The Effects of Medical Expenditures and Smoking on the Incidence of Myocardial Infarction

June 10, 2024

Catharine Bowman

Repositories:

Repositories: **Original Cohort Data Repository** and **Final Project Repository with Data, Analysis, README, and Quarto Report Files**.

Introduction

Cardiovascular disease is a major contributor to global morbidity and mortality. It is estimated that the number of cardiovascular disease cases has double over the last thirty years with 271 million cases reported in 1990 and 523 million reported in 2019 Roth et al. (2020). In the United States alone, it was estimated that cardiovascular disease contributed to 168.2 per 100,000 deaths in 2020 and $320 billion in healthcare spending "Mortality in the United States, 2020" (2021); Birger et al. (2021). Myocardial infarction, one of the leading killers among cardiovascular disease patients, is estimated to increase in its prevalence by 30.1% by 2060, emphasizing the need to better understand causes, associated risk factors, and potential points of intervention Mohebi et al. (2022). Hence, this analysis explores the relationships between myocardial infarction, smoking, and annual out-of-pocket healthcare expenditure.

Furthermore, this report also demonstrates the various skills learned in EPI 203 Methods for Reproducible Population Health and Clinical Research. Throughout this project, I aimed to demonstrate file folder organization, version control with Github, principles of effective coding, and the generation of a reproducible and dynamic Quarto document. Peer-to-peer code review would also be an important and anticipated step outside of the context of a written assignment.

Methods

Packages

The following packages were used for this report: **table1() and methods**, **arsenal()**, **tidyverse()**, **dplyr()** and **rmeta()**.

Study Design and Dataset

This is a retrospective cohort study with four assumptions (see section below, *Assumptions*).

This study assesses the relationship between annual out-of-pocket costs of medical care, smoking history and myocardial infarction, as defined by variables, `cost` and `cardiac`, in the [**cohort data set**](https://github.com/MethodsForReproducibleHealthResearch/Assignment7.git "cohort dataset"). This data set includes 5,000 observations from adult aged 18-65 years and 5 variables, including `smoke`, `female`, `age`, `cardiac` and `cost`.

File Read in and Initial Exploration

The variables `smoke`, `female`, and `cardiac` are binary. A value of 1 represents an observation who reported female (`sex`), has previously smoked (`smoke`) or has a history of myocardial infarction (`cardiac`). A value of 0 represents a non-female participant (coded as `male` for the purposes of this analysis), no smoking history, or no incident myocardial infarction. The variables `age` (measured in years) and `cost` (measured in US dollars) are continuous variables.

Statistical Analyses

Assumptions

We assumed the `cost` variable represents annual cost of out-of-pocket care measured in US dollars and `cardiac` represents the presence of incident myocardial infarction. Furthermore, we assumed time ordering was upheld between exposures and outcome, such that annual cost of out-of-pocket care and smoking were measured prior to the myocardial infarction outcome. Finally, we recognize several confounding factors are not accounted for in this analysis.

Data Cleaning Procedures

Data cleaning began with an initial exploration of the available date. The mean, median, and range of all continuous variables were calculated to detect extreme values (or values that are not biologically plausible, such as age=200 years). Furthermore, histograms of continuous variable were produced to assess distributional characteristics (e.g. normality). Similarly the crude count data and proportions of binary variables were calculated.

Data processing included the transformation of categorical variables into factors and renaming for data visualization. For instance, the variable `cardiac` was re-labelled as `CardiacEvent` with categories of 'No Myocardial Infarction' for `cardiac=0` or 'Myocardial Infarction' for `cardiac=1`. The variable `female` was converted into a factor and renamed so that `female=0` indicated a male observation and `female=1` indicated a female observation. The variable `smoke` was re-labelled `Smoking` with `smoke=0` indicating a non-smoker and `smoke=1` indicated a smoker. Finally, the variable `age` was renamed `Age` for presentation.

Analytic Plan

Because we are using a binary outcome variable, myocardial infarction, logistic regression was used to evaluate its relationship with both annual out-of-pocket medical care and smoking. Cost of care was transformed from a continuous to categorical variable due to practical considerations for policy development and organization. Categories were defined as <$9800, $9801-$9950, and >$9950 to minimize data sparsity and optimize equal distribution of observations.

Table 1 Overview

A table 1 was created using the **table1() package** The absolute count and percentage of each exposure category was presented for annual out-of-pocket care cost categories, (<$9800, $9801-$9950, and >$9950), smoking history (non-smoker and smoker), and sex (male and female). The mean (with standard deviation) and median (with minimum and maximum) were reported for age. Each metric was presented for the overall cohort (n=5000) and stratified by myocardial infarction status.

Regression Models

Crude and adjusted odds ratios with 95% confidence intervals were generated using logistic regression (`glm()`). Adjusted models included age and sex. The outcome of interest was incident myocardial infarction and the exposure for the first model was annual out-of-pocket cost of care (USD) category and the second model was smoking history (history present/absent).

Data Visualization

Forest plots were generated using `ggplot()` and methods outlined by **Dayimu A. (2024)**. Crude and adjusted odds ratios and 95% confidence intervals were stored in a new dataset, which was used in `ggplot()` for visualization.

Results

Cohort Description and Demographics

This dataset includes 5,000 observations, 4810 of which did not have a myocardial infarction reported. Approximately, 66% of individuals who had a cardiac event reported annual out-of-pocket care costs of >$9950, compared to 21.4% of individuals without a cardiac event documented. The majority of individuals with a cardiac event did not have a history of smoking (59.5%); a finding that was shared with the non-cardiac event group (91.0%). There were more male participants in the cardiac event group than females (90.0%). Finally, the mean age across both cardiac and non-cardiac event groups was approximately 40 years.

Table 1: Demographic and clinical characteristics of cohort, overall and stratified by myocardial infarction status

	No Myocardial Infarction	Myocardial Infarction	Overall
	(N=4810)	(N=190)	(N=5000)
CostCategories			
<$9800	3161 (65.7%)	42 (22.1%)	3203 (64.1%)
>$9950	1031 (21.4%)	125 (65.8%)	1156 (23.1%)
$9801-$9950	618 (12.8%)	23 (12.1%)	641 (12.8%)
Sex			
Male	2394 (49.8%)	171 (90.0%)	2565 (51.3%)
Female	2416 (50.2%)	19 (10.0%)	2435 (48.7%)
Age			
Mean (SD)	41.5 (13.5)	39.5 (13.7)	41.5 (13.5)
Median [Min, Max]	41.0 [18.0, 65.0]	39.0 [18.0, 64.0]	41.0 [18.0, 65.0]

Association Between Cost of Care and Myocardial Infarction

The results of this investigation revealed that increased annual out-of-pocket costs of care (as defined by categories of the `cost` variable) were associated with an increased odds of having experiencing a myocardial infarction (as defined by the variable `cardiac`).

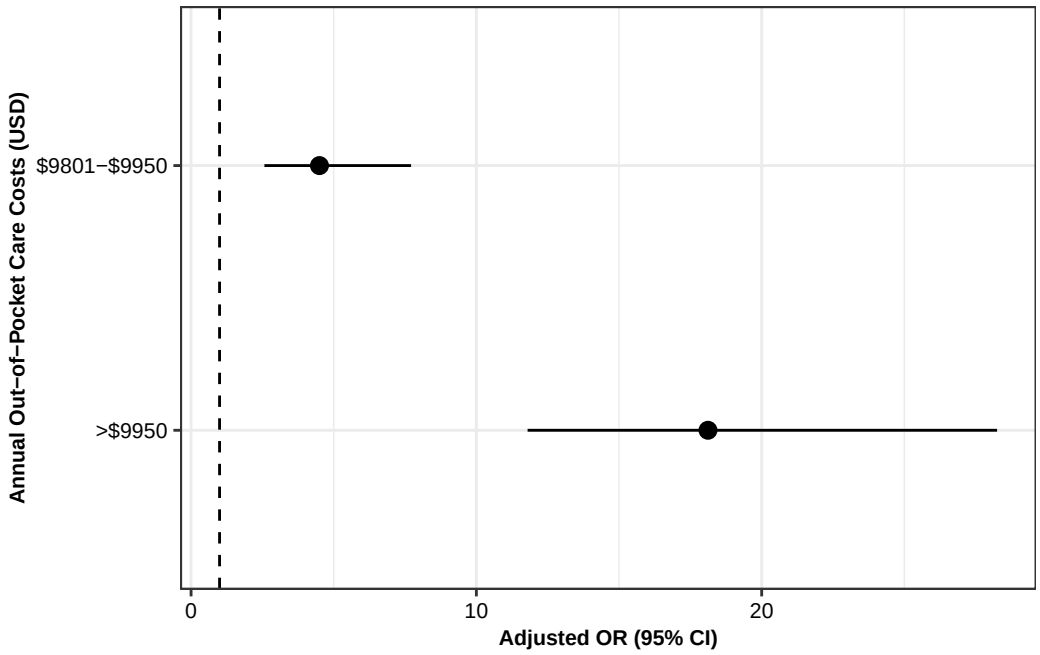
Crude Model

The results of the crude model indicated that having annual out-of-pocket care costs between $9800 and $9950 was associated with an increased odds of myocardial infarction of 2.80 [95% CI 1.65, 4.65] when compared to the reference cost of less than $9800. Furthermore, participants with care costs greater than $9950 had an increased odds of myocardial infarction of 9.12[95% CI 6.45,13.17] relative to the reference group (Table 2a).

Adjusted Model

After adjustment for sex and age having annual out-of-pocket care costs between $9800 and $9950 was associated with an odds of myocardial infarction of 4.50[95% CI 2.57, 7.71] when compared to the reference cost of less than $9800. Furthermore, participants with care costs greater than $9950 had an increased odds of myocardial infarction of 18.12 [95% CI 11.80, 28.25] relative to the reference group (Figure 1, Table 2a).

Figure 1: Adjusted Odd Ratios of Myocardial Infarction by Annual Cost of Care (USD)



Association Between Smoking History and Myocardial Infarction

The relationship between smoking history (as defined by the variable `smoke`) and myocardial infarction (as defined by the variable `cardiac`) were also explored.

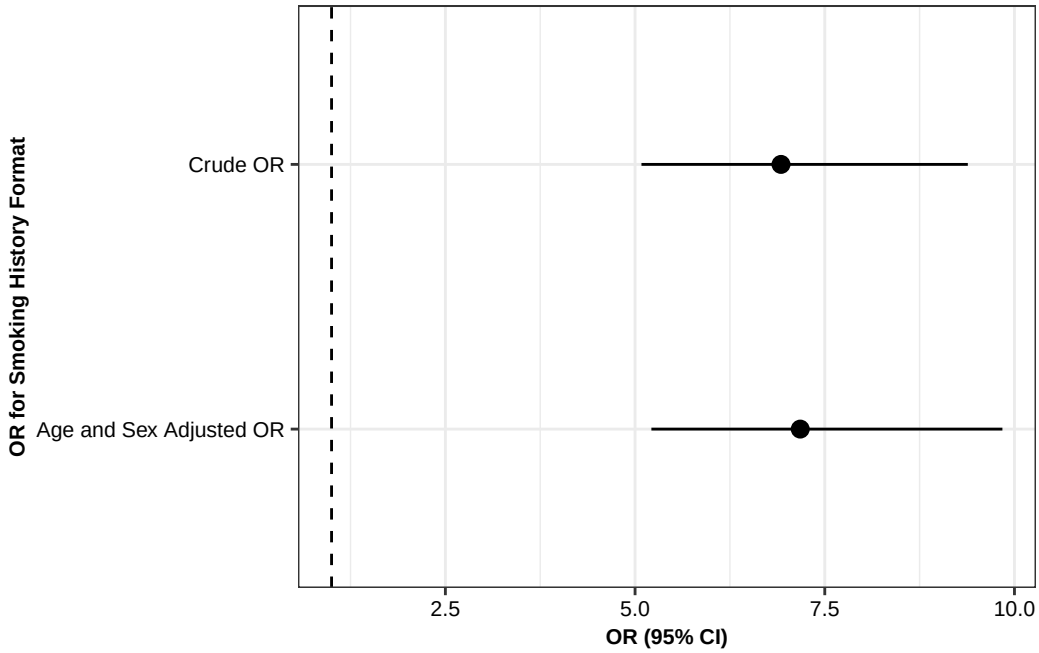
Crude Model

The results of the crude model indicate that having a smoking history is associated with an odds of myocardial infarction of 6.92[95% CI 5.08, 9.39] when compared to non-smokers (Table 2b).

Adjusted Model

After adjustment for sex and age, the results continued to demonstrate that having a smoking history was associated with an odds of myocardial infarction of 7.18[95% CI 5.21, 9.84] when compared to non-smokers (Figure 2, Table 2b).

Figure 2: Odds Ratios of Myocardial Infarction by Smoking Status



Data Summary and Conclusion

This analysis demonstrated evidence of increased odds of myocardial infarction among adults with higher annual out-of-pocket costs of care (relative to annual expenditure <$9800) and among smokers (compared to non-smokers). This finding was consistent across crude and age- and sex-adjusted models (Tables 2a and 2b).

Table 2a: Summary Odds Ratios For Main Effect of Expenditure in Relationships Between Out-of-Pocket Care Expenditure and Myocardial Infarction

Measure	Parameter	Odds Ratio	95% Confidence Interval
Crude Model: Intercept	Intercept	0.01	0.01, 0.02
Crude OR	Expenditure: $9801-$9950	2.80	1.65, 4.65
Crude OR	Expenditure: >$9950	9.12	6.44, 13.17
Adjusted Model: Intercept	Intercept	0.18	0.12,0.30
Adjusted OR	Expenditure: $9801-$9950	4.50	2.57, 7.71
Adjusted OR	Expenditure: >$9950	18.12	11.80, 28.25
Adjusted OR	Age	0.94	0.93, 0.95
Adjusted OR	Sex (Ref: Male)	0.22	0.13, 0.36

Table 2b: Summary Odds Ratios For Main Effect of Smoking in Relationships Between Smoking and Myocardial Infarction

Measure	Parameter	Odds Ratio	95% Confidence Interval
Crude Model: Intercept	Crude Intercept	0.03	0.02, 0.03
Crude OR	Smoking	6.92	5.08, 9.39
Adjusted Model: Intercept	Adjusted Intercept	0.07	0.04, 0.11
Adjusted OR	Smoking	7.18	5.21, 9.84
Adjusted OR	Age	0.99	0.98, 1.00
Adjusted OR	Sex (Ref: Male)	0.11	0.06, 0.17

Given that both smoking and increased care expenditure are associated with an increased odds of myocardial infarction, this analysis presents a potential avenue for future intervention aimed towards reducing the overall burden of cardiovascular disease.

References

Birger, Maxwell, Alexander S. Kaldjian, Gregory A. Roth, Andrew E. Moran, Joseph L. Dieleman, and Brandon K. Bellows. 2021. "Spending on Cardiovascular Disease and Cardiovascular Risk Factors in the United States: 1996 to 2016." *Circulation* 144 (4): 271–82. https://doi.org/10.1161/circulationaha.120.053216.

Mohebi, Reza, Chen Chen, Nasrien E. Ibrahim, Cian P. McCarthy, Hanna K. Gaggin, Daniel E. Singer, Emily P. Hyle, Jason H. Wasfy, and James L. Januzzi. 2022. "Cardiovascular Disease Projections in the United States Based on the 2020 Census Estimates." *Journal of the American College of Cardiology* 80 (6): 565–78. https://doi.org/10.1016/j.jacc.2022.05.033.

"Mortality in the United States, 2020." 2021. https://doi.org/10.15620/cdc:112079.

Roth, Gregory A., George A. Mensah, Catherine O. Johnson, Giovanni Addolorato, Enrico Ammirati, Larry M. Baddour, Noël C. Barengo, et al. 2020. "Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019." *Journal of the American College of Cardiology* 76 (25): 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010.