The Effect of Age, Smoking Status, and Gender on Incident Myocardial Infarction, Stroke and Diabetes: Data from the Framingham Heart Study

Group 4: Xuan Pu, Tyler Richter, Abhishek Sathe, Meghna Singh, Mary Whipple, Peixuan Zhang

Abstract

Age is a common risk factor for many diseases, yet as an independent variable it cannot be manipulated or changed systematically by investigators. This study examined the effect of age as modified by smoking status and gender on the odds of incident myocardial infarction (MI)/stroke and diabetes. We generated two separate datasets (one for MI/stroke and one for diabetes) from the Framingham Heart Study's publicly available data. We used a case-control sampling method to deal with the rare outcomes of interest by including all participants with the outcome (MI/stroke=1 or diabetes=1) and randomly sampled an equal number of controls from the remaining participants without the outcome. Correlated outcomes from the same participants were addressed by fitting separate Generalized Estimating Equations (GEE) to each dataset (MI/stroke and diabetes). Our analysis suggested that age was associated with increased odds for incident MI/stroke and the odds of developing diabetes after adjusting for body mass index, total cholesterol, and systolic blood pressure. Current smoking status did not modify the effect of age on MI/stroke or diabetes. There was evidence to suggest that the effect of age on development of MI/stroke varied across gender, where females showed 6% higher odds of MI/stroke for each additional year of age (compared to males). Major limitations of the analysis were the exclusion of participants who had less than three recorded timepoints (may lead to systematic bias if missing not at random) and lack of data on other known risk factors related to our outcomes of interest, such as arrhythmias, vascular disease, and hormone alterations. Future research should consider these important variables and others that may be modifiable risk factors for cardiovascular disease.

Introduction

Cardiovascular disease (CVD) was the cause of almost fifty percent of deaths in America by the year 1940, yet the prevention and treatment of CVD were poorly understood. In addition, the death of President Franklin Roosevelt from heart failure and the lack of understanding of his treatment and CVD risk factors, helped spur the creation of the Framingham heart study (FHS) [1] in 1948. FHS is a longitudinal study of CVD of community-dwelling adults in Framingham, Massachusetts. Started in 1948 with 5,209 initial participants (2336 men, 2873 women) who were asked to complete follow-up measures every 2 years. FHS was initially funded by the National Heart Act that allocated $500,000 for a 20-year epidemiological study of heart disease. Between 1966-1971 (around the end of the original 20-year tenure of the FHS), the study was briefly funded by private donors including some life-insurance companies before the National Heart Institute provided additional funding and recruitment of children of the original participants into a new off-spring cohort [2]. In 2002, the third generation cohort (grandchildren of the original cohort) was initiated to explore genetic contributions to CVD. In the past 50 years, FHS has produced more than 3000 articles in leading medical journals.

Analysis of FHS data found that high systolic blood pressure and high cholesterol were major risk factors of CVD and heart failure. Findings from FHS have helped identify other risk factors for CVD and have influenced clinical practice. Obese and overweight body mass index (BMI)

has been associated with increased CVD risk [3]. A number of epidemiologic studies have shown a strong association between cigarette smoking (both active and passive smoking) and CVD. Gender-based differences in the risk of CVD has also been shown by several studies. Compared to men, CVD develops in women 7 to 10 years later on average [4].

Age is a common risk factor for many diseases, yet as an independent variable it cannot be manipulated or changed systematically by investigators. Using data from the FHS study, the purpose of the present study was to evaluate the effect of age on the risk of CVD (myocardial infarction (MI) or stroke) as the primary outcome and risk of diabetes as a secondary outcome, after adjusting for systolic blood pressure, total cholesterol and BMI. We also aimed to assess if smoking status and gender modified the effect of age on the outcomes of interest.

Methods

Data Source

Data were obtained from the publicly available teaching data from the FHS. The teaching dataset used in this particular study is a subset of the data collected as part of the Framingham study and includes laboratory, clinic, questionnaire, and adjudicated event data on 4,434 participants. Data were collected during three examination periods, approximately 6 years apart, between 1956 and 1968. Each participant was followed for a total of 24 years for the outcome of the following events: angina pectoris, MI, stroke, or death.

Exclusion Criteria

To have a balanced dataset, participants who only have 1 or 2 repeated measurements were excluded, which left us a dataset with 3 observations in each cluster. Since our interest outcome is MI/stroke and diabetes, those who had those conditions at study entry time point were also excluded. A total of 1326 participants were excluded using this step. Finally, participants with missing values of body mass index (BMI) and total cholesterol measurements were excluded for available data analysis.

Statistical Analysis

Before removing missing values of BMI and total cholesterol, two-sample t-test or proportion tests were performed to compare the baseline characteristics of the missing value group and non-missing value group for both BMI and total cholesterol variables. After the exclusion with criteria stated before, there were 2785 participants in the final dataset. Considering a large number of clusters and rare events, where only 146 people (5%) developed MI/Stroke and 173 people (6%) developed diabetes during the study, a case-control sampling method was applied to generate datasets for analysis. For each interested outcome, MI/Stroke and diabetes, all the cases were kept and an equal number of control participants were randomly sampled from the remaining control observations. Finally, two datasets were generated, one for MI/stroke with 292 clusters and one for diabetes with 346 clusters in total. The baseline characteristics of case and control groups were also compared, as shown in **Table 1**.

To estimate the effect of age on the primary outcome (MI/stroke) and secondary outcome (diabetes), Generalized Estimating Equations (GEE) model with logistic link function was fit to account for correlation of repeated measurements. The correlation structure was set as auto-regressive (AR-1) because all the measurements were recorded at equally spaced time intervals. The base model to estimate the effect of age on outcomes of interest (primary or secondary) had covariates BMI, total cholesterol level (mg/dL) and systolic blood pressure

included in the model. To test if the effect of age would be modified by gender or current smoking status, interaction terms were separately added to the base model. This led to each outcome having three different GEE models. Sensitivity analysis was performed by using exchangeable correlation structure in the model. All the analyses were performed using R, version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) and SAS, version 9.4 (SAS Institute, Inc, Cary, NC). Statistical significance determined a priori was set at $p < 0.05$.

Results

The decision process used to obtain the datasets for analysis is shown in **Figure 1**. After excluding participants with less than three timepoints, those with an MI, stroke, or diabetes at baseline, and those with missing data for variables included in our analysis, we were left with n=146 participants with MI/stroke at follow-up and n=173 with diabetes at follow-up. An equal number of participants without each outcome were randomly selected for analysis. As indicated in **Table 1A**, participants in MI/Stroke cases had significantly higher systolic blood pressure (mean value of 140.84 vs. 130.70 in control), higher BMI (mean value of 26.93 vs. 25.78 in control), and there were fewer females (39.70% vs. 61.00% control) when compared to the MI/Stroke control group. A similar conclusion was obtained in diabetes case and control groups as shown in **Table 1B**.

Effect of age on the incidence of MI or stroke

Fitting a Binomial GEE model with logistic link and auto-regressive (AR-1) working correlation structure, we concluded that with each increasing year of age, the odds of having MI or stroke increased by 15.7% after adjusting for BMI, total cholesterol, and systolic blood pressure (95% Confidence Interval (CI): 12% to 19% increase in odds, $p < 0.0001$) (**Table 2**). Next, we examined if the effect of age differed by smoking status, by adding current smoking status and an interaction between current smoking status and age to the previous model. The results of this model indicated that for each increasing year of age, current smokers had 5% higher odds of having MI or stroke than non-smokers, with the 95% CI of 5% lower odds to 12% higher odds ($p = 0.073$). The results were not statistically significant and hence there was no evidence to suggest that the effect of age on MI or stroke differed by smoking status. Finally, we examined if there was evidence of an interaction between age and gender. Adding gender and an interaction term between gender and age to the first model, the results indicated that for each increasing year of age, females had 6% higher odds of having MI or stroke when compared to males, with the 95% CI indicating 0% lower odds to 14% higher odds ($p = 0.045$). The results using SAS yielded similar 95% CI values, with $p = 0.0526$. Although these results were borderline significant, it does suggest a potential interaction and further investigation beyond this cohort. Similar results were obtained when an exchangeable working correlation structure was used.

Effect of age on the incidence of diabetes

Fitting a Binomial GEE model with logistic link and AR-1 working correlation structure, we concluded that with each increasing year of age, the odds of having diabetes increased by 11% after adjusting for BMI, total cholesterol and systolic blood pressure (95% CI: 8% to 13% increase in odds, $p < 0.0001$) (**Table 3**). In a second model where we added current smoking status and an interaction term for current smoking and age, there was no evidence that the effect of age differed by smoking status. For each increasing year of age, current smokers had 3% higher odds of having diabetes than non-smokers, but the 95% CI indicated 1% lower odds to 8% higher odds ($p = 0.13$). Finally, we added gender and an interaction term between gender

and age to the first model. The model indicated that for each increasing year of age, females had 2.5% lower odds of having a stroke or MI than males, with the 95% CI indicating 7% lower odds to 2% higher odds ($p = 0.23$). Since the results were not statistically significant, we concluded that there was no evidence to suggest that the effect of age on the development of diabetes differed by gender. Use of an exchangeable working correlation structure instead of AR-1 did not substantially alter the model results.

Discussion

The results of this study suggest that age is associated with both increased odds of having a MI or stroke and odds of developing diabetes, after adjusting for BMI, total cholesterol, and systolic blood pressure which is in line with published findings [5]. We did not observe a statistically significant interaction between age and current smoking status for either outcome, but did observe a possible interaction between gender and age on the development of MI/stroke, with females having 6% higher odds of an MI/stroke for each additional year of age, when compared to males. Replication of these methods using a different random sample of controls from this study would be useful to determine if similar results are obtained. Another study on the same dataset [6] has concluded that gender differences do exist for stroke incidence and women have "higher lifetime risk of stroke at all ages".

There are several limitations to this study. First, as we limited our analysis to participants who had outcome measures at all three time points, the results may be biased. Outcomes for participants who had less than three timepoints may be missing not at random, potentially because these participants were sicker and thus were unable to attend a study visit or may have died prior to data collection. Additionally, as this study was based on secondary data analysis of a teaching dataset, we were unable to examine the effects of other potentially important variables, including lifelong smoking history (including pack years), diagnosed peripheral vascular disease (e.g., carotid, subclavian, lower extremities), cardiac arrhythmias, and menopausal/hormonal status. Previous research has demonstrated that peripheral artery disease is a coronary heart disease risk equivalent in both men and women [7], and thus, would be important to include in a future analysis of risk for incident MI, stroke, and diabetes. Atrial fibrillation is also a known risk factor for MI and stroke that may have altered our results, as risk for these conditions also increase with age [8,9]. Additionally, we did not control for the effect of antihypertensive medications in the analysis.

Conclusion

In summary, the results of this study add to the available information on the link between age and development of diabetes, MI, and stroke, independent of important cardiovascular disease risk factors. Additionally, our analysis supports a potential link between being female and increased risk of MI/stroke. Further research is needed to examine these relationships in a larger sample with additional potentially important variables.

References

1. Framingham Heart Study (FHS) Retrieved from https://www.nhlbi.nih.gov/science/framingham-heart-study-fhs

2. Mahmood, S. S, Levy, D., Vasan, R.S., Wang, T.J. (2013). The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 383 (9921): 999–1008. doi:10.1016/S0140-6736(13)61752-3

3. Khan SS, Ning H, Wikins JT, Allen N, Carnethion M, Berry JD, Sweis RN, Lloyd-Jones DM, Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity, JAMA Cardiol. 2018 Apr 1;3(4):280-287. doi: 10.1001/jamacardio.2018.0022.

4. Maas AH, Appelman YE. Gender differences in coronary heart disease. *Neth Heart J*. 2010;18(12):598-602.

5. Niccoli, T., & Partridge, L. (2012). Ageing as a Risk Factor for Disease. Current Biology, 22(17), R741–R752. doi.org/10.1016/j.cub.2012.07.024

6. Petrea, R. E., Beiser, A. S., Seshadri S, Kelly-Hayes, M., Kase, C. S., Wolf, P. A. (2009) Gender differences in stroke incidence and post-stroke disability in the Framingham Heart Study. Stroke 2009, 40:1032–1037.

7. Subherwal, S., Patel, M. R., Kober, L, Peterson, E. D., Bhatt, D. L., Gislason, G. H., Olsen, A. M., Jones, W. S., Torp-Pedersen, C., & Fosbol, E. L. (2015). Peripheral artery disease is a coronary heart disease risk equivalent among both men and women: Results from a nationwide study. *Eur J Prev Cardiol, 22*(3), 317-25. doi: 10.1177/2047487313519344.

8. Soliman, E. Z. Safford, M. M., Muntner, P., Khodneva, Y., Dawood, F. Z., Zakai, N. A., Thacker, E. L., Judd, S., Howard, V. J., Howard, G., Herrington, D. M., & Cushman, M. (2014). Atrial fibrillation and the risk of myocardial infarction. *JAMA Intern Med, 174(*1), 107-14. doi: 10.1001/jamainternmed.2013.11912.

9. Sanders, G. D., Lowenstern, A., Borre, E., Chatterjee, R., Goode, A., Sharan, L., LaPointe, N. M. A., Raitz, G., Shah, B., Yapa, R., Davis, J. K., Lallinger, K., Schmidt, R., Kosinski, A., & Al-Khatib, S. (2018). Stroke prevention in patients with atrial fibrillation: A systematic review update. Agency for Healthcare Research and Quality (US); 2018 Oct. Report No.:18-EHC018-EF. Rockville, MD..

***Figure 1*. Flow diagram of the process to obtain case-control cohorts for analysis.**

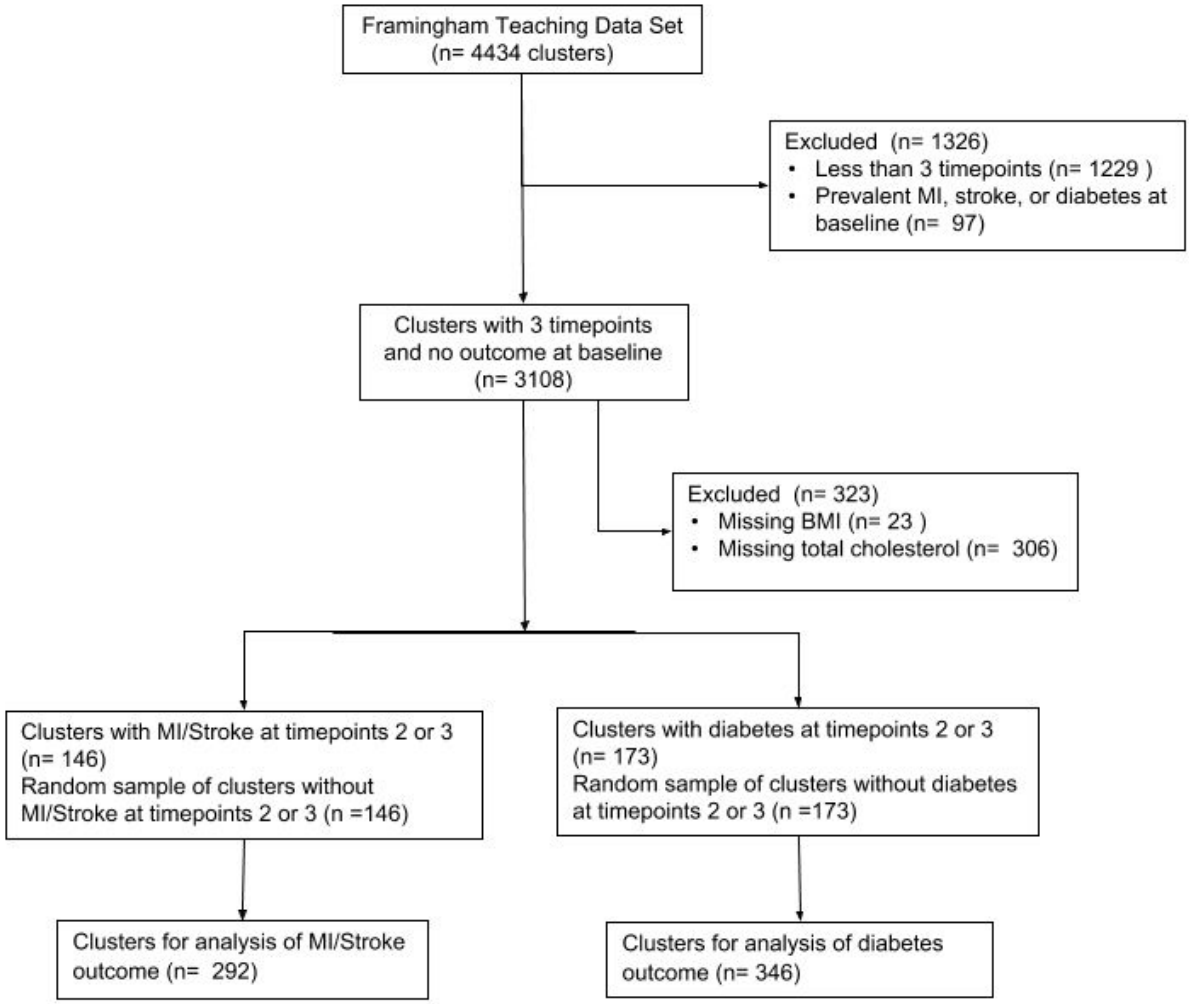

***Table 1*. Comparison of baseline characteristics between cases and controls.**

A. Comparison between MI/Stroke cases and controls

Characteristic	MI/Stroke Cases (n=146) Mean (SD) or N [%]	MI/Stroke Controls (n=146) Mean (SD) or N [%]
Age	53.12 (7.54)	48.48 (7.85)
Gender (female)	58 [39.70]	89 [61.00]
Current smoker	74 [50.70]	73 [50.00]
BMI (kg/m^2)	26.93 (4.28)	25.78 (3.83)
Total cholesterol (mg/dL)	252.64 (45.85)	239.45 (58.10)
Systolic BP (mmHg)	140.84 (22.60)	130.70 (20.02)

B. Comparison between diabetes cases and controls.

Characteristic	Diabetes Cases (n=173) Mean (SD) or N [%]	Diabetes Controls (n=173) Mean (SD) or N [%]
Age	51.03 (8.10)	48.68 (8.50)
Gender (female)	91 [52.60]	102 [59.00]
Current smoker	79 [45.70]	76 [43.90]
BMI (kg/m^2)	28.14 (5.30)	25.59 (3.80)
Total cholesterol (mg/dL)	246.78 (46.40)	237.52 (43.30)

***Table 2*. Binomial GEE model with logit link for effect of age on the incidence of MI or stroke**

		95% Confidence Interval		
Coefficient*	Estimate	Lower	Upper	p-value
Age	1.16	1.12	1.19	<0.0001
BMI (kg/m^2)	0.96	0.91	1.01	0.14
Total Cholesterol (mg/dL)	1.00	0.99	1.00	0.33
Systolic BP (mmHg)	1.00	0.99	1.01	0.70

*Odds ratios.

Table 3. Binomial GEE model with logit link for effect of age on the incidence of diabetes

		95% Confidence Interval		
Coefficient*	Estimate	Lower	Upper	p-value
Age	1.11	1.08	1.13	<0.0001
BMI (kg/m^2)	1.02	0.98	1.06	0.34
Total Cholesterol (mg/dL)	0.99	0.99	1.00	0.01
Systolic BP (mmHg)	1.01	1.00	1.02	0.01

*Odds ratios.