The Impact of Aspirin Daily Treatment in 8 Chronic Disease on Healthcare Utilization

Matthew Onimus[1]

[1] Merck & Co C/O Jefferson Univserity

Author Note

Abstract

 Daily aspirin treatment has been prescibed for any number of different chronic diagnoses ranging from heart failure to cancer. This study attempts to understand the impact of healthcare utlization in total costs and utilizer category in regards to aspirin treatment on 8 different chronic conditions. The chronic conditions or diseases of interest are Alzheimer's, asthma, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, diabetes, hypertension, and osteoarthritis. The total healthcare utlization costs are defined as the sum of inpatient, primary care physician, and emergency costs within the period window either pre or post treatment. The utilizer category is defined in 3 groups: low, moderate, or high utilizer within the period window either pre or post treatement.

Aspirin is a low cost treatment option that has already been shown to prolong life possibly due the anti-inflammorty properties(Vane & Botting, 2003). To date, and this author did not search all that extensively, no study has been performed to understand the impact of daily aspirin treatment on healthcare utilization and/or utilizer category.

This researcher believes that the daily aspirin treatment will lead to a decrease total healthcare utilization per each chronic condition or disease but the decrease will not be enough to change the patients utilizer category.

Differences were assessed on each chronic condition or disease category using the standard difference-in-difference method. The change in utilizer category was first created and then either binomial or multinomial logistic regression was applied to understand the change in category based upon treatment for each disease or condition.

Keywords: Aspirin, Healthcare Utilization, Difference in Difference, Logistic Regression

The Impact of Aspirin Daily Treatment in 8 Chronic Disease on Healthcare Utilization

Data and Analysis Overview

All of the data used to complete this analysis was obtained from the AHE505 data files located in SAS. Any patients without a preexisting disease or condition were removed from the main data set.

To evaluate differences in total health care utilization per member per month, a difference-in-difference approach was used. Either a binomial or multinomial regression was used to assess utilizer category changes change depending on the number of changes present within the subsetted data.

The pre-treatment data begin at some point and end at some point later. The post-treatment data begin at some later point and ended a little later than that.

To prepare the data set for analysis, individual data sets were created for each disease or condition. A summary of the resulting descriptive statistics for age and sex is provided in Table 1.

Difference in Difference Analysis

To estimate the effect of daily aspirin treatment on health care utilization, each disease or condition type was subset. To each subset of data a linear model was fit with a gamma distribution with a log link using the time, treatment, and interaction of the two terms. Once a model was fitted, the $Wald - \chi^2$ test was performed on the interaction term (treatment*time) to determine statistical significance. Finally, a difference-in-difference is calculated based on the time and treatment groups. A summary of these results is presented in Table 2.

Binomal or Multinomial Regression for Utilizer Category

To prepare the data for utilizer category analysis, a new variable, utilizer change, was created based on the difference of the post utilizer category and the pre utilizer category. There are 6 possible utilizer change outcomes. The reference level, 0, signifies no utilizer change. If the utilizer category decreased, determined by a -1, -2, or -3 in the utilizer change variable, these were marked as small decrease, moderate decrease, and large decrease, respectively. If the utilizer category increased, determined by a 1, 2, or 3 in the utilizer change variable, these were marked as small increase, moderate increase, or large increase, respectively. In the dataset presented here, only small decrease, no change, and small increase were possible outcomes.

To estimate the effect of the daily aspirin treatment on utilizer category, each disease or condition type was subset and to each subset of data the appropriate regression model was used based upon the number of category change outcomes and the treatment variable. Once a model was fitted, the p-value for the treatment variable was calculated to determine statistical significance (refer to Table 3). Finally, relative risks were calculated for each model based on treatment (refer to Table 4).

Results

Difference in Difference Results

For the difference-in-difference analysis, the $Wald - \chi^2$ p-value for Alzheimer's, asthma, coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, diabetes, and osteoarthritis were all > 0.1, indicating the results are not statistically significant. Hypertension, which had the largest sample size, was the only condition which had a $Wald - \chi^2$ p-value of 0.09 indicating the result is statistically significant at a significance level of 0.1.

Additionally, all of the difference-in-difference results were negative numbers indicating

the treatment (daily aspirin dosing) actually increased the health care utilization of the treatment group. The increases ranged from ~$21.00 to ~$49.00 per member per month.

Binomal or Multinomial Regression for Utilizer Category

For the binomal or multinomial regression analysis, the p-value was only significant (at a 0.05 signficance level) for 6 of the 8 disease types; asthma, coronary artery disease, chronic obstructive pulmonary disease, diabetes, hypertension, and osteoarthritis.

For the 6 statistically significant disease and condition types, relative risk ratios were calculated. As a reminder, the reference level was no change (0) and each of the relative risks are presented as changing from no treatment to treatment. For example, the relative risk of a patient with asthma switching from no treatment to treatment (daily aspirin) is 1.75 for being in the no change category compared to the small decrease category. For that sample asthma patient, the relative risk of switching from no treatment to treatment is 2.08 for being in the no change category to the small increase category.

Discussion

The difference-in-difference and the utilizer category results were opposite of the expectations. Daily aspirin treatment resulted in an increase in health care utilizer per member per month by modestly large amount for some disease or condition categories despite not being statistically significant. The regression analysis for utilizer change category resulted in more statistically significant outcomes, showing that treatment resulted the possibility of both small increases or decreases in utilizer change category.

R Packages

We used R (Version 3.6.3; R Core Team, 2020) and the R-packages *dplyr* (Version 1.0.5; Wickham et al., 2021), *forcats* (Version 0.5.0; Wickham, 2020), *ggplot2* (Version 3.3.3; Wickham, 2016), *haven* (Version 2.3.1; Wickham & Miller, 2020), *kableExtra* (Version 1.3.1; Zhu, 2020), *MASS* (Version 7.3.51.5; Venables & Ripley, 2002a), *nnet* (Version 7.3.15; Venables & Ripley, 2002b), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *readr* (Version 1.4.0; Wickham & Hester, 2020), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.1; Müller & Wickham, 2021), *tidyr* (Version 1.1.3; Wickham, 2021), and *tidyverse* (Version 1.3.0; Wickham, Averick, et al., 2019) for all our analyses.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames.* Retrieved from https://CRAN.R-project.org/package=tibble

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Vane, J. R., & Botting, R. M. (2003). The mechanism of action of aspirin. *Thrombosis Research, 110*(5), 255–258. https://doi.org/https://doi.org/10.1016/S0049-3848(03)00379-7

Venables, W. N., & Ripley, B. D. (2002a). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Venables, W. N., & Ripley, B. D. (2002b). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from https://www.stats.ox.ac.uk/pub/MASS4/

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2020). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2021). *Tidyr: Tidy messy data.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Hester, J. (2020). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Wickham, H., & Miller, E. (2020). *Haven: Import and export 'spss', 'stata' and 'sas' files.* Retrieved from https://CRAN.R-project.org/package=haven

Zhu, H. (2020). *KableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra

Table 1

Descriptive Statistics by Disease/Condition

dis	meanAge	sdAge	female	male
ALZ	55.46	15.12	135	181
AST	54.32	12.19	4621	3486
CAD	57.92	9.80	4145	1752
CHF	57.28	10.61	877	499
COPD	55.34	11.74	1869	2254
DIAB	55.28	11.02	6935	5323
HTN	55.44	10.78	16844	14513
OSTEO	60.47	8.27	83	1660

Table 2

Difference in Difference Result Summary

disease	sampleSize	diffInDiff($)	wald
ALZ	316.00	-27.10	0.89
AST	8,107.00	-31.53	0.40
CAD	5,897.00	-36.86	0.48
CHF	1,376.00	-49.08	0.77
COPD	4,123.00	-28.17	0.56
OSTEO	1,743.00	-21.93	0.68
HTN	31,357.00	-23.41	0.09
DIAB	12,258.00	-27.94	0.31

Table 3

Logistic Regression Summary

disease	regressionType	treatmentPvalue
ALZ	binomial	0.44
AST	multinomial	0.00
CAD	multinomial	0.00
CHF	multinomial	0.07
COPD	multinomial	0.00
OSTEO	multinomial	0.00
HTN	multinomial	0.00
DIAB	multinomial	0.00

Table 4

Relative Risks Summary

disease	smallDecrease	smallIncrease
AST	1.75	2.08
CAD	3.20	2.00
COPD	2.59	2.08
OSTEO	1.10	2.14
HTN	1.55	2.08
DIAB	2.23	1.90