Biometrics 0, 1–7 DOI: 0000-0000-0000
January 2023

Average treatment effect of cholesterol-lowering medication and average systolic blood pressure (SBP), mm Hg.

Jackson Gazin

Department of Statistics, Wake Forest University

*email: gazij22@wfu.edu

and

Ashley Mullan

Department of Statistics, Wake Forest University

*email: mullae22@wfu.edu

and

Anh Nguyen

Department of Statistics, Wake Forest University

*email: nguyp22@wfu.edu

Summary: We investigate the average treatment effect among the treated (ATT) of cholesterol-lowering medication on the mean systolic blood pressure (mm Hg). Using data from the National Health and Nutrition Examination Survey (NHANES), we fit a propensity score model to estimate the ATT among adults living in the United States of America.

Key words: cholesterol-lowering medicationsystolic blood pressureblood pressure control.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Controlling blood pressure (BP) reduces the risk for cardiovascular disease. However, the prevalence of BP control (i.e., systolic BP < 140 and diastolic BP < 90) among US adults with hypertension has decreased since 2013. We invite teams to analyze publicly available data from US adults to help identify potential causes or correlates of worsening BP control among US adults with hypertension over the past decade, as this may allow for development of effective interventions to help control BP and prevent cardiovascular disease.

2. Materials and methods

2.1 *Data*

The National Health and Nutrition Examination Survey (NHANES) combines interviews and physical examinations to assess the health and nutritional status of adults and children in the United States of America. The program started in the early 1960s and has been conducted every two years since 1999. The survey samples from a nationally representative 5,000 persons each year. The participants are located in counties across the country, 15 of which are visited each year. The interview asks questions about demographic, socioeconomic, dietary, and health-related questions. The examination consists of medical, dental, physiological measurements, and laboratory tests.

The NHANES dataset we are using can be downloaded from **cite** . The dataset contains information from the survey from 1999 to 2020 with a sample of 59,799 rows and 111 chosen columns. Each row is a noninstitutionalized US adults participated in the survey between 1999 and 2020. The columns contain information about demographics, blood pressure levels, hypertension status, antihypertensive medication usage, and co-morbidities.

For this analysis, we had 38977 rows with NA values. We decided to deal with this by

removing all the na values. We decided this was preferable to removing certain columns since we were still left with 20,822 data points which is still an extremely large data set.

2.2 *Statistical methods*

We fitted a propensity score model with "Total cholesterol" in mg/dL as our explanatory variable since this was the only variable, and "taking cholesterol medication" as our response variable, as it represents our exposure variable. We employed a Logistic Regression model for the propensity score model.

Subsequently, we used the propensity score model to fit our outcome model, estimating the average treatment effect among the treated. We utilized a linear regression model with "cholesterol medication" as our explanatory variable and "Systolic blood pressure (SBP)" in mm Hg as our response variable.

To assess the appropriateness of our propensity score model and proceed with our final model, we employed weighted mirrored histograms, ECDF plots, and Love Plots.

2.2.1 *Exploratory data analysis.*

2.2.2 *Modeling.*

3. Results {$results}

3.1 *Study population*

We aim answer our causal question by fitting an average treatment effect among the treated. Our causal question is as follows: Among those who take cholesterol lowering medication, does taking this cholesterol lowering medication change their systolic blood pressure?

3.2 *Propensity score model and Diagnostics*

We fitted a propensity score model with "total cholesterol" in mg/dL as our explanatory variable and "taking cholesterol medication" as our response variable. We employed a Logistic

Regression model for this purpose. Subsequently, we examined a Mirrored Histogram of our propensity scores for both exposure groups. The table below demonstrates significant overlap in propensity scores across both groups, indicating very little evidence of a positivity violation in our model, which is promising.

Next, we used our propensity score model to generate weights for the average treatment effect among the treated (ATT), aligning with our causal question. To assess the appropriateness of our propensity score model and the resulting weights, we created a Weighted Mirror Histogram of our propensity scores. As shown below, we achieved sufficient balance between the exposed and unexposed groups. Furthermore, the distribution of the unexposed group now resembles that of the exposed group, which is the desired outcome when using ATT weights.

We also generated a love plot, displaying the standardized mean difference changes for the exposed and unexposed groups regarding our "total cholesterol" variable in both unweighted and weighted data. As expected, our weighted data exhibit a standardized mean difference of 0 for the "total cholesterol" variable, which is ideal and allows us to proceed.

Finally, we created a weighted empirical cumulative distribution function (eCDF) plot for our continuous variable, "total cholesterol." As shown below, the Weighted ECDF plot indicates balance not only in the mean across exposure groups but also across their respective distributions. No changes are required for our model, and we can now proceed to estimate our average treatment effect among the treated.

3.3 *Average treatment effect among the treated*

We estimated the average treatment effect of taking cholesterol-lowering medication. Our findings indicate that, on average, individuals who take cholesterol-lowering medication experience an increase of 9.26 mm Hg in their Systolic Blood Pressure (SBP). We are 95 percent confident that the average effect of taking this medication on SBP falls within the range of at

least 8.578 mm Hg to at most 9.941 mm Hg. Additionally, we conducted a sensitivity analysis to account for the possibility that our Directed Acyclic Graph (DAG) might not include all potential confounding variables. In this analysis, we calculated the necessary relationship between an unmeasured confounder and the change in weight required to shift the lower bound of the confidence interval to the null hypothesis level (5%). We considered exposure confounder effects of sizes 0.05, 0.10, and 0.15, which would necessitate confounder-outcome effects of 172, 85, and 57, respectively, to influence the interval. The smallest of these values, 57, is over 6 times greater than the estimated treatment effect. Therefore, it is reasonable to conclude that the Average Treatment Effect (ATT) we computed is robust against potential confounding factors.

4. Discussion

5. Supplementary information

The data can be downloaded from GitHub or accessed via the cardioStatsUSA R package. For both the file and information about the R package, see `https://github.com/jhs-hwg/cardioStatsUSA`.

All code for the analysis can be accessed at **link github**

6. Acknowledgement

7. Section title

Text with citations by Heagerty et al. (2000), (Pepe, 2003).

7.1 *Subsection title*

as required (Hoerl and Kennard, 1970; Zou and Hastie, 2005). Don't forget to give each section and subsection a unique label (see Sect. 7).

Paragraph headings. Use paragraph headings as needed.

7.2 *Equations*

Here is an equation:

$$f_X(x) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; \alpha, \beta, x > 0$$

Here is another:

$$a^2 + b^2 = c^2 \tag{1}$$

Inline equations: $\sum_{i=2}^{\infty}\{\alpha_i^{\beta}\}$

8. Figures and tables

8.1 *Figures coming from R*

Normal figure embedded in text.

```
## Warning in plot.formula(runif(25) ~ runif(25)): the formula 'runif(25) ~
## runif(25)' is treated as 'runif(25) ~ 1'
```

[Figure 1 about here.]

8.2 *Tables coming from R*

```
print(xtable::xtable(head(mtcars)[,1:4],
caption = "Caption centered under table", label = "tab1"),
comment = FALSE, timestamp = FALSE, caption.placement = "top")
```

[Table 1 about here.]

Table 1 shows these numbers. Some of those numbers are plotted in Figure **??**.

```
head(mtcars[,1:4])
```

```
##                   mpg cyl disp  hp
## Mazda RX4         21.0   6  160 110
## Mazda RX4 Wag     21.0   6  160 110
## Datsun 710        22.8   4  108  93
## Hornet 4 Drive    21.4   6  258 110
## Hornet Sportabout 18.7   8  360 175
## Valiant           18.1   6  225 105
```

References

Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics* **56,** 337–344.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12,** 55–67.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction.* Oxford University Press.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67,** 301–320.

Received Dec 2023

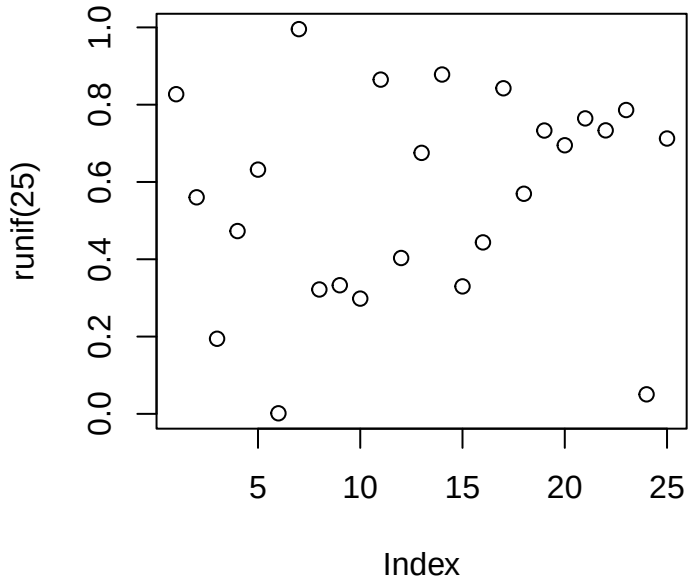

Figure 1. Output from `pdf()`

Table 1
Caption centered under table

	mpg	cyl	disp	hp
Mazda RX4	21.00	6.00	160.00	110.00
Mazda RX4 Wag	21.00	6.00	160.00	110.00
Datsun 710	22.80	4.00	108.00	93.00
Hornet 4 Drive	21.40	6.00	258.00	110.00
Hornet Sportabout	18.70	8.00	360.00	175.00
Valiant	18.10	6.00	225.00	105.00