**Project Proposal**

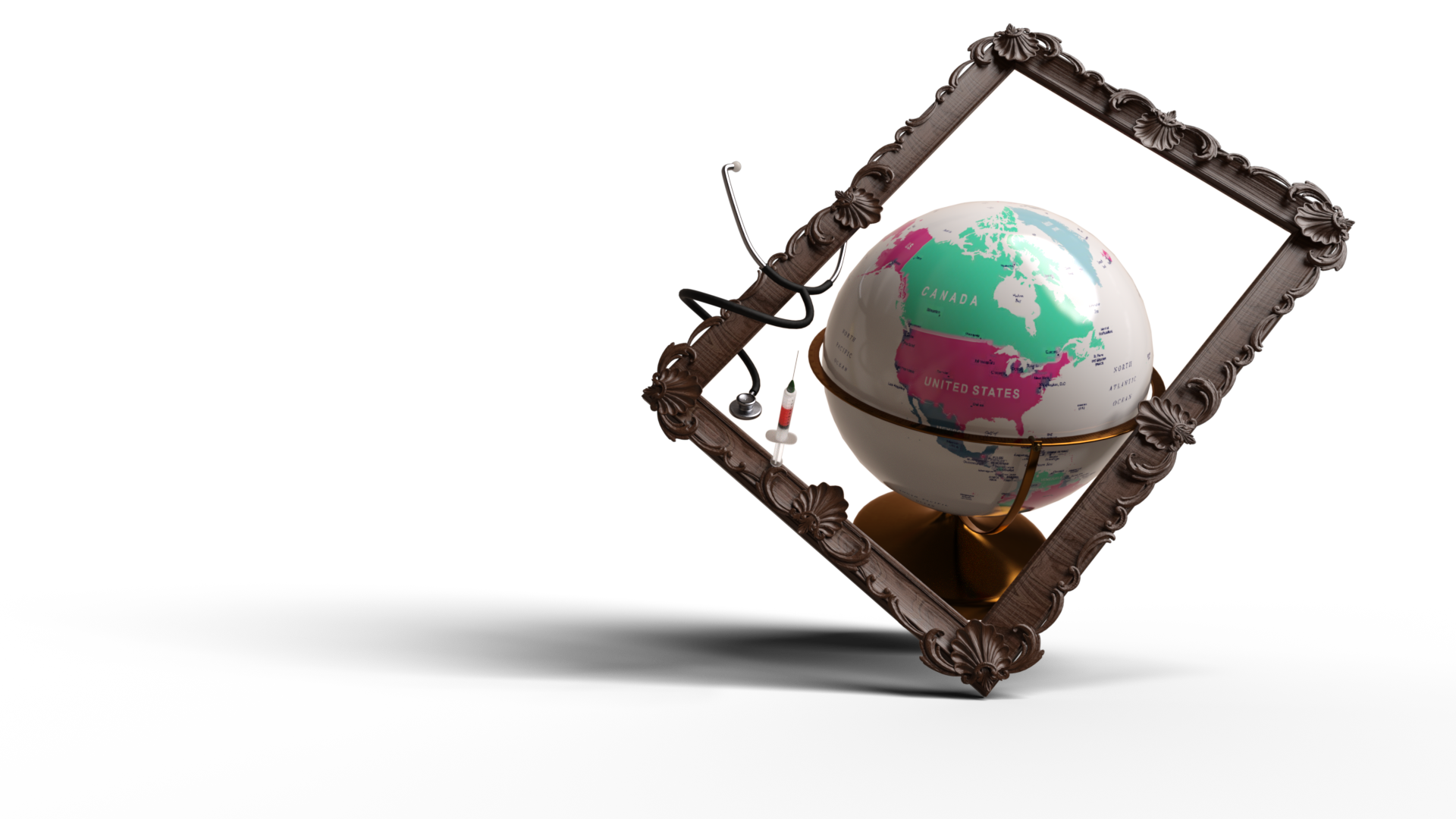
Group 5

Zaid Abouqamar

Shakil Asfaque

April Mattinson

Gary Murray

Hari Sankaran Nair

**Table of Content**

[Abstract 1](#_Toc44720701)

[Introduction 1](#_Toc44720702)

[Competitive Analysis 2](#_Toc44720703)

[Proposed Plan 4](#_Toc44720704)

[Business Problem 4](#_Toc44720705)

[Data Set 4](#_Toc44720706)

[Planned Analysis/Visualization 4](#_Toc44720707)

[Data Cleansing 6](#_Toc44720708)

[Cost/Benefit Analysis 7](#_Toc44720709)

[Conclusion 7](#_Toc44720710)

[Citations 8](#_Toc44720711)



# **Abstract**

Healthcare is expensive in the USA, for both governments and individuals. Roughly one third of the working age population has a medical debt. Chronic conditions like heart and lung disease are the most costly to society - both financially and in terms of human suffering - and tend to disproportionately affect lower income households. Charity is one way to help minimize medical costs, and charitable giving is big business. Frequently donors have insufficient information about the impact of their giving on society. Often, they are attracted by big budget marketing campaigns and ‘sexy’ causes. This has resulted in certain conditions, like breast and prostate cancers, attracting roughly $22.22 in charitable giving per person affected, while COPD attracts only $0.13 per person affected.

In this study we will analyze the root causes of these very costly chronic conditions. Using correlation, multiple regression and various visualization techniques, we seek to understand what preventative measures can be used, and what unhealthy behaviours can be avoided in order to minimize the incidence of costly conditions in the USA. Once the drivers of health outcomes are understood, this can help donors who wish to maximize the ‘societal impact’ of their giving, to determine what types of programs and charities to donate to.

# **Introduction**

The United States spent roughly 3.25 trillion dollars on healthcare in 2017 [1]. This includes money from government, private health insurers and individuals. Given that many individuals in the US either have no health insurance or insurance where coverage is inadequate, many forgo critical preventative procedures, leading to future health issues. Many individuals go into debt in order to pay for life saving procedures such as surgery and hospital care. A 2020 study indicates that 32% of the American working population has a medical debt [2], and the effects are even more devastating in lower income households [3]. As a result, many charities exist to support Americans with insufficient income and medical care needs.

The two most expensive diseases to the healthcare system include cardiovascular issues like coronary heart disease and stroke and smoking and environmentally related issues such as COPD and asthma [4]. These diseases also tend to disproportionately affect people in lower income brackets [5], further highlighting the importance of preventative care and lifestyle choices that support better health.

In this investigation we will use data from the CDC and Robert Wood Johnson Foundation’s ‘500 Cities Project’. This project collected data from a large set of individuals in almost 500 cities across the USA. Data collected included 14 independent variables (9 health related preventative procedures and 5 unhealthy behaviours) and 13 dependent variables or health outcomes including the diseases discussed above. The data measures the prevalence of each of these 27 variables as a percentage of the population across the almost 500 cities. Using several statistical and visualization techniques we will study which independent variables have the largest impact on health outcomes. The results of the study may be used to provide guidance to individuals looking to donate to a health-related cause that will maximize the number of people positively affected by their spend.

# **Competitive Analysis**

Charitable giving is big business in the United States. In 2017 Americans gave $427.71 Billion in charitable donations, of which 9% (approximately $42.6 Billion) went to health care related charities [6]. When making donations, individuals and corporations are influenced by several factors. These include amongst other factors:

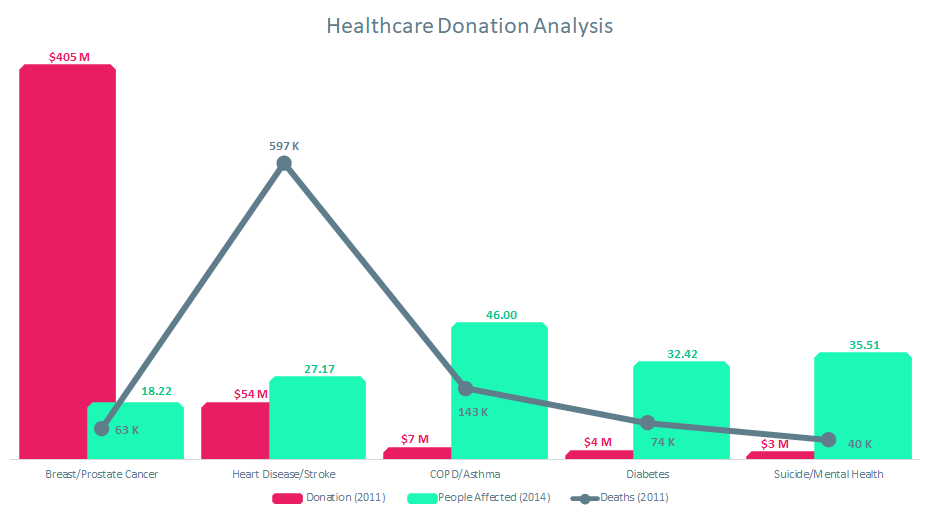
● What cause is currently popular?

● Is there a personal or business connection with the charity?

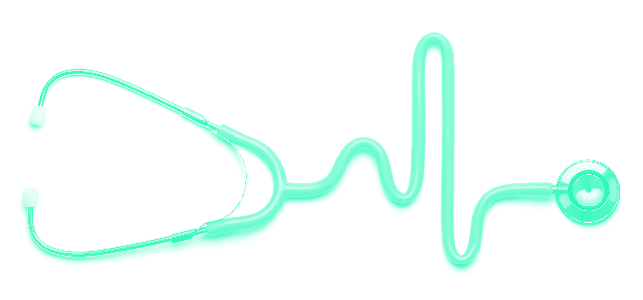
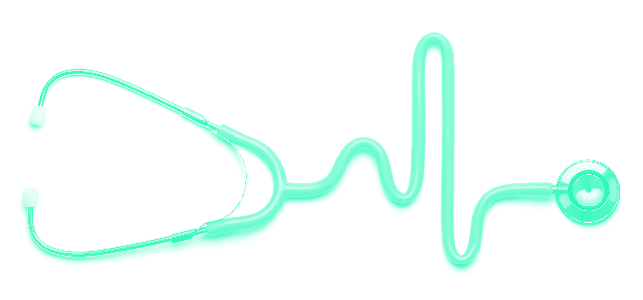
● Is the charity well known and easily accessible?

● What does the individual or corporation receive in return for their donation?

While charitable giving is a personal choice; donating based on the factors above does not always lead to the most efficient allocation of donation dollars. As the chart below shows, money donated to charity does not always match the cost to society of these conditions. Donations to breast and prostate cancer charities far exceeds giving to COPD/Asthma, despite these being the most widespread [7] and costly [8] conditions in the USA. Deaths from heart and stroke disease exceed all other causes in the US and yet donations to charities that support heart and stroke amount to only 13% of the donations raised for just two forms of cancer - breast and prostate. (Refer to Figure 1)



**Figure 1: Healthcare Donation Analysis**



This disparity is magnified when viewed from the perspective of dollars donated per person affected. Per the chart below, charities benefitting people with COPD and Asthma received only $.13 in charitable spending per person affected. The comparable figure for breast and prostate cancers was $22.22. (Refer to Figure 2)



**Figure 2: Donated Per Person Affected**

# 

# 

# 

# **Proposed Plan**

### **Business Problem**

The goal of our study is to understand how various unhealthy behaviours plus the use of preventative services contribute to various health outcomes in a cross-section of US cities. This information can be used to help donors make decisions on efficient allocation of charitable giving if their goal is to reduce the societal cost of disease.

### **Data Set**

The CDC collected high-quality, small-area epidemiologic data for 27 measures in almost 500 U.S. cities. This covered a total population of over 103 million or 33.4% of the U.S. population. Population measures in each city were based on the 2010 U.S. Census. [9]

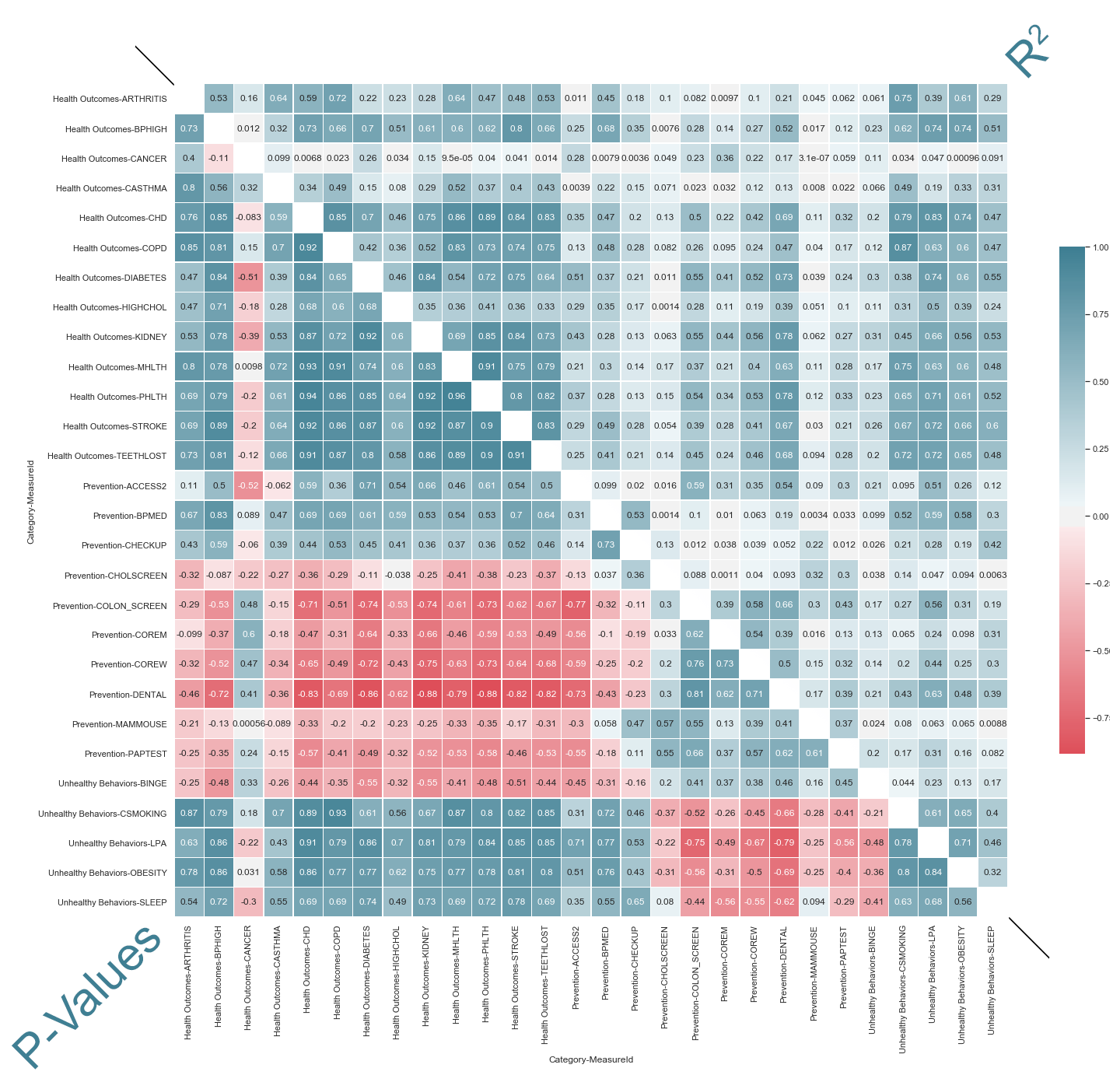
Participants were asked to answer a “Yes” or “No” question to measure annual prevalence for the major categories: Unhealthy Behaviours, Prevention Methods and Health Outcomes. Projection to the population was accurate with a 95% confidence interval. Annual prevalence was measured with demographic characteristics, when feasible, for adults aging greater than or equal to 18 years old (except for teeth loss - 65 years and older).[10]

The CDC dataset provided two types of measures: crude prevalence and age-adjusted prevalence. In a nutshell, crude prevalence measures the prevalence against the entire population, while age-adjusted prevalence is measured against age distribution (thus being more accurate). Since it was more accurate, age-adjusted prevalence was used in our study for all analyses. [10]

### **Planned Analysis/Visualization**

We plan to use two modelling techniques as well as data visualizations. Given the large number of independent and dependent variables, our first model will run correlations for each variable against each other variable. We will examine the correlation coefficient values to determine where relationships exist amongst variables. These will be displayed in correlation heatmap (see below) to guide us on which variables to investigate next. We also plan to use multiple regression to analyze a number of independent variables against our key dependent variables (health outcomes) to try to find a model that best predicts which preventative measures and unhealthy behaviours are the best determinants of health outcomes. Our visualizations will include the correlation heatmap and some linear graphing of select variables to check the validity of our regression modelling.

As a next step we may use our regression model to predict health outcomes in a different country. This will test our model and assess whether other countries display similar characteristics to the USA. This is dependent on our ability to find a robust dataset that provides incidence for both the independent variables and the health outcomes which we find relevant in our model. (Refer to Figure 3).

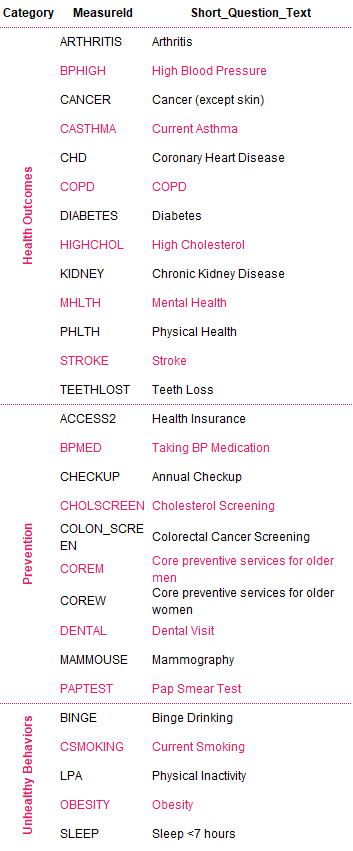


**Figure 3: Correlation Heat Map**

### 

### **Data Cleansing**

For the final project we will cleanse the data from the 500 Cities dataset. Cleansing includes the removal of extraneous columns (such as empty, duplicate or irrelevant columns), filter that dataframe to obtain age-adjusted city level prevalence values, and pivot the table to obtain a pragmatic data structure format for our plots to generate. Also, the data contains abbreviated codes for each measure; Table 1 contains description of each measure (Refer to Table 1)



**Table 1: Mapping Table**

# **Cost/Benefit Analysis**

A proverb states that an ounce of prevention is worth a pound of cure. This statement is exceptionally true when it comes to healthcare. It is estimated that 60% of Americans are living with a chronic condition and 42% are living with more than one [11]. These conditions are costly to the healthcare system and to individuals, costly in terms of lost productivity; and most importantly they are costly in terms of the quality of life for those affected.

Prevention of chronic disease is the best way to avoid the social and financial costs mentioned above. Prevention can come in the form of preventative health care measures and also through individual lifestyle choices.

The largest charities in the USA (The Y, Goodwill, Catholic Charities, the United Way and others) are all complex social services charities supporting multiple programs, populations and issues. Given this complexity it is difficult to find data that shows charitable giving that is directed to prevention versus spending that is directed to cures. This is beyond the scope of this project. Nonetheless, based on the results of our analysis we intend to determine the key drivers of better health outcomes, particularly for those conditions that are chronic and costly. Using this information, charitable givers can research their charity of choice and specify that their dollars be directed towards a particular program which meets the objective of maximizing value of charitable dollars spent.

# **Conclusion**

Donors want to make a positive impact on society. Sometimes their goals are specific, like supporting a cause that is personally important to them, and other times they just want to make a difference. Unfortunately, donors are often uninformed and do not take the time to research the charities to which they are contributing. Charities with big marketing budgets, smart marketing tactics and causes that are deemed ‘sexy’ can end up targeting donor dollars to causes that do not generate the greatest good to society. A prime example of this was the ALS ice bucket challenge that went viral in 2014. By encouraging people to take a challenge and nominate their friends to do the same, this generated significant social media attention and earned the charity over $23 million in revenue - or roughly $1,450 per person living with the disease - in the first year in which it was introduced [12],[13].

This does not mean that ALS is not an important cause to donate to (the disease is debilitating and terminal), nor does it suggest that people shouldn’t donate to the charity; however in terms of sheer numbers of people affected, the impact is significantly less than for charities that help prevent some of the most chronic and costly diseases faced by Americans.

It is our hope that the outcome of this study will aid in an understanding of what factors most influence negative health outcomes and will help donors make better informed choices on where to target their charitable giving.

# **Citations**

1. Fay, Bill. Hospital and Surgery Costs, Debt.org, America's Debt Help Organization, Retrieved June 28, 2020 from <https://www.debt.org/medical/hospital-surgery-costs/>
2. Leonhardt, Megan. Become Debt-Free, CNBC, CNBC Make It, Retrieved June 29, 2020 from <https://www.cnbc.com/2020/02/13/one-third-of-american-workers-have-medical-debt-and-most-default.html>
3. Survey: 79 Million Americans Have Problems with Medical Bills or Debt, Common Wealth Fund, The Commonwealth Fund Publication, Retrieved June 28, 202 from <https://www.commonwealthfund.org/publications/newsletter-article/survey-79-million-americans-have-problems-medical-bills-or-debt#:~:text=In%20fact%2C%2041%20percent%20of,medical%20bill%20or%20debt%20problems>.
4. Beaton, Thomas. Top 10 Most Expensive Chronic Diseases for Healthcare Payers, Health Payer Intelligence, Health Payer Intelligence, Retrieved June 19, 2020 from <https://healthpayerintelligence.com/news/top-10-most-expensive-chronic-diseases-for-healthcare-payers>
5. Shaw KM, Theis KA, Self-Brown S, Roblin DW, Barker L. Chronic Disease Disparities by County Economic Status and Metropolitan Classification, Behavioral Risk Factor Surveillance System, 2013. Prev Chronic Dis 2016;13:160088. DOI: http://dx.doi.org/10.5888/pcd13.160088 from <https://www.cdc.gov/pcd/issues/2016/16_0088.htm#tables>
6. Charitable Giving Statistics, Nptrust.org, National Philanthropic Trust, Retrieved June 30, 2020 from <https://www.nptrust.org/philanthropic-resources/charitable-giving-statistics/#:~:text=General%20Philanthropy,a%200.7%25%20increase%20from%202017.&text=Corporate%20giving%20in%202018%20increased,a%205.4%25%20increase%20from%202017.&text=Foundation%20giving%20in%202018%20increased,a%207.3%25%20increase%20from%202017>.
7. 500 Cities: Local Data for Better Health, 2016 release, Data.gov, Centers for Disease Control and Prevention, Retrieved June 19, 2020 from <https://catalog.data.gov/dataset/500-cities-local-data-for-better-health-2016-release-3f265>
8. Beaton, Thomas. Top 10 Most Expensive Chronic Diseases for Healthcare Payers, Health Payer Intelligence, Health Payer Intelligence, Retrieved June 19, 2020 from <https://healthpayerintelligence.com/news/top-10-most-expensive-chronic-diseases-for-healthcare-payers>
9. Programs And Interventions, CDC, Centers for Disease Control and Prevention, Retrieved June 28, 2020 from <https://www.cdc.gov/500cities/programs-interventions.htm>
10. Crude and Age-Adjusted Rate, Health.Mo, Missouri Department of Health and Senior Services, Retrieved June 30, 2020 from <https://health.mo.gov/data/documentation/crude-aarate.php>
11. Levine S, Malone E, Lekiachvili A, Briss P. Health Care Industry Insights: Why the Use of Preventive Services Is Still Low. Prev Chronic Dis 2019;16:180625. DOI: http://dx.doi.org/10.5888/pcd16.180625 from <https://www.cdc.gov/pcd/issues/2019/18_0625.htm>
12. Belluz, Julia. The truth about the Ice Bucket Challenge: Viral memes shouldn't dictate our charitable giving, Vox.com, Vox, Retrieved June 30, 2020 from <https://www.vox.com/2014/8/20/6040435/als-ice-bucket-challenge-and-why-we-give-to-charity-donate>
13. ALS Association, Alsa.org, Who Gets ALS?, Retrieved June 29, 2020 from <http://webco.alsa.org/site/PageServer/?pagename=CO_1_WhoGets.html#:~:text=Approximately%205%2C000%20people%20in%20the,ALS%20at%20any%20given%20time>