Group AA Milestone 6

2022-12-11

Problem statement

Smoking cigarettes can damage the body's organs, increase the risk of numerous illnesses, and generally harm people's health. The most common preventable cause of mortality in the U.S. is smoking cigarettes. Smoking is linked to numerous cancers, heart conditions, strokes, respiratory conditions such chronic obstructive pulmonary disease, diabetes, and other disorders. However, some smokers are more vulnerable to these negative effects than others. The question we explored is "what characteristics and behaviors among smokers in California lead to adverse health outcomes?"

At any age, even among heavy and longtime smokers, quitting smoking can have a significant positive impact on both short- and long-term health. For the prevention and cessation of smoking, a range of therapies are successful. With the use of data from the 2011 California Smokers' Cohort (CSC), the California Tobacco Surveys (CTS), we may better understand the tobacco use and smoking habits of Californian smokers and put initiatives into place in high-risk areas to encourage quitting.

Methods

In the 2011 CSC study, 1,000 smokers completed the follow up survey and were included in the data set for analysis. We cleaned the data for analysis of 5 selected health outcomes, as well as select demographic variables. Our data were originally split in two data sets, which we combined by merging on the patient ID variable. Once that was finalized we cleaned and analyzed the data by factoring all variables that were characters, combining all race variables into one column, accounting for missing variables, and calculating pack years. The study population did not have a very diverse racial/ethnic spread, so we decided to exclude race in our analysis. We decided to focus on health outcomes and the stratification by income level, and subset our data accordingly to include variables that would be of importance in our visualizations and analysis. We also created new variables to categorize income into low, medium, and high income, as well as one to categorize pack years to "average", "above average", and "below average" pack years.

Table 1: Pack Years and Health Outcomes Stratified by Above/Below Average Pack Years

	level	above average	average	below average	p	test
n		317	5	503		
Pack Years (mean (SD))		39.50 (16.86)	22.00 (0.00)	10.44 (6.10)	<0.001	
Asthma (%)	No	254 (80.1)	2 (40.0)	403 (80.1)	0.083	
	Yes	63 (19.9)	3 (60.0)	100 (19.9)		
Diabetes (%)	No	274 (87.3)	5 (100.0)	473 (94.0)	0.003	
	Yes	40 (12.7)	0 (0.0)	30 (6.0)		
Heart Disease (%)	No	274 (87.3)	5 (100.0)	473 (94.0)	0.003	
	Yes	40 (12.7)	0 (0.0)	30 (6.0)		
Other Mental Health (%)	No	255 (81.7)	4 (80.0)	406 (81.2)	0.979	
	Yes	57 (18.3)	1 (20.0)	94 (18.8)		

Visualizations

Visual 1: Table One of Pack Years and Health Outcomes

```
##                           Stratified by Above/Below Average Pack Years
##                            level above average average        below average
##   n                                  317             5            503
##   Pack Years (mean (SD))          39.50 (16.86) 22.00 (0.00)   10.44 (6.10)
##   Asthma (%)                No      254 (80.1)      2 ( 40.0)    403 (80.1)
##                             Yes      63 (19.9)      3 ( 60.0)    100 (19.9)
##   Diabetes (%)              No      274 (87.3)      5 (100.0)    473 (94.0)
##                             Yes      40 (12.7)      0 (  0.0)     30 ( 6.0)
##   Heart Disease (%)         No      274 (87.3)      5 (100.0)    473 (94.0)
##                             Yes      40 (12.7)      0 (  0.0)     30 ( 6.0)
##   Other Mental Health (%)   No      255 (81.7)      4 ( 80.0)    406 (81.2)
##                             Yes      57 (18.3)      1 ( 20.0)     94 (18.8)
##                           Stratified by Above/Below Average Pack Years
##                            p      test
##   n
##   Pack Years (mean (SD))  <0.001
##   Asthma (%)               0.083
##
##   Diabetes (%)             0.003
##
##   Heart Disease (%)        0.003
##
##   Other Mental Health (%)  0.979
##
```

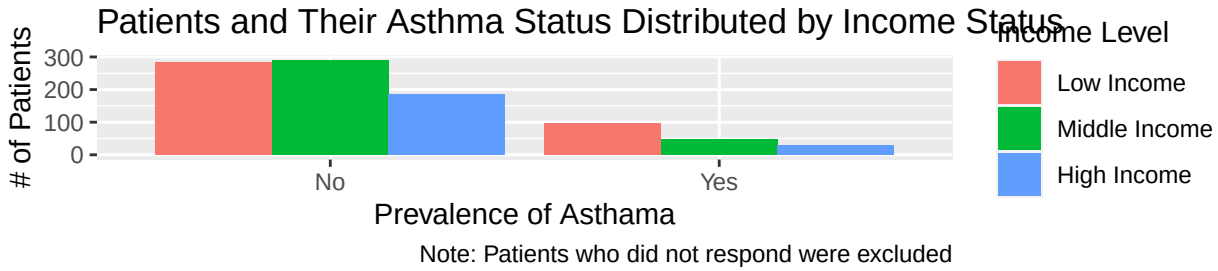
Cross Tab of Asthma Status x Income Level				
	Income Level			
Asthma_status	High Income	Middle Income	Low Income	Total
No	187	290	285	762
Yes	30	49	97	176
Total	217	339	382	938

Cross Tab of Heart Disease Status x Income Level					
		Income Level			
	Heartdis_status	High Income	Middle Income	Low Income	Total
2	No	202	319	337	858
3	Yes	14	20	43	77
4	Total	216	339	380	935

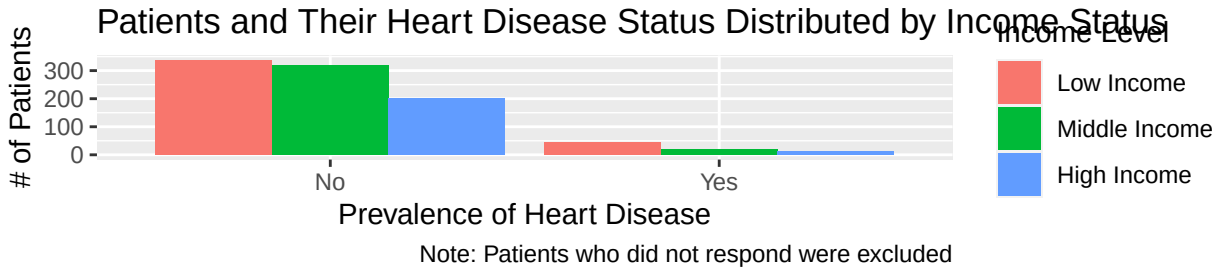
Cross Tab of Diabetes Status x Income Level					
		Income Level			
	Diabetes_status	High Income	Middle Income	Low Income	Total
2	No	202	319	337	858
3	Yes	14	20	43	77
4	Total	216	339	380	935

Visual 2: Outcome and Income Cross Tabulations

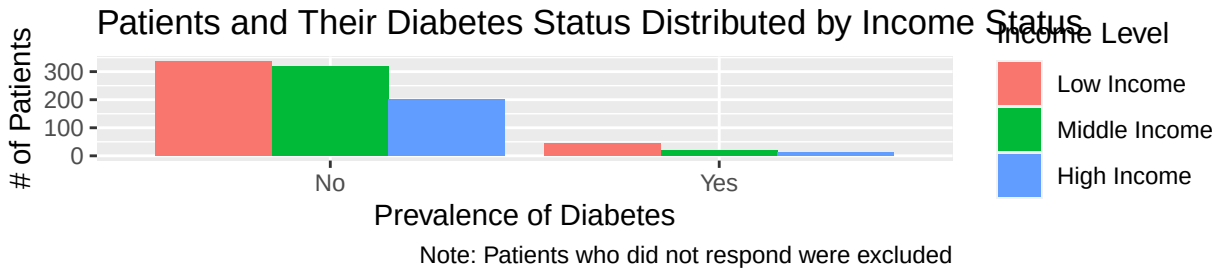
Visual 3: Of those who do have health outcomes, how are they distributed by income status. How are those that do not have health outcomes distributed by income? Prevalence of Asthma, Heart Disease and Diabetes Stratified by Income Status



Note: Patients who did not respond were excluded



Note: Patients who did not respond were excluded



Note: Patients who did not respond were excluded

Results

In our visualizations, when comparing all bar-graphs of health outcomes off asthma, heart disease, and diabetes stratified by income status, we can observe that the study population has a larger proportion of low-income patients, therefore we cannot assume that those with health outcomes are disproportionately low-income households. In patients who have heart disease, it is observed that they are the exact same patients that have diabetes, which we can interpret that heart disease and diabetes have some biological correlation.

In our outcomes table, those who were social smokers were about 75% of the study population, which may allude to a need for improved community public health intervention to reduce the prevalence on social smoking. Our outcomes table also includes prevalence percentages for the health outcomes of asthma, heart disease and other mental illnesses. If we wanted to investigate further, then we could you these percentages if we wanted to see if they are statistically significant enough for this study population.

In our distribution table, we can see that of those in the study population that did smoke, smoked an average of 13 cigarettes per day and have an average 21 pack years. Considering that 20-40 pack years are categorized as moderate smokers, these groups have higher odds in health outcomes and there could be effect modification on other confounders that we have not identified within the study.

Discussion

Based on our results, it seems that income is a considerable mediator for our health outcomes of interest related to smoking. The majority of people who reported adverse health outcomes were in the lowest income group, and the least amount of people reporting negative health outcomes were in the highest income group. For each negative health outcome, those in the lower income bracket were more likely to report the outcome than those in the middle income group. Those in the middle income group were more likely to report the negative health outcome than those in the highest income group.

Additionally, we found that a person reporting average, below average, or above average pack years does not strongly correlate with disease outcome. Those who reported above-average pack years were no more likely to report asthma (19.9%) or mental health issues (18.3%) than those with below average pack years (19.9% and 18.8%). However, those with above average pack years were more likely to report diabetes (12.7%) than those with below average pack years (6%) and more likely to report heart disease (12.7%) than those with below average pack years (6%).

Based on these findings, our recommendation for CDPH would be to prioritize quitting and smoking prevention campaigns and programs among lower income communities. We would not recommend campaigns to decrease smoking (and therefore pack years) since pack years did not seem to have as large of an effect on disease outcome. This recommendation for interventions targeted towards certain populations is aimed at attempting to reduce health disparities since based on our results, those in lower income groups are more likely to report negative health outcomes than those in higher income groups.