**Impact of Race on Preventable Deaths in the United States**

Peyton Mosbaugh

**Executive Summary:**

The purpose of this study is to see race’s role in preventable deaths. Furthermore, are Black Americans at a higher risk of dying from a preventable death than White Americans? If so, could a lack of healthcare and a higher poverty level be a potential cause of that? Health risk factors such as obesity and tobacco-use will be examined for each race as well as socio-economic factors such as being uninsured and unemployed. Preventable deaths for each race will also be examined and the following are those that the study will focus on: HIV-related mortalities, cancer mortalities, deaths due to influenza, heart disease mortalities, diabetes mortalities, homicides and suicides. This helps to show if the races are putting themselves at risk with the risk factors, if they are put at a disadvantage due to socio-economic factors, and if there really are significant differences in preventable deaths between these two races. If it is found that Black Americans are at a higher risk and that they are put at a disadvantage due to socio-economic factors, then maybe new efforts can be put forth to help diminish the health care gap between races.

**Background:**

Still in today’s society, there are many inequalities that put American citizens at a disadvantage due to their gender, race, ethnicity, or religion. Unfortunately, the American Healthcare System is a place where these gaps occur. Due to inequalities in the country’s healthcare system, there has been a disproportionate impact for people of color, causing gaps in health insurance coverage, uneven access to services, and poorer health outcomes. More than ninety percent of twenty-something Black Americans have their college degree yet are still paid significantly less for the same job as their White American co-workers. This income gap, of course, has its own ramifications. With healthcare being a very pricey expense for many Americans, this inequality in pay makes it harder for Black Americans to get high-quality health insurance. Although more have health insurance and longer life expectancies than several decades ago, Black Americans are still suffering from illness and deaths at much higher rates than other races in America. Although there are many different factors that play into healthcare disparities, the bottom line is that Black Americans are not receiving the same quality of care as their white counterparts. As a result of this, they are substantially more likely to die from early health issues that are at least partially treatable or preventable than White Americans. In fact, the preventable death rate for Black citizens is double that of White citizens. Even though recent legislature like the Affordable Care Act has given more access to affordable healthcare to uninsured Americans, the United States still have a very long way to go in reducing these health disparities.

Genetics, lifestyle, and risk factors do, of course, play a large role in premature deaths. In some cases, the healthcare system just cannot simply save a person from passing unexpectedly or earlier than anticipated. However, in the case of preventable deaths, this is not necessarily the case. There is an abundant amount of screenings and testing that can be done as preventable measures in routine check-ups to catch things like diabetes, heart diseases, certain cancers, and even suicidal thoughts or depression. A lack of healthcare accessibility and coverage prevents Americans from partaking in these preventive tests. Two-thirds of the difference between Black Americans and White Americans are currently due to preventable or curable deaths. Some of these preventable deaths include: diabetes, the flu, certain cancers, infectious diseases (especially sexually-transmitted diseases), and even suicide. The objective of this project is to look into potential risk factors such as obesity and tobacco use to see if the number of people putting themselves at a heightened risk differs between the two races. Differences in socio-economic factors such as being uninsured or unemployed is also examined between the two groups. The number of deaths that could have been prevented is also analyzed to see if indeed there is a gap between Black Americans and White Americans. The goal is to use data to potentially highlight these disparities; and as a result, possibly the data can be used to push forward policy changes that could help to eliminate the racial gap in America’s healthcare system.

**Methods:**

The CRISP-DM was used in order to analyze the data. The first step of the process is business understanding, or properly defining the problem and objectives of the project, relative to the business at hand. Then there is the data understanding phase, where the types of data that are necessary for the project are determined and are then collected. Next, the data is prepared by cleaning and integrating the data. Then, statistical calculations are performed on the data. After this step, an evaluation of the modeling and analysis is done to ensure that nothing has been overlooked, including any important key factors. Then the findings can be published and used in the future.

As mentioned earlier, the Affordable Care Act has been a tremendous push in the right direction in terms of making healthcare more accessible and affordable to those who need it. To dive deeper in an understanding of the Affordable Care Act, or ACA, specifically, it allows states to decide whether or not they want to opt into a Medicaid expansion. Although this sounds like a win in helping get care for citizens, twelve states have still not passed or opted into this Medicaid expansion: Wyoming, Texas, South Dakota, Wisconsin, Mississippi, Tennessee, Alabama, Georgia, North Carolina, South Carolina, Kansas, and Florida. A handful of these states are located in the South, that has a higher proportion of Black Americans than the national average. If these states decided to pass the Medicaid expansion that is a part of the ACA, then this would be a big push in helping the racial disparities in healthcare. The business goal of this project is to find data that highlights the disparities to show these states that the ACA could help save lives, especially Black Americans.

In order to perform this analysis, the proper data was needed. Data from the CDC was used for the project, since they have very up-to-date and accurate numbers being the nation’s health protection agency. In addition, datasets for unemployment were found from the United States Bureau of Labor. The timeframe for the datasets is 2005-2017. Risk factors such as obesity and tobacco use were looked at, as these could be causes for preventable deaths that people often bring upon themselves and put themselves at a higher risk. Socio-economic factors of being uninsured and unemployed also were looked at, as these impact a person’s ability to get healthcare and preventative screenings. Then, of course, the statistics on the number of deaths due to the preventable causes of diabetes, cancer, HIV, heart disease, flu, homicide, and suicide for both races were also found. All of these datasets were collected as individual datasets and then the sliced data that was of primary interest were compiled into CSV files. One CSV file contained all of the data for Black Americans; another CSV file contained the data for White Americans; and the third and final CSV file contained all of the data for the project.

Once the CSV files were created, they were implemented into a Python program, where the data analysis began. The first step taken in the data understanding phase of the project was to create boxplots for each of the variables for both the Black American and White American datasets. The separate CSV files were used for this part of the project in case the numbers differed drastically between the two races, as this step was just to check for any outliers in the data. After this, histograms were created for each of the variables within the dataset containing all of the data for both races. This allowed for the data to be grouped by race, so that it could be compared how the values of the two races differ from one another.

In terms of data preparation and cleaning for the project, it had to be ensured that all of the data was in a consistent format between the two races and among the variables. The data also had to be checked for any missing values, which would be filled in with average values. Here, irrelevant observations were removed. This is also the phase in the project where further actions were taken with any outliers that were found.

When performing statistical analysis on the data, correlation matrices were first created for the overall data, the Black American data, and the White American data. A correlation matrix shows the correlations between all of the numeric variables within the dataset. This will help in pointing out if the risk factors of obesity and being a smoker are highly correlated to any of these preventable diseases. The correlations between socio-economic factors of being uninsured and unemployed and the preventable deaths were also looked at through the correlation matrices.

**Results:**

While creating the boxplots of the variables to find any outliers, four in total were found, two for each race. The two outliers with the Black American data were for the diabetes and suicide variables whereas the White American outliers were for flu-related disease and diabetes. Because the outliers did not line up between the two datasets, the outliers were left unchanged in the data. One of the main processes for this project was comparing the data between the two races over time, and if the outliers were removed or changed, this task would be quite difficult.

When looking at the histograms for each of the variables to compare the data between the two races, the trends are quite interesting. For unemployment, the percentage of unemployed Black Americans is significantly higher than for White Americans over the time frame of 2005-2017. On average, the percent of unemployed Black Americans is close to double the percent of unemployed White Americans. That can be seen in the following histogram:

Chart, bar chart

Description automatically generated

The percent of uninsured Black Americans over the years was also higher than the percent of uninsured White Americans. Although the difference in those percentages was not as drastic as that of unemployment, there is still a consistent difference over these thirteen years. The following is a histogram showing that difference:

Chart, bar chart

Description automatically generated

Looking into the risk factors of obesity and tobacco-use are very important because these are two somewhat “controllable” factors that people can be aware of to lower their risk of diseases. Obviously, this is not the case and not controllable for all people, but generally speaking. The health risk factors had much different results than the socio-economic factors. In terms of tobacco-users, the percentage of smoking White Americans was higher than that of Black Americans from 2005-2013. From 2014-2017, the percentage of smoking Black Americans was slightly higher or about the same as the percent of White Americans who use tobacco. Five to ten percent more of White Americans were obese from 2005-2017 than Black Americans. Obesity in white citizens was consistently much higher. Here it can be seen that Black Americans were at a higher risk of preventable deaths due to socio-economic factors, but White Americans were at a higher risk because of health factors.

All of the preventable death data was represented in the number of deaths per 100,000 deaths. The histograms comparing death rates for HIV, flu-related, cancer, heart disease, and diabetes all showed very similar trends in that more Black American deaths were due to these preventable deaths than those of White Americans. This was the case for each year included in this project. In terms of diabetes deaths, Black American deaths were almost twice as high as those of White American deaths due to this health condition. Deaths due to homicide followed a similar trend, but was nearly four times higher for Black Americans than White Americans. Deaths from suicide was the only preventable death that was higher for White Americans and was close to three times higher than for Black Americans.

Looking at the correlation matrix for each race assists in understanding which risk factors are correlated to the preventable deaths. For each race, the correlations to the health risk factors of obesity and tobacco-use and the deaths as well as the correlations between being uninsured or unemployed and the deaths were closely examined. The correlation between White uninsured Americans and White Americans deaths due to cancer, with a correlation coefficient of 0.78, was very strong. A similar strong correlation was seen with Black Americans and a correlation coefficient of 0.74. There was also a very strong correlation between White Americans who used tobacco and those who were uninsured, with a coefficient of 0.89. Surprisingly, unemployment did not show any strong correlations with any other variable from either dataset. The percentage of White American tobacco-users was highly correlated to cancer (0.93) and HIV deaths (0.83), as well as being obese (0.89). For Black Americans, the correlation of tobacco users was highly correlated to diabetes (0.90), heart disease (0.88), cancer (0.91), flu-related (0.80), and HIV deaths (0.90). There were several more strong correlations for Black tobacco-users than White. Although obesity in Black Americans did not have any strong correlations within the dataset, this was not the case for obese White Americans, which had a correlation coefficient of 0.89 with heart disease deaths, 0.93 with cancer deaths, 0.93 with flu deaths, and 0.9 with HIV deaths.

**Discussion:**

An important thing to note with this project relates to the term preventable death. It is known that not all instances of cancer or diabetes deaths could have been cured or prevented, because even the world of healthcare has its limitations. However, the term preventable simply is referring to conditions that have screenings, tests, and other preventative measures to help catch these conditions early and make swift initiatives to treat them. Health insurance is so vital and important in being able to afford these preventative measures, and so that is why being uninsured can really put a person at risk of dying from these preventable or somewhat preventable diseases. With the results of this data, it was expected that more Black Americans would have more deaths from these conditions than White Americans, based on the idea of healthcare disparities in the United States that were touched on in the background of this project. It was also decided to exclude homicides from the findings of this project in terms of healthcare. It is not really in the realm of healthcare to prevent homicides, other than working to save lives once they happen. Although it is a preventable death, it is not as deeply rooted in healthcare as the other preventable deaths and since this project was centered on the inequalities in healthcare, it was decided that the homicide data was not relevant. One surprising aspect of the findings was that there were higher percentages of White Americans who put themselves at a heightened risk by smoking or being obese, yet they had less deaths from the preventable conditions, aside from suicide, than Black Americans. Based on the background, it is known that being obese and being a smoker drastically increase a person’s risk of developing most of these conditions and dying from them, especially since these two factors weaken a person’s immune system. With this knowledge and seeing that these two factors dramatically increase risk of disease, it shows that there has to be another factor playing into why more Black Americans are dying despite being less obese and smoking less. It was also expected that more Black Americans would be unemployed and uninsured than White Americans, since so much of the background for this project touched on this issue. Although there was a consistent difference in the percentage of Black Americans uninsured compared to White Americans, the difference on average was within two percent. Considering how strong the trend is of Black Americans dying from these preventable diseases and the fact that they smoke less and are less obese, one would think that this divide in healthcare coverage would be greater. Moving forward in order to further explain this difference, the quality of the health care and health insurance should be looked at. Because unemployment differs so greatly between the two races, it could be impacting the quality and coverage of health insurance that they are able to afford. Going along with this, having high deductibles with insurance can also prevent someone from going to receive preventative screenings and tests since most of it would most likely have to be paid out of pocket. Since there is a great divide in unemployment between the two races, this might be a way to dive deeper and further explain why there is this racial inequality in today’s healthcare and how more affordable health insurance and health care could aid in fixing it and saving lives.

**References:**

Bureau, U. (2020, September 16). Income, Poverty and Health Insurance Coverage in the United States. Retrieved December 06, 2020, from https://www.census.gov/newsroom/press-releases/2020/income-poverty.html

Byrnes, H. (2020, May 07). This is the racial divide among cancer deaths in every state. Retrieved December 06, 2020, from https://www.usatoday.com/story/money/2020/05/07/the-racial-divide-in-cancer-deaths-in-every-state-and-dc/111637122/

Goldman, D. (2020, November 16). Obesity, second to smoking as the most preventable cause of US deaths, needs new approaches. Retrieved December 06, 2020, from https://theconversation.com/obesity-second-to-smoking-as-the-most-preventable-cause-of-us-deaths-needs-new-approaches-129317

Homepage - Health, United States - Products. (2019, October 30). Retrieved December 18, 2020, from https://www.cdc.gov/nchs/hus/index.htm

Maxwell, C., & Carratala, S. (2020). Health Disparities by Race and Ethnicity. Retrieved December 06, 2020, from https://www.americanprogress.org/issues/race/reports/2020/05/07/484742/health-disparities-race-ethnicity/

Radley, D. (2019, August 06). Diving into the Data: How Many Americans Die Prematurely from Treatable Causes? Retrieved December 06, 2020, from https://www.commonwealthfund.org/blog/2019/americans-die-prematurely-treatable-causes

STDs Continue to Rise in the U.S. Press Release. (2019, October 08). Retrieved December 06, 2020, from https://www.cdc.gov/nchhstp/newsroom/2019/2018-STD-surveillance-report-press-release.html

Unemployment Rate by Race. (n.d.). Retrieved December 18, 2020, from https://www.macrotrends.net/2508/unemployment-rate-by-race

Vital Signs: Preventable Deaths from Heart Disease and Stroke|DHDSP|CDC. (2019, March 13). Retrieved December 06, 2020, from https://www.cdc.gov/dhdsp/vital\_signs.htm

Young, C. L. (2020, February 20). There are clear, race-based inequalities in health insurance and health outcomes. Retrieved December 06, 2020, from https://www.brookings.edu/blog/usc-brookings-schaeffer-on-health-policy/2020/02/19/there-are-clear-race-based-inequalities-in-health-insurance-and-health-outcomes/

**Ten Possible Presentation Questions:**

1. What other factors do you think could potentially be impacting these healthcare inequalities?
   1. I believe that the quality of care and the quality of affordable insurance also plays a huge role in these inequalities. Like I mentioned, if a policy has an extremely high deductible in order to keep the monthly costs down, then it might only cover part or no part of a preventative screening, which are vital in early detection. If these screenings are not done early, then the condition may escalate to being too far gone for assistance and cures. Also, if a doctor is not of the highest quality, then they may miss something and not catch something that they should have with a patient.
2. Why did you only decide to look at Black and White Americans when there are many other races?
   1. In the future, I would love to look into the other races that make America the melting pot. However, in the short amount of time that I had for this project, I decided to focus on the two most populous races.
3. Why do you think that there are more Black Americans unemployed than White Americans?
   1. Although I would have to do further research on this topic, during my background study for the project I found that Black Americans are more likely to be arrested and incarcerated during their lifetime than White Americans. Whether this is from systemic racism or not, the bottom line is that being incarcerated or having a criminal record can severely impair someone’s possibilities of employment.
4. What would be the next steps for this project in your opinion?
   1. Moving forward, I would love to begin looking at the quality of health insurance and care and how this differs between races. I would also want to expand this to look into disparities with other races as well.
5. What other preventable deaths do you think would be beneficial to look at in the future?
   1. I think it would be beneficial to look into pregnancy-related mortalities as this is a common one in terms of preventative medicine and deaths. I also believe it would be beneficial to look into more specific cancers, as obviously some have better outlooks than others.
6. Why did you choose the deaths that you did?
   1. I chose the included preventable deaths because they were among some of the top deaths for Americans and were also listed as having preventive medicines associated with them. Many articles stated that despite taking so many lives each year, that there were things that people and doctors could do to drastically lower the numbers.
7. What other risk factors other than smoking and obesity do you think could impact preventable deaths?
   1. I think also looking into family history, other health conditions, as well as possibly alcohol abuse could be helpful in looking further into preventable deaths.
8. Why do you think suicide was higher in White Americans than Black Americans?
   1. I believe that there are many different factors that could play into it, but in my brief research for this project, I came across the idea that played into the idea that actually reflected back to the poverty divide. One cause could be that White Americans are surrounded by more intelligent and wealthier peers, putting a great pressure on them. I would have to do more research into this to get a better and more accurate answer, however.
9. Why do you think there was the shift in smoking habits in 2014?
   1. One theory I have for this shift has to do with the use of e-cigarettes. I have heard many first-hand accounts of people beginning to use e-cigarettes or “vaps”, but the users do not believe that it is a tobacco product because it is synthetic. This might be an issue of misclassification, but again, I would have to do more research on the topic to be sure.
10. What would you want to see from policy makers?
    1. I think one thing that could help these inequalities in healthcare would be more affordable and higher quality health insurance and care. It is one thing to make health insurance more affordable, but with this lower cost, we also need to make sure that care and insurance is of a good quality.