Correlation between tobacco use, dental health, and factors of chronic health in US residents

Eliza Chai

Abstract

Aims: Tobacco use is a leading cause of preventable disease, disability, and death in the United States. Studies on the association between tobacco use, chronic disease risk, and the oral health risk of adults could be beneficial to learning about the trend of health objectives. The objective of this study was to examine tobacco usage as a potential predictor of chronic health and oral health of the US residents and potentially study the association of chronic disease conditions, oral health, and tobacco usage in gender subgroups and race/ethnicity subgroups.

Methods: This cross-sectional study analyzed population-based data from the 2020 US Behavioral Risk Factor Surveillance System (BRFSS) available on the CDC website. A sample of 282,071 adults was examined in the study. Multivariable logistic regression analyses were conducted to determine the association between tobacco use and health exposure variables (chronic health and oral health) and interactions between tobacco use, gender subgroup, and race/ethnicity subgroups, controlling for age, gender, race/ ethnicity, income level, health insurance status, BMI, education level, and other health diagnoses.

Results: Tobacco use or smoking status was significantly associated with oral health status. The odds of having good oral health status are estimated to be 42% lower for smokers compared to non-smokers, adjusting for all the other variables (95% CI: 20.7 - 57.5% lower). There is no evidence for interaction between smoking status and gender (p=0.211), but there is evidence for interaction between smoking status and race/ ethnicity group (p=0.0478). However, there is no evidence that tobacco use or smoking status was significantly associated with chronic health status. The odds of having good chronic health status are estimated to be 28.3% lower for smokers compared to non-smokers, adjusting for all the other variables (95% CI: 49.1% lower - 1.1% higher). There is no evidence for interaction between smoking status and race/ ethnicity group (p=0.702), but there is evidence for interaction between smoking status and gender (p=0.0000138).

Conclusions: Tobacco use is correlated with worsened oral health status among US residents. However, tobacco use is not correlated with chronic health status with the selected health variables and indicators used in the study.

Introduction

Tobacco use is a leading cause of preventable disease, disability, and death in the United States[1]. To further understand the association between tobacco use, chronic disease risk, and the oral health risk of adults, we would be using behavioral health risk data to study the pattern, interactions, and associations of the health objectives. In this study, we examine tobacco usage as a potential predictor of chronic health and oral health of US residents and potentially study the association of chronic disease conditions, oral health, and tobacco usage in gender subgroups and race/ethnicity subgroups.

The scientific questions of interest include:

- Is there an association between current tobacco use and oral health among residents in the United State in 2020? Does the association differ between gender? Does it differ between race/ ethnicity?
- Is there an association between current tobacco use and indicators of chronic heart diseases among residents in the United State in 2020? Does the association differ between gender? Does it differ between race/ ethnicity?

Data Set Description

The data set used for this study is the Behavioral Risk Factor Surveillance System (BRFSS)[2], which is the nation's premier system of health-related telephone surveys in all 50 states as well as the District of Columbia and three U.S. territories. Specifically, the BRFSS collects state data about U.S. residents regarding their health-related risk behaviors and events, chronic health conditions, use of preventive services, emerging health issues (vaccine shortage and influenza-like illness), and others. The data are collected using Random Digit Dialing (RDD) techniques on both landlines and cell phones by in-house interviewers or contracts with telephone call centers or universities throughout the year. The survey utilizes a standardized core questionnaire, optional modules, and state-added questions that allow analysis on population-based data cross-sectionally and longitudinally.

In this study, the 2020 BRFSS data from the CDC website is used for the cross-sectional analysis. A sample of 282,071 adults was examined in the dataset. Details regarding the variables' names and CDC BRFSS Codebook[3] are in the appendices.

Response Variables:
Oral health status: frequent dental visits (within the past year) and teeth extraction due to gum disease (< 5 teeth)
Chronic health status: heart disease, stroke, asthma, cancer, COPE, kidney, diabetes
Predictor Variable: tobacco use/ smoking status
Variables adjusted:

- Age: dichotomize, > 65 years old

[1] "Data and Statistics | Smoking and Tobacco Use | CDC." *Centers for Disease Control and Prevention*
[2] "CDC - About BRFSS." *Centers for Disease Control and Prevention*
[3] "Codebook LLCP 2020." *Centers for Disease Control and Prevention*

- Gender: Male and female
- Race/ ethnicity: categorize, Asian, Non-Hispanic; Black, Non-Hispanic; Hispanic; American Indian/ Alaskan Native Non-Hispanic; White, Non-Hispanic; Other race, Non-Hispanic
- Income level: dictomize, > $50000 per year
- Health insurance status
- BMI: dichotomize, overweight/ obesse = 1 (ref)
- Education level: categorize, little/none education, some education, advanced education

Statistical Methods

Descriptive statistics in terms of the weighted percentage of each variable of interest are presented (see **Figure 4** in the Table & Figures section). The figure includes oral health variables (teeth extraction, dental visits), tobacco use/ smoking status, and chronic health variables (heart disease, stroke, asthma, cancer, COPE, kidney, and diabetes). The statistical analysis adjusting for variables such as age, gender, race/ ethnicity, income level, health insurance status, BMI, and education level help determine whether assumptions are supported by the BRFSS data and provide further information on possible confounding by these factors. The analysis will examine the effect modification of the following (1) gender as an effect modifier on the association between smoking status and oral health status, (2) race/ ethnicity as an effect modifier on the association between smoking status and oral health status, (3) gender as an effect modifier on the association between smoking status and chronic health status, and (4) race/ ethnicity as an effect modifier on the association between smoking status and chronic health status.

Histograms of the number of observations for predictor variables and response variables stratified by gender and race/ethnicity are provided for data visualization (see **Figures 1 - 3**). A histogram of the adjusted odds ratio will be presented to help visualize the association among response variables (oral health), controlling variables (age, gender, race/ ethnicity, income level, health insurance status, BMI, education level, heart disease, asthma, cancer, chronic pulmonary disease, arthritis, kidney disease, diabetes), and predictor variables (smoking status) (**Figure 7**). Further, a histogram of the adjusted odds ratio will be presented to help visualize the association among response variables (chronic health), controlling variables (age, gender, race/ ethnicity, income level, health insurance status, BMI, education level, heart attack, arthritis), and predictor variables (smoking status) (**Figure 8**). THe results of exploratory missing data analysis will be shown by histograms of number and percentages of missing values for each variable included in the study (**Figure 5, 6**).

The oral health-related variables are categorical, current tobacco use, and all chronic health variables are dichotomous (binary). Since the question of interest for this study is assessing the association between (1) tobacco use and oral health variables, and (2) tobacco use and chronic health indicators, adjusting for other covariates. To accomplish this goal, two fitted models with multivariate logistic regressions using robust variance estimators will be utilized. The two regression models will include covariates: age, gender, race/ ethnicity, income level, health insurance status, BMI, and education level. These variables are all viewed as potential confounders or effect modifiers so will be accounted for in analyses. The

frequencies and weighted percentages of each response will be examined. The weights generated by the CDC will be used to correct for the different probability of sample selection and account for design effects (R *survey* package[4]). The adjusted odds ratio, their 95% CI, and p-values will be reported in the result (**Table 1- 4**). Instead of using ANOVA method for testing the interactions terms, we will use the function *regTermTest*[5] in the survey package to test sets of regression terms, since the models are not fitted by maximum likelihood with *svyglm*[6].

Model 1: To examine the association between current tobacco use and oral health among residents in the United States in 2020
The model used will be multivariate logistic regression, including the covariates stated below and the estimated odds ratio as the outcome. Wald tests (with *regTermTest*) will be performed to test for the overall significance of oral health variables adjusting for demographics (at a 0.05 significance level). From the testing results, we can determine if the categories of variables are significant after accounting for the other baseline variables.

Model 2: To examine the association between current tobacco use and chronic health indicators among residents in the United States in 2020
The model used will be multivariate logistic regression, including the covariates stated below and the estimated adjusted odds ratio as the outcome. Wald tests (with *regTermTest*) will be performed to test for the overall significance of chronic health variables adjusting for demographics (at a 0.05 significance level). From the testing results, we can determine if the categories of variables are significant after accounting for the other baseline variables.

Results

In **Figures 1 - 4**, we see that sex/gender is fairly balanced in terms of weighted proportion in the dataset. However, there is an unbalanced proportion of different races and ethnicities, which could induce bias in our estimates and inference. In **Figures 5 - 6**, we see that the annual income variable has the highest number and percentage of missing values. The BMI (overweight/obese) status has the second highest number and percentage of missing values. These missing data patterns are expectable since the BRFSS dataset is survey data and personalized information on income and high BMI might be sensitive for participants to share voluntarily. Overall, the majority of the variables are present for analysis and fitting models of interest. Thus, we use the available data approach to handle the missing data in the data set.

Tobacco use or smoking status was significantly associated with oral health status.
The odds of having good oral health status are estimated to be 42% lower for smokers compared to non-smokers, adjusting for all the other variables (95% CI: 20.7 - 57.5% lower). We have sufficient evidence that the odds ratio is different from 1 at the 0.05 level (p=0.000634). In addition, there is no evidence for interaction between smoking status and gender (p=0.211), but there is evidence for interaction between smoking status and race/ethnicity group (p=0.0478).

[4] Lumley, Thomas. *survey: Analysis of Complex Survey Samples*
[5] "R: Wald test for a term in a regression model"
[6] "R: Survey-weighted generalized linear models."

When looking at the relationship between tobacco use and chronic health status, we find no evidence that tobacco use or smoking status was significantly associated with chronic health status. The odds of having good chronic health status are estimated to be 28.3% lower for smokers compared to non-smokers, adjusting for all the other variables (95% CI: 49.1% lower - 1.1% higher). There is no evidence for interaction between smoking status and race/ethnicity group (p=0.702), but interestingly there is evidence for interaction between smoking status and gender (p=0.0000138).

Discussion

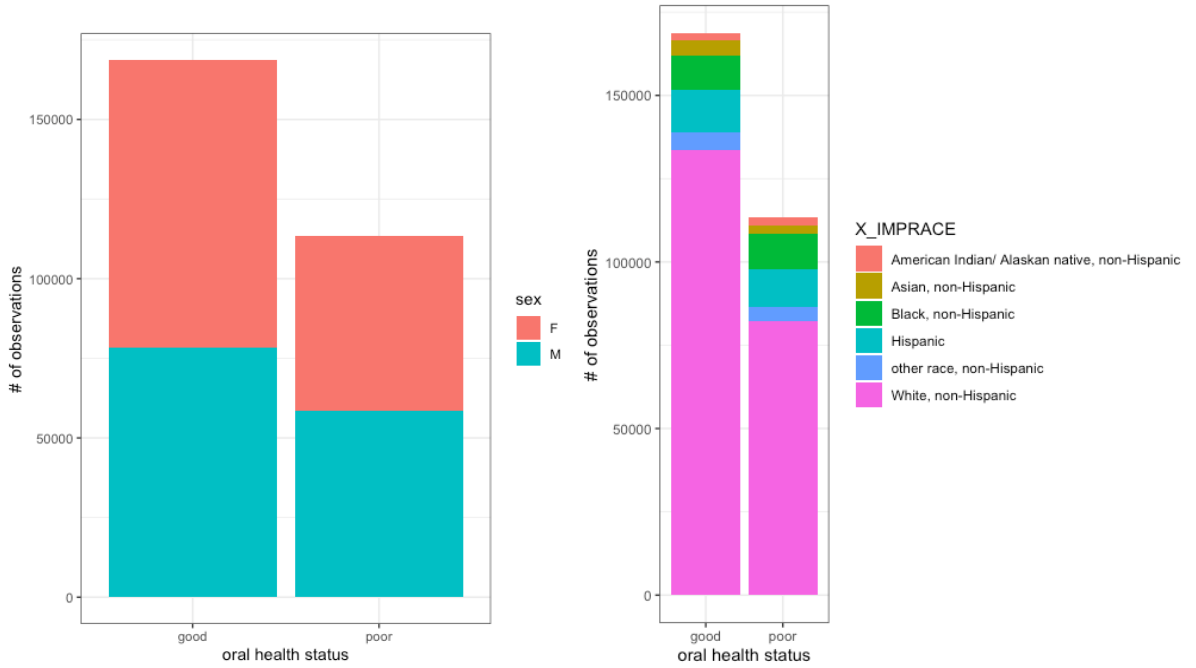
Tobacco use is correlated with worsened oral health status among US residents. However, tobacco use is not correlated with chronic health status with the selected health variables and indicators used in the study. From our analysis, it is surprising to see evidence for effect modification of sex/gender for the association between smoking status and chronic health. Potentially, there are variables to consider in the model to investigate the effect modification of sex/gender on smoking status and chronic health. In addition, we see effect modification of race/ethnicity for the association between smoking status and oral health. Since we have disproportional groups of race/ethnicity in the dataset, the inferential results might be affected and require further care.

For future studies and analyses, we could look at cross-sectional data in different years of the BRFSS dataset and investigate the trend over time in longitudinal studies. It would also be intriguing to study the effect modification of other variables, including regions, income status, education level, and health coverage status using the BRFSS dataset to help plan health programs relevant to smoking prevention.

References

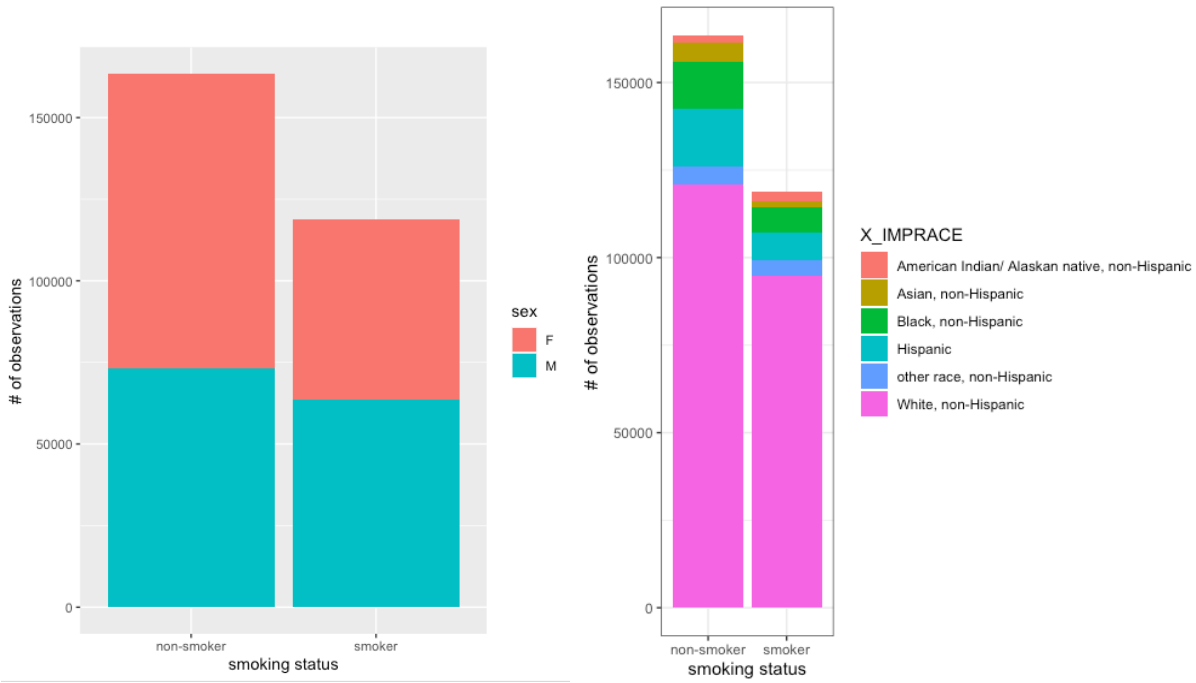
1. “Data and Statistics | Smoking and Tobacco Use | CDC.” Centers for Disease Control and Prevention, https://www.cdc.gov/tobacco/data_statistics/index.htm. Accessed 8 June 2022.
2. “CDC - About BRFSS.” Centers for Disease Control and Prevention, https://www.cdc.gov/brfss/about/index.htm. Accessed 8 June 2022.
3. “Codebook LLCP 2020.” Centers for Disease Control and Prevention, 6 August 2021, https://www.cdc.gov/brfss/annual_data/2020/pdf/codebook20_llcp-v2-508.pdf. Accessed 8 June 2022.
4. Lumley, Thomas. survey: Analysis of Complex Survey Samples, 19 July 2021, https://cran.r-project.org/web/packages/survey/survey.pdf. Accessed 8 June 2022.
5. “R: Wald test for a term in a regression model.” r-survey, https://r-survey.r-forge.r-project.org/survey/html/regTermTest.html. Accessed 8 June 2022.
6. “R: Survey-weighted generalized linear models.” r-survey, https://r-survey.r-forge.r-project.org/survey/html/svyglm.html. Accessed 8 June 2022.

Tables and Figures



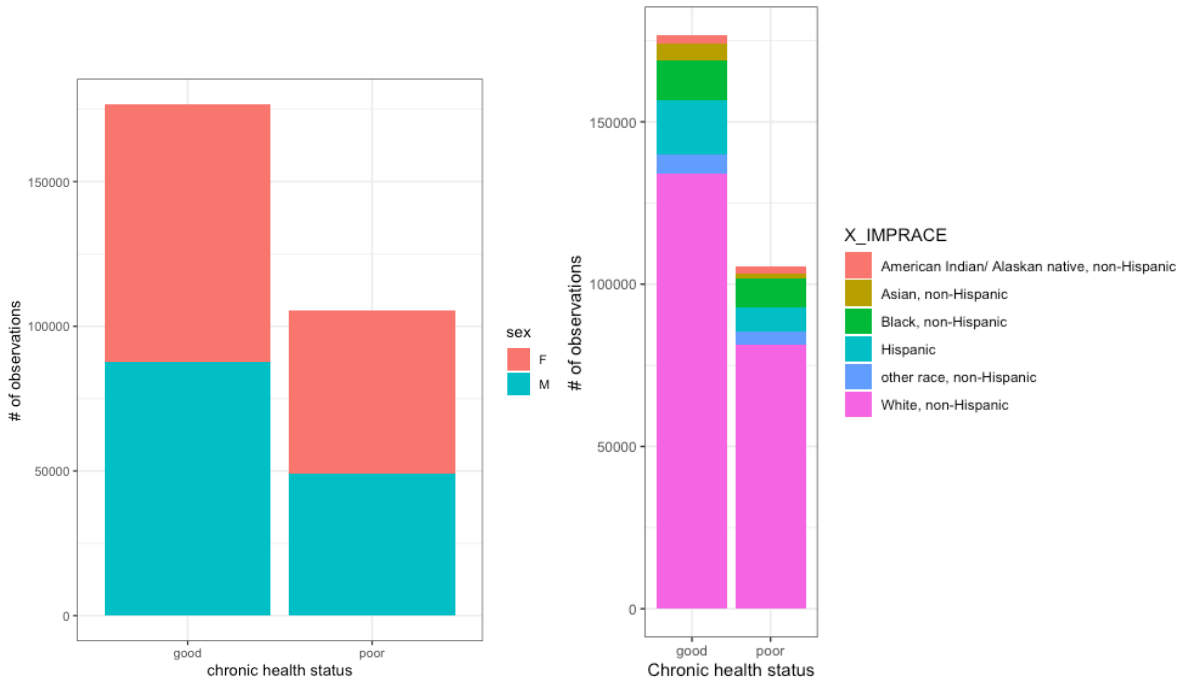
(a) (b)

Figure 1 (a): Histograms of the number of observations of oral health status stratified by sex/ gender. **(b):** Histograms of the number of observations of oral health status stratified by race/ ethnicity.



(a) (b)

Figure 2 (a): Histograms of the number of observations of smoking status stratified by sex/ gender. **(b):** Histograms of the number of observations of smoking status stratified by race/ ethnicity.



(a) (b)

Figure 3 (a): Histograms of the number of observations of chronic health status stratified by sex/ gender. **(b):** Histograms of the number of observations of chronic health status stratified by race/ ethnicity.

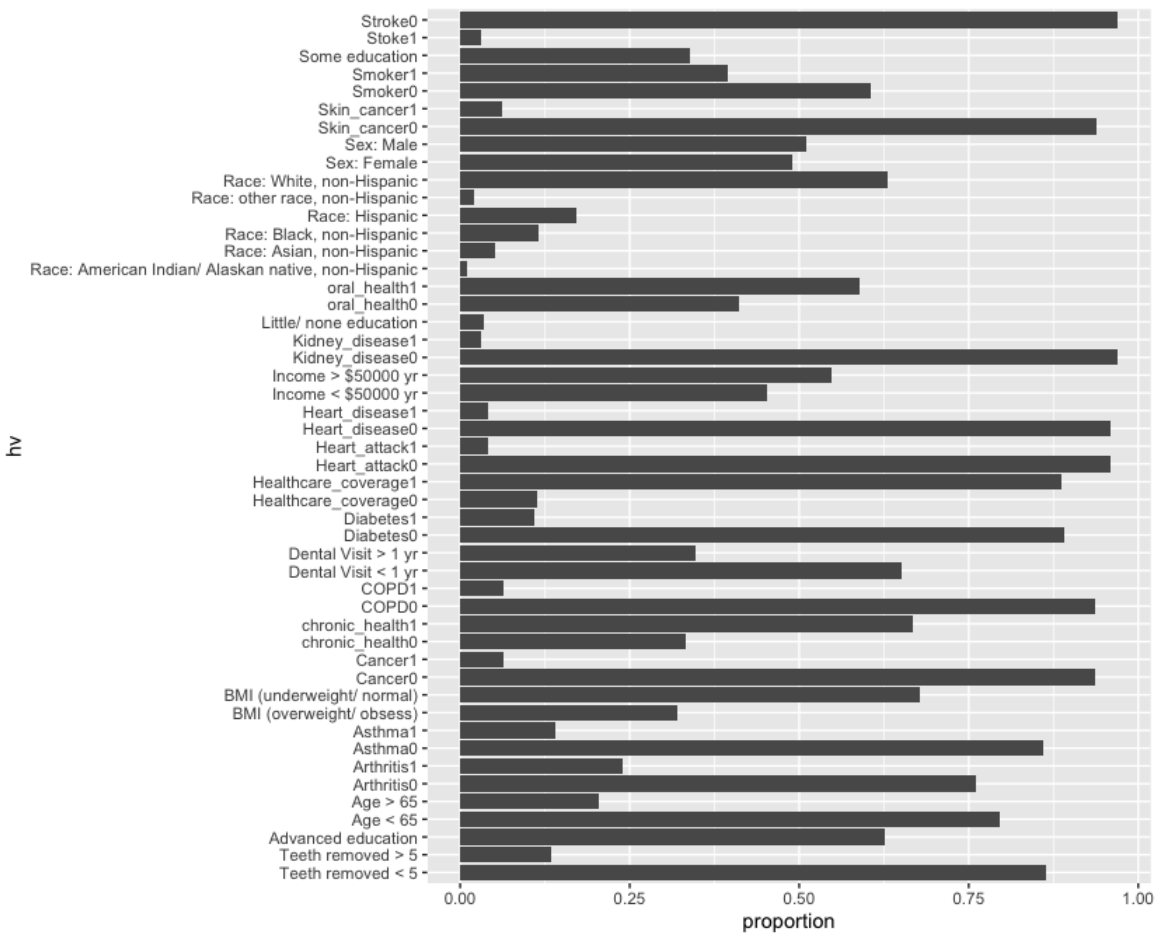


Figure 4: Histograms of the weighted prevalence for variables included the study.

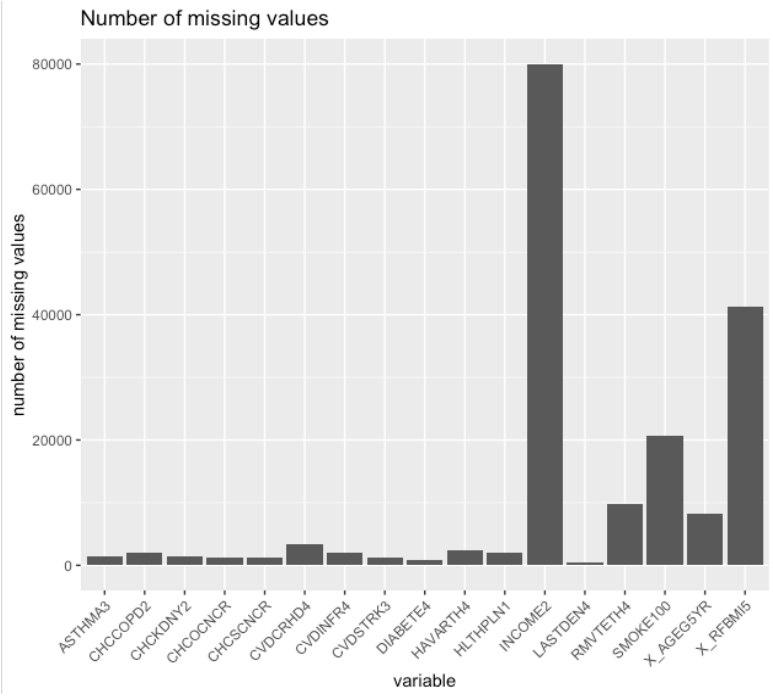


Figure 5: Histograms of the number of missing values for each variable included in the study.

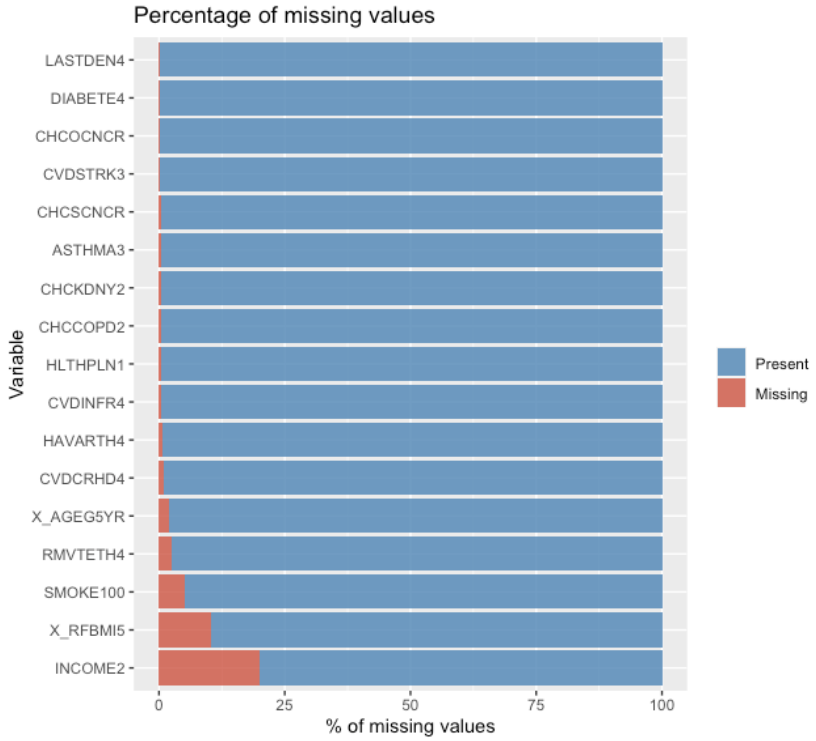


Figure 6: Percentage of missing values for each variable included in the study.

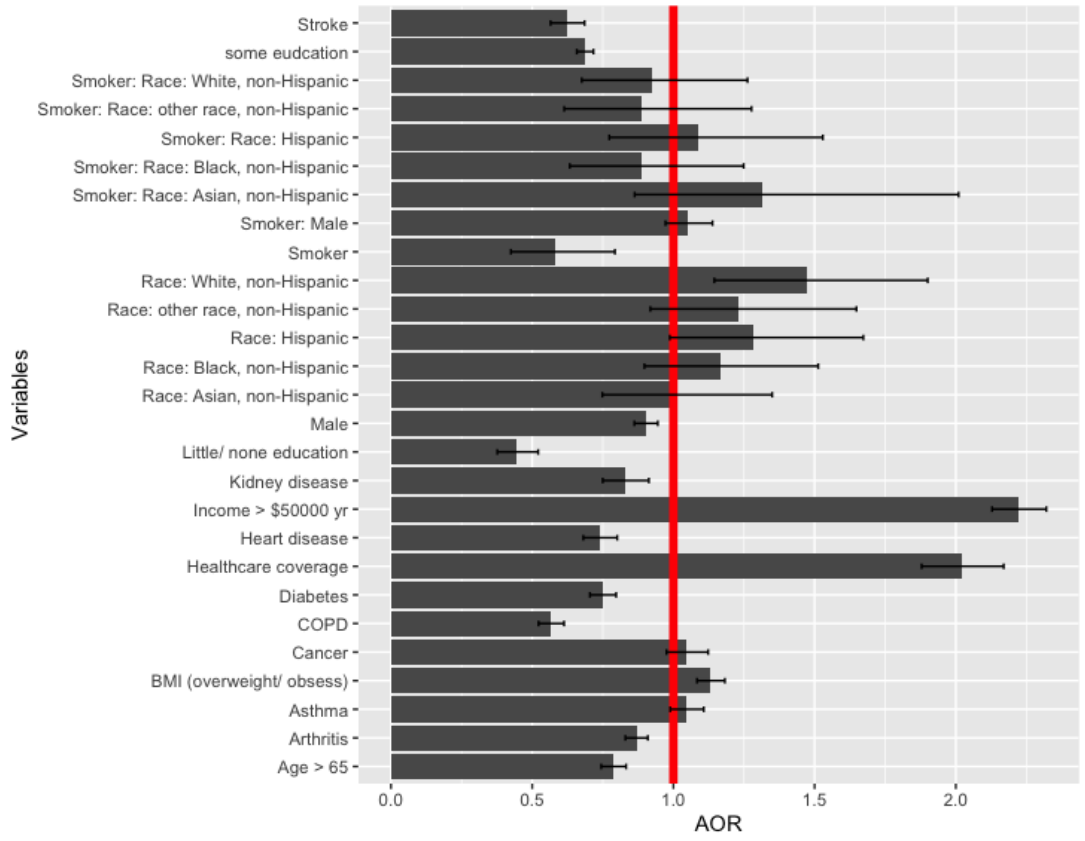


Figure 7: Histograms of adjusted odds ratio for logistic regression of oral health status on tobacco use/smoking status.

	Estimate	Std. Error	t value	Pr(>|t|)	2.5 %	97.5 %
(Intercept)	-0.36695	0.1333	-2.7525	5.91e-03	-0.6282	-0.1057
Smoker	-0.54444	0.1593	-3.4167	6.34e-04	-0.8568	-0.2321
Sex: Male	-0.23927	0.0285	-8.3854	5.08e-17	-0.2952	-0.1833
Race: Asian, non-Hispanic	0.00494	0.1505	0.0328	9.74e-01	-0.2900	0.2999
Race: Black, non-Hispanic	0.15280	0.1332	1.1468	2.51e-01	-0.1084	0.4140
Race: Hispanic	0.25073	0.1345	1.8638	6.24e-02	-0.0129	0.5144
Race: other race, non-Hispanic	0.20726	0.1492	1.3889	1.65e-01	-0.0852	0.4997
Race: White, non-Hispanic	0.38869	0.1293	3.0059	2.65e-03	0.1352	0.6421
Age > 65	-0.10282	0.0234	-4.3854	1.16e-05	-0.1488	-0.0569
Income > $50000 yr	0.79860	0.0220	36.2263	1.09e-286	0.7554	0.8418
Healthcare coverage	0.70261	0.0367	19.1497	1.10e-81	0.6307	0.7745
BMI (overweight/ obese)	0.12382	0.0223	5.5645	2.63e-08	0.0802	0.1674
Education: little/ none education	-0.81434	0.0830	-9.8138	9.91e-23	-0.9770	-0.6517
Education: some education	-0.37573	0.0218	-17.2071	2.54e-66	-0.4185	-0.3329
Heart disease	-0.30308	0.0414	-7.3130	2.62e-13	-0.3843	-0.2219
Stroke	-0.47435	0.0493	-9.6161	6.89e-22	-0.5710	-0.3777
Asthma	0.04545	0.0289	1.5711	1.16e-01	-0.0112	0.1021
Cancer	0.04552	0.0360	1.2663	2.05e-01	-0.0249	0.1160
COPD	-0.56947	0.0407	-13.9952	1.73e-44	-0.6492	-0.4897
Arthritis	-0.14059	0.0233	-6.0280	1.66e-09	-0.1863	-0.0949
Kidney disease	-0.18948	0.0503	-3.7661	1.66e-04	-0.2881	-0.0909
Diabetes	-0.28837	0.0316	-9.1277	7.04e-20	-0.3503	-0.2265
Smoker * Sex: Male	0.05055	0.0404	1.2521	2.11e-01	-0.0286	0.1297
Smoker * Race: Asian, non-Hispanic	0.27522	0.2157	1.2757	2.02e-01	-0.1476	0.6981
Smoker * Race: Black, non-Hispanic	-0.11760	0.1734	-0.6783	4.98e-01	-0.4574	0.2222
Smoker * Race: Hispanic	0.08353	0.1741	0.4798	6.31e-01	-0.2577	0.4247
Smoker * Race: other race, non-Hispanic	-0.12231	0.1870	-0.6540	5.13e-01	-0.4888	0.2442

Table 1: Estimated coefficients for adjusted log odds ratios of oral health status on tobacco use/smoking status with corresponding standard errors, t-values, p-values, and 95% confidence intervals (see model 1).

	AOR	2.5 %	97.5 %
Smoker	0.580	0.425	0.793
Age > 65	0.787	0.744	0.832
Race: Asian, non-Hispanic	1.005	0.748	1.350
Race: Black, non-Hispanic	1.165	0.897	1.513
Race: Hispanic	1.285	0.987	1.673
Race: other race, non-Hispanic	1.230	0.918	1.648
Race: White, non-Hispanic	1.475	1.145	1.901
Male	0.902	0.862	0.945
Income > $50000 yr	2.222	2.128	2.321
Healthcare coverage	2.019	1.879	2.170
BMI (overweight/ obsess)	1.132	1.084	1.182
Education: Little/ none education	0.443	0.376	0.521
Education: some education	0.687	0.658	0.717
Heart disease	0.739	0.681	0.801
Stroke	0.622	0.565	0.685
Asthma	1.046	0.989	1.108
Cancer	1.047	0.975	1.123
COPD	0.566	0.522	0.613
Arthritis	0.869	0.830	0.909
Kidney disease	0.827	0.750	0.913
Diabetes	0.749	0.704	0.797
Smoker * Male	1.052	0.972	1.138
Smoker * Race: Asian, non-Hispanic	1.317	0.863	2.010
Smoker * Race: Black, non-Hispanic	0.889	0.633	1.249
Smoker * Race: Hispanic	1.087	0.773	1.529
Smoker * Race: other race, non-Hispanic	0.885	0.613	1.277
Smoker * Race: White, non-Hispanic	0.923	0.675	1.263

Table 2: Estimated coefficients for adjusted odds ratios of oral health status on tobacco use/smoking status with 95% confidence intervals.

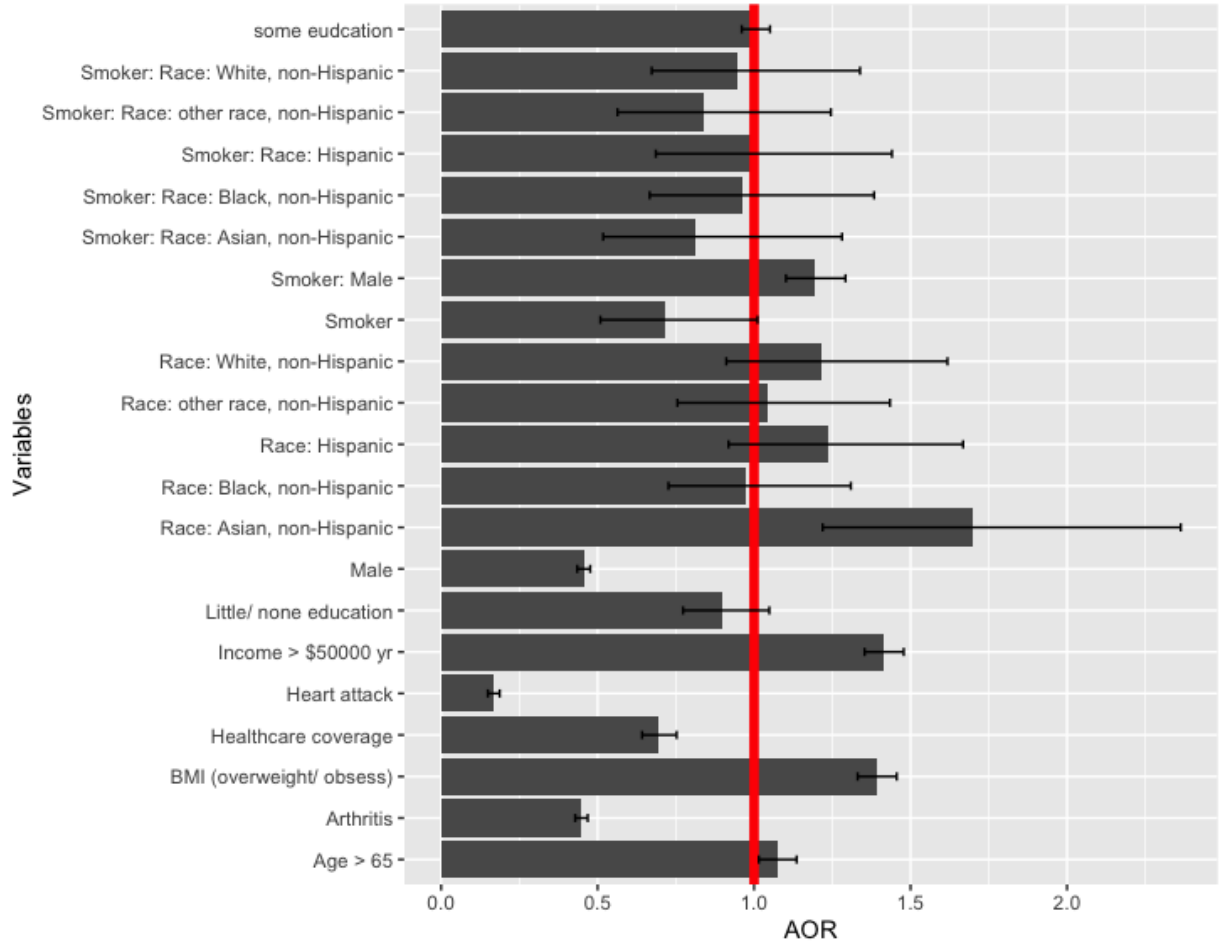


Figure 8: Histograms of adjusted odds ratio for logistic regression of chronic health status on tobacco use/ smoking status.

	Estimate	Std. Error	t value	Pr(>|t|)	2.5 %	97.5 %
(Intercept)	1.10726	0.1502	7.372	1.69e-13	0.8129	1.4017
Smoker	-0.33233	0.1751	-1.898	5.77e-02	-0.6755	0.0108
Sex: Male	0.07139	0.0287	2.487	1.29e-02	0.0151	0.1276
Race: Asian, non-Hispanic	0.52908	0.1689	3.133	1.73e-03	0.1981	0.8601
Race: Black, non-Hispanic	-0.02505	0.1503	-0.167	8.68e-01	-0.3196	0.2695
Race: Hispanic	0.21319	0.1522	1.401	1.61e-01	-0.0852	0.5115
Race: other race, non-Hispanic	0.03990	0.1637	0.244	8.07e-01	-0.2809	0.3607
Race: White, non-Hispanic	0.19414	0.1464	1.326	1.85e-01	-0.0929	0.4812
Age > 65	-0.78719	0.0233	-33.751	3.27e-249	-0.8329	-0.7415
Income > $50000 yr	0.34659	0.0225	15.375	2.51e-53	0.3024	0.3908
Healthcare coverage	-0.36356	0.0404	-8.988	2.52e-19	-0.4428	-0.2843
BMI (overweight/ obese)	0.33075	0.0229	14.462	2.18e-47	0.2859	0.3756
Education: little/ none education	-0.10496	0.0778	-1.348	1.78e-01	-0.2575	0.0476
Education: some education	0.00478	0.0229	0.208	8.35e-01	-0.0401	0.0497
Heart attack	-1.78907	0.0575	-31.127	2.37e-212	-1.9017	-1.6764
Arthritis	-0.80313	0.0228	-35.196	9.06e-271	-0.8479	-0.7584
Smoker * Male	0.17674	0.0407	4.347	1.38e-05	0.0971	0.2564
Smoker * Race: Asian, non-Hispanic	-0.20540	0.2311	-0.889	3.74e-01	-0.6584	0.2476
Smoker * Race: Black, non-Hispanic	-0.04066	0.1864	-0.218	8.27e-01	-0.4060	0.3247
Smoker * Race: Hispanic	-0.00586	0.1893	-0.031	9.75e-01	-0.3768	0.3651
Smoker * Race: other race, non-Hispanic	-0.17707	0.2024	-0.875	3.82e-01	-0.5737	0.2195
Smoker * Race: White, non-Hispanic	-0.05224	0.1754	-0.2998	7.66e-01	-0.3961	0.2916

Table 3: Estimated coefficients for adjusted log odds ratios of chronic health status on tobacco use/smoking status with corresponding standard errors, t-values, p-values, and 95% confidence intervals (see model 2).

	AOR	2.5 %	97.5 %
Smoker	0.717	0.509	1.011
Age > 65	1.074	1.015	1.136
Race: Asian, non-Hispanic	1.697	1.219	2.363
Race: Black, non-Hispanic	0.975	0.726	1.309
Race: Hispanic	1.238	0.918	1.668
Race: other race, non-Hispanic	1.041	0.755	1.434
Race: White, non-Hispanic	1.214	0.911	1.618
Male	0.455	0.435	0.476
Income > $50000 yr	1.414	1.353	1.478
Healthcare coverage	0.695	0.642	0.753
BMI (overweight/ obsess)	1.392	1.331	1.456
Education: little/ none education	0.900	0.773	1.049
Education: some education	1.005	0.961	1.051
Heart attack	0.167	0.149	0.187
Arthritis	0.448	0.428	0.468
Smoker * Male	1.193	1.102	1.292
Smoker * Race: Asian, non-Hispanic	0.814	0.518	1.281
Smoker * Race: Black, non-Hispanic	0.960	0.666	1.384
Smoker * Race: Hispanic	0.994	0.686	1.441
Smoker * Race: other race, non-Hispanic	0.838	0.563	1.246

Table 4: Estimated coefficients for adjusted odds ratios of chronic health status on tobacco use/smoking status with 95% confidence intervals.

Appendices

BRFSS CodeBook 2020:

Oral health sections

- Variable 1: Not including teeth lost for injury or orthodontics, how many of your permanent teeth have been removed because of tooth decay or gum disease?
- Variable 2: Including all types of dentists, such as orthodontists, oral surgeons, and all other dental specialists, as well as dental hygienists, how long has it been since you last visited a dentist or a dental clinic for any reason?

Tobacco use

- Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes]

Chronic health sections

- Variable 1: Ever Diagnosed with Angina or Coronary Heart Disease
- Variable 2: Ever Diagnosed with a Stroke
- Variable 3: Ever Told Had Asthma
- Variable 4: (Ever told) you had any other types of cancer?
- Variable 5: (Ever told) (you had) chronic obstructive pulmonary disease, C.O.P.D., emphysema, or chronic bronchitis?
- Variable 6: Ever told you have kidney disease?
- Variable 7: (Ever told) you had diabetes?

Variable name

Health Indicator	Variable name	Value recode to 1 (ref)
Dental visit	LASTDEN4	1
Number of permanent teeth removed	RMVTETH4	1,8
Tobacco Use	SMOKE100	1
Heat attack	CVDINFR4	1
Heart disease	CVDCRHD4	1
Stroke	CVDSTRK3	1
Asthma	ASTHMA3	1
Skin cancer	CHCSCNCR	1
Cancer	CHCOCNCR	1
Chronic obstructive pulmonary disease	CHCCOPD2	1
Arthritis	HAVARTH4	1
Kidney disease	CHCKDNY2	1
Diabetes	DIABETE4	1
Age	X_AGEG5YR	1 (> 65 yr)
Gender	SEXVAR	1 (male)
Race	X_IMPRACE	category
Income level	INCOME2	7 & 8 (>$50000 yr)
Have healthcare coverage	HLTHPLN1	1
BMI (overweight/obese)	X_RFBMI5	1
Education level	EDUCA	category