Leiden University - Causal Inference Report (Group 21)

March 2024

All authors contributed equally to this group project.

1 Abstract

In this report we investigate the causal relationship between anti-hypertensive medication (BPMEDS) and risk of death among participants in the Framingham Heart Study [1]. Our research aims to estimate the average treatment effect (ATE) of taking BPMEDS via G-computation modeling. We found that there is a positive ATE of 8.2% points. However, important assumptions of our model are believed to be violated, raising doubts about the validity of the estimate.

2 Introduction

The Framingham Heart Study [1] is a long-term cohort study, directed by the National Heart, Lung, and Blood Institute (NHLBI) in the United States to identify the underlying characteristics that contribute to cardiovascular disease (CVD). The study began in 1948 with 5,209 participants from the Framingham community in Massachusetts. The study is still running in 2024 with the initial participant's children and grandchildren being the new active participants. Due to its longevity and comprehensive design, the Framingham study is often regarded as one of the pioneering studies in the fields of CVD research and epidemiology.

For this group assignment, we were given an amended version of the original Framingham data with 4,434 participants instead of 5,209. The data contains information about the participants at baseline in 1956. The timeline of the given data ends in 1980, so 24 years after the start of our study, where various outcome variables such as CVD, angina pectoris, myocardial infarction, stroke, hypertension, and death were recorded.

3 Research Question

For this assignment we were asked to investigate the causal relationship between BPMEDS and the risk of death. Thus, our research question is:

What is the ATE of using BPMEDS on the risk of death in 1980 for participants in the Framingham study?

$$\mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] \tag{1}$$

Equation 1 represents our research question rewritten using potential outcome notation. Y is an indicator for death, i.e. dead or alive. $Y(1)$ is the potential outcome under the exposure, so any outcome of Y given that the participant did take BPMEDS. $Y(0)$ represents the counterpart where the participant did **not** take BPMEDS, i.e. the potential outcome under no exposure. Finally, we take the expectation of both potential outcomes and subtract them to compute the population level treatment effect of taking blood pressure medication on the probability of dying. Said probability is also often referred to as the risk of death. As Y is binary the ATE can also be described as the risk difference of dying between the two exposure levels.

4 Research Protocol

As the design of Framinhgam study is observational and not experimental it is beneficial for our analysis to formulate the protocol components of the target trial we are trying to emulate. To this end, we can use the framework proposed by Hernan et al [2]. Table 1 summarises our target trial components.

Table 1: Emulation of Target Trial

Protocol Component	**Description**
Eligibility criteria	Individuals living in the Framingham community who were present during the first examination period in 1956
Treatment strategies	Intervention group - Use of anti-hypertensive medication (BPMEDS) Control group - No use of anti-hypertensive medication (BPMEDS)
Assignment procedures	The given information does not provide much context about the assignment procedure. It is clear though that assignments have not been randomised as the Framingham study is not experimental. Thus, it is important to adjust for confounders at baseline to correctly emulate the target trial.
Follow-up period	Baseline measurements in 1956 with final measurement in 1980.
Outcome	Binary outcome - Is the participant dead or alive in 1980
Causal contrasts of interest	Average treatment effect for the risk of death between the two exposure levels
Analysis plan	Adjustment for baseline confounders as intention-to-treat effect cannot be computed in this observational study

5 Methodology

5.1 Assumptions

For this section, we will examine the main assumptions underlying our causal inference analysis: consistency, exchangeability, and positivity. These assumptions are fundamental in ensuring the validity of our causal estimations. By meeting these conditions, we ensure the **identifiability** of the average treatment effect which allows us to correctly infer the causal effect of the BPMEDS.

For **consistency**, we assumed that each individual in the treatment group is exposed to the same BPMEDS in order to maintain uniformity in the treatment assignment. In other words, we assumed that multiple versions (e.g. brands, dosages, timing) of the BPMEDS do not exist and instead, everyone is given the exact same medication. As the data at hand is observational we assumed that **exchangeability** can only be met if we adjust for baseline confounders. Additionally, we assumed that all confounders are observable and have been recorded in the given dataset. This ensures, that treated and untreated individuals are comparable. We also assumed that all individuals have a non-zero probability of receiving either treatment, i.e. that the **positivity** assumption is met. All three assumptions are addressed in more detail later on when we refer to G-computation and propensity score weighting. If any of the assumptions mentioned above are violated it would not be possible to correctly identify the causal effect of interest.

5.2 Missing Data

In our analysis, we examined the participants with missing values to understand if the missingness would affect our results. We compared the explanatory variables of the participants with missing data against those with no missing data. Whether or not there are significant differences between the characteristics of these two groups will help to indicate which missing mechanism is shown in the data. Under the mechanism of missing completely at random, a complete case analysis can be performed as the missingness is not related to the observed or unobserved data. However, if there are clear differences between the groups then data imputation is required as the mechanism is missing at random.

5.3 DAG

A directed acyclic graph (DAG) consists of a set of nodes and directed arrows between nodes. The nodes represent variables in the analysis. The directed arrows show both the associations between variables and the direction of the association. By specifying the correct DAG, then a causal relationship can be estimated by adjusting for confounding variables. The minimal adjustment set of variables is chosen to meet the requirements of d-separation and backdoor criterion [3]. Given that there are 15 variables in our dataset, there are many possible associations to specify, so we have primarily focused on the associations that will affect our minimal adjustment set. We have constructed our DAG using both our expert medical knowledge and the findings from our exploratory data analysis. In our exploratory data analysis, we inspected whether explanatory variables are collinear, and if they have associations with either the exposure or outcome. After fitting our DAG, we checked that our minimal adjustment set was correct by inspecting the conditional dependencies remaining in the data. Our check confirmed that the chosen minimal adjustment set agrees with our data, and therefore can be used in G-computation and propensity score weighting.

5.4 G-computation

For our statistical analysis, we used G-computation as our primary method for estimating the causal effect between BPMEDS and risk of death. G-computation is more statistically efficient compared to propensity score weighting, under the assumption that the model is correctly specified. We employed a G-computation model to adjust for the confounders identified from our DAG analysis. Initially, we fit a logistic regression model with death as our outcome variable of interest against the explanatory variables being our treatment (BPMEDS) along with the minimal adjustment set of confounders. This adjustment helps isolate the causal effect of BPMEDS on the risk of death, accounting for the influence of potential confounders. We first estimated the counterfactual outcomes by setting the exposure variable to different values (BPMEDS == 1, BPMEDS == 0) while holding the confounders constant at their observed values. This allowed us to assess the causal effect of the treatment by comparing the observed outcome with the hypothetical outcome that would have occurred under different treatment scenarios. We then tested for causal effects by comparing these expected outcomes under the different treatment groups to obtain the average treatment effect. To verify the assumptions for conditional exchangeability and positivity, we compared the distribution of the explanatory variables between the treatment groups through the use of boxplots. Similar shapes and spreads of the boxplots indicate that the groups are comparable and hence, exchangeable, whilst considerable overlap supports positivity.

5.5 Propensity Score Weighting

Propensity score weighting was conducted as a sensitivity analysis to reinforce the findings from G-computation. Even though we believe propensity score weighting to be less statistically efficient compared to G-computation, the method is still useful to check the causal inference assumptions, as they are easier to verify. We fitted a propensity score model with BPMEDS as the outcome variable and the minimal adjustment set as the explanatory variables. We then used the standardized mean difference to check for exchangeability and the overlapping propensity score plots for positivity. There also exists propensity score matching. However, only 3% of the dataset received BPMEDS and so matching on this small sample would substantially reduce our dataset and therefore lead to a loss of accuracy. As a result, this approach was not be followed and instead, we used propensity score weighting for our sensitivity analysis.

6 Results

6.1 Missing Data

We found that there is a total of 193 rows containing missing values, of which BPMEDS had 61 missing values. This accounts for 4.4% of our entire dataset and we deemed the proportion of this missing data to be relatively low. We also compared the distribution of the explanatory variables between the full dataset and a subset containing only missing values. There were no extreme differences observed, therefore we concluded that the missingness mechanism is likely missing completely at random (MCAR). We proceeded with a complete case analysis by removing any rows with missing values, as this approach is expected to yield valid results given the assessed missing mechanism is MCAR.

6.2 DAG

Our finalised DAG is shown in Figure 1. The key to understanding our DAG is how the explanatory variables and outcome variables relate to BPMEDS and DEATH. The explanatory variables: AGE, BMI and SYSBP are all part of forks between the exposure and outcome, and so need to be included in the adjustment set to remove confounding bias. The event variables are all mediators in a chain between the exposure and outcome. Therefore, adjusting for these variables would introduce overcontrolling bias as it would remove the indirect effect of BPMEDS on DEATH. In our exploratory data analysis, we found that systolic and diastolic blood pressure have a correlation of 78.6% so only one variable should be included to remove multicollinearity from the analysis. When fitting both models we compared AIC for a model including systolic blood pressure versus a model including diastolic blood pressure. For both G-computation and propensity score methods the model with systolic blood pressure was a better fit. We also found that education, heart rate and smoking status have no direct relationship with the exposure or outcome and therefore we have not included these variables in the analysis.

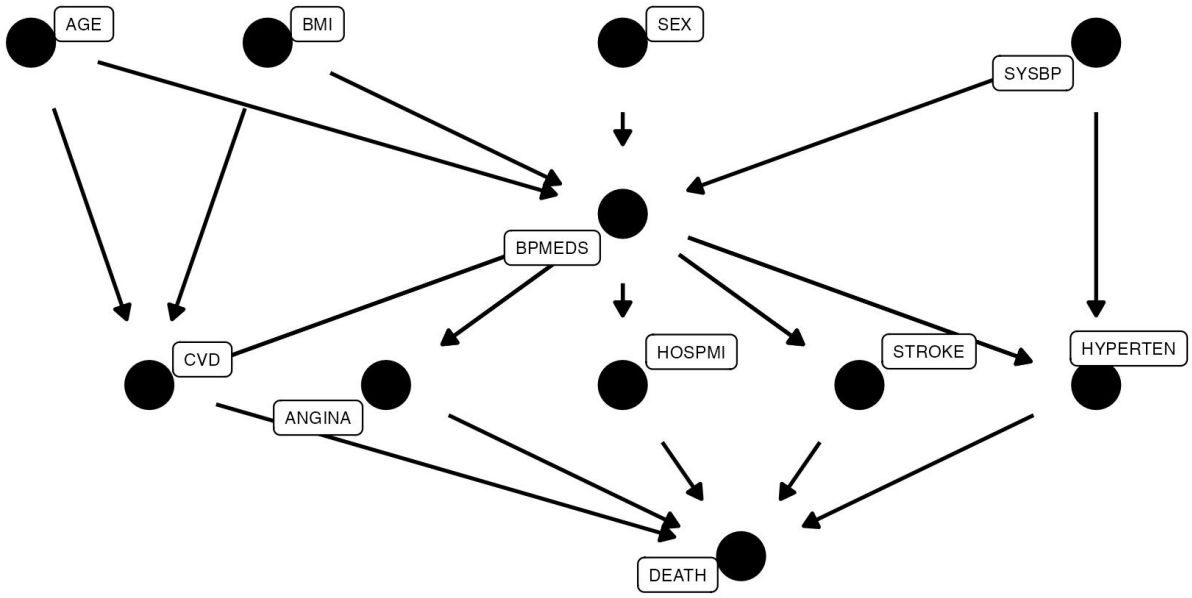
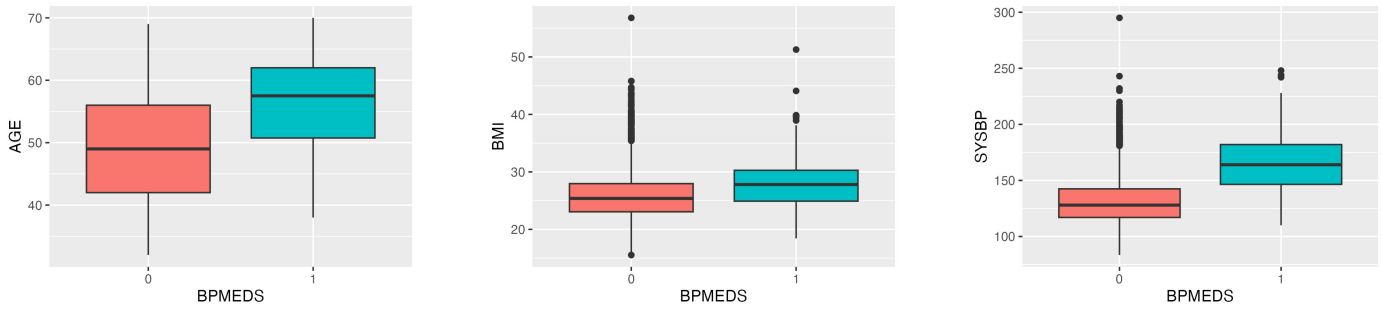


Figure 1: DAG where exposure = BPMEDS and outcome = DEATH (with BPMEDS → CVD)

6.3 G-computation

The G-computation model (Figure 3) estimated the ATE of BPMEDS on the risk of death. Our results revealed that the predicted outcome under treatment was 42.4% points whereas the predicted outcome under no treatment was 34.2% points. This yielded an ATE of 8.2% points which indicates that taking BPMEDS increases an individual's risk of death by said percentage points. The 95% confidence interval for the ATE is (-0.002, 0.165).



(a) Age vs BPMEDS (b) BMI vs BPMEDS (c) SYSBP vs BPMEDS

Figure 2: Checking for conditional exchangeability (G-computation)

The boxplots in Figures 2 illustrate the distribution of the confounders between the treatment groups. It appears that there are notable differences in the distribution of age. The treated individuals exhibit a wider interquartile range, spanning from 50 to 62 years old with a mean of 56.29. In contrast, the untreated group has a slightly narrower interquartile age range, between 42 and 56 years old with a mean of 49.61. A similar trend is illustrated for systolic blood pressure. The treated group exhibits a higher interquartile range spanning from about 146 to 182mmHg, with a mean of 165.6 mmHg. Whereas, the untreated group ranges from approximately 117 to 142.5mmHg, with a mean of 131.7 mmHg. The BMI distributions have slightly smaller differences regarding the interquartile ranges and means. The treated group spans from 24.91 to 30.28 with a mean of 28.03, whereas the untreated group spans from 23.06 to 27.96 with a mean of 25.75. Overall, there is substantial overlap among all the boxplots given the confounding variables.

6.4 Propensity Score

We fitted the propensity score model (Figure 4) using the adjustment set identified from the DAG. Thus, we used BMI, systolic blood pressure and age as the explanatory variables. We calculated an ATE of 16% points, with the expected risk

of death under the exposure being 50.7% points and the expected risk of death under no exposure being 34.7% points. The 95% confidence interval for the ATE was (0.024, 0.296).

The plots in the appendix are used to check whether the assumptions of positivity and exchangeability hold. Figure 5 depicts the standardised mean difference for the unadjusted and adjusted data. Where unadjusted relates to the data prior to propensity score weighting and adjusted relates to the weighted data. As a rule of thumb 0.1 is taken as a threshold for well-adjusted covariates. The covariates in our adjustment set (BMI, age and systolic blood pressure) all fall short of this threshold. To verify positivity we have plotted the density of the propensity scores split by the exposure groups in Figure 6, which illustrates that there is sufficient overlap.

7 Discussion

In our G-computation analysis, we computed an ATE of 8.2% points. Implying that taking BPMEDS increases an individual's risk of death by said amount. The positive ATE from our propensity score analysis supports the result we found in our G-computation model. This finding is unexpected based on our medical understanding prior to the analysis. However, the confidence interval for the G-computation ATE includes 0 indicating that the estimate of our primary analysis is not statistically significant. This raises questions about the validity of the underlying assumptions required for correct causal inference.

Firstly, there is insufficient information to assume consistency is met. The provided information simply states "use of anti-hypertensive medication" and therefore the branding, dosage, and timing of the taken treatment is unknown. Thus, we cannot infer that the exposure is consistent across participants without more information. Secondly, we believe that the assumption of conditional exchangeability is violated. In our G-computation analysis, the distribution of the exposure and unexposure groups differ for each confounder in our adjustment set. This is demonstrated by the substantial difference in quartile values for age and systolic blood pressure depicted in Figure 2. These discrepancies suggest that the treatment groups may not be comparable. This conclusion is further supported by the imbalance in group proportions since only approximately 3% of participants received the medication, given that the sample is so small it may not be representative of the population. Our sensitivity analysis reinforces this conclusion, as shown by the adjusted standardised mean differences being larger than the 0.1 threshold value in Figure 5. The aforementioned boxplots demonstrate that the positivity assumption holds for the G-computation model, as there is considerable overlap among the values of each confounder per treatment level. This is also supported by the overlap in propensity values shown in the density plots in Figure 6. Given that consistency does not appear to be met and conditional exchangeability is violated, the identifiability of the causal estimand cannot be guaranteed.

Alongside the violations of the required assumptions, there are some limitations to the dataset. The purpose of the Framingham study is to evaluate causal relationships with longitudinal data. However, our dataset has been simplified to a cross-sectional study where the given explanatory variables are recorded from the first clinical visit. Removing the data from following clinical visits will reduce the accuracy of our estimates. Additionally, we created the DAG from our own medical knowledge which could be incorrect. We also assumed that the dataset includes all confounders. It is possible that there are unmeasured confounders such as other health indicators like diabetes or cholesterol levels. Perhaps adjusting for these variables would give us conditional exchangeability and more accurate estimates.

8 Conclusion

To summarise, we found a positive ATE indicating an increase in the risk of death for the treatment group. This is contrary to our initial belief. However, we believe that the identifiability conditions are not met due to the limitations of the given dataset. Consequently, our estimate for the ATE is not an accurate reflection of the causal effect of interest.

References

[1] *About the Framingham Heart Study.* The National Heart, Lung, and Blood Institute. URL: `www.framinghamheartstudy.org/fhs-about/` (visited on 03/20/2024).

[2] Miguel A Hernán and James M Robins. "Using big data to emulate a target trial when a randomized trial is not available". In: *American journal of epidemiology* 183.8 (2016), pp. 758–764.

[3] Judea Pearl. *Causality.* Cambridge university press, 2009.

A Appendix

A.1 Models

```
glm(formula = DEATH ~ BPMEDS + BMI + SYSBP + AGE, family = binomial, data = cc_fram)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.96514    0.34374  -23.17   <2e-16 ***
BPMEDS       0.43327    0.20600    2.10    0.035 *
BMI          0.00348    0.00929    0.37    0.708
SYSBP        0.01580    0.00186    8.50   <2e-16 ***
AGE          0.09995    0.00473   21.12   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5464.7  on 4240  degrees of freedom
Residual deviance: 4572.0  on 4236  degrees of freedom
AIC: 4582

Number of Fisher Scoring iterations: 4
```

Figure 3: G-computation model

```
glm(formula = BPMEDS ~ BMI + SYSBP + AGE, family = binomial,
    data = cc_fram)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.21834    0.82420  -14.82   <2e-16 ***
BMI           0.02883    0.01853    1.56   0.1197
SYSBP         0.03992    0.00353   11.32   <2e-16 ***
AGE           0.04204    0.01240    3.39   0.0007 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1230.38  on 4240  degrees of freedom
Residual deviance:  991.89  on 4237  degrees of freedom
AIC: 999.9

Number of Fisher Scoring iterations: 7
```

Figure 4: Propensity score model

A.2 Figures

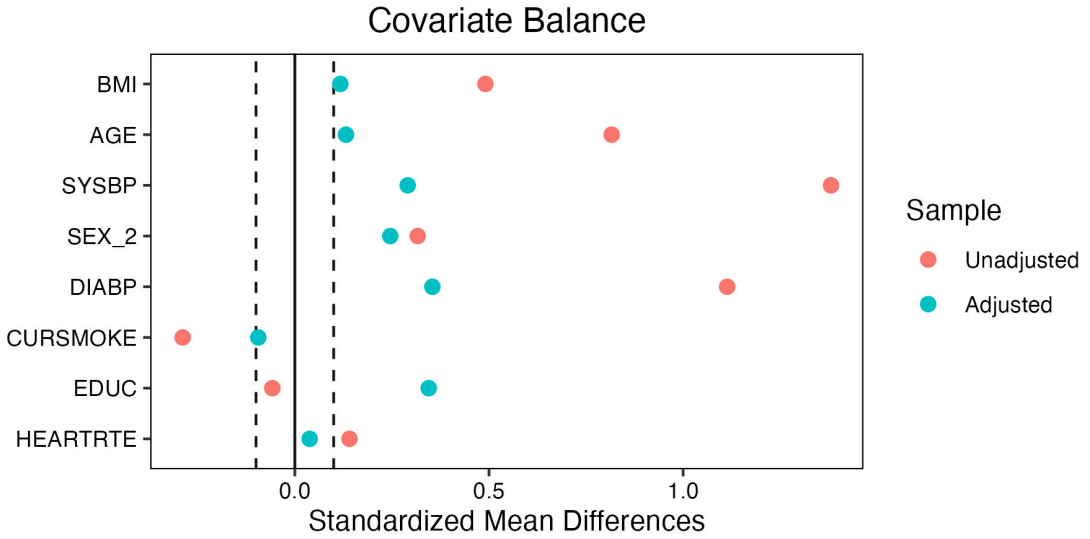


Figure 5: Exchangability check (propensity score model)

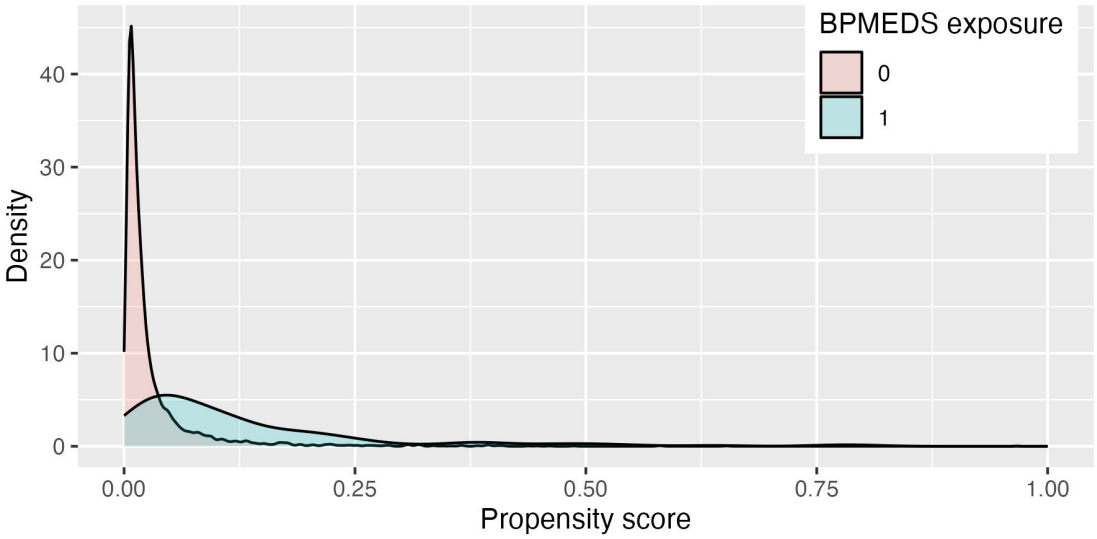


Figure 6: Positivity check (propensity score model)