1

The Causal Effect of Health Expenditure on Life Expectancy

Nate Lentz NPL1@WILLIAMS.EDU
Computer Science
W3076099

1. Introduction

The COVID-19 pandemic has shown us that mortality rates are highly variable even in the face of a common source, such as the coronavirus. Why is that? Some countries have better healthcare, some have more strict rules in response to COVID, and a plethora of other factors that effected COVID rates. One factor that seems to explain many other of these factors is the health expenditure of a country, or more specifially, the percent of country's total expenditure that is designated to healthcare. This is because the more a country spends on healthcare, we can expect them to have better healthcare for sick individuals and a better control on their population in regards to health guidelines. This then begs the question if this logic applies more broadly. Does health expenditure effect life expectancy, which can be used as a marker for overall health and longevity in a country's population? Thus, some insight into how the status of a country effects some of the aspects of it's people's health could give some insight into the value of monetary resources and the true health of a country's population

To research this question, I utilized a World Health Organization data set on Life expectancies across the world (WHO Dataset). The question that I sought to answer was, to what extent does health expenditure influence life expectancy. To proceed answering this question, I first examined the data and determined which variables would be effective to use. Then, I cleaned and winsorized the data such that I could perform proper analysis. This analysis consisted of following the process of causal influence, wherein I drew the causal graph for the data set, followed by calculating the average causal effect of health expenditure on life expectancy.

This analysis lead me to believe that the total health expenditure of a country heavily contributes to an increased life expectancy. This is when we adjust for confounding variables and make assumptions about unmeasured confounders in the graph.

2. Preliminaries

Directed Acyclic Graphs, or DAGs, are defined as some graph, $\mathcal{G} = (V, E)$, which is "simple," meaning that they have at most one edge between any pair of vertices, V_i, V_j, all edges in the graph are directed, and that for any $V_i \in V)$ there is no directed path from $V_i....,V_i$. In words, a DAG is simply a graph with only directed edges and no cycles. Causal models of a DAG $\mathcal{G}$ defined over a set of variables V may be interpreted as a tuple consisting of the DAG itself, and a system of non-parametric structural equations with independent errors equipped with the do-operator (Pearl, 2009). Each variable is determined as a function of its parents and an independent error term. This induces a distribution $p(V)$ that factorizes according to the DAG $\mathcal{G}$ as follows,

$$p(V) = \prod_{V_i \in V} p(V_i \mid \text{pa}_{\mathcal{G}}(V_i)),$$

where $\text{pa}_{\mathcal{G}}(V_i)$ denotes the parents of V_i in $\mathcal{G}$. Under this interpretation, a directed edge $V_i \rightarrow V_j$ may be interepreted as saying that V_i is potentially a direct cause of V_j. Conditional independences in $p(V)$ can be read off from the DAG via d-separation, i.e., $(X \perp\!\!\!\perp Y \mid Z)_{\text{d-sep}} \implies (X \perp\!\!\!\perp Y \mid Z)_{\text{in } p(V)}$. To facilitate structure learning, I will restrict my analysis to the set of *faithful* distributions where $(X \perp\!\!\!\perp Y \mid Z)_{\text{d-sep}} \iff (X \perp\!\!\!\perp Y \mid Z)_{\text{in } p(V)}$. In addition, I make the simplifying assumption that the relations between my variables are linear.

An unmeasured confounder, U, is a node representing some information that we are not privy to in a given graph. We use an acyclic directed mixed graph, or ADMG, to represent a graph which has unmeasured confounders. This is because ADMGs preserve causal and statistical information from the DAG that it is derived from. In an ADMG, we preserve causal information because we state that V_i is a cause of V_j in $\mathcal{G}(V)$ if and only if V_i is a cause of V_j in $\mathcal{G}(V \cup U)$. Furthermore, an ADMG preserves statistical information because in for all disjoint subsets X, Y, Z, of the observed variables V,

$$X \perp\!\!\!\perp_{m-sep} Y|Z \text{ in } \mathcal{G}(V) \leftrightarrow X \perp\!\!\!\perp_{d-sep} Y|Z \text{ in } \mathcal{G}(V \cup U). \tag{1}$$

In words, $X \perp\!\!\!\perp Y|Z$ in the model implied by $\mathcal{G}(V)$ if and only if $X \perp\!\!\!\perp Y|Z$ in the model implied by $\mathcal{G}(V \cup U)$. As such, for any unmeasured confounders that are the parents that are the parents of two nodes, V_i, V_j, we can draw a bidirected edge between them in the graph $V_i \leftrightarrow V_j$.

From some DAG or ADMG, $\mathcal{G}$, we can construct a single world intervention graph, or SWIG of the original, $\mathcal{G}(a)$ which illustrates conditional independences between factuals and counterfactuals when we modify some treatment variable, A, in $\mathcal{G}$ (Richardson and Robins, 2013). To construct the SWIG, we use our original graph as the base of $\mathcal{G}(a)$, and we split some treatment variable, A, from $\mathcal{G}$ into a "random" vertex A and a "fixed" vertex a. A inherits all of the incoming edges into A and a inherits all of the outgoing edges. Following this, all vertices V_i which are descendants of a in our SWIG, $\mathcal{G}(a)$ are now converted to potential outcomes $V_i(a)$ to maintain logical equivalence counterfactuals.

2.1 Assumptions

For my research, I made a few assumptions based on background knowledge about what edges could or could not be present, as well as what the causal order of variables should be. Furthermore, my analysis only covers the set of faithful distributions of DAG $\mathcal{G}$. This means that my analysis only covers interactions in $\mathcal{G}$ where $(X \perp\!\!\!\perp Y|Z)_{d-sep} \leftrightarrow (X \perp\!\!\!\perp Y|Z)_{\text{in } p(V)}$.

3. Methods

3.1 Data Processing

My data set, WHO Life Expectancy Data, comes from Kaggle through user Kumar Rajarshi, who put the data together from the World Health Organization and United Nations websites. Initially, the data contained 22 columns, including information that I deemed unnecessary for my analysis such as country, year, status, vaccination information, and weight information. Most of these categories I cut out because I did not believe that they were causally relevant to the question I was asking, but I had to cut the body mass index category because much of the data in the column did not make realistic sense. That is, many of the values were below 10 or above 50, which are extremely unrealistic for normal BMI's. This left me with nine variables to work with: Life expectancy (the outcome variable), total expenditure (the treatment variable), percentage expenditure, adult mortality, under-five deaths, alcohol consumption, schooling, gross domestic product (GDP), human developmement index (HDI). With this data, I dropped any and all rows which contained a missing data in the form of a n/a or nan (not-actual-number).

3.2 Data Cleaning

While searching through my data, I found that there was some data which fell far outside the range of even a normal outlier. As such, I turned to a process called winsorizing to make my data more accurate and usable, as the large outliers could skew my analysis, which I do not want, especially if said data was incorrect. Winsorizing essentially censors the data to limit the effects caused by possible spurious outliers. Instead of trimming or dropping the data that falls outside the range that we want to winsorize, which usually falls around the 90-95th percentile, values outside that range are replaced by the values at the end of the usable range. The table below represents the variables used in my final analysis.

3.3 Graph Construction

To extract the casual graph from the data set that I had constructed, I utilized the program Tetrad to perform a causal discovery on the data (Richard Scheines, 1998). I added some background knowledge to the graph prior to its construction, as I placed the variables in three tiers, wherein variables in later tiers cannot cause variables in prior tiers. In tier one I had the economic factors: GDP, percentage expenditure, total expenditure, HDI. In tier two I had the intermediate factors: adult mortality, schooling, under five deaths, and alcohol. In the third and final tier I just had the output variable, life expectancy. The logic behind these tiers stem from conventional and background knowledge. For tier one, I placed economic factors because in macroeconomic terms, GDP = consumption + investment + goverment spending + net exports, and thus should only be causally effected by other economic factors, otherwise there are too many unmeasured confounders that could effect it. The same logic follows for the two health expenditure variables as well as HDI. The second tier I had the variables that were directly effected by goverment spending, which were all variables that described the health and education of a population. Finally, I had life expectancy in its own tier because it ultimately has many unmeasured confounders like

Graph Variables	
Variable	Description
Life Expectancy	The outcome variable in this analysis.This represents the average life expectancy in a country from a given year.
Total Expenditure	The treatment variable in this analysis. This variable measures the general government expenditure on health as a percentage of total government expenditure.
Percentage Expenditure	This variable measures the government health expenditure as a percentage of GDP.
Alcohol	This variable represents the alcohol consumption by liters per capita (15+).
Adult Mortality	This variable represents the mortality rate for individuals between the ages of 15 and 60 (per 1000 people).
GDP	The gross domestic product (per captia) of a given country, measured in USD.
Under Five Deaths	This variable represents the mortality rate for individuals under the age of 5 (per 1000 people)
Schooling	This variable the average years of schooling for a country's population.
HDI	Represents the Human Development Index of a given country

Table 1: Causal Variables

genetics that effect life expectancy as well as different aspects of the graph, so for the sake of simplicity in my analysis, I assumed life expectancy only had incoming edges.

I then used the Greedy Fast Causal Inference algorithm which implements Greedy Equivalence Search (GES), which is a method used for learning DAGs (Chickering, 2002). This was done to complete a causal search of the data set to learn an equivalence class of possible causal structures with no unmeasured confounders. For scoring the graph, I used the Conditional Gaussian Bayesian Information Criterion (CG-BIC) score (Haughton, 1988; Schwarz, 1978). All of my variables were continuous, so I left all of the parameters for the search at default, except I increased the number of bootstraps to 50. The bootstrapping method used re-sampling with preservation.

I made some alterations to the graph generated by these methods in Tetrad. These alterations were all supported by background information. This may have skewed my graph, and by extension my analysis, but it was necessary given that some of the relationships found by Tetrad were not founded in truth given some conventional knowledge.

A non-parametric conditional independence test known as the Fast Conditional Independence Test was used to check for conditional independences during sensitivty testing (Chalupka et al., 2018).

3.4 Causal Identification

To identify the causal relationships in the graph, I utilized the backdoor criterion. A set of variables, Z, which satisfies the backdoor criterion with respect to some treatment A and an outcome Y in a DAG $\mathcal{G}$ if $A \perp\!\!\!\perp Y(a)|Z$ in the SWIG $\mathcal{G}(a)$ cite (Pearl 1998). This way I could use an appropriate backdoor adjustment set to discover the causal relationship between health expenditure and life expectancy.

I used an ADMG, because my graph contained bidirected edges. By m-seperation in the SWIG, with a treatment A, valid adjustment set C, and outcome Y we have:

$$\mathbb{E}[Y(a)] = \sum_{c} p(C) \times \mathbb{E}[Y|A = a, C]. \tag{2}$$

Because of the local Markov property, we know that the proper adjustment set, C, is the Markov Blanket of A. The blanket is the district and parents of A, without A itself. Observe:

$$mb_{\mathcal{G}}(A) \equiv dis_{\mathcal{G}}(A) \cup pa_{\mathcal{G}}(dis_{\mathcal{G}}(A)) \backslash A \tag{3}$$

From this, we can write the average causal effect, or ACE, formula as defined by Pearl (1995):

$$ACE \equiv \mathbb{E}[Y(a) - Y(a')] = \sum_{c} p(C) \times \mathbb{E}[Y|A = a, C] - \sum_{c} p(C) \times \mathbb{E}[Y|A = a', C] \tag{4}$$

3.5 Causal Estimation

With a graph generated and a process for causal identification outlined, I proceeded with calculating the ACE. My treatment, total expenditure, contained continuous values and I thus used a python implementation of the backdoor mean function to calculate the ACE for my data. The backdoor mean function utilizes backdoor adjustment over the entire range of continuous values of a given data set. This means that we predict the data for A' at each and every data point in the range of our treatment. For these predictions I utilized the statsmodels package for python, specifically predicting using the Gaussian model because my data was continuous (Seabold and Perktold, 2010). I used a random seed of 100 to generate the data. Because I had to iterate over the entire range of the treatment, 0.337-11.66 at an interval of 0.01, I got a point estimate of the ACE for the predictions at each data point, rather than one single value that represented the ACE of total health expenditure on life expectancy. Calculating the confidence intervals had to be done the same way, with a 95% confidence interval being calculated for each point estimate we calculated. For these confidence intervals, I used a bootstrap of n = 10, because we were calculating so many bootstraps, it would have been inefficient to have a higher number of bootstraps per point estimate.

4. Results

4.1 The Graph

The graph below (Figure 1) illustrates the final ADMG generated after using Tetrad and background knowledge to edit the graph that came from Tetrad. Any edge that was considered a double ensemble due to unmeasured confounders was converted to a bidirected edge, as we can see in red. In the case where there was a possible unmeasured confounder affecting one side of an edge, I used background knowledge to assume that the edge was simply directed. The cases where there were unmeasured confounders that effected both sides of an edge, I made bidirected. The background knowledge I used to make these assumptions was mostly centered on what variables in reality could effect each other (see section 3.3 paragraph 1), as well as was based on the p-values of edges as given to me by Tetrad.

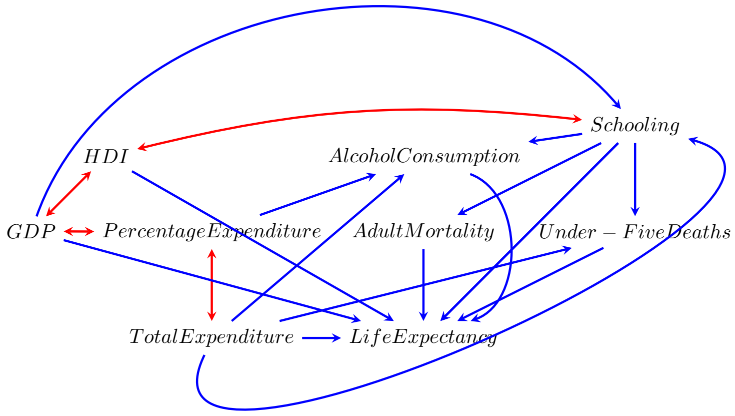

Figure 1: The ADMG, $\mathcal{G}$, learned over the life expectancy data

4.2 Causal Identification and Estimation

Given the ADMG above, I could draw a SWIG, where total expenditure split into a random varible and a fixed variable. As such, life expectancy, under-five deaths, alcohol consumption, and schooling all become potential outcomes of total expenditure. I found that the set $Z = \{\text{GDP, Percentage Expenditure}\}$ forms a valid backdoor adjustment set for the effect of total expenditure on life expectancy. This is because the Markov blanket for total expenditure is comprised of percentage expenditure and GDP, and thus:

$$\text{Total Expenditure} \perp\!\!\!\perp \text{Life Expectancy} \mid \{\text{Percentage Expenditure, GDP}\} \tag{5}$$

Now, we can calculate the point estimates for our backdoor mean, or ACE, using the valid backdoor adjustment set Z. After iterating over the all of the possible values of total expenditure, we get the point estimates, which we iterate over to get the confidence interval at each point estimate. The results of this code is illustrated in Figure 2.

While this plot is useful to see a visual representation of the point estimates of the ACE, it we can also take the slope of the plotted line. The slope of this line gave me the ACE estimate to be ≈ 0.567. We cannot get an exact estimate for the confidence intervals because there is not a linear slope which tracks the confidence intervals that we can calculate from the graph. However, visually we can see that there is a fairly tight spread amongst the lower and upper bounds of the confidence intervals, which trends upwards. Thus, we can infer that the slope is positive along the lower bounds of the confidence intervals, and as such, we must have a positive ACE.

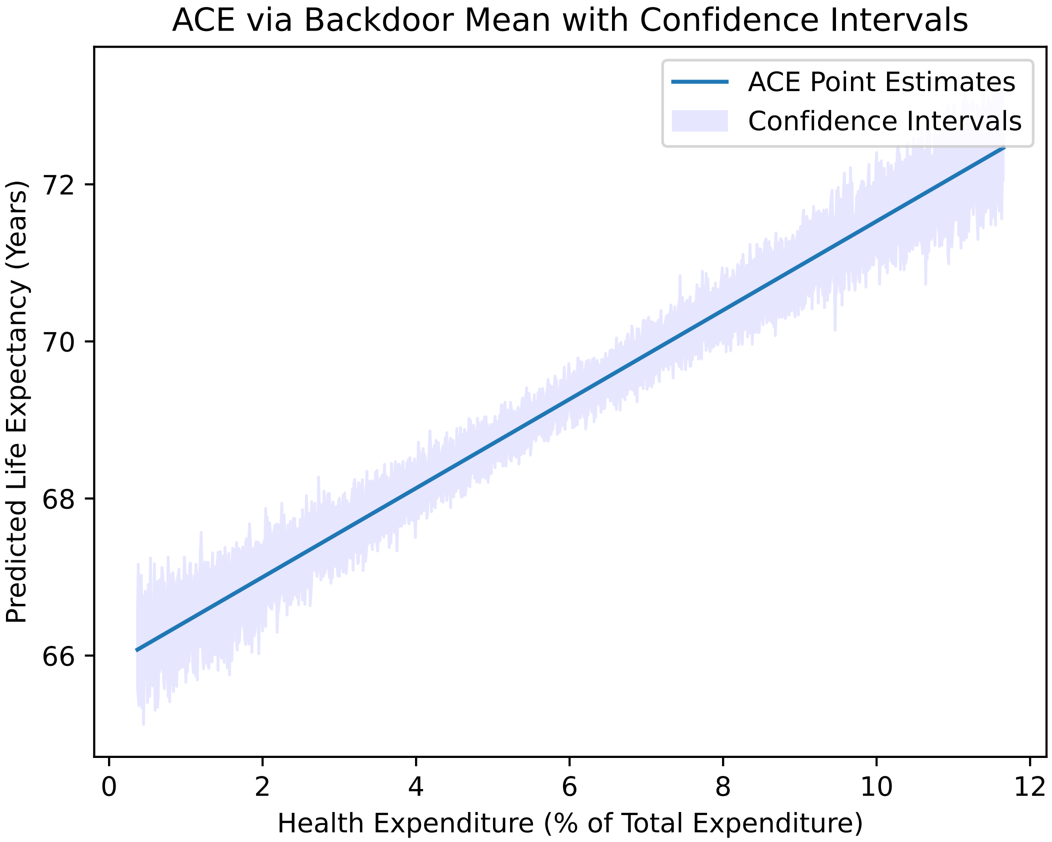

Figure 2: This plots the point estimates for the ACE of total health expenditure on life expectancy along with the confidence intervals for said point estimates

4.3 Sensitivity Analysis

Given that I made some assumptions with the data and in constructing my graph, I proceeded with sensitivity analysis. For this, I used the Fast Conditional Independence Test (FCIT), which yields a value from 0 to 1 which indicates edge presence. A value closer to 0 means that an edge is likely present, whereas an edge with a value closer to 1 is likely not present. I decided to test the most important edges to my analysis, which included the edge Total Expenditure → Life Expectancy, and the Markov blanket edges. These edges include Total Expenditure ↔ Percentage Expenditure, and Percentage Expenditure ↔ GDP. I used the FCIT package on python, testing the presence of each of these edges. For bidirected edges, I tested the presence of an edge pointing left and an edge pointing right. Observe:

Table 2: Sensitivity Testing.

Edge	**Presence (p-value)**
Total Expenditure → Life Expectancy	9.785e-05
Total Expenditure → Percentage Expenditure	5.209e-05
Total Expenditure ← Percentage Expenditure	2.403e-07
Percentage Expenditure → GDP	2.419e-07
Percentage Expenditure ← GDP	2.294e-06

The table above shows that each of the important edges are present in the graph. This illustrates that there is certainly a causal relationship between total expenditure and life expectancy. Furthermore, it proves that the bidirected edges that I assumed were there, based on background knowledge, are indeed present. As such, the Markov blanket that I used for backdoor adjustment should be accurate.

5. Discussion

Based on results presented by my analysis in Section 4, I found that increasing total government health expenditure can significantly increase the life expectancy of a country. This is evident in the ACE estimate of 0.567 that I calculated from the slope of the point estimates from the plot in Figure 2. In words, this tells us that for each additional percent of total government expenditure that goes to healthcare, we can expect the life expectancy of the given country to increase by about 0.567 years. While my estimates of the 95% confidence interval for this ACE could not be directly measured, the trend of them still suggested that the effect is, most likely, at least positive, and thus that increasing total health expenditure should increase life expectancy.

This analysis was conducted under some assumptions, with background information being supplemented in place of unmeasured confounders. Death, and by extension life expectancy, is caused by a plethora of factors which were not measured in my analysis. As such, my observation is as robust as it can be given the information that I was privy to, but is not necessarily complete.

6. Conclusion

This project consisted of following the causal inference pipeline to discover if, and if so, how much total government expenditure on health effects a given country's life expectancy. Using a backdoor mean, I calculated an ACE of 0.567, which indicates that total health expenditure has a significant positive impact on life expectancy. While this result was produced by accurate methods, these methods were built on background knowledge and assumptions of faithfulness that may not be entirely accurate. Furthermore, although I completed sensitivity analysis, significant edits were made to the graph learned by Tetrad to find the Markov blanket, which was necessary to complete my causal analysis. In the future, I would want to find a way to incorporate the variables in my data which I ended up dropping, such as vaccination data, to see how, and if, they would effect my analysis. I also think testing the causal effect of different variables in my data set, such as alcohol consumption, on life expectancy could yield interesting results.

References

Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Fast conditional independence test for vector variables with large sample sizes. *arXiv preprint arXiv:1804.02747*, 2018.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002.

Dominique MA Haughton. On the choice of a model to fit data from an exponential family. *The annals of statistics*, pages 342–355, 1988.

Judea Pearl. *Causality.* Cambridge University Press, 2009.

Clark Glymour Christopher Meek Thomas Richardson Richard Scheines, Peter Spirtes. The tetrad project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, 33(1):65–117, 1998.

Thomas S Richardson and James M Robins. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality, 2013.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. *In 9th Python in Science Conference*, 2010.