Heart Disease and Risk Factor Prevalence in Washington State Adults in 2020

Rachel Flodin

Abstract

The prevalence of heart disease and seven risk factors for heart disease in Washington state adults in 2020 is examined using survey methods, including whether the prevalence varies by urban/rural residence, income level, or education level. In addition, I examine which risk factors have the highest association with heart diesase, after controlling for age, sex, and race, using logistic regression. Overall, approximately 5.47% of Washington state adults had heart disease in 2020. Obesity and smoking were the two risk factors which had high prevalence, and also high associations with heart disease, so focusing on reducing obesity and smoking could have a positive impact on reducing heart disease in the future.

Introduction

Heart disease is the leading cause of death for men and women in the U.S., with an estimated 659,000 deaths occurring each year, and an estimated cost of $363 billion per year from 2016-2017, which includes costs related to health care (including medications) and lost productivity.[11] It is the second leading cause of death in Washington state, with higher rates seen in people who live in areas with lower levels of income and education.[23] Heart disease includes many different heart conditions including coronary artery disease (CAD), the most common type, and heart attack.[10] A heart attack occurs when blood flow to the heart is blocked; the longer the blockage occurs, the greater the damage that can occur to the heart.[17,8] Coronary artery disease (CAD) occurs when the arteries supplying the heart are damaged or diseased, which can be caused by a buildup of plaque (deposits which contain cholesterol).[16]

There are many risk factors for heart disease, including age, smoking, physical activity, diabetes, obesity, poor mental health and excessive alcohol use, among others.[7,22] In addition, lack of health care coverage and inability to pay for care can lead to less/no access to preventative care, and delayed treatment, which can then lead to worse health outcomes.[2] In 2016, an estimated 15% of adults in Washington state were 65 or older, and it is expected that this proportion will grow to 22% by 2030.[23] Since heart disease death rates are highest in this age group, and since this age group is estimated to grow in size in the next eight years, knowing the current prevalence of heart disease and of specific risk factors for heart disease will aid in planning for the future.[23]

Since 1984, the Behavior Risk Factor Surveillance System (BRFSS) survey has collected health-related information from U.S. residents, beginning with only 15 states it reached nationwide coverage in 1993.[5] This study will look at the overall prevalence of heart disease, specifically heart attacks and coronary artery disease, and risk factors for heart disease in adults (age 18+) living in Washington state in 2020 using BRFSS survey data. Specifically, the following risk factors will be examined: diabetes, tobacco use, alcohol use, poor mental health, obesity, and lack of health care coverage or cost preventing care. It will also examine whether the prevalence of heart disease and risk factors differ between urban and rural residents, income level (less than $15,000, $15,000 to less than $25,000, $25,000 to less than

$35,000, $35,000 to less than $50,000, and $50,000 or more), and education level (did not graduate high school, graduated high school, some college or technical school, and graduated from college or technical school) in Washington state. Finally, I will examine which risk factors have the highest association with heart disease, after controlling for age, sex, and race.

Methods

Data Description

The data for this study come from the Washington State BRFSS survey, which is made available for download by request from the WA State Department of Health Center for Health Statistics at https://fortress.wa.gov/doh/opinio/s?s=BRFSSDataOrderForm. The 2020 BRFSS survey for Washington State contains 12,902 rows in total. The data contains 600 variables, which includes variables for the survey characteristics and for each survey question, variables imputed by the CDC, and additional variables containing weights needed for correct population calculations due to the survey design.

However, the Washington State BRFSS survey data does not contain the urban/rural designation variable which is present in the CDC 2020 BRFSS survey data available for download at https://www.cdc.gov/brfss/annual_data/annual_2020.html. So, I merged the urban/rural variable from the 2020 CDC BRFSS data into the 2020 Washington State BRFSS data.

Some variables have separate levels for "don't know/not sure" and for "refused", and some have a single variable for "don't know/refused/missing". In order to keep this consistent across all variables, I treated all 3 of these levels as missing values (NA).

The binary variable for heart disease in the data is `michd`, which contains values 1 (has been told by a health professional that they have heart disease), 0 (has *not* been told by a health professional that they have heart disease), and NA. I also used the heart attack (myocardial infarction or MI) and coronary artery disease (CAD) variables, denoted as `cvdinfr4` and `cvdcrhd4`, respectively. These two binary variables contain the values 1 (yes), 0 (no), NA.

The variable for diabetes is `diabete4`. There is evidence that having diabetes during pregnancy only[12], and/or having pre- or borderline diabetes[4] may also increase the risk of heart disease. I kept this variable unchanged to see prevalence for diabetes, diabetes during pregnancy, and pre/borderline diabetes. I also created a binary variable named `diabetesYN` from the `diabete4` variable, which contained values 1 = yes ("yes"), 0 = no ("yes, pregnancy", "no", and "no, pre or borderline diabetes"), and NA. I chose not to include the people with diabetes only during pregnancy and the people with pre/borderline diabetes in the 1=yes level since the risk of heart disease from these two conditions is lower than for people living with diabetes.

For tobacco use, I will look at smoking using variable `smoker3`, which indicates what level of smoker a person is. Being a former smoker also has a higher risk of heart disease, but to a lesser extent than for a current smoker.[1] While the amount a person smokes/smoked (measured in pack years) is important to consider, this was not available in the survey data. Based on this, I recoded the variable to use 1 = current smoker, 2 = former smoker, 3 = never smoker, and NA. I also created a binary variable named `smokerYN` from the `smoker3` variable, which contained values 1 = yes (current smoker), 0 = no (former and never smoker), and NA. I chose not to include former smoker in the 1=yes level since the risk of heart disease for former smokers is generally considered less than for current smokers.

There are two variables of interest for alcohol use. The first variable, `rfdrhv7`, indicates heavy alcohol consumption, defined as 14 or more drinks per week for men and as 7 or more drinks per week for women. The second variable, `rfbing5`, indicates whether or not a person reported binge drinking.

Both heavy alcohol use and binge drinking are risk factors for heart disease[13,18], so I created a new combined binary variable, `hvyalc`, indicating whether or not a person was a heavy alcohol user or a binge drinker since a person could be categorized as both. The values for the `hvyalc` variable are 0 = no, 1 = yes.

The variable for mental health is `ment14d`. I recoded this variable into a binary variable to indicate poor mental health (defined as 14 or more days in the past 30 days that the respondent said their mental health was not good[20]). The new values for the `ment14d` variable are 0 = no, 1 = yes.

The variable for BMI is `bmi5cat`. Being either overweight or obese are both risk factors for heart disease[3], so I recoded this variable to indicate whether a person is 1 = overweight ($25.0 \leq$ BMI < 30.0), 2 = obese (BMI ≥ 30.0), 0 = BMI < 25.0, and NA. I also created a binary variable named `obeseYN` from the `bmi5cat` variable, which contained values 1 = yes (BMI ≥ 30.0), 0 = no (BMI < 30.0), and NA. I chose not to include overweight ($25.0 \leq$ BMI < 30.0) in the 1=yes level since the risk of heart disease for individuals with an overweight BMI is generally considered less than for an obese BMI.

The variable for health care coverage, including health insurance, prepaid plans such as HMOs, government plans such as Medicare or Indian Health Service, is `hlthpln1`, which contains values to 0 (yes), 1 (no), and NA. There is an additional variable, `medcost`, indicating whether the respondent needed to see a doctor, but was unable to due to cost in the past 12 months. This variable takes the values 1 (yes), 0 (no), and NA. Since both lack of ability to pay and lack of health care coverage can be risk factors for heart disease, I included both variables in my analysis.

The variable for urban/rural designation is `urbstat`. The urban/rural designation is made at the county level, where urbal and rural are defined by the 2013 National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties.[6]

I also used the income variable `incomg`, the education variable `educag`, and the imputed variables for age (`impage`), sex (`impsex`), and race (`imprace`).

Survey Design

The BRFSS is a survey, mostly conducted by telephone, of adults (18+) living in the United States. It is coordinated by the CDC, but is administered by each state, the District of Columbia, Guam, and Puerto Rico (for 2020). For the telephone survey, households are selected from lists of potential landline and cell phone numbers, which includes unlisted numbers. A random adult in each household is selected to be interviewed for landline calls, and the adult who answers the cell phone call is interviewed. The survey questions are designed to gather data on chronic health conditions (including heart disease), risks, and prevention.[21,9]

The sample for this survey is a disproportionate stratified sample design for the landline phones. The landline phone numbers are divided into two strata, one for listed numbers and one for unlisted numbers, and samples are drawn from each strata separately. For cell phones, the sampling is done from a list of dedicated cell phone numbers. Washington State also does geographic stratification, where they take disproportionate samples from strata of sub-state regions. This allows for better sample sizes of populations of interest defined by smaller geographic regions.[9]

Sampling Weights

The BRFSS uses weighting in order to make the sample more representative by incorporating both the survey design and also some demographic adjustments. The weighting scheme incorporates both design weights and raking. The design weights help adjust for bias due to nonresponse and errors from non-coverage, and the raking helps adjust for demographic differences between the sample and the true population.

The design weights are calculated by multiplying the geographic stratum weight by the inverse of the number of phones in the household and by the number of adults (age 18+) in the household, where the geographic stratum weight is calculated for each combination of geographic strata and listed/unlisted strata as the number of available phone numbers divided by the number of phone numbers selected. For cell phones, the number of phones and the number of adults are each set to 1.

The final weight to be used in survey analysis (variable `llcpwt`) comes after raking the design weight to margins which include gender by age group, race/ethnicity, education, marital status, tenure, gender by race/ethnicity, age group by race/ethnicity, and phone ownership. The margins may also include region, region by age group, region by gender, and region by race/ethnicity, if geographic strata were used. And, may include county, county by age group, county by gender, and county by race/ethnicity, if at least 1 county had 500 or more respondents.

Statistical Methods

I plan to use the `survey` package by Thomas Lumley in `R` version 4.1.0 for my analysis.[19,14] The CDC recommends that this data be analyzed using sampling procedures, using both the `llcpwt` variable for survey weights and the `ststr` variable for stratification. It is important to use the weighting scheme in order to help reduce selection bias and noncoverage, and to ensure that the total number of cases are equal to the population estimates for the state.

I will be using the `svymean` function to calculate the prevalence of heart disease and the six heart disease risks, and their standard errors. The `confint` function can be used to generate 95% confidence intervals for these proportions. I will use the `svyby` function with the `svymean` argument to calculate the prevalence of heart disease and the six heart disease risks, their standard errors, and 95% confidence intervals, grouping by the levels of each of the 3 variables I am interested in (urban/rural, income level, and education level). Both the `svymean` and `svyby` functions weight each observation by the sampling weights and include effects of any stratification and clustering, as specified in the svydesign object.[15]

I will also perform logistic regression using the `svyglm` function (with the binomial family and logit link) to determine which risk factors have the highest association with heart disease, after controlling for sex and race. The `svyglm` function uses the sampling weights for both the parameter estimation and the standard errors. The standard errors are similar to robust standard errors, but they also incorporate the stratification in the survey design.[15] The response variable is the binary variable `michd`, and the predictor variables will be the binary versions of the risk factor variables discussed above in the data description section. I will use an $\alpha = 0.05$-level to determine whether the predictor variable association is significant.

In order to test whether the prevalence of heart disease and the seven (binary) heart disease risks differ by group (urban/rural, income level, and education level), I will perform Wald tests of the multi-level factor variables `urbstat`, `incomg`, and `educag` using the `regTermTest` function[15], both without and with controlling for age, sex, and race.

I will remove missing values using the na.rm=TRUE argument in all of the functions listed above.

Results

Missing Values

The indicator variable for heart disease has 102 missing values, which is less than 0.79% of all of the survey responses. However, the variable for income level has almost 20% missing, and the variable for BMI has almost 12% missing. All other variable have less than 10% missing, with urban/rural, and the imputed variables for age, sex, and race having no missing values.

Heart Disease Prevalence

Heart disease prevalence in Washington state in 2020 was about 5.5% (95% CI: $(5.01\%, 5.93\%)$), with heart attack prevalence being 3.42% $(3.06\%, 3.78\%)$, and coronary artery disease prevalence 3.42% $(3.34\%, 4.11\%)$. Note: a person can be diagnosed with both heart attack and coronary artery disease.

	Prevalence	SE	95% CI
Urban	0.053	0.0023	(0.0484, 0.0576)
Rural	0.1222	0.0188	(0.0853, 0.1591)
< $15,000	0.1172	0.0146	(0.0887, 0.1457)
$15,000 to < $25,000	0.084	0.0094	(0.0655, 0.1025)
$25,000 to < $35,000	0.0674	0.0089	(0.05, 0.0848)
$35,000 to < $50,000	0.0818	0.0094	(0.0634, 0.1003)
> $50,000	0.0406	0.0029	(0.035, 0.0463)
Did not graduate high school	0.073	0.0109	(0.0516, 0.0944)
Graduated high school	0.0609	0.005	(0.0511, 0.0707)
Some college/technical school	0.0607	0.0044	(0.052, 0.0694)
Graduated college/technical school	0.0385	0.0027	(0.0332, 0.0438)

Table 1: Heart disease prevalence by urban/rural, income level, and education level, with standard error and 95% confidence intervals.

There was higher prevalence of heart disease seen in rural residents (12.22% $(8.53\%, 15.91\%)$). There is statistically significant evidence at the $\alpha = 0.05$-level that the prevalence of heart disease differs between urban and rural residents (Wald test, p-value < 0.001), and it remained a statistically significant difference after controlling for age, sex, and race (Wald test, p-value 0.019).

There was also a higher prevalence of heart disease seen in three of the five levels of income (see red highlighted rows in Table 1), but a lower prevalence in the highest level of income (4.06% $(3.5\%, 4.63\%)$, highlighted in yellow). There is statistically significant evidence that the prevalence of heart disease is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant after controlling for age, sex, and race (Wald test, p-value < 0.001).

There was a lower prevalence of heart disease in the highest level of education (3.85% $(3.32\%, 4.38\%)$, highlighted in yellow). There is statistically significant evidence that the prevalence of heart disease is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant after controlling for age, sex, and race (Wald test, p-value < 0.001).

Risk Factor Prevalence

Diabetes:

Diabetes prevalence in Washington state in 2020 was about 8.6% (95% CI: (8.02%, 9.24%)), with much lower prevalence of diabetes during pregnancy only at 1.0% (0.78$, 1.23%), and pre or borderline diabetes at 1.83% (1.52%, 2.14%).

	Prevalence	SE	95% CI
Urban	0.0848	0.0031	(0.0787, 0.091)
Rural	0.1414	0.0203	(0.1015, 0.1812)
< $15,000	0.1482	0.017	(0.1149, 0.1814)
$15,000 to < $25,000	0.1247	0.0123	(0.1006, 0.1489)
$25,000 to < $35,000	0.0973	0.0121	(0.0736, 0.1209)
$35,000 to < $50,000	0.1125	0.0105	(0.0919, 0.1331)
> $50,000	0.0725	0.004	(0.0647, 0.0802)
Did not graduate high school	0.1211	0.0158	(0.0902, 0.1520)
Graduated high school	0.089	0.0064	(0.0763, 0.1016)
Some college/technical school	0.0971	0.0056	(0.0862, 0.108)
Graduated college/technical school	0.0636	0.0039	(0.056, 0.0713)

Table 2: Diabetes prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There was higher prevalence of diabetes seen in rural residents (14.14% (10.15%, 18.12%)). There is statistically significant evidence at the $\alpha = 0.05$-level that the prevalence differs between urban and rural residents (Wald test, p-value < 0.001). However, the difference failed to be statistically significant after controlling for age, sex, and race (Wald test, p-value 0.139).

There was also a higher prevalence of diabetes seen in the lowest two of the five levels of income (see red highlighted rows in Table 2), but a lower prevalence in the highest level of income (7.25% (6.47%, 8.02%)). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

There was a lower prevalence of diabetes in the highest level of education (6.36% (5.6%, 7.13%)). There is statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

Smoking:

The prevalence of current smokers in Washington state in 2020 was about 11.5% (95% CI: $(10.72\%, 12.24\%)$), with over double the prevalence of former smokers at 25.93% $(24.93\%, 26.92\%)$.

	Prevalence	SE	95% CI
Urban	0.1137	0.0039	(0.1060, 0.1214)
Rural	0.1551	0.021	(0.114, 0.1963)
< $15,000	0.2551	0.0218	(0.2124, 0.2978)
$15,000 to < $25,000	0.2193	0.0165	(0.1869, 0.2516)
$25,000 to < $35,000	0.1681	0.0174	(0.1341, 0.2022)
$35,000 to < $50,000	0.1535	0.0130	(0.128, 0.179)
> $50,000	0.0756	0.0046	(0.0666, 0.0845)
Did not graduate high school	0.1783	0.0181	(0.1428, 0.2138)
Graduated high school	0.1657	0.0095	(0.1471, 0.1843)
Some college/technical school	0.1348	0.0071	(0.1209, 0.1487)
Graduated college/technical school	0.0442	0.0035	(0.0375, 0.051)

Table 3: Smoking prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There may be slightly higher prevalence of smoking in rural residents (15.51% $(11.4\%, 19.63\%)$). We do see evidence that there is a statistically significant difference in prevalence between urban and rural residents at the $\alpha = 0.05$-level (Wald test, p-value 0.03), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value 0.035).

There was a higher prevalence of smoking seen in all but the highest level of income (see red highlighted rows in Table 3), but a lower prevalence in the highest level of income (7.56% $(6.66\%, 8.45\%)$). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

There was a higher prevalence of smoking seen in people who did and did not graduate high school (see red highlighted rows in Table 3), but a lower prevalence of smoking in the highest level of education (4.42% $(3.75\%, 5.1\%)$). There is statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

Heavy Alcohol Use/Binge Drinking:

The prevalence of heavy alcohol use/binge drinking in Washington state in 2020 was 17.07% (95% CI: $(16.18\%, 17.96\%)$).

	Prevalence	SE	95% CI
Urban	0.1715	0.0047	(0.1624, 0.1806)
Rural	0.1406	0.0182	(0.105, 0.1763)
< \$15,000	0.1162	0.0168	(0.0833, 0.1492)
\$15,000 to < \$25,000	0.1438	0.0141	(0.1162, 0.1714)
\$25,000 to < \$35,000	0.1532	0.017	(0.1198, 0.1866)
\$35,000 to < \$50,000	0.1783	0.0138	(0.1512, 0.2054)
> \$50,000	0.1964	0.0066	(0.1834, 0.2094)
Did not graduate high school	0.146	0.0177	(0.1114, 0.1806)
Graduated high school	0.179	0.0097	(0.1599, 0.1981)
Some college/technical school	0.179	0.0084	(0.1626, 0.1957)
Graduated college/technical school	0.1639	0.0066	(0.1509, 0.1769)

Table 4: Heavy alcohol use prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There may be slightly lower prevalence of heavy alcohol use/binge drinking in rural residents (14.06% $(10.5\%, 17.63\%)$). However, we see no evidence that there is a statistically significant difference in prevalence between urban and rural residents at the $\alpha = 0.05$-level (Wald test, p-value 0.127), and we continue to see no statistically significant evidence after controlling for age, sex, and race (Wald test, p-value 0.479).

There was a higher prevalence of heavy alcohol use seen in the highest level of income(19.64% $(18.34\%, 20.94\%)$), but a lower prevalence in the lowest level of income (11.62% $(8.33\%, 14.92\%)$). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

There was no real difference in prevalence of heavy alcohol use seen in any education level (see Table 4). And, we see no statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value 0.21), and we continue to see no statistically significant evidence after controlling for age, sex, and race (Wald test, p-value 0.054).

Poor Mental Health:

Poor mental health prevalence in Washington state in 2020 was 13.73% (95% CI: (12.94%, 14.52%)).

	Prevalence	SE	95% CI
Urban	0.1374	0.0041	(0.1293, 0.1455)
Rural	0.1343	0.019	(0.0972, 0.1715)
< $15,000	0.2817	0.0233	(0.2361, 0.3273)
$15,000 to < $25,000	0.2344	0.016	(0.203, 0.2659)
$25,000 to < $35,000	0.1774	0.0184	(0.1413, 0.2135)
$35,000 to < $50,000	0.1278	0.0119	(0.1045, 0.1512)
> $50,000	0.1046	0.0052	(0.0944, 0.1148)
Did not graduate high school	0.1598	0.0175	(0.1255, 0.1941)
Graduated high school	0.1507	0.0086	(0.1338, 0.1677)
Some college/technical school	0.1649	0.0077	(0.1499, 0.1799)
Graduated college/technical school	0.095	0.0051	(0.0851, 0.105)

Table 5: Poor mental health prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There was no real difference in the prevalence of poor mental health in rural (13.43% (9.72%, 17.15%)) vs urban (13.74% (12.93%, 14.55%)) residents. And, we see no statistically significant evidence that the prevalence is not the same between urban and rural residents at the $\alpha = 0.05$-level (Wald test, p-value 0.876), and we continue to see no statistically significant evidence after controlling for age, sex, and race (Wald test, p-value 0.742).

There was a higher prevalence of poor mental health seen in the lowest two of the five levels of income (see red highlighted rows in Table 5), but a lower prevalence in the highest level of income (10.46% (9.44%, 11.48%)). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

There was a higher prevalence of poor mental health seen in people who completed some college/technical school (16.49% (14.99%, 17.99%)), but a lower prevalence in the highest level of education (9.5% (8.51%, 10.5%)). There is statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

Obesity:

Prevalence of obesity in Washington state in 2020 was about 28.03% (95% CI: (26.96%, 29.09%)), with even higher prevalence of overweight BMI at 35.98% (34.81%, 37.15%).

	Prevalence	SE	95% CI
Urban	0.2797	0.0055	(0.2689, 0.2905)
Rural	0.303	0.026	(0.2521, 0.3539)
< \$15,000	0.3661	0.0257	(0.3156, 0.4164)
\$15,000 to < \$25,000	0.3033	0.0183	(0.2673, 0.3392)
\$25,000 to < \$35,000	0.3085	0.0214	(0.2665, 0.3504)
\$35,000 to < \$50,000	0.3128	0.0163	(0.2808, 0.3448)
> \$50,000	0.2657	0.0073	(0.2515, 0.28)
Did not graduate high school	0.3143	0.0238	(0.2677, 0.3608)
Graduated high school	0.3185	0.0118	(0.2954, 0.3415)
Some college/technical school	0.3174	0.01	(0.2977, 0.337)
Graduated college/technical school	0.2067	0.0071	(0.1929, 0.2205)

Table 6: Obesity prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There may be slightly higher prevalence of obesity in rural residents (30.3% (25.21%, 35.39%)). However, we see no statistically significant evidence that the prevalence is not the same between urban and rural residents at the $\alpha = 0.05$-level (Wald test, p-value 0.369), and we continue to see no statistically significant evidence after controlling for age, sex, and race (Wald test, p-value 0.963).

There was a higher prevalence of obesity seen in the lowest level of income (36.61% (31.56%, 41.64%)). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value 0.013).

There was a higher prevalence of obesity seen in people who did graduate high school and who had some college/technical school (see red highlighted rows in Table 6), but a lower prevalence in the highest level of education (20.67% (19.29%, 22.05%)). There is statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

No Health Care Coverage:

Prevalence of adults without health care coverage/insurance in Washington state in 2020 was about 9.4% (95% CI: $(8.68\%, 10.12\%)$), with higher prevalence seen in urban residents.

	Prevalence	SE	95% CI
Urban	0.0948	0.0038	(0.0874, 0.1022)
Rural	0.0637	0.0156	(0.0332, 0.0942)
< \$15,000	0.1508	0.0197	(0.1122, 0.1894)
\$15,000 to < \$25,000	0.2078	0.0173	(0.1739, 0.2417)
\$25,000 to < \$35,000	0.1701	0.0184	(0.134, 0.2062)
\$35,000 to < \$50,000	0.1064	0.0108	(0.0853, 0.1275)
> \$50,000	0.0462	0.0038	(0.0387, 0.0537)
Did not graduate high school	0.3368	0.0223	(0.2931, 0.3806)
Graduated high school	0.1135	0.008	(0.0979, 0.1291)
Some college/technical school	0.0705	0.0054	(0.0599, 0.081)
Graduated college/technical school	0.0345	0.0034	(0.0279, 0.0411)

Table 7: No health care coverage prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There may be slightly lower prevalence of adults without health care coverage/insurance in rural residents (6.37% $(3.32\%, 9.42\%)$). However, we see no statistically significant evidence that the prevalence is not the same between urban and rural residents at the $\alpha = 0.05$-level (Wald test, p-value 0.103), and we continue to see no statistically significant evidence after controlling for age, sex, and race (Wald test, p-value 0.68).

There was a higher prevalence of adults without health care coverage seen in the lowest three of the five levels of income (see red highlighted rows in Table 7), but a lower prevalence in the highest level of income (4.62% $(3.87\%, 5.37\%)$). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

There was a higher prevalence of adults without health care coverage seen in people who did not graduate high school (33.68% $(29.31\%, 38.06\%)$), but a lower prevalence seen in people who had some college/technical school and those who graduated college/technical school (see red highlighted rows in Table 7). There is statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

Cost Prevented Care:

Prevalence of cost preventing a health care visit in Washington state in 2020 was 8.86% (95% CI: $(8.2\%, 9.53\%)$)

	Prevalence	SE	95% CI
Urban	0.0885	0.0034	(0.0818, 0.0953)
Rural	0.0934	0.0164	(0.0613, 0.1255)
< \$15,000	0.1948	0.0214	(0.1528, 0.2368)
\$15,000 to < \$25,000	0.169	0.0143	(0.1409, 0.1971)
\$25,000 to < \$35,000	0.1654	0.0185	(0.129, 0.2017)
\$35,000 to < \$50,000	0.1109	0.011	(0.0893, 0.1325)
> \$50,000	0.0561	0.004	(0.0483, 0.0638)
Did not graduate high school	0.1727	0.017	(0.1392, 0.2061)
Graduated high school	0.0991	0.0071	(0.0851, 0.113)
Some college/technical school	0.0936	0.006	(0.0819, 0.1053)
Graduated college/technical school	0.0517	0.0041	(0.0437, 0.0599)

Table 8: Cost prevented care prevalence by urban/rural, income, and education, with standard errors and 95% confidence intervals.

There was no real difference in the prevalence of cost of health care preventing care in rural (9.34% $(6.13\%, 12.55\%)$) vs urban (8.85% $(8.18\%, 9.53\%)$) residents. And, we see no statistically significant evidence that the prevalence is not the same between urban and rural residents at the $\alpha = 0.05$-level (Wald test, p-value 0.766), and we continue to see no statistically significant evidence after controlling for age, sex, and race (Wald test, p-value 0.191).

There was a higher prevalence of cost preventing care seen in the lowest three of the five levels of income (see red highlighted rows in Table 8), but a lower prevalence in the highest level of income (5.61% $(4.83\%, 6.38\%)$). There is statistically significant evidence that the prevalence is not the same across all income levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value 0.001).

There was a higher prevalence of cost preventing care seen in people who did not graduate high school (17.27% $(13.92\%, 20.61\%)$), but a lower prevalence in adults who graduated college/technical school (5.17% $(4.37\%, 5.99\%)$). There is statistically significant evidence that the prevalence is not the same across all education levels (Wald test, p-value < 0.001), and it remained statistically significant even after controlling for age, sex, and race (Wald test, p-value < 0.001).

Overall we see that overweight BMI, obese BMI, and former smoker had the highest prevalence rates (above 20%) of the risk factors, with overweight BMI being above 35%. Heavy alcohol use, poor mental health, and current smoker all had prevalence rates of $10 - 20\%$.

Highest Associated Risk Factors

A logistic regression model with binomial family and logit link was fit with the `svyglm` function, with heart disease as the response and all seven (binary) risk factors, age, sex, and race as predictors. As we can see in Table 10 below, nine of the variables were significant at the $\alpha = 0.05$-level (highlighted in red). Of the seven risk factors, five were significant: diabetes, smoking, heavy alcohol use, obesity, and cost prevented care. Only no health care coverage and poor mental health were not significantly associated with heart disease.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-6.024	0.417	-14.462	< 0.001
Diabetes	0.986	0.133	7.445	< 0.001
Smoking	0.544	0.167	3.259	0.001
Heavy Alcohol Use	-0.582	0.197	-2.959	0.003
Poor Mental Health	0.12	0.171	0.701	0.483
Obesity	0.321	0.121	2.647	0.008
No Health Insurance	-0.402	0.296	-1.358	0.174
Cost Prevented Care	0.543	0.213	2.550	0.011
Urban/Rural	0.261	0.221	1.178	0.239
Income 15-25	-0.319	0.227	-1.403	0.161
Income 25-35	-0.585	0.249	-2.350	0.019
Income 35-50	-0.385	0.249	-1.550	0.121
Income >50	-0.6	0.214	-2.797	0.005
Education Grad HS	-0.145	0.227	-0.637	0.524
Education Some College	-0.206	0.227	-0.908	0.364
Education Grad College	-0.477	0.234	-2.033	0.042
Age	0.069	0.005	14.474	< 0.001
Sex Female	-0.799	0.116	-6.914	< 0.001
Race Black	0.018	0.452	0.039	0.969
Race Asian	-0.573	0.435	-1.316	0.188
Race Native American	0.007	0.389	0.019	0.985
Race Hispanic	-0.07	0.276	-0.253	0.8
Race Other	-0.042	0.272	-0.155	0.876

Table 9: Logistic regression log odds estimates with robust standard errors and 95% confidence intervals.

We can see that diabetes has the highest association with heart disease. Having diabetes, compared to not having diabetes, increases the odds of having heart disease by 168.2% (95% CI: $(2.068, 3.477)$), holding all other variables constant. Smoking (OR: 1.724, 95% CI: $(1.242, 2.392)$), cost prevented care (OR: 1.722, 95% CI: $(1.134, 2.614)$), obesity (OR: 1.379, 95% CI: $(1.087, 1.749)$), and age (OR: 1.072, 95% CI: $(1.062, 1.082)$) all had higher associations with heart disease.

Discussion

It was not unexpected that the prevalence of heart disease differed between urban and rural residents, since rural communities often tend to have lower levels of income and education, which are known to have higher rates of heart disease. The urban/rural designation is made at the county level, and for Washington state, only nine counties were classified as rural. This feels like a very narrow definition since much of eastern Washington does not contain large towns or cities, especially when compared to western Washington. It seems like this should be reexamined to see if other counties should be reclassified as rural, and results should be recomputed to see if they are greatly impacted by the change.

Almost all risk factors had higher prevalence among low income and low education levels, and lower prevalence among high income and high education levels. However, it was very unexpected to see that heavy alcohol use had higher prevalence among people making over $50,000 and lower prevalence among people making less than $15,000. Although, alcohol is a very expensive product, and one that is expendable, which may explain why adults with very low income may not drink as heavily. It was also unexpected that there was such a difference between the poor mental health prevalence between adults who completed some college/technical school compared to adults who graduated college/technical school, with those who did and did not graduate high school having prevalence rates comparable to the overall prevalence. Perhaps this is because adults who have not graduated college/technical school may still be in school and under stress. Or perhaps it is because they may have stopped college/technical school before graduating, leaving with debt from student loans, but lacking the degree that may have increased their earning power.

It was also unexpected that race was not shown to be a significant predictor of heart disease, since it has been shown that the risk of heart disease is higher in certain races. However, a very large majority of the responders were Caucasian, which may explain why this occurred in the analysis.

Overall we saw that overweight BMI, obese BMI, and former smoker had the highest prevalence rates, followed by heavy alcohol use, poor mental health, and current smoker. Since obesity was one of the top three risk factors when we looked at the overall prevalence, and it also has a high association with heart disease. Along with the knowledge that obesity is a known risk factor for heart disease, a good way to reduce heart disease prevalence would also be to reduce obesity prevalence. The only other variable with a prevalence above 10%, and a high association with heart disease is smoking. Continuing to work to reduce smoking rates in the state should also then help reduce heart disease prevalence.

Some limitations of this analysis are that we may have survey bias based on who is willing to respond to a survey, and missing values for some individual questions may be biasing our results, especially in regards to questions about income, BMI, and other sensitive questions, which had very high rates missing data. Another limitation is that the answers were self-reported, which could increase the rate of incorrect responses due to forgetfulness, misunderstanding the question, or internal pressure to lie over sensitive topics. Another limitation is that I do not have data for two well-known risk factors (high cholesterol level and high blood pressure) because they were not included in the 2020 survey year, so the associations of the risk factors, age, sex, and race are most likely incorrect due to this omission.

For future work, it would be interesting to look at the prevalence rates over time for heart disease and the risk factors to get a clearer picture on how we are doing with heart disease prevention.

References

1. Ahmed, Patel, K., Nyaku, M. A., Kheirbek, R. E., Bittner, V., Fonarow, G. C., Filippatos, G. S., Morgan, C. J., Aban, I. B., Mujib, M., Desai, R. V., Allman, R. M., White, M., Deedwania, P., Howard, G., Bonow, R. O., Fletcher, R. D., Aronow, W. S., & Ahmed, A. (2015). Risk of Heart Failure and Death After Prolonged Smoking Cessation: Role of Amount and Duration of Prior Smoking. Circulation. Heart Failure, 8(4), 694–701. https://doi.org/10.1161/CIRCHEARTFAILURE.114.001885.
2. American Heart Association. (2018, May 18). Access to Care. https://www.heart.org/en/get-involved/advocate/federal-priorities/access-to-care.
3. Bogers, Bemelmans, W. J. E., Hoogenveen, R. T., Boshuizen, H. C., Woodward, M., Knekt, P., van Dam, R. M., Hu, F. B., Visscher, T. L. S., Menotti, A., Thorpe, R. J., Jamrozik, K., Calling, S., Strand, B. H., & Shipley, M. J. (2007). Association of Overweight With Increased Risk of Coronary Heart Disease Partly Independent of Blood Pressure and Cholesterol Levels: A Meta-analysis of 21 Cohort Studies Including More Than 300,000 Persons. Archives of Internal Medicine (1960), 167(16), 1720–1728. https://doi.org/10.1001/archinte.167.16.1720.

4. Cai, Zhang, Y., & Li, M. (2020). Association between prediabetes and risk of all cause mortality and cardiovascular disease.(Research / Original Research / Updated Meta-Analysis). British Medical Journal, 370(8254), 101.
5. Centers for Disease Control and Prevention. (2014, July 31). About the Behavioral Risk Factor Surveillance System (BRFSS). https://www.cdc.gov/brfss/about/about_brfss.htm.
6. Centers for Disease Control and Prevention. (2017, June 1). NCHS Urban-Rural Classification Scheme for Counties. https://www.cdc.gov/nchs/data_access/urban_rural.htm#2013_Urban-Rural_Classification_Scheme_for_Counties.
7. Centers for Disease Control and Prevention. (2019, December 19). Know Your Risk for Heart Disease. https://www.cdc.gov/heartdisease/risk_factors.htm.
8. Centers for Disease Control and Prevention. (2021, January 11). Heart Attack Symptoms, Risk, and Recovery. https://www.cdc.gov/heartdisease/heart_attack.htm.
9. Centers for Disease Control and Prevention. (2021, July 7). Behavioral Risk Factor Surveillance System Overview: BRFSS 2020. https://www.cdc.gov/brfss/annual_data/2020/pdf/overview-2020-508.pdf.
10. Centers for Disease Control and Prevention. (2022, January 28). Heart Disease. https://www.cdc.gov/heartdisease/index.htm.
11. Centers for Disease Control and Prevention. (2022, February 7). Heart Disease Facts. https://www.cdc.gov/heartdisease/facts.htm.
12. Gunderson, Sun, B., Catov, J. M., Carnethon, M., Lewis, C. E., Allen, N. B., Sidney, S., Wellons, M., Rana, J. S., Hou, L., & Carr, J. J. (2021). Gestational Diabetes History and Glucose Tolerance After Pregnancy Associated With Coronary Artery Calcium in Women During Midlife: The CARDIA Study. Circulation (New York, N.Y.), 143(10), 974–987. https://doi.org/10.1161/CIRCULATIONAHA.120.047320.
13. Johns Hopkins Medicine. (2022). Alcohol and Heart Health: Separating Fact from Fiction. https://www.hopkinsmedicine.org/health/wellness-and-prevention/alcohol-and-heart-health-separating-fact-from-fiction.
14. Lumley, T. (2020). "survey: analysis of complex survey samples". R package version 4.0.
15. Lumley, T. (2010). Complex surveys: a guide to analysis using R. John Wiley.
16. Mayo Clinic. (2020, June 5). Coronary Artery Disease. https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613.
17. Mayo Clinic. (2022, April 7). Heart attack. https://www.mayoclinic.org/diseases-conditions/heart-attack/symptoms-causes/syc-20373106.
18. Napoli, Nicole. (2013, April 23). Binge Drinking in College Can Lead to Heart Disease Later in Life. American College of Cardiology. https://www.acc.org/about-acc/press-releases/2013/04/24/10/59/binge-drinking-study.
19. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
20. United Health Foundation. (2022). Annual Report: Frequent Mental Distress. https://www.americashealthrankings.org/explore/annual/measure/mental_distress.
21. Washington State Department of Health. (n.d.) Collecting Data. https://doh.wa.gov/data-statistical-reports/data-systems/behavioral-risk-factor-surveillance-system-brfss/brfss-collecting-data.
22. Washington State Department of Health. (2018). Relationship Between Behavioral Health and Chronic Diseases. https://waportal.org/sites/default/files/resources/Relationship-between-BH-and-CD_180508.pdf.
23. Wiesman, J., & Lofy, K. (2018). Washington State Health Assessment. Washington State Department of Health. https://doh.wa.gov/sites/default/files/legacy/Documents/1000//2018SHA_FullReport.pdf.