The Healthcare Costs of Smoking

Betsy Thorne

Advisor: Professor Rennhoff

Table of Contents

1. Introduction

With mounting healthcare costs in the United States, coupled with the declining health of the overall population, healthcare policy has become a hotly debated and widely divisive topic (Manchikanti, et. al. 2011). The passage of the Patient Protection and Affordable Care Act, more commonly known as "Obamacare," is the latest policy aimed at reforming a significant sector of the American economy. Due to the legislation, many more people have health insurance coverage, leading to increased concerns about who is actually paying for the healthcare. The latest debate includes the question – who should be paying for our healthcare? For those individuals who make risky health choices with full awareness of the consequences, and subsequently get sick – should healthy individuals shoulder the burden of healthcare costs for those individuals who make risky health choices with full awareness of the consequences?

Health insurance attempts to lower costs by pooling risk. The nature of pooling risk, though, means that some people are paying higher premiums than they would spend if their premium were based on their anticipated usage. On the other hand, others benefit because the amount of healthcare they consume is worth more than the premiums they pay. Healthy enrollees, who have lower healthcare costs than their insurance premiums, subsidize sick enrollees.

Historically, health insurance companies have tried to avoid these subsidies. This was done either by not covering high-risk consumers, or by charging higher risk consumers (including the elderly and those with pre-existing health conditions) much higher premiums. Due to the Patient Protection and Affordable Care Act, insurers are no longer permitted to deny anyone health insurance (referred to as guaranteed issue), as well as engage in these risk ratings, which tend to push those who have the highest healthcare costs out of the health insurance market. Guaranteed issue, along with community ratings – the idea that people in similar locations have access to similarly priced health insurance plans – make health insurance much more accessible to everyone in a given community. With the implementation of the Patient Protection and Affordable Care Act, adjusted community ratings are now used, requiring that providers offer most people in relative locations similar insurance premiums, with a few notable exceptions. These exceptions include groups that traditionally have substantially higher expected healthcare costs, such as the elderly population and those who smoke. This means that these few groups are not being offered similar healthcare premiums. Guaranteed issue, as well as adjusted community ratings, mean that the Patient Protection and Affordable Care Act increases the extent in which the healthy population is subsidizing the sick population.

With the increased number of individuals covered by health insurance and the increased amount of subsidization, it is natural to think about which risky behaviors are likely to increase costs for everyone. One of the most widespread and widely acknowledged risky behaviors is smoking. Not only is it widely known that smoking increases the risk of serious, life threatening diseases, but the American Cancer Society also estimated that between 2000 and 2004, tobacco-related healthcare costs were $96 billion – almost six times that of France, the country with the second highest healthcare costs related to tobacco. Because smoking is a voluntary act (in comparison to growing old, which is not), the debate as to who should pay these mounting costs is highly controversial.

Because smokers choose to engage in the risky act of smoking, causing increased probabilities of becoming severely ill, there is then a potential for healthy people to pay healthcare costs of these individuals who choose to take part in risky health decisions. By understanding the additional costs we are likely to incur, it will help us to assess the value of measures aimed at reducing smoking.

2. Research Question

This paper will estimate the difference in health expenditures associated with lung cancer, coronary heart disease, chronic obstructive lung disease and strokes between smokers and non-smokers. These specific four diseases are the focus because they are the most common and most serious diseases associated with smoking as determined by the Centers for Disease Control and Prevention (CDC 2014).

3. Literature Review

Lung Cancer

Lung cancer is the most widely studied and well-known disease associated with smoking. In 2003, it was estimated that 28% of all cancers reported in the United States were a type of lung cancer (Surgeon General 2004). While not all of these lung cancer diagnoses may be due to smoking, the Surgeon General's 2004 report shows strong causality between lung cancer and smoking and claims that a significant portion of these diagnoses can be attributed to the voluntary act of smoking. It was found that "cigarette smoking is by far the largest cause of lung cancer, and the worldwide epidemic of lung cancer is attributable largely to smoking" (Surgeon General 2004, p 43). In conclusion, the Surgeon General states that if a smoker chooses to smoke, there is a high probability of that individual contracting lung cancer.

What exactly are the probabilities of any person getting lung cancer if they smoke? Estimates depend on multiple variables, including "duration of smoking, average number of cigarettes smoked per day, duration of abstinence [from smoking], and age" (Memorial Sloan Kettering 2003). In a study done in order to determine probabilities, Villeneuve and Mao compared individuals who smoked routinely with those who never smoked. Using a life table method, they estimated that 172 out of 1,000 males who smoked routinely would develop lung cancer compared to 96 out of 1,000 males who never smoked. Similarly, they estimated that 116 out of 1,000 females who smoked would develop lung cancer compared to 43 out of 1,000 females who never smoked (Villeneuve and Mao 1994). In contrast, the United States Surgeon General's Report from 1990 compiled multiple cohort and case-control studies to find that "compared with the risk among never smokers, the risk of lung cancer for smokers may be increased twentyfold or more for heavy smokers" (Surgeon General 1990, p 107). The Surgeon General's 1990 report also reported differences in lung cancer risk among smokers after the cessation of smoking, relative to current smokers and current non-smokers. In summary, it was found that past smokers were anywhere from 10-800% more likely to be diagnosed with lung cancer than a non-smoker, but was also 20-90% less likely to be diagnosed with lung cancer compared to current smokers (Surgeon General 1990).

Coronary Heart Disease (CHD)

Researchers have found that cigarette smoking is a significant predictor of CHD in the United States for individuals of both sexes and that smoking could account for up to 30 percent of all deaths from coronary heart disease (CHD) (Surgeon General 1983). The Surgeon General's 1983 and 1990 reports rely on a mixture of cross-sectional and longitudinal studies in order to determine prevalence of CHD in smokers versus non-smokers. One study focused on autopsies of deceased individuals. This study found that current cigarette smokers had prevalence of CHD ranging from 11.7 to 23.4%, while non-smokers had a prevalence of 5.3%. Another study included in the Surgeon General's 1990 report followed more than 1 million men and women for 6 years. Out of the 12,724 male deaths due to CHD, it was estimated that 5,358 of these would not have taken place if smokers had the same death rates from CHD as that of non-smokers. Among females, 40% of CHD deaths could have been avoided had smokers have the same rates of death from CHD as non-smokers. An additional study included in the Surgeon General report "estimated that 24 percent of first major coronary events were cigarette related and independent of other risk factors" (Surgeon General 1983, p 65). It is believed that this number should be adjusted for younger individuals, who tend to have relatively fewer risk factors for CHD. Lastly, estimates find that up to 75 percent of CHD cases in young women can be credited to habitual cigarette smoking

(Surgeon General 1983, p 66). In all of the studies reported by the Surgeon General, there appears to be overwhelming evidence that smoking has a correlation with higher rates of CHD (1983).

Chronic Obstructive Lung Disease (COLD)

Strong positive relationships have also been made between chronic obstructive lung disease (COLD) and smoking (Surgeon General 2004).

In a study done in 1973 by Buist and Ross, 1,073 people were administered a single breath N_2 test (Surgeon General 1984). This test is commonly used in studies in order to measure the effectiveness of the respiratory system. Out of the 524 smokers tested, 64 percent had abnormal outputs in a number of categories such as lung volume and vital lung capacity, which are markers for COLD. Only 11 percent of those who reported being non-smokers had abnormal outputs. Among ex-smokers, 61 percent had irregular results. While abnormal outputs may not guarantee a COLD diagnosis, the two have a positive relationship (Surgeon General 1984). The Surgeon General's report concludes that 80 to 90 percent of the cases of COLD in the United States should be credited to smoking (1984).

It is also important to note that habitual and sustained smoking can lead to a higher likelihood of developing COLD (Surgeon General 1990). In a longitudinal study done in Denmark, researchers followed the smoking habits and respiratory health of 8,045 subjects, ranging in age from 30-60 years of age, over a 25-year span (Lokke, et. al. 2006). At each check-in, lung function was measured and smoking habits were recorded. At the end of 25 years, data was analyzed. Findings showed that 96 percent of men who never smoked had normal lung function, compared to 59 percent of habitual smokers (Lokke, et. al. 2006). In women, 91 percent of non-smokers had normal lung function and 69 percent of smokers had normal lung function. Rates of incidence of

COLD were 35.5 percent for regular smokers, and 7.8 percent for non-smokers. Overall, the authors estimated that the "absolute risk of developing [COLD] among continuous smokers is at least 25%" (Lokke, et. al. 2006).

Stroke

Lastly, strokes occur at seemingly higher proportions in smokers than non-smokers. Data has shown that in younger populations, especially women, cigarette smoking can lead to a higher risk of stroke (Surgeon General 1983). While this may be the case, there have been some longitudinal studies that have not found enough significant evidence to come to the conclusion that strokes are directly due to smoking.

Even though there are some inconclusive results, there has also been plenty of data and evidence to show that there indeed may be a positive relationship between strokes and smoking. In a retrospective study done by Molgaard and coworkers, with cases taken from multiple sources in southern California, associations with long-lasting habitual smoking were found to directly associate with strokes. Even when "cigarette smoking was dichotomously coded into categories of low and high lifetime exposure, consistent significant positive associations were found with strokes occurring in both bivariate and multivariate analysis when controlling for blood pressure" (Molgaard, et al.). Another study done by Kahn, followed subjects for 8.5 years and found that stroke mortality was 1.4 times greater in smokers than in non-smokers (Surgeon General 1983).

4. Methods

In order to determine the healthcare costs of a smoker in comparison of those for a non-smoker, I will find the difference in probabilities of smokers and non-smokers contracting each of the four diseases discussed. In order to estimate probabilities, I will be using a binomial logit model such that:

$$y_i^* = \mathbf{X_i}\beta + \varepsilon_i \quad \text{where } y_i = 1 \text{ when } y_i^* > 0 \text{ and } y_i = 0 \text{ when } y_i^* \leq 0$$

where X_i is a vector of characteristics of person *i*, ε_i is an error term for individual *i*, and y_i is a binary health indicator variable. $Y_i = 1$ if the individual has a positive diagnosis and $y_i = 0$ otherwise. I will evaluate the diagnoses of lung cancer, CHD, COLD and strokes separately. Y_i^* is a latent variable. To determine the probability of an individual becoming ill from this model, we know $P(y_i = 1) = P(y_i^* > 0) = P(\mathbf{X_i}\beta + \varepsilon_i > 0)$. Given the assumptions of the logit model, these probabilities can be expressed as $P(y_i = 1) = G(\mathbf{X_i}\beta) = \frac{e^{X_i\beta}}{1+e^{X_i\beta}}$ and $P(y_i = 0) = 1 - G(\mathbf{X_i}\beta) = \frac{1}{1+e^{X_i\beta}}$.

In order to estimate the β parameters, I will use maximum likelihood such that:

$$L = \prod_{i=1}^{n} G(X_i\beta)^{y_i}(1 - G(X_i\beta))^{1-y_i}$$

Using the logit model, I will estimate the marginal effect that smoking has on the probability of contracting each of the four diseases. I will then multiply the marginal effect by the cost of treating each of the four diseases in order to estimate the difference in the cost associated with lung cancer, CHD, COLD and strokes between smokers and non-smokers.

5. Data

This study uses data from the adult sample of the 2012 National Health Interview Survey (NHIS). The survey is conducted by the Centers for Disease Control and Prevention on an annual basis and surveys individuals across the United States about their health background and health behaviors. The adult sample of the survey randomly selects one adult from each family surveyed and collects information for that individual. While in the past, all adults in a family had an equal probability of being selected for the sample adult data, the survey now oversamples adults over 65 years old age, who are black, Hispanic or Asian. The survey covers basic demographic information as well as a diverse range of health topics that vary from year to year. This includes information on topics such as doctor visits, hospitalizations, health insurance coverage, and overall access to healthcare. Responses to the survey are confidential, and an individual's response is based off of personal interpretation.

Included in the NHIS is information on a broad variety of an individual's health diagnoses such as dementia, different types of cancer and blunt injuries. The four diseases of interest with regard to smoking – lung cancer, CHD, COLD and stroke – were all included in the survey. Included in my data *lung cancer*, *CHD*, *COLD*, and *stroke* are dummy variables indicating whether an individual has or has not been diagnosed by a medical professional with the specific disease.

An individual's propensity to develop one of these four diseases is likely linked to a number of factors beyond smoking. Therefore, I included other variables that were a part of the NHIS. *Male* indicates the individual's gender – either male or female. *Anxious* indicates whether the individual reported routinely feeling anxious over the past 12 months. *Married* specifies is if an individual is married or not and *white* denotes those who are considered Caucasian or white in race. *BMI* (body mass index) is an indicator of body shape, which is calculated by dividing an individual's mass (in kilograms) by squared height (in meters). *Age* indicates an individuals age in years, with the youngest potential respondent being 18, and the oldest being 85. The survey did not allow those who were over the age of 85 to indicate their specific age; they could only respond as 85+, so I made the assumption that all individuals that responded 85+ were 85. *Days5plus* specifies the number of days in the past year that an individual drank 5 or more alcoholic beverages in one day. Lastly, the variable *active* is a dummy variable that utilizes the answers to multiple survey questions. The NHIS has multiple questions regarding amount of exercise an individual participates in. This includes the frequency of strength activities, the frequency of light or moderate activities and the frequency of vigorous activities. All three of the frequencies are measured in times per week. Any responses that were not definitive, such as, "don't know" or "not sure" I considered to be no activity. In order to form the new variable *active*, I added together each of the three frequencies, letting total activity be the summation of how many times an individual engaged in vigorous, light/moderate or strength activities. Based off of recommendations from the CDC, I made the assumption that exercising three times a week is the least amount of exercise needed for significant beneficial health effects ("How Much Physical Activity Do Adults Need?"). Therefore, for the purpose of constructing the *active* variable, an individual was considered active (*active* = 1) if their total was 3 or greater, and inactive (*active* = 0) if it was less than 3.

The main explanatory variable of interest was *cigsday*, which reports the number of cigarettes a person smokes per day. This variable included only people who indicated in other survey questions that they are currently active smokers.

The full 2012 adult sample of the NHIS contains responses from 34,575 adults. Due to missing information, incomplete observations and erroneous data I was forced to delete 25,946 observations, lowering the total useful observations to 8,629. Any individual that did not answer any part of a variable included in my model was deleted. Examples of erroneous data that was deleted included individuals that responded with more than 365 days in which they drank more than five alcoholic beverages and those who reported having body mass indices greater than 75. Also included in erroneous data were individuals that said they did one type of activity more than 28 times a week. Lastly, if a person confirmed smoking behaviors, but did not offer the number of cigarettes they smoke a day, they were deleted from the dataset.

After cleaning the data I ran summary statistics, which can be found in Table 1 below.

Table 1 – Comparison of summary statistics for full sample and those diagnosed with lung cancer, CHD, COLD and stroke (means)

	Full Sample	Lung Cancer	CHD	COLD	Stroke
Male	0.5396	0.4762	0.7267	0.5014	0.5617
Married	0.4146	0.3333	0.4837	0.3816	0.3915
White	0.7848	0.9524	0.8134	0.8747	0.7404
Age	48.401	69.857	65.384	62.262	63.417
Anxious	24.09	23.81	25.38	40.67	34.47
BMI	27.709	27.889	28.913	28.061	27.219
Number of cigs/day (all smokers)	11.754	16.400	14.940	16.811	13.321
Number of cigs/day (entire population)	5.747	-	-	-	-
Number of days 5+ drinks, last year	17.410	2.951	16.082	18.274	17.339
Freq strength activity (times/week)	1.126	0.476	1.069	0.584	0.803
Freq light/moderate activity (times/week)	3.089	2.261	2.735	2.492	2.459
Freq vigorous activity (times/week)	1.846	2.333	1.332	0.828	0.972
Active	0.596	0.571	0.505	0.408	0.455

Total number of observations = 8,629 for no illness. Total number of observations for lung cancer = 21. Total number of observations for CHD = 461. Total number of observations for COLD = 359. Total number of observations for stroke = 235.

In order to check that the percentage of people who contracted each disease in my dataset was representative of the entire United States population, I compared them to actual prevalence rates of each disease in the United States. These can be found in Table 2. Most percentages are relatively equal, therefore meaning that my sample is representative of the overall American population.

Table 2 – Percentage of Individuals with Disease

	Actual	**Calculated**
Lung Cancer	0.12%[1]	0.21%
Coronary Heart Disease	6.0%[2]	5.56%
Chronic Obstructive Lung Disease	4-5%[3]	4.28%
Stroke	3.0%[4]	2.79%

Calculated values from dataset are weighted.

Some of the summary statistics were relatively unsurprising, with those who contracted any one disease having an average of more cigarettes smoked, less physical activity and more years of life. The average number of cigarettes smoked by people who claimed they smoked in the survey was about 11.75, while people who were diagnosed with any of the four diseases averaged from about 13 cigarettes per day for those diagnosed with stroke, up to 17 cigarettes per day for those diagnosed with COLD. The average age of people included in my data was 48, while the average age of people who were diagnosed with one of the diseases was in the range of 62-70. Finally, the

1 The American Lung Association has estimated that 399,431 Americans are living with lung cancer. Assuming that there are about 310 million Americans, the percentage of Americans living with lung cancer is 0.12.% (http://www.lung.org/lung-disease/lung-cancer/resources/facts-figures/lung-cancer-fact-sheet.html)

2 CDC - http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6040a1.htm

3 CDC - http://www.cdc.gov/copd/data.htm

4 AHA - http://circ.ahajournals.org/content/125/1/e2.full

statistics show that on average, people diagnosed with one of the four diseases exercised slightly less in all types of activities – strength, light/moderate and vigorous.

There were also a few parts of the summary statistics that were surprising. For example, the average number of days that someone drank more than five alcoholic beverages was about 17.5, while the average for someone diagnosed with lung cancer was less than 3. There is also a much higher percentage of men that get CHD as compared to women. Of the 461 individuals that were positively diagnosed with CHD, approximately 73% of them were men. Body mass indices were relatively equal for all categories, staying between 27 and 29. About a quarter of the entire population responded that they experience anxiety. For those diagnosed with COLD, about 4 out of 10 said they experienced routine anxiety. Lastly, the mean of the dummy variable, *active*, for the overall population was 0.596, which means that more than one half of the population is participating in some form of exercising 3+ times a week. This was surprising to me, as I would have expected this number to be lower. For those diagnosed with one of the four diseases though, the mean for *active* was always lower. Individuals diagnosed with COLD had the lowest *active* mean, at 0.408. Those with lung cancer had the highest mean, 0.571. Overall, I was intrigued by how high all of the means were.

Overall, it seems that one of the largest difference in summary statistics appears to be cigarettes smoked per day. Therefore, it should show some effect on the probability of an individual contracting one of the four diseases.

Finally, I ran the logistic regression using dummy variables as dependent variables that represent whether the individual was diagnosed with lung cancer, CHD, COLD or a stroke. A weighted variable was included in the NHIS dataset, and included in my model, in order to make sure that the results were representative of the entire population.

6. Results

I estimated the parameters in STATA using maximum likelihood. Because logit parameters do not have a direct interpretation, I follow the common convention and report the marginal effects of each variable. These marginal effects are presented in Table 3 below.

Table 3 – Marginal Changes with Standard Errors with Pseudo Log-Likelihoods

	Lung Cancer	**Coronary Heart Disease**	**Chronic Obstructive Lung Disease**	**Stroke**
cigsday	0.00000956 (0.00001)	0.0001865 (0.00017)	0.0008837*** (0.00014)	0.0003233*** (0.00011)
bmi	0.000023 (0.00002)	0.0011284*** (0.00024)	0.0004368* (0.0026)	-0.0003716* (0.0022)
male†	-0.0000528 (0.00014)	0.019367*** (0.00317)	-0.0035436 (0.00304)	0.001513 (0.00268)
white†	0.0002771 (0.00022)	-0.0019934 (0.00395)	0.009001*** (0.00324)	-0.0105993*** (0.00393)
married†	-0.0000801 (0.0017)	0.0003297 (0.00272)	-0.0035399 (0.00285)	0.0001643 (0.00233)
active†	0.0001635 (0.00014)	-0.0051326* (0.00291)	-0.0100747*** (0.00332)	-0.0065616*** (0.00242)
anxious†	0.0001416 (0.0025)	0.0108037*** (0.00408)	0.0239099*** (0.00517)	0.0127712*** (0.00382)
days5plus	-0.0000142* (0.00001)	-0.00000349 (0.00002)	0.0000227 (0.00002)	0.00000223 (0.00002)
age	0.0000267 (0.00002)	0.0017407*** (0.00012)	0.0012965*** (0.0001)	0.0008462*** (0.00008)
N	8590	8590	8590	8590
Pseudo Log-Likelihood	-775293.51	-9633022.5	-8180519.5	-5938337.9

† Marginal change is for discrete change of dummy variable from 0 to 1.
** Significant at the 90% confidence level.*
*** Significant at the 95% confidence level.*
**** Significant at the 99% confidence level.*

The main variable that was used to determine differences between smokers and non-smokers was the *cigsday* variable. As shown by the output, results were highly statistically significant for stroke. Therefore, it can be concluded that the number of cigarettes a person smokes per day has significant effect on the probability that he or she will be diagnosed with a stroke. Results for lung cancer and CHD were not statistically significant. Lung cancer may not have had significant results because of the small sample of people within the data that had been diagnosed. While the percentage of people diagnosed with lung cancer within the sample was close enough to the overall population, only 21 observations out of the 8,629 adults in the sample reported that they were diagnosed with lung cancer. With so few individuals diagnosed with lung cancer in the sample, it is difficult to draw any strong conclusions related to lung cancer. CHD was likely not statistically significant because of other major factors that can contribute to a person being diagnosed. The 1983 Surgeon General Report determined that factors such as diet and exercise were also major contributing factors to the diagnosis of CHD. Overall, marginal effects for each of the diseases appeared small but all were positive. For lung cancer, the marginal effect was 0.00000956 percentage points, 0.00019 percentage points for CHD, 0.00088 percentage points for COLD, and 0.00032 percentage points for a stroke. This means, for example, that by increasing the number of cigarettes smoked per day by one unit, the probability of an individual getting lung cancer becomes increases by 0.00000956 percentage points. This means that an increase in cigarettes smoked per day will lead to an individual have the greatest vulnerability to being diagnosed with COLD, followed by stroke. No concrete conclusions can be drawn about the impact of *cigsday* on the vulnerability of being diagnosed with CHD and lung cancer due to the lack of statistical significance in my data.

Other variables also had marginal effects that were statistically significant. Body mass index, which has long been used as an indicator of healthy weight ranges, had positive correlations with the diagnoses of CHD and COLD, meaning that an increase in a person's BMI leads to a higher probability of contracting one of the two diseases. On the other hand, BMI had a negative relationship with the diagnosis of strokes – meaning that an increase in BMI leads to a decrease in the probability of having a stroke. This is slightly intriguing, as an increase in BMI tends to be considered a risk for many health conditions.

The next variable, *male*, has a positive statistically significant marginal effect for CHD, denoting that men are more likely to be positively diagnosed with CHD than women are. This was one of the largest marginal effects in the data, which was not expected. It is possible that this variable is confounded with others – for example, maybe males smoke more than females and instead of the gender being the main factor behind increased diagnoses of CHD, the increased smoking is the driving factor.

W*hite* had a positive marginal effect on the diagnosis of COLD, but a negative marginal effect on the diagnosis of stroke. Therefore, a Caucasian individual has a higher chance of being positively diagnosed with COLD, in comparison to an African-American individual, and vice versa for strokes.

Activity levels, *active*, have negative correlations with the probabilities of positive diagnoses, which indicate that a less active person is more likely to get sick. These marginal effects were statistically significant for CHD, COLD and stroke. This is a relatively intuitive result, as the benefits of physical activity tend to be well known, and are recognized to lower the probability of developing a variety of different health problems. Activity levels had the largest marginal effect on the diagnosis of COLD, and the smallest marginal effect on the diagnosis of CHD.

Next, anxiety levels, *anxious*, interestingly had highly significant marginal effects on the probability of being diagnosed with CHD, COLD and stroke. All were positive relationships, meaning that the more an individual perceives being anxious, the more likely he or she is to get any one of these three diseases. Anxiety has the largest marginal effect on the probability of diagnosis of COLD.

Days5plus was only significant for the probability of being diagnosed with lung cancer. This was intriguing, as I would have expected excessive alcohol intake to have significant marginal effects on the diagnoses of other diseases as well. Even more interesting is that the correlation is negative, meaning the more times per year a person drinks 5+ alcoholic beverages, the less likely that person is to be diagnosed with lung cancer. The marginal effect is small, but was significant at the 90% confidence level. I would attribute this surprising result more than anything to the small number of people, 21, within my data that had been positively diagnosed with lung cancer.

Lastly, *age*, had positive relationships with the probability of diagnoses of CHD, COLD, and stroke. This is not unexpected, and is also reflected in the summary statistics, as the average age of people diagnosed with these conditions is much greater.

In order to better illustrate the effect that smoking has on one's propensity to get ill, I use the parameter estimates to compare the probability of contracting each disease as a function of cigarette consumption. In finding these probabilities, I will assume two individuals are identical with the exception of the number of cigarettes smoked per day. Summary statistics from the data showed the mean number of cigarettes smoked per day among those who smoke was between 11 and 12. Because of this, I will choose to assume that those who smoke cigarettes are smoking about 11 cigarettes per day. This is approximately half a pack of cigarettes. I also calculate the same probabilities for someone that smokes 1 pack of cigarettes (20 cigarettes) per day. All other characteristics are equal to their sample mean. Any differences are, therefore, driven solely by smoking behavior. These probabilities are reported in Table 4 below.

Table 4 – Probability of contracting disease given number of cigarettes smoked per day

	Lung Cancer	**Coronary Heart Disease**	**Chronic Obstructive Lung Disease**	**Stroke**
0 cigarettes/day	0.0002249	0.0200973	0.0147279	0.0112367
11 cigarettes/day	0.0147823	0.0221455	0.0244360	0.0147823
20 cigarettes/day	0.0184854	0.0239718	0.0368058	0.0184854

Table 5 shows the percentage change in the probability of contracting one of the four diseases that results from an individual increasing smoking from 0 cigarettes per day to 11 cigarettes per day and 0 cigarettes per day to 20 cigarettes per day.

Table 5 – Percent increase in probability of contracting disease given change in cigarettes smoked per day from 0

	Lung Cancer	**Coronary Heart Disease**	**Chronic Obstructive Lung Disease**	**Stroke**
11 cigarettes/day	46.65%	10.19%	65.92%	31.55%
20 cigarettes/day	100.59%	19.28%	149.91%	64.51%

Results show that, as expected, increasing the number of cigarettes per day will have a positive effect on the probability that the smoker will contract each of the four diseases. While the levels are rather small in Table 4, the percentage changes in Table 5 are quite dramatic. The largest effect is with COLD, as someone who smokes 11 cigarettes per day is about 67% more likely to be diagnosed with COLD as compared to someone who does not smoke at all. CHD seems to be the disease least affected by smoking, as someone who smokes 11 cigarettes a day is only about 10% more likely to be diagnosed as compared to a non-smoker. Not surprisingly, increasing the number of cigarettes smoked leads to increased probability of diagnoses of each of the diseases. Someone who smokes 20 cigarettes per day is about 150% more likely to be diagnosed with COLD, 100% more likely to be diagnosed with lung cancer, and 20% more likely to be diagnosed with CHD as compared to a non-smoker.

In order to come up with medical costs for treating the four medical conditions, I used the chronic disease cost calculator, as compiled by the Center for Disease Control and Prevention. Table 6 shows estimates of the costs of treating each of the four conditions in the United States.

Table 6 – Estimated medical costs for treating each of four diseases

	Estimated Cost per person
Cancer	$11,140
Coronary Heart Disease	$7,690
Chronic Obstructive Lung Disease	$2,090
Stroke	$16,600

I was unable to find a cost directly for lung cancer, although that CDC has estimated that costs are somewhat similar for most types of cancer, therefore the estimate for cancer in general seems reasonable to use for lung cancer. These estimated costs are medical costs that combine all direct payouts, including from government programs, such as Medicare and Medicaid, as well as private insurance payments. These costs include "expenditures for office based visits, hospital outpatient visits, emergency room visits, inpatient hospital stays, dental visits, home health care, vision aids, other medical supplies and equipment, prescription medicines, and nursing homes" (CDC). Lastly, it should be noted that the CDC expects significant increases in medical costs for each of these diseases in the future.

In conclusion, based off of calculated probabilities and estimated medical costs, Table 7 below summarizes the expected differences in medical costs for the four diseases between a smoker and a non-smoker. Probabilities of being positively diagnosed with each disease given the number of cigarettes smoked per day were used from Table 4, while medical costs were taken from Table 6.

Table 7 – Expected healthcare costs (in U.S. dollars) per person for treating each disease depending on number of cigarettes smoked per day

	Lung Cancer	**Coronary Heart Disease**	**Chronic Obstructive Lung Disease**	**Stroke**
0 cigarettes/day	$2.51	$154.55	$30.78	$186.53
11 cigarettes/day	$164.67	$170.30	$51.07	$245.39
20 cigarettes/day	$205.93	$184.34	$76.92	$306.86

Results show that smoking increases the expected medical costs of an individual. This was not unexpected. Because the probability that a non-smoker will be diagnosed with lung cancer is so low, the expected healthcare costs for lung cancer in a non-smoker was about $2.50. This expected healthcare cost due to lung cancer for a non-smoker increases by more than 60 times to $164.67 for an individual who smokes an average of 11 cigarettes per day, showing that clearly the number of cigarettes smoked has an impact on the expected healthcare costs related to lung cancer. The expected healthcare costs least affected by smoking was CHD. This is shown by the small range of the expected healthcare costs – $155-$185. This is likely because many other factors have been linked to a positive diagnosis of CHD. In general, expected healthcare costs were overall very low for COLD because the cost of treating it is relatively low compared to the other diseases. Lastly, strokes had the highest overall expected healthcare costs (ranging from $185 to about $300) due to high healthcare costs, but also a high impact of the number of cigarettes smoked on the probability of diagnosis.

In order to fully capture the scope of expected costs for each of these diseases, I multiplied the expected healthcare costs per person by the number of people living in the United States. These results can be found in Table 8, shown in millions of dollars.

Table 8 – Total expected healthcare costs (in millions of U.S. dollars) for treating each disease depending on number of cigarettes smoked per day

	Lung Cancer	**Coronary Heart Disease**	**Chronic Obstructive Lung Disease**	**Stroke**
0 cigarettes/day	$778.10	$47,910.50	$9,541.80	$57,824.30
11 cigarettes/day	$51,047.70	$52,793.00	$15,831.70	$76,070.90
20 cigarettes/day	$63,838.30	$57,145.40	$23,845.20	$95,126.60

As shown, total expected healthcare costs for each of these diseases are very large. Even though without smoking, there is still a financial burden, it is obvious that increasing the number of cigarettes smoked greatly increases the overall expected healthcare costs for each of the four diseases.

7. Discussion/Conclusions

This paper investigates the differences in expected healthcare costs between smokers and non-smokers due to lung cancer, coronary heart disease, chronic obstructive lung disease and stroke. With the use of the newest release of the National Health Interview Survey, as well as the data from the CDC's chronic disease cost calculator, I have found that the expected healthcare costs associated with an average smoker are indeed greater than that of a non-smoker. In controlling for multiple other variables that could lead to the diagnosis of lung cancer, CHD, COLD or a stroke, I found that the probability of a smoker being positively diagnosed with one of the four diseases increases, in varying degrees, with the number of cigarettes smoked. Because smoking leads to higher probabilities of becoming sick, smokers therefore have greater expected healthcare costs for each of these four diseases than a non-smoker. The disease with the largest relative difference in expected healthcare costs between smokers and non-smokers was COLD. On the other hand, expected healthcare costs for CHD due to smoking were relatively the smallest.

As healthcare costs in the United States continue to be a growing concern, the economic burden of a voluntary activity, which affects overall health, must fall somewhere. While, for reasons of simplicity, my model does not estimate the expected healthcare costs due to smoking in their entirety, the scope of the four diseases included are enough to show substantial differences in costs between smokers and non-smokers. Health policy must answer as to whom these costs should fall to. Should we subsidize individuals voluntarily engaging in a risky health behavior? As findings have shown, as the number of cigarettes a person smokes increases, the probability that the individual will get sick, and therefore have higher expected healthcare costs, increases. Does this mean that healthcare insurers should charge premiums to smokers based on how many cigarettes they smoke?

Further research into this topic could prove beneficial. Improving survey techniques could lead to better estimates of the probability of a person becoming sick. Newer data could provide new estimates for the costs of treating specific diseases. A very important drawback to this study includes assumed fixed costs for treating healthcare conditions. Healthcare clearly does not operate with fixed costs, but for the simplicity of my model, this assumption was necessary. Deeper investigation could examine differences in healthcare costs depending on the source of the payout – the patient, the government, or a private health insurance provider.

Investigations could also include looking at lifetime healthcare costs for smokers compared to non-smokers. Because smokers are more likely to get sick, are they more likely to die earlier in life and therefore have similar lifetime healthcare costs as non-smokers? With changes in healthcare policy, the manner in which healthcare is consumed and paid for has the potential to change in the coming years. New and updated data could provide for meaningful results.

References

Centers for Disease Control and Prevention. “Health Effects of Cigarette Smoking.” (2014).

"How Much Physical Activity Do Adults Need?" *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 03 Mar. 2014. Web.

Lokke, A., P. Lange, H. Scharling, P. Fabricius, and J. Vestbo. "Developing COPD: A 25 Year Follow up Study of the General Population." *Thorax* (2006): 935-39.*Thorax*. British Thoracic Society, 22 May 2006.

Manchikanti, Laxmaiah, M.D., David Caraway, M.D., Allan T. Parr, M.D., Bert Fellows, M.A., and Joshua Hirsch, M.D. "Patient Protection and Affordable Care Act of 2010: Reforming the Health Care Reform for the New Decade." *Pain Physician* (2011): E35-67.

"Risk for Lung Cancer Varies Among Smokers." *Memorial Sloan Kettering Cancer Center*. Memorial Sloan Kettering Cancer Center, 19 Mar. 2003.

The Association between Cerebrovascular Disease and Smoking: A Case-Control Study Molgaard C.A., Bartok A., Peddecord K.M., Rothrock J. Divisions of Epidemiology and Biostatistics and Health Services Administration, Graduate School of Public Health, San Diego State University. Department of Neurology, University of California at San Diego, Calif., USA Neuroepidemiology 1986;5:88–94.

"Tobacco: The True Cost of Smoking." *American Cancer Society*. (2014). American Cancer Society.

U.S. Department of Health and Human Services. *The Health Consequences of Smoking – Cardiovascular Disease: A Report of the Surgeon General*. Rockville, MD. U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health, 1983.

U.S. Department of Health and Human Services. *The Health Consequences of Smoking – Chronic Obstructive Lung Disease: A Report of the Surgeon General*. Rockvillr, MD. U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health. 1984.

U.S. Department of Health and Human Services. *The Health Benefts of Smoking Cessation*. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. DHHS Publication No. (CDC) 90-8416, 1990.

U.S. Department of Health and Human Services. *The Health Consequences of Smoking: A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004.

Villeneuve, PJ, and Y. Mao. "Lifetime Probability of Developing Lung Cancer, by Smoking Status, Canada." *National Center for Biotechnology Information*. U.S. National Library of Medicine. 1994 Nov-Dec;85(6):385-8.