Predicting Medical Charges

CONNOR DIGGINS, Marquette University

ACM Reference Format:
Connor Diggins. 2019. Predicting Medical Charges. 1, 1 (December 2019), 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Insurance companies in the United States are legally allowed to discriminate their medical charges based off factors that contribute to added risks in health [8]. This project seeks to figure out which factors determine these charges and speculate on why this may be to decide whether the discrimination truly is due to added health risks. The potential influences looked at are age, gender, region, children, smoking, and body mass index (BMI). The data consists of 1,338 people. The ages range from 18-64 and were spread out fairly evenly. Gender was split almost evenly as 51% of people in the study were males. Regions were split up by southwest, southeast, northwest, and northeast instead of the traditional separating of United States territories that includes a Midwest section. Children represents the number of children covered on one's health insurance or dependents. Smoking is whether or not someone officially discloses that they smoke to their insurance provider. BMI is a ratio of height to weight (kg/m^2) that is used to provide an understanding of someone's weight that takes into consideration how tall they are. Someone that is 6'2 150 pounds has a much lower BMI than someone that is 5'2 150 pounds even though they have the same weight. BMI has a few commonly accepted categories. A BMI below 18.5 is considered underweight, 18.5 - 25 is normal/healthy, 25 - 30 is overweight, and more than 30 is considered obese. There are also three classes of obesity; Class 1 which is a BMI from 30-35, Class 2 which is a BMI from 35 to 40, and Class 3, considered severely obese, which is a BMI greater than 40 [2]. To put into some perspective, a BMI under 15 or over 55 is very uncommon. Before looking at any of the data, there were some ideas of what would seem fair to charge someone more for. It would make sense that someone out of the healthy zone for BMI would pay more for medical costs because they would be at a higher risk of health complications. The same goes for smoking as it is widely advertised in the United States that smoking causes health issues. Growing older also tends to have more complications with health. Having children would mean more people that could potentially have a health issue, so that could also be a factor.

2 LITERATURE REVIEW

BMI, smoking, and age all have data on how they affect the physical health of people in the United States. Gender and region of United States resided in did not have any strong evidence linking them with health issues. Number of dependents does not mean someone is more likely to have any

Author's address: Connor Diggins, Marquette University.

XXXX-XXXX/2019/12-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Fig. 1. Chicago Bears Linebacker, Khalil Mack, during the first half of an NFL football game in Chicago, Nov. 10, 2019], via The Dallas Morning News

health faults but intuitively more people on your insurance, the more possibility for one of those people to have an issue.

2.1 BMI

BMI is not a perfect science. Since it only takes into account height and weight, those that are muscular or pregnant can be falsely seen as unhealthy. Pregnant women have a higher weight since they have a fetus inside of them that counts towards their weight when looking at a scale. The problem with muscular people is that body fat is not taken into account in the measurement so you could think a healthy person is really overweight or even obese. In a similar sense, those that are in the healthy section for BMI could really not be healthy at all if they have a high body fat or are not very active [4]. For example, let's look at All-Pro NFL Linebacker, Khalil Mack. Khalil Mack is 6'3" 247 lbs, which puts him at a BMI of 30.9. This falls under the obese category, but looking at Fig 1, we can see he is in incredible physical condition. His body fat is so low, and he is so muscular, that a simple measurement of his weight to height ratio can not truly convey his physical health because there is no distinction between fat and muscle in the weight. However, there are documented health complications that arise with those that are not a healthy weight. Type 2 diabetes, high blood pressure, heart disease, sleep apnea, osteoarthritis, fatty liver disease, and kidney disease have all been linked to obesity [7]. A 2016 study found that a little under half of those in the overweight category of BMI had a healthy cardiometabolic profile, which includes cholesterol, blood pressure, and blood sugar. This is a bit lower than the two-thirds of those in the healthy BMI category that had a healthy cardiometabolic profile [4]. There are also health complications that arise from being underweight. Some of these include malnutrition, vitamin deficiencies, anemia, osteoporosis, decreased immune functions, and fertility issues (in women) [6]. The increase in health risks of those outside of the normal BMI category could potentially be used as justification for charging those with an unhealthy BMI more money. If this was done, it should be expected that both overweight and underweight people get charged extra being that there are plenty of health risks involved for both. A 2014 study also found that while the risks of some diseases increase with BMI, those with a higher BMI tend to live a bit longer on average [9]. This further complicates the use of BMI in determining insurance charges since those in the obese category do not seem to pass away early, so the increase in diseases does not necessarily line up with a higher risk of early death.

2.2 Smoking

As opposed to BMI, smoking is a very simple measurement: one either smokes or they do not. Smoking is much less controversial than BMI as there are many known health complications that

arise from smoking, but there is no conflicting data that could possibly suggest that smoking is beneficial to physical health in any sense. Smoking causes about 90% of all lung cancer deaths and about 80% of all deaths from chronic obstructive pulmonary disease (COPD). Estimates show smoking increases the risk of coronary heart disease and strokes by 2-4 times, and can cause cancer almost anywhere in the body including bladder, cervix, esophagus, liver, pancreas, and stomach. The risks of smoking are almost endless, and you can find information on smoking affecting almost any part of the body [1]. This makes smoking seem like the most likely candidate for being a component of any potential increases in insurance charges.

2.3 Age

There are many health concerns with someone as they get older. Arthritis, heart disease, osteoporosis, Alzheimer's, pneumonia, and influenza are some complications that are more likely the older you are [5]. This is quite conclusive and intuitive as since everyone dies, the older you get, eventually you have to have something go wrong physically. However, since people are unable to control their age, it could be questionable if age plays a significant factor in medical charges.

3 RESEARCH QUESTIONS

1. Do insurance companies discriminate against overweight/obese people? This is not about whether it is moral to make people with an increase in risk of health complications pay more, but whether everyone with a discrepancy in BMI has to pay more. If those that are overweight are charged more, those that are underweight should also see an increase in charges since they both have added health risks.

2. How do the affects of smoking compare to other factors when predicting insurance charges? From the literature review, smoking has the highest effect on physical health and is a choice, so it would be expected to play the biggest role in determining health insurance cost.

3. Is BMI valued too highly by insurance companies? Based on research, BMI is not an exact science for determining physical health, so using it too much when determining insurance costs may be unfair.

4. Is there discrimination against age even though it can not be controlled? While there are health risks with being older, no one can just stop aging like someone can stop smoking (even though it can be difficult to stop smoking).

4 METHODS

There was not a lot of cleaning to do as the data was clean to begin with. This is known because of three reasons:

1) The rows represent one of the 1,338 observations.

2. The columns represent a variable (age, gender, smoking, etc.).

3. Each type of observational unit forms a table; all of the cells represent data that is from a specific observation and a unique variable.

However, there was a little manipulation done with the data. When looking at the variable for children, the original data used a number for how many children were under the observation's

insurance (min 0, max 5). This made visualizing the data confusing and not as simple to analyze. To resolve this, the children column was changed to a Boolean where someone either had dependents or did not. This made visualizing the data look much cleaner and easy to understand.

5 RESULTS

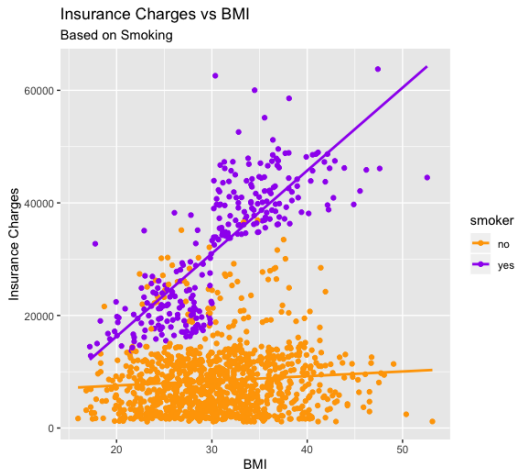

Fig. 2

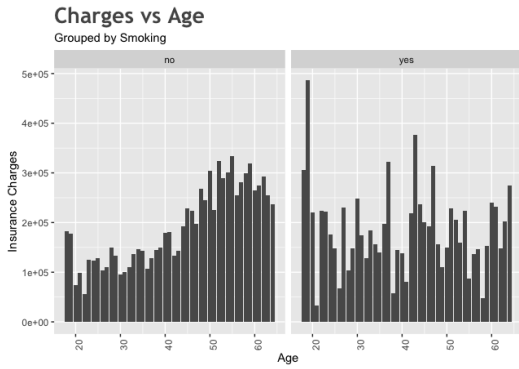

Fig. 3

Taking a look at Fig. 2, there seems to be an increase in insurance charges as BMI increases. If you take a closer look, insurance charges look heavily influenced by BMI only if one smokes. Let's take two linear regressions where y is Insurance Charges and x is BMI:

Smokers: $y = -9050 + 1381x$

Non-Smokers: $y = 7600 + 48x$

Smokers' slope of 1381 is interpreted as every increase of 1 BMI would increase charges by 1,381

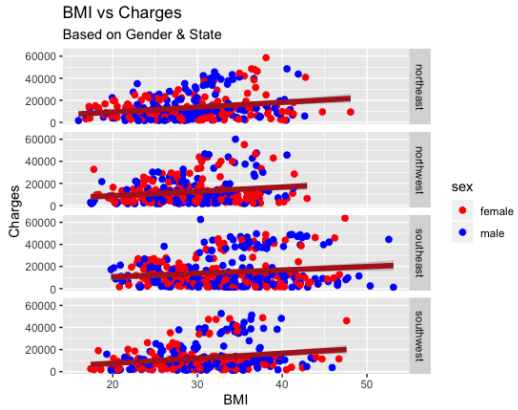

Fig. 4

dollars. Non-Smokers' slope of 48 is interpreted as every increase of 1 BMI would increase charges by 48 dollars. This shows how BMI plays a much larger role in smokers than non-smokers. Also, the linear regression equations' slopes are both positive, showing an increase in BMI means increase in charges but more interestingly, a decrease in BMI means a decrease in charges. Someone that is in the underweight category of BMI, regardless of whether or not they smoke, would be predicted to have lower charges than someone in the normal category of BMI. This does not look to be a flaw in the regression equations as Fig 2 shows data points with very low charges for the very low BMIs.

Fig 3 shows insurance charges based on age and whether someone smokes. From this graph, we see a steady increase in insurance charges when age increases for those that do not smoke. For those that do smoke, there is no type of pattern in the graph that would suggest age plays a factor. This shows how large of an impact smoking has on amount paid and shows how age also plays a factor. When someone smokes, the extra charges are so much it makes it difficult to even see any type of pattern with the age. At the same time, when someone does not smoke, you can see that there is a correlation between age and insurance charges.

Fig 4 shows insurance charges based on BMI, but this time it is grouped by gender and region. Insurance charges increase slightly with BMI for all four regions, and gender is scattered quite randomly throughout all four regions, suggesting that both region and gender do not play a significant role in determining insurance charges.

The p-values for each factor shows the probability of the data under the assumption that the component does not play a role in determining insurance costs. Traditional alpha values, thresholds for whether a p-value is small enough to be considered significant, are 0.01 and 0.05. The p-values for the variables are as follow:

BMI, Age, and Smoking: less than $2 * 10^{-16}$
Children: 0.00297
Gender: 0.70456

This backs up our graphs as BMI, age, and smoking all play a significant role in determining charges while children do as well (but not as much), and gender does not play a significant role.

6 DISCUSSION

We saw from Fig 2 that BMI plays a much larger role when smoking is also present. This could be a clever way at solving the problem high muscle, low fat people like Khalil Mack introduce. If someone takes care of their body so much where their BMI is so high because of all of the muscle, it is intuitively less likely that they would smoke since they care so much about their body, and smoking is so detrimental to physical health. Conversely, if someone does smoke, you would think that it is less likely they take care of their body enough to have that high of a BMI because putting a ton of effort into your health but also smoking is an oxymoron. In the literature review, we saw plenty of health concerns for both overweight and underweight people. However, Fig 2 shows that underweight people are actually getting charged less than normal BMI people, while overweight people get charged more than normal BMI people. This does not seem to be a mistake either due to lack of data because the graph shows points that are very low BMI and very low insurance charges. It is debatable on whether insurance companies should be allowed to discriminate based on weight. However, if they discriminate against overweight people, they should do the same with underweight people because they both have plenty of health concerns associated with them. Fig 3 and the p-value shows how age plays a factor, especially when smoking is not involved. This means that insurance companies discriminate against something that while it raises concerns for health, is out of the control of people. No one can control aging like you can control smoking, yet it is still a large factor in determining costs. Fig 4 shows that there does not seem to be any increases in insurance costs based off of region or gender, which intuitively makes sense and was predicted. Based on how BMI is a much larger factor with smoking and how age does not even look like a factor when controlled for smoking, smoking seems to be the largest predictor of insurance costs, which makes sense based on the literature review and how you can control whether or not you smoke.

7 SUMMARY

The three largest factor when determining insurance charges are smoking, BMI, and age. Smoking is the greatest, which seems fair due to all of the complications that can come with it and how it is a choice to smoke. BMI perhaps plays too large of a role due to the incompleteness of using a weight to height ratio to measure physical health, but using it more so with smokers makes it seem much more fair. Age is a big factor, especially when smoking is not involved, because it causes many health complications although it can not be controlled. Gender and region did not play a significant factor, as expected. Insurance companies are legally allowed to discriminate if it is based off of something that increases likelihood of physical health issues, and it is important to see what factors they use and whether or not they are valid in using them.

8 CITATIONS

[8],[1],[4], [3], [6], [7], [9], [2], [5].

REFERENCES

[1] Centers for Disease Control and Prevention. 2017. Defining Adult Overweight and Obesity. (11 April 2017). https://www.cdc.gov/obesity/adult/defining.html

[2] Centers for Disease Control and Prevention. 2019. Health Effects of Cigarette Smoking. (17 Jan 2019). https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/index.htm

[3] Carl J. Lavie MD. 2014. Obesity and Cardiovascular Diseases: Implications Regarding Fitness, Fatness, and Severity in the Obesity Paradox. (2014). https://www.sciencedirect.com/science/article/pii/S0735109714003349

[4] Robert H. Shmerling MD. 2016. How useful is the body mass index (BMI)? (30 March 2016). https://www.health.harvard.edu/blog/how-useful-is-the-body-mass-index-bmi-201603309339
[5] Judy Meleliat. 2018. 11 Common Aging Health Issues. (21 Sep 2018). https://www.aegisliving.com/resource-center/11-common-aging-health-issues/
[6] Janna Young MPH. 2017. 6 Health Risks of Being Underweight. (17 April 2017). https://www.healthline.com/health/underweight-health-risks
[7] National Institute of Diabetes, Digestive, and Kidney Diseases. 2015. Health Risks of Being Overweight. (Feb 2015). https://www.niddk.nih.gov/health-information/weight-management/health-risks-overweight#problems
[8] Kyle D. Logue Ronen Avraham and Daniel Benjamin Schwarz. 2013. Understanding Insurance Anti-Discrimination Laws. *University of Michigan Law School Scholarship Repository* 22 (Jan. 2013), 4–11. https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1163&context=law_econ_current
[9] AJ Tomiyama. 2016. Misclassification of cardiometabolic health when using body mass index categories in NHANES 2005–2012. (4 Feb 2016). https://www.nature.com/articles/ijo201617