictive model

In this section we use the factors in this research to build the oral health condition predictive model. The input dataset uses the following labels:

1) **population_density** means population density which equals population divided by area in miles at the state level.
2) **dental_100000** means dentist density equal to total number of dentists divided by population at the state level.

3) **sales_tax** means sales tax rate at the state level.

4) **water_risk** stands for surface water quality at the state level.

5) **Rank** means the rank of fluoride content in drinking water in all 50 states.

6) **percent** means the value of fluoride content in drinking water in all 50 states.

7) **temperature** means average temperature in the year 2015 at the state level.

8) **OHCS_all** means oral health condition score for all income data at the state level.

9) **OHCS_low** means oral health condition score for low-income data at the state level.

10) **OHCS_med** means oral health condition score for median-income data at the state level.

11) **OHCS_high** means oral health condition score for high-income data at the state level.

Using these factors, we will attempt to build a predictive model that aims to predict oral health condition.

### 4.2.1 Correlation Coefficient Check

In the first step we check the variable relationship by calculating the correlation coefficients among the selected factors. We have collected the following basic statistical information.

Table 3. Correlation Coefficient Check

| Variable | N | Mean | Std Dev | Sum | Minimum | Maximum |
|---|---|---|---|---|---|---|
| **population_density** | 50 | 168.67470 | 207.77862 | 8434 | 1.11223 | 1025 |
| **dental_100000** | 50 | 58.33911 | 10.86221 | 2917 | 40.93553 | 81.73090 |
| **sales_tax** | 50 | 0.05066 | 0.01987 | 2.53300 | 0 | 0.07250 |
| **water_risk** | 50 | 2.40000 | 1.04978 | 120.00000 | 1.00000 | 4.00000 |
| **rank** | 50 | 25.50000 | 14.57738 | 1275 | 1.00000 | 50.00000 |
| **percent** | 50 | 0.72066 | 0.22823 | 36.03300 | 0.11700 | 0.99900 |
| **temperature** | 50 | 53.55400 | 9.05924 | 2678 | 30.10000 | 78.70000 |

Then we calculate the correlation coefficients; one excludes percent and the other excludes rank. The following tables show the analysis results.

Table 4. Correlation Coefficients Excluding Percent

| | population_density | dental_100000 | sales_tax | water_risk | Rank | Temperature |
|---|---|---|---|---|---|---|
| **population_density** | 1.00000 | 0.42253 | 0.22492 | -0.12702 | -0.04320 | 0.06883 |
| **dental_100000** | 0.42253 | 1.00000 | -0.13675 | -0.03580 | 0.27204 | -0.32545 |
| **sales_tax** | 0.22492 | -0.13675 | 1.00000 | 0.07148 | -0.32320 | 0.25139 |
| **water_risk** | -0.12702 | -0.03580 | 0.07148 | 1.00000 | 0.27339 | 0.05133 |
| **Rank** | -0.04320 | 0.27204 | -0.32320 | 0.27339 | 1.00000 | 0.01010 |
| **Temperature** | 0.06883 | -0.32545 | 0.25139 | 0.05133 | 0.01010 | 1.00000 |

*Note*. Exclude percent (fluoride in drinking water)

Table 5. Correlation Coefficients Excluding Rank

| | population_density | dental_100000 | sales_tax | water_risk | Percent | Temperature |
|---|---|---|---|---|---|---|
| **population_density** | 1.00000 | 0.42253 | 0.22492 | -0.12702 | -0.01744 | 0.06883 |
| **dental_100000** | 0.42253 | 1.00000 | -0.13675 | -0.03580 | -0.36421 | -0.32545 |
| **sales_tax** | 0.22492 | -0.13675 | 1.00000 | 0.07148 | 0.33831 | 0.25139 |
| **water_risk** | -0.12702 | -0.03580 | 0.07148 | 1.00000 | -0.23630 | 0.05133 |
| **Percent** | -0.01744 | -0.36421 | 0.33831 | -0.23630 | 1.00000 | -0.04352 |
| **Temperature** | 0.06883 | -0.32545 | 0.25139 | 0.05133 | -0.04352 | 1.00000 |

*Note*. Exclude rank (fluoride in drinking water)

The correlation coefficients are between -0.32545 (dentist density and temperature) and 0.42253 (dentist density and population density). The correlation coefficient for dentist density and population density is 0.42253, which is the only one greater than 0.4. This is because both factors consider population in their calculation. Dentist density equals number of dentists divided by total population at the state level, while population density equals population divided by area in miles. Therefore, these two factors have relatively higher correlation coefficients. Since the cutoff correlation coefficient in this study is 0.5, all factors are included in the predictive model design.

### 4.2.2 Collinearity Check

The second step is the collinearity check. As mentioned in Chapter 3, this study uses variance inflation factor (VIF) to evaluate the degree of collinearity.

Table 6. Collinearity Check

| Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Tolerance | Variance Inflation |
|---|---|---|---|---|---|---|---|
| **Intercept** | 1 | 28.32142 | 3.06854 | 9.23 | <.0001 | . | 0 |
| **population_density** | 1 | 0.00029034 | 0.00073654 | 0.39 | 0.6954 | 0.69081 | 1.44757 |
| **dental_100000** | 1 | 0.04683 | 0.01628 | 2.88 | 0.0063 | 0.51735 | 1.93292 |
| **sales_tax** | 1 | 14.45547 | 7.45912 | 1.94 | 0.0594 | 0.73674 | 1.35732 |
| **water_risk** | 1 | -0.10891 | 0.12986 | -0.84 | 0.4064 | 0.87059 | 1.14865 |
| **Percent** | 1 | -0.70524 | 2.02672 | -0.35 | 0.7296 | 0.07562 | 13.22436 |
| **Rank** | 1 | -0.02004 | 0.03020 | -0.66 | 0.5106 | 0.08349 | 11.97817 |
| **Temperature** | 1 | -0.02391 | 0.01657 | -1.44 | 0.1566 | 0.71777 | 1.39320 |

Since tolerance = 1/VIF, as shown in the analysis result table above, all of the tolerance values in this result except rank and percent are more than 0.25, which indicates that not all factors are considered collinear. After excluding one of the percent and rank, tolerance for all other factors is >0.25.

Table 7. Tolerance Values Excluding Percent

| Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Tolerance | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 27.36938 | 1.37508 | 19.90 | <.0001 | . | 0 |
| population_density | 1 | 0.00029029 | 0.00072897 | 0.40 | 0.6924 | 0.69081 | 1.44757 |
| dental_100000 | 1 | 0.04897 | 0.01492 | 3.28 | 0.0020 | 0.60340 | 1.65728 |
| sales_tax | 1 | 14.02632 | 7.28089 | 1.93 | 0.0607 | 0.75745 | 1.32022 |
| water_risk | 1 | -0.10956 | 0.12851 | -0.85 | 0.3986 | 0.87077 | 1.14841 |
| Rank | 1 | -0.01016 | 0.01019 | -1.00 | 0.3243 | 0.71808 | 1.39261 |
| Temperature | 1 | -0.02222 | 0.01568 | -1.42 | 0.1638 | 0.78504 | 1.27383 |

*Note*. Exclude percent (fluoride in drinking water)

Table 8. Tolerance Values Excluding Rank

| Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Tolerance | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 26.64747 | 1.73592 | 15.35 | <.0001 | . | 0 |
| population_density | 1 | 0.00030493 | 0.00073140 | 0.42 | 0.6788 | 0.69143 | 1.44628 |
| dental_100000 | 1 | 0.04961 | 0.01563 | 3.17 | 0.0028 | 0.55392 | 1.80531 |
| sales_tax | 1 | 14.24325 | 7.40361 | 1.92 | 0.0610 | 0.73810 | 1.35483 |
| water_risk | 1 | -0.11937 | 0.12806 | -0.93 | 0.3564 | 0.88361 | 1.13172 |
| percent | 1 | 0.55913 | 0.68654 | 0.81 | 0.4199 | 0.65041 | 1.53749 |
| Temperature | 1 | -0.02161 | 0.01610 | -1.34 | 0.1867 | 0.75060 | 1.33226 |

*Note*. Exclude rank (fluoride in drinking water)

As a result, the conclusion drawn from the collinear analysis and correlation coefficient check is that all factors are independent and can be used in the predictive model design.

### 4.2.3 Predictive model Design

This study uses multilinear regression to build the model, with a classic stepwise regression method. Since the dataset is limited, cutoff = 0.2 was chosen, as mentioned in chapter 4.

For all income data, we get the following model:

Table 9. Income Data Model

| Variable | Parameter Estimate | Standard Error | F Value | Pr > F |
|---|---|---|---|---|
| **Intercept** | **26.91380** | **1.23146** | **477.65** | **<.0001** |
| **sales_tax** | **13.60550** | **6.87978** | **3.91** | **0.0541** |
| **dental_100000** | **0.05333** | **0.01268** | **17.69** | **0.0001** |
| **Rank** | **-0.01358** | **0.00947** | **2.06** | **0.1586** |
| **Temperature** | **-0.02042** | **0.01517** | **1.81** | **0.1849** |

There are four factors considered in the final all-income predictive model in the analysis. They are sales tax, dentist density, and fluoride in drinking water by rank and temperature.

The significant level for all parameters is within the cutoff of 0.2. Sales tax is 0.0541, dentist density is 0.0001, fluoride in drinking water by rank is 0.1586, and fluoride in drinking water by temperature is 0.1849.

In summary, the prediction function for all-income data is:

OHCS (oral health condition score) = 26.9138 + 0.05333 × dentist density + 13.60550 × sales tax − 0.01358 × fluoride in drinking water by rank − 0.02042 × temperature

The following figure shows the distribution of predicted and real values.

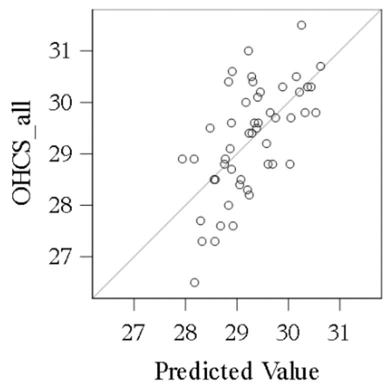


Figure 8. Distribution of
Predicted and Real Values

After getting these results, we also performed more analyses for residual values to evaluate the model and determine whether the residuals have normal distribution. When we check the residual of this model, we get the following two results:

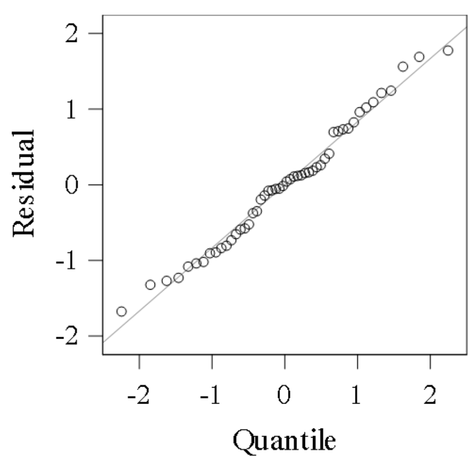


Figure 9. Distribution of Differences
Between Predictive and True Values

The plot above shows the distribution of all differences between the predictive value and the true value. When plotted in the following chart, they are shown to obey normal distribution.

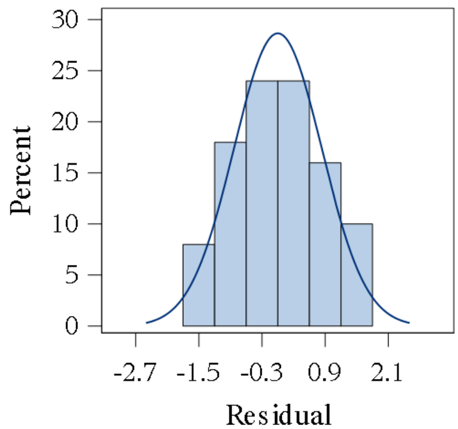


Figure 10. Normal Distribution
of Predictive and True Values

As mentioned before, the study has a second regression step that considers income level to adjust the different income levels at the state level.

1) Low- income group:

   The regression between OHCS_all and OHCS_low is analyzed to obtain the function for the low-income group:

Table 10. Low-Income Group Function

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | 1 | -4.21313 | 4.79601 | -0.88 | 0.3841 |
| **OHCS_all** | OHCS_all | 1 | 0.99854 | 0.16372 | 6.10 | <.0001 |

The function for the low-income group is:

OHCS-low (oral health condition score for low-income) = $-4.21313+0.99854\times$ [26.9138+0.05333× dentist density + 13.60550× sales tax $-$ 0.01358 × fluoride in drinking water by rank $-$ 0.02042× temperature]

After calculation:

OHCS-low (oral health condition score for low-income) = 22.66138 + 0.05325 × dentist density + 13.58564 × sales tax −0.01356 × fluoride in drinking water by rank − 0.02039 × temperature

2) Median-income group:

The regression between OHCS_all and OHCS_med is analyzed to obtain the function for the median-income group:

Table 11. Median-Income Group Function

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | 1 | 7.84872 | 4.01009 | 1.96 | 0.0561 |
| **OHCS_all** | OHCS_all | 1 | 0.71447 | 0.13689 | 5.22 | <.0001 |

The function for the low-income group is:

OHCS-median (oral health condition score for median-income) = 7.84872 +0.71447 × [26.9138+0.05333× dentist density + 13.60550× sales tax − 0.01358 × fluoride in drinking water by rank − 0.02042× temperature]

After calculation:

OHCS-median (oral health condition score for median-income) = 27.0778 + 0.03810 × dentist density + 9.72072 × sales tax – 0.00970 × fluoride in drinking water by rank −0.01459 × temperature

3) High-income group:

The regression between OHCS_all and OHCS_high is analyzed to obtain the function for the high-income group:

Table 12. High-Income Group Function

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | 1 | 7.82177 | 4.32958 | 1.81 | 0.0771 |
| **OHCS_all** | OHCS_all | 1 | 0.84028 | 0.14780 | 5.69 | <.0001 |

The function for the high-income group is:

OHCS_high (oral health condition score for high-income) = 7.82177 +0.84028 × [26.9138 + 0.05333 × dentist density + 13.60550 × sales tax −0.01358 × fluoride in drinking water by rank −0.02042 × temperature]

After calculation:

OHCS_high (oral health condition score for high-income) = 29.42817 + 0.04481 × dentist density + 11.43243 × sales tax −0.01141 × fluoride in drinking water by rank −0.01716 × temperature

## 4.3 Summary

In this chapter, we presented detailed analysis for all the prepared factors and built the oral health condition predictive model based on the analysis results. To evaluate oral health condition, we calculated the Oral Health Condition Score for every record in the whole dataset. We analyzed distribution for every factor, and the relationship between the factor and Oral Health Condition Score.

Before we built the predictive model, we also checked the correlation coefficient and collinearity among the factors. We found that all the factors can be considered as independent factors because they are under the cutoff for correlation coefficient <0.5 and VIF <4 (equally, tolerance >0.25), except fluoride in drinking water by rank and fluoride in drinking water by percentage.  As a result, we used one of them to build the model and decided that using rank gives a better prediction result.

In the predictive model for all income levels we built a model which contains four parameters: sales tax, dentist density, and fluoride in drinking water by rank and temperature based on the analysis result of the multilinear regression stepwise method.

Since income is an individual factor based on the individual's income situation, the study separated the data into three income levels and used a second step regression to consider the effect of income.  In the end we got three predictive models: the low-income group model, the median-income group model, and the high-income group model.  Now we have a predictive model which can be used to predict oral health condition based on the selected factors.

# CHAPTER 5. CONCLUSION AND DISCUSSION

The last chapter discusses the model validation and presents a conclusion from this study. Further topics that can be explored in future studies are identified. The following three sections discuss these topics step by step.

## 5.1 Predictive model Validation and Discussion

As a way of validation, we calculate the 95% confidence limits line for the model validation and the 95% prediction limits line based on the prediction results plus or minus 5% for the three final models: OHCS low-income group, OHCS median-income group and OHCS high-income group.

1) OHCS predictive model for the low-income group:

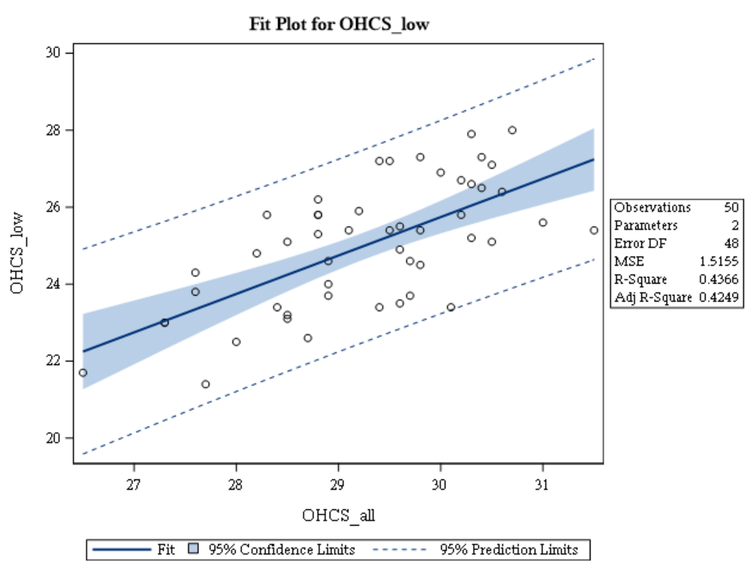


Figure 11. OHCS Predictive model for Low-Income Group

In the low-income group model, we find that although R square is 0.4366 and the MSE (mean square error) is 1.5155, the statistical results show that the parameters are independent and the MSE is not too large to be used in the model. From the chart above we can conclude that all prediction results are included in the 95% prediction limits, which indicates that all the prediction results are within 5% error for the actual results. Although the 95% confidence limits area does not cover all the actual data points, the model can predict Oral Health Condition Score for the low-income group reasonably well.

2) OHCS predictive model for the median-income group:

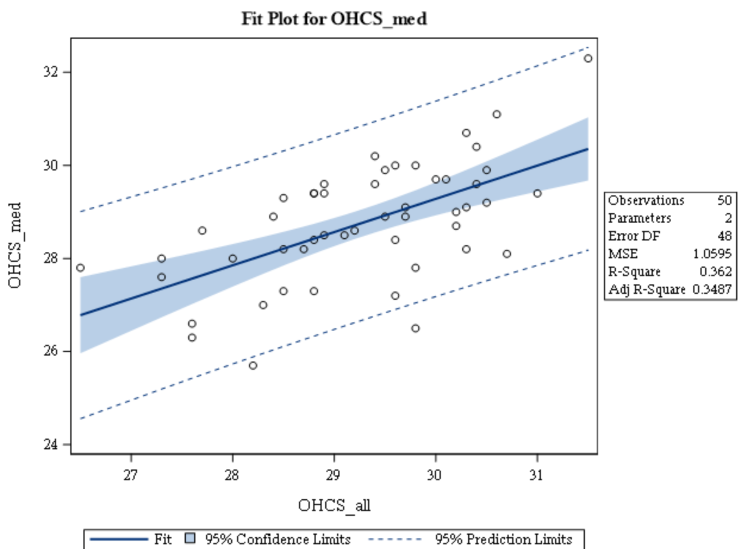


Figure 12. OHCS Predictive model for Median-Income Group

The model for the median-income group OHCS predictive model shows results similar to the low-income group model. However, two points are outside of the 95% prediction limits. As a result, R square (equals 0.362) is lower than R square for the low-income group. Also the data distribution is more spread out than the low-income group distribution. The performance of this model is not as good as the low-income group model performance.

We do want to point out that the MSE in this model (equals 1.0595) is lower than that of the low-income group model. This observation indicates that although two data points are outside of the 95% prediction limits, the majority of data are more concentrated in this group.

3) OHCS predictive model for the high-income group:

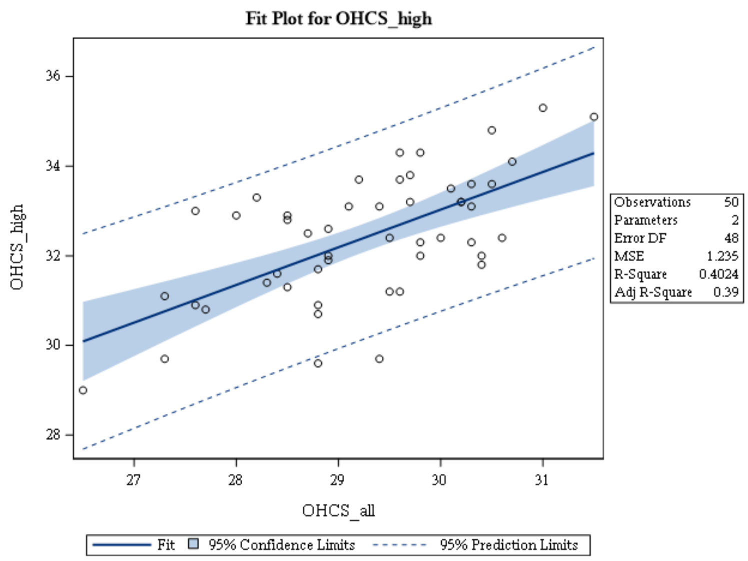


Figure 13. OHCS Predictive model for High-Income Group

In the validation of high-income group, the high-income group predictive model performance is between the low-income and median-income groups. Two of 50 predictions are outside of the 95% prediction limits. R-square and MSE for the high-income group are both between the low-income and median-income groups.

In conclusion, all three models perform fairly well according to the validation. There are two possible reasons that there are still some outliers in the dataset. The first is that the dataset is of limited size. The obsevation in the dataset is only 50, since there are 50 states in the United States. The second is that the model contains data for only one year. To further improve the performance of these models, we need more data from a longer time period. In addition, if we reduce the confidence limits to 90%, the majority of the data points will be included in the confidence limit areas.

## 5.2 Conclusion and Prospect

Based on the model validation and discussion of the model performance above, we can conclude that these three Oral Health Condition Score predictive models can predict OHCS reasonably well based on the proposed factors in this study. From the model equation, we observed that some factors such as income, dentist density, water quality, and sales tax provide a positive effect on oral health condition, while some other factors such as temperature have a negative effect.

This study can be utilized as a helpful reference for research in related areas, including but not limited to insurance companies, research institutes that work on improving oral health, and government agencies responsible for oral health insurance policy.

To further improve the model in future studies, the following steps could be included:

1) Collect more data from more years. If the models include more data, the performance of the predictive models will very likely be improved.
2) Adjust the research level from state level to county level, or even account level. This could give the study better accuracy in analysis and prediction. However, this type of further study would require more resource support and legal permits.
3) Consider more factors. Use of factors such as age, disease history, etc. that could potentially be related to OHCS, model performance is likely to be improved as well.

# REFERENCES


ACT, P. L. (2010). Patient protection and affordable care act. Senate and House of Representatives of the United States of America, 111, 48.

American Dental Association. (2015). *Oral health and well-being in the United States*.

Amstutz, R. D., & Rozier, R. G. (1995). Community risk indicators for dental caries in schoolchildren: An ecologic study. *Community Dentistry and Oral Epidemiology, 23*(3), 129-137.

Andrews, D. F., & Pregibon, D. (1978). Finding the outliers that matter. Journal of the Royal Statistical Society. Series B (Methodological), 85-93.

Cassel, J. (1971). Health consequences of population density and crowding. In National Academy of Sciences Office of the Foreign Secretary Study Committee, *Rapid population growth: consequences and policy implications*. Baltimore: Johns Hopkins Press.

CDC Water Fluoridation Reporting System. (2016). Fluoridation Statistics [Web log post]. Retrieved from https://www.cdc.gov/fluoridation/statistics/2014stats.htm.

Centers for Medicare and Medicaid Services, U. S. (2015). *National health expenditures 2015 highlights*.

DeStefano, F., Anda, R. F., Kahn, H. S., Williamson, D. F., & Russell, C. M. (1993). Dental disease and risk of coronary heart disease and mortality. *BMJ, 306*(6879), 688-691. doi:10.1136/bmj.306.6879.688

Devore, J. L. (2015). Probability and statistics for engineering and the sciences. *Cengage Learning*.

Drenkard, S., & Walczak, J. (2016). *State and local sales tax rates in 2016*. Washington: The Tax Foundation, March.

Frost, J. (2013). Regression analysis: How do I interpret R-squared and assess the goodness-of-fit [Web log post]. The Minitab Blog, 30 .

Geiger, R. L. (2008). *Research and relevant knowledge: American research universities since World War II*. Transaction Publishers.

Gratrix, D., & Holloway, P. (1994). Factors of deprivation associated with dental caries in young children. *Community Dental Health, 11*(2), 66-70.

Huberfeld, N., Leonard, E. W., & Outterson, K. (2013). Plunging into endless difficulties: Medicaid and coercion in national federation of independent business v. Sebelius. BUL Rev., 93, 1.

Locker, D. (2000). Deprivation and oral health: A review. *Community Dentistry and Oral Epidemiology, 28*(3), 161-169. doi:10.1034/j.1600-0528.2000.280301.x

Moeller, J., Starkel, R., Quionez, C., & Vujicic, M. (2017, June). Income inequality in the United States and its potential effect on oral health. *Journal of the American Dental Association, 148*(6), 361-368.

Manski, R. J., Moeller, J. F., & Maas, W. R. (2001). Dental services: An analysis of utilization over 20 years. *The Journal of the American Dental Association, 132*(5), 655-664.

McGeehin, M. A., & Mirabelli, M. (2001). The potential impacts of climate variability and change on temperature-related morbidity and mortality in the United States. *Environmental Health Perspectives, 109* (Suppl 2), 185.

McClure, F. J. (1943). Ingestion of fluoride and dental caries: quantitative relations based on food and water requirements of children one to twelve years old. *American Journal of Diseases of Children*, 66(4), 362-369.

Medjedovic, E., Medjedovic, S., Deljo, D., & Sukalo, A. (2015). Impact of fluoride on dental health quality. *Materia Socio-Medica, 27*(6), 395.

Michaud, D. S., Joshipura, K., Giovannucci, E., & Fuchs, C. S. (2007). A prospective study of periodontal disease and pancreatic cancer in US male health professionals. *JNCI: Journal of the National Cancer Institute, 99*(2),171. doi:10.1093/jnci/djk021

Munson, B., & Vujicic, M. (2014, October). *Supply of dentists in the United States is likely to grow* (Research brief). Health Policy Institute, American Dental Association.

NOAA National Centers for Environmental Information. (2016). State of the Climate: National Climate Report for Annual 2015 [Web log post]. Retrieved from https://www.ncdc.noaa.gov/sotc/national/201513.

Nowak, A. J., & Casamassimo, P. S. (2002). The dental home: A primary care oral health concept. *The Journal of the American Dental Association, 133*(1), 93-98. doi:https://doi.org/10.14219/jada.archive.2002.0027

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673-690.

Paradise, J. (2009). *Oral health coverage and care for low-income children: The role of Medicaid and CHIP*. Henry J. Kaiser Family Foundation.

Pelican Water System. (2016). 3 Major Water Issues for Each State[Web log post]. Retrieved from https://www.pelicanwater.com/water-quality-issues-by-state.php.

Powell, L. M., Chriqui, J. F., Khan, T., Wada, R., & Chaloupka, F. J. (2013). Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: A systematic review of prices, demand and body weight outcomes. *Obesity Reviews, 14*(2), 110-128. doi:10.1111/obr.12002

Riggs, B. L., Wahner, H. W., Dunn, W. L., Mazess, R. B., Offord, K. P., & Melton, L. 3. (1981). Differential changes in bone mineral density of the appendicular and axial skeleton with aging: relationship to spinal osteoporosis. *The Journal of Clinical Investigation*, 67(2), 328-335.

Rumsey, D. J. (2015). U Can: statistics for dummies. John Wiley & Sons.

Sturm, R., Powell, L. M., Chriqui, J. F., & Chaloupka, F. J. (2010). Soda taxes, soft drink consumption, and children's body mass index. *Health Affairs, 10*, 1377.

Suominen, A. L., Helminen, S., Lahti, S., Vehkalahti, M. M., Knuuttila, M., Varsio, S., & Nordblad, A. (2017). Use of oral health care services in Finnish adults: Results from the cross-sectional health 2000 and 2011 surveys. *BMC Oral Health, 17*(1), 78. doi:10.1186/s12903-017-0364-7

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography*, 6(1), 35-39.

US Census Bureau. (2016). Annual estimates of the resident populations: April 1, 2010 to July 1, 2015.

U.S. Department of Health and Human Services. (2000). *Oral health in America: A report of the surgeon general* (Executive summary). Rockville, MD: National Institute of Dental and Craniofacial Research, National Institutes of Health.

Wu, B., Luo, H., Flint, E., & Qin, L. (2015). Dental care utilization among caregivers who care for older adults. *Research on Aging*, *37*(4), 388-412.