Role of Smoking and Cholesterol in CVD, CHD, and Stroke in the Framingham Heart Study

by Irene Hsueh

Irene Hsueh

Introduction

The Framingham Heart Study is a longitudinal prospective cohort study of the etiology of cardiovascular disease among a population of free-living subjects in the community of Framingham, Massachusetts. Established in 1948, the Framingham Heart Study now has data on three generations of participants.

The purpose of this report is to examine the association between incidence of heart disease with cholesterol levels and smoking using data collected on the Framingham Study population. The dataset contains 4434 unique participants between the ages of 32 and 81, with 1944 men and 2490 women. The data was provided in a longitudinal format for up to 3 follow-up periods and contained information about incidence and time in days to the first cardiovascular disease (CVD), coronary heart disease (CHD), and stroke event.

The Center for Disease Prevention and Control (CDC) states that smoking is a major cause of CVD and contributes to approximately 1 in 4 of deaths from CVD due to smoking's potential to raise triglycerides, lower good cholesterol (HDL), and increase the buildup of plaque[1]. The Framingham data shows that smokers generally have an almost two-fold increased risk of CHD or myocardial infarction[2]. Similarly, unhealthy levels of cholesterol can increase the risk for heart disease[3]. The risk presents itself around 200 mg/dL, and there doesn't appear to be a threshold where the risk is not present at all or a threshold where the risk suddenly begins. In addition, risk factors such as age, sex, history of heart disease in the family, obesity, diabetes, an unhealthy diet, and clinical depression can increase the potential for an individual to develop heart disease[4]. Levy et al. found evidence to support that age and sex play major roles in the development of coronary heart disease (CHD)[5]. Among men between the ages 30 and 62, there is a linear increase in the incidence of coronary heart disease as age increases. Women also have a greater risk for coronary heart disease after menopause, though their risk always remains lower than the risk for men.

As a result of the literature that highlighted the risk of smoking, age, triglycerides, and other factors that can increase the risk of heart disease, incidence and time to CVD, stroke, and CHD were chosen as the outcomes. Quantitative explanatory variables that were provided include age, systolic blood pressure (mmHg), diastolic blood pressure (mmHg), number of

---

[1] "Heart Disease and Stroke." *Health Effects*, Centers for Disease Control and Prevention, 28 Apr. 2020, www.cdc.gov/tobacco/basic_information/health_effects/heart_disease/index.htm.

[2] Levy, Daniel, et al. "Stratifying the Patient at Risk from Coronary Disease: New Insights from the Framingham Heart Study." The American Heart Journal, vol. 119, no. 3, 1990, pp. 712–717.

[3] "High cholesterol." *Disease and Conditions,* Mayo Clinic , 13 Jul. 2019, https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/symptoms-causes/syc-20350800

[4] *Causes and Risks of Heart Disease*, Erica Roth, 17 Sep 2018, https://www.healthline.com/health/heart-disease/causes-risks#risk-factors

[5] Levy, Daniel, et al. "Stratifying the Patient at Risk from Coronary Disease: New Insights from the Framingham Heart Study." The American Heart Journal, vol. 119, no. 3, 1990, pp. 712–717.

cigarettes smoked per day, serum total cholesterol (mg/dL), high-density lipoprotein cholesterol (mg/dL), low-density lipoprotein cholesterol (mg/dL), BMI (kg/m^2), serum glucose (mg/dL), and heart rate (beats per minute). Binary yes/no variables include use of antihypertensive medication, whether they were a current smoker, and if they had diabetes (serum glucose of over 200 mg/dL). Lastly, sex was a binary categorical variable.

Data

Descriptive characteristics about the study population as a whole and for each sex are presented in Table 1. Only baseline values from Exam 1 were used. Total cholesterol was divided into four categories (<200 mg/dL, 200-219, 220-259 mg/dL, and 260+ mg/dL) since there was evidence of a nonlinear association between total cholesterol levels and mortality[6]. Cigarettes smoked per day were also separated into four categories (0 per day, 1-10 per day, 11-20 per day, and 21+ per day).

The only variables that were evenly distributed between sexes was age, serum glucose, diabetes, and incidence of stroke. For the rest of the predictor and outcome variables, there was a significant difference in the proportion or means between males and females.

Table 1: Baseline Characteristics of the Framingham Dataset

Baseline Characteristics				
	Male (n=1,944) 43.84%	**Female** (n=2,490) 56.16%	**Total** (n=4,434)	**p-value**
Age (mean, SD)	49.79 (8.72)	50.03 (8.64)	49.93 (8.68)	0.3450
Systolic BP (mean, SD)	131.74 (19.44)	133.82 (24.46)	132.91 (22.42)	0.0016
Diastolic BP (mean, SD)	83.71 (11.44)	82.60 (12.50)	83.08 (12.06)	0.0020
Use of BP Meds (n,%)				
No BP Meds	1,880 (97.81%)	2,349 (95.84%)	4,229 (96.71%)	0.0003
BP Meds	42 (2.19%)	102 (4.16%)	144 (3.29%)	
Heart Rate (mean, SD)	74.40 (11.90)	77.06 (12.15)	75.89 (12.11)	<0.0001

[6] D'Agostino RB, Belanger AJ, Kannel WB, Higgins M. Role of smoking in the U-shaped relation of cholesterol to mortality in men. The Framingham Study. Am J Epidemiol. 1995 May 1;141(9):822-7. doi: 10.1093/oxfordjournals.aje.a117517. PMID: 7717358.

Total Cholesterol (mean, SD)	233.58 (42.36)	239.68 (46.22)	236.98 (44.65)	<0.0001
Cholesterol Category (n,%)				
<200	392 (20.16%)	530 (21.29%)	922 (20.79%)	<0.0001
200 - 219	364 (18.72%)	369 (14.82%)	733 (16.53%)	
220 - 259	710 (36.52%)	817 (32.81%)	1,527 (34.44%)	
260+	478 (24.59%)	774 (31.08%)	1,252 (28.24%)	
Smoking Status (n,%)				
Not Smokers	769 (39.56%)	1,483 (59.60%)	4,229 (96.71%)	<0.0001
Smokers	1,175 (60.44%)	1,006 (40.40%)	144 (3.29%)	
Cigarettes Per Day (mean, SD)	13.23 (13.78)	5.65 (8.96)	8.97 (11.93)	<0.0001
Cigarette Category (n, %)				
0 Cigarettes	769 (39.89%)	1,484 (59.98%)	2,253 (51.18%)	<0.0001
1-10 Cigarettes	195 (10.11%)	460 (18.59%)	655 (14.88%)	
11-20 Cigarettes	583 (30.24%)	433 (17.50%)	1,016 (23.08%)	
21+ Cigarettes	381 (19.76%)	97 (3.92%)	478 (10.86%)	
BMI (mean, SD)	26.17 (3.41)	25.59 (4.56)	25.85 (4.10)	<0.0001
Serum Glucose (mean, SD)	82.32 (24.72)	82.07 (24.14)	82.19 (24.40)	0.7468
Diabetes (n,%)				
No Diabetes	1,885 (96.97%)	2,428 (97.51%)	4,313 (97.27%)	0.2690
Diabetes	59 (3.03%)	62 (2.49%)	121 (2.73%)	
Cardiovascular Disease (n,%)				
No CVD	1,258 (64.71%)	2,019 (81.08%)	3,277 (73.91%)	<0.0001
CVD	686 (35.29%)	471 (18.92%)	1,157 (26.09%)	

Coronary Heart Disease (n,%)				
No CHD	1,234 (63.48%)	1,960 (78.71%)	3,194 (72.03%)	<0.0001
CHD	710 (36.52%)	530 (21.29%)	1,240 (27.97%)	
Stroke (n,%)				
No Stroke	1,751 (90.07%)	2,268 (91.08%)	4,019 (90.64%)	0.2508
Stroke	193 (9.93%)	222 (8.92%)	415 (9.36%)	
Days to CVD (mean, SD)	6,273.70 (3,015.18)	7242.86 (2,549.01)	6,817.95 (2,804.32)	<0.0001
Days to CHD (mean, SD)	6,155.88 (3,066.85)	7064.61 (2,656.32)	6,666.20 (2,878.82)	<0.0001
Days to Stroke (mean, SD)	7,003.43 (2,509.10)	7539.78 (2,262.43)	7,304.63 (2,388.34)	<0.0001

Methods

Cox proportional hazard models were used to model survival time to CVD, CHD, and stroke. ZPH tests using Schoenfeld residuals were used to check the proportional hazards assumption. Model fit was determined using the Schwarz Information Criterion (SBC).

Previous investigations using the Framingham Heart Study adjusted for age, sex, total cholesterol, high density lipoprotein, systolic blood pressure, use of antihypertensive medications, smoking status, diabetes, and BMI in their Cox proportional hazards models[7,8,9,10]. Levels of low- and high-density lipoprotein were not provided in the first exam, so the other 8 covariates were included. The first step of the model selection process was to determine whether cholesterol and smoking should be coded as continuous or categorical variables. Table 5 in the Appendix shows the SBC values of the varying models for the three outcomes CVD, CHD, and stroke. The model chosen had total cholesterol as a continuous variable and cigarettes smoked as a categorical variable due to the low SBC values. In addition, cigarette

---

[7] Peter W.F. Wilson, William P. Castelli, William B. Kannel, "Coronary risk prediction in adults (The Framingham Heart Study)", The American Journal of Cardiology, vol. 59, no. 14, 1987, pp G91-G94

[8] O'Donnell CJ, Elosua R. Factores de riesgo cardiovascular. Perspectivas derivadas del Framingham Heart Study [Cardiovascular risk factors. Insights from Framingham Heart Study]. Rev Esp Cardiol. 2008 Mar; 61(3):299-310. Spanish. PMID: 18361904.

[9] Levy, Daniel, et al. "Stratifying the Patient at Risk from Coronary Disease: New Insights from the Framingham Heart Study." The American Heart Journal, vol. 119, no. 3, 1990, pp. 712–717.

[10] D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008 Feb 12;117(6):743-53. doi: 10.1161/CIRCULATIONAHA.107.699579. Epub 2008 Jan 22. PMID: 18212285.

category did not violate the proportional hazards assumption, while cigarettes smoked per day did.

The next step was to determine if systolic blood pressure, diabetes, use of antihypertensive medication, diabetes, and BMI all needed to be included in the final model. From the SBC values (Table 6 in the Appendix), dropping any of those variables did not improve model fit. D'Agostino et al. logarithmically transformed their continuous variables[11]. In this investigation, logarithmic transformation of the continuous variables (total cholesterol, age, systolic blood pressure, and BMI) did not improve model performance (Table 7 in Appendix). Wilson et al. included an interaction term between total cholesterol and age, so various interaction terms were considered[12]. For this dataset, including interaction terms did not improve the model fit (Table 8 in Appendix).

Lastly, the data was stratified by sex and Cox proportion hazards models were run separately for males and females due to earlier findings that most covariates differed significantly between sexes. The 95% hazards ratio confidence intervals for cholesterol and cigarette category overlapped, so the effects were not significantly different for between sexes (Tables 9-14 in Appendix).

The final model chosen adjusted for total cholesterol, cigarette category, sex, age, SBP, blood pressure medications, diabetes, and BMI, without any interaction terms or stratification. The reference coding were 0 cigarettes smoked, female, no blood pressure medications, and no diabetes.

Results

Table 2: Cox proportional hazard model regression coefficients for CVD

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	zph test p-value	p-value
Total Cholesterol	0.00306	1.003	1.002-1.004	0.0907	<0.0001
1-10 Cigarettes	0.19715	1.218	1.013-1.465	0.0737	0.0363
11-20 Cigarettes	0.41644	1.517	1.302-1.766	0.8954	<.0001
21+ Cigarettes	0.34842	1.417	1.159-1.728	0.4318	0.0007
Sex (Male)	0.88603	2.425	2.128-2.765	0.1703	<0.0001

[11] D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008 Feb 12;117(6):743-53. doi: 10.1161/CIRCULATIONAHA.107.699579. Epub 2008 Jan 22. PMID: 18212285.

[12] Peter W.F. Wilson, William P. Castelli, William B. Kannel, "Coronary risk prediction in adults (The Framingham Heart Study)", The American Journal of Cardiology, vol. 59, no. 14, 1987, pp G91-G94

Age	0.05847	1.060	1.052-1.069	0.0543	<0.0001
Systolic Blood Pressure	0.01533	1.015	1.013-1.018	0.5499	<0.0001
BP Meds	0.36172	1.436	1.103-1.868	0.1807	0.0071
Diabetes	0.82074	2.272	1.777-2.906	0.0513	<0.0001
BMI	0.02754	1.028	1.013-1.043	0.8981	0.0003

Table 3: Cox proportional hazard model regression coefficients for CHD

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	zph test p-value	p-value
Total Cholesterol	0.00424	1.004	1.003-1.005	0.9061	<.0001
1-10 Cigarettes	0.07421	1.077	0.900-1.289	0.3348	0.4184
11-20 Cigarettes	0.20775	1.231	1.060-1.430	0.0772	0.0065
21+ Cigarettes	0.26895	1.309	1.081-1.584	0.1531	0.0058
Sex (Male)	0.77762	2.176	1.922-2.464	0.1452	<0.0001
Age	0.04513	1.046	1.038-1.054	<0.0001	<0.0001
Systolic Blood Pressure	0.01177	1.012	1.009-1.015	0.4850	<0.0001
BP Meds	0.35662	1.428	1.105-1.846	0.4430	0.0065
Diabetes	0.64325	1.903	1.469-2.464	0.0547	<0.0001
BMI	0.03603	1.037	1.022-1.052	0.7864	<0.0001

Table 4: Cox proportional hazard model regression coefficients for stroke

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	zph test p-value	p-value
Total Cholesterol	-0.00213	0.998	0.996-1.000	0.5234	0.0764
1-10 Cigarettes	0.21597	1.241	0.915-1.683	0.2316	0.1646
11-20 Cigarettes	0.62455	1.867	1.456-2.396	0.7705	<0.0001
21+ Cigarettes	0.19017	1.209	0.812-1.800	0.5039	0.3489

Sex (Male)	0.37031	1.448	1.166-1.799	0.5493	0.0008
Age	0.08563	1.089	1.075-1.104	0.0202	<0.0001
Systolic Blood Pressure	0.01889	1.019	1.015-1.024	0.0870	<0.0001
BP Meds	0.72410	2.063	1.448-2.938	0.3536	<0.0001
Diabetes	0.88064	2.412	1.629-3.572	0.5560	<0.0001
BMI	0.01663	1.017	0.993-1.041	0.3227	0.1626

CVD

Schoenfeld residuals showed the proportional hazards assumption was met for all variables (plots in SAS output). All covariates were significantly associated with time to CVD.

The hazard ratio for total cholesterol was 1.003 (95% confidence interval: 1.002, 1.004), meaning that for each 1 unit increase in mg/dL of total cholesterol, the hazard for CVD increased 1.003 times.

Subjects who smoked 1-10 cigarettes a day had 1.218 (95% confidence interval: 1.013-1.465) times the hazard of CVD compared to those who did not smoke. Subjects who smoked 11-20 cigarettes a day had 1.517 (95% confidence interval: 1.302-1.766) times the hazard of CVD compared to those who did not smoke. Subjects who smoked 21+ cigarettes a day had 1.417 (95% confidence interval: 1.159-1.728) times the hazard of CVD compared to those who did not smoke.

The hazard ratio for age was 1.060 (95% confidence interval: 1.013-1.018), meaning for each additional year older a subject was, their hazard for CVD increased by 1.060. The hazard ratio for SBP was 1.015 (95% confidence interval: 1.013-1.018), meaning for each 1 unit increase in mmHg of SBP, the hazard for CVD increased 1.015 times. The hazard ratio for BMI was 1.028 (95% confidence interval: 1.013-1.043), meaning for each 1 unit increase in kg/m^2, the hazard for CVD increased 1.028 times.

Males had 2.425 (95% confidence interval: 2.128-2.765) times the hazard of CVD compared to females. Patients who took antihypertensive medication had 1.436 (95% confidence interval: 1.103-1.868) times the hazard of CVD compared to those who did not take blood pressure medication. Subjects with diabetes had 2.272 (95% confidence interval: 1.777-2.906) times the hazard of CVD compared to those without diabetes.

CHD

Schoenfeld residuals showed the proportional hazards assumption was met for all variables except age (plots in SAS Output). All covariates were significantly associated with time to CHD, except for 1-10 cigarettes smoked per day.

The hazard ratio for total cholesterol was 1.004 (95% confidence interval: 1.003, 1.005), meaning that for each 1 unit increase in mg/dL of total cholesterol, the hazard for CHD increased 1.004 times.

Subjects who smoked 1-10 cigarettes a day had 1.077 (95% confidence interval: 0.900-1.289) times the hazard of CHD compared to those who did not smoke, but this effect was not significant. Subjects who smoked 11-20 cigarettes a day had 1.231 (95% confidence interval: 1.060-1.430) times the hazard of CHD compared to those who did not smoke. Subjects who smoked 21+ cigarettes a day had 1.309 (95% confidence interval: 1.081-1.584) times the hazard of CHD compared to those who did not smoke.

The hazard ratio for age was 1.046 (95% confidence interval: 1.038-1.054), meaning for each additional year older a subject was, their hazard for CHD increased by 1.046. The hazard ratio for SBP was 1.012 (95% confidence interval: 1.009-1.015), meaning for each 1 unit increase in mmHg of SBP, the hazard for CHD increased 1.012 times. The hazard ratio for BMI was 1.037 (95% confidence interval: 1.022-1.052), meaning for each 1 unit increase in kg/m^2, the hazard for CHD increased 1.037 times.

Males had 2.176 (95% confidence interval: 1.922-2.464) times the hazard of CHD compared to females. Patients who took antihypertensive medication had 1.428 (95% confidence interval: 1.105-1.846) times the hazard of CHD compared to those who did not take blood pressure medication. Patients with diabetes had 1.903 (95% confidence interval: 1.469-2.459) times the hazard of CHD compared to those without diabetes.

Stroke

Schoenfeld residuals showed the proportional hazards assumption was met for all variables except age (plots in SAS Output). All covariates were significantly associated with time to stroke, except for total cholesterol, 1-10 cigarettes smoked per day, 21+ cigarettes smoked per day, and BMI.

The hazard ratio for total cholesterol was 0.998 (95% confidence interval: 0.996, 1.000), meaning that for each 1 unit increase in mg/dL of total cholesterol, the hazard for stroke decreased by a factor of 0.998. However, this effect was not significant.

Subjects who smoked 1-10 cigarettes a day had 1.241 (95% confidence interval: 0.915-1.683) times the hazard of stroke compared to those who did not smoke, but this effect was not significant. Subjects who smoked 11-20 cigarettes a day had 1.867 (95% confidence interval: 1.456-2.396) times the hazard of stroke compared to those who did not smoke. Subjects who smoked 21+ cigarettes a day had 1.209 (95% confidence interval: 0.812-1.800) times the hazard of CHD compared to those who did not smoke, but this effect was not significant.

The hazard ratio for age was 1.089 (95% confidence interval: 1.075-1.104), meaning for each additional year older a subject was, their hazard for stroke increased by 1.089. The hazard ratio for SBP was 1.019 (95% confidence interval: 1.015-1.024), meaning for each 1 unit

increase in mmHg of SBP, the hazard for stroke increased 1.019 times. The hazard ratio for BMI was 1.017 (95% confidence interval: 0.993-1.041), meaning for each 1 unit increase in kg/m^2, the hazard for stroke increased 1.017 times. However, this effect was not significant.

Males had 1.448 (95% confidence interval: 1.166-1.799) times the hazard of stroke compared to females. Patients who took antihypertensive medication had 2.063 (95% confidence interval: 1.448-2.938) times the hazard of stroke compared to those who did not take blood pressure medication. Patients with diabetes had 2.412 (95% confidence interval: 1.629-3.572) times the hazard of stroke compared to those without diabetes.

Discussion

The unavailability of LDL and HDL levels for early periods may decrease the accuracy of the models, as Leo et al. found significant associations of LDL and HDL with heart disease[13]. The proportional hazards assumption was violated for age in the CHD and stroke model. However, since the sample size was large (4273 observations), and age wasn't the explanatory variable of interest, the violation most likely did not make a large difference in the parameter estimates.

Pearson and Spearman correlation coefficients were calculated between the included covariates to identify any collinearity issues (correlation chart in SAS output). The largest correlations were between SBP and age (Pearson coefficient = 0.39849, Spearman coefficient = 0.39597) and SBP and BMI (Pearson coefficient = 0.32809, Spearman coefficient = 0.32403). Those correlation coefficients were below 0.4, so any linear relationships between the variables would be considered very weak.

Box and whiskers plots were examined to identify outliers in continuous variables. There were no outliers for age. For total cholesterol, 115 observations were considered outliers, having values of 331 mg/dL or more. For systolic blood pressure, values greater than 183.75 mmHg were considered outliers. Only 150 observations exceeded that threshold, which is a minor portion of the whole dataset. For BMI, there was only 1 observation below the lower outlier threshold of 15.59 kg/m^2, and 103 observations exceeding the upper limit of 35.59 kg/m^2. Compared to the large sample size of 4434, those couple hundred outliers would not affect the distribution of the dataset in any noticeable way (plots in SAS output).

Conclusion

Time to CVD was significantly associated with total cholesterol and smoking category. For each 1 unit increase in mg/dL of total cholesterol, the hazard for CVD increased 1.003 times (95% confidence interval: 1.002, 1.004). Subjects who smoked 1-10 cigarettes a day had 1.218

---

[13] Leo J. S., et al. Lipoprotein(a)-Cholesterol and Coronary Heart Disease in the Framingham Heart Study. Rev Esp Cardiol. 2008;61(3):299-310.

(95% confidence interval: 1.013-1.465) times the hazard of CVD compared to those who did not smoke. Subjects who smoked 11-20 cigarettes a day had 1.517 (95% confidence interval: 1.302-1.766) times the hazard of CVD compared to those who did not smoke. Subjects who smoked 21+ cigarettes a day had 1.417 (95% confidence interval: 1.159-1.728) times the hazard of CVD compared to those who did not smoke. The hazard ratio for 21+ cigarettes was lower than the hazard ratio for 11-20 cigarettes. The confidence intervals for 11-20 cigarettes and 21+ cigarettes overlapped, which indicates that there is no significant difference for hazard of CVD between those who smoke 11-20 and 21+ cigarettes a day.

Time to CHD was significantly associated with total cholesterol and smoking category. For each 1 unit increase in mg/dL of total cholesterol, the hazard for CVD increased 1.004 times (95% confidence interval: 1.003, 1.005). Subjects who smoked 11-20 cigarettes a day had 1.231 (95% confidence interval: 1.060-1.430) times the hazard of CHD compared to those who did not smoke. Subjects who smoked 21+ cigarettes a day had 1.309 (95% confidence interval: 1.081-1.584) times the hazard of CHD compared to those who did not smoke. The confidence intervals for 11-20 cigarettes and 21+ cigarettes overlapped, which indicates that there is no significant difference for hazard of CHD between those who smoke 11-20 and 21+ cigarettes a day.

Lastly, time to stroke was significantly associated with smoking 11-20 cigarettes a day. Subjects who smoked 11-20 cigarettes a day had 1.870 (95% confidence interval: 1.458-2.398) times the hazard of CHD compared to those who did not smoke. In this population, incidence and time to stroke was not explained by the predictors of interest, total cholesterol and smoking.

CVD, CHD, and stroke were all significantly associated with sex, age, systolic blood pressure, use of antihypertensive medications, and diabetes. Males had more than twice the hazard for CVD and CHD than females, and an almost 50% increase in hazard for stroke than females. The hazard ratio for all three outcomes increased each year of age and mmHg of systolic blood pressure. Subjects taking blood pressure medications had an almost 50% increase in hazard for CVD and CHD, and more than twice the hazard for stroke. Diabetics had more than twice the hazard for CVD and stroke than non-diabetes, while almost twice the hazard for CHD than non-diabetes.

Appendix

Table 5. SBC values for Cox proportional hazards models testing cholesterol category, total serum cholesterol, smoking status, cigarettes per day, and cigarette category

Model	Covariates	SBC Values		
		CVD	CHD	Stroke
A1	Cholesterol Category, Smoking Status, Sex, Age, SBP, BP_Meds, Diabetes, BMI	17614.792	19070.508	6240.337
A2	Total Cholesterol, Smoking Status, Sex, Age, SBP, BP_Meds, Diabetes, BMI	17395.850	18798.445	6171.734
A3	Cholesterol Category, Cigarettes, Sex, Age, SBP, BP_Meds, Diabetes, BMI	17438.741	18884.722	6198.189
A4	Total Cholesterol, Cigarettes, Sex, Age, SBP, BP_Meds, Diabetes, BMI	17218.240	18611.223	6129.658
A5	Cholesterol Category, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI	17432.678	18883.124	6196.626
A6	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI	17221.371	18623.770	6129.222

Table 6. SBC values for Cox proportional hazards models testing combinations of covariates SBP, blood pressure medication taken, diabetes, and BMI

Model	Covariates	SBC Values		
		CVD	CHD	Stroke
B1	Total Cholesterol, Cigarette Category, Sex, Age, SBP, Diabetes, BMI	17561.616	19022.870	6239.411
B2	Total Cholesterol, Cigarette Category, Sex, Age, BP_Meds, Diabetes, BMI	17323.587	18683.627	6189.006
B3	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, BMI	17248.935	18636.664	6138.645
B4	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes	17323.721	18720.604	6174.435
B5	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds	17354.297	18736.694	6184.307
B6	Total Cholesterol, Cigarette Category, Sex, Age, SBP	17692.309	19135.546	6292.885
B7	Total Cholesterol, Cigarette Category, Sex, Age, BP_Meds	17498.227	18832.719	6259.920
B8	Total Cholesterol, Cigarette Category, Sex, Age, Diabetes, BMI	17686.540	19099.878	6321.740
B9	Total Cholesterol, Cigarette Category, Sex, Age, Diabetes	17823.262	19230.967	6382.562
B10	Total Cholesterol, Cigarette Category, Sex, Age, BMI	17719.230	19117.457	6331.885
B11	Total Cholesterol, Cigarette Category, Sex, Age	17865.377	19256.106	6396.214

Table 7: SBC values for Cox proportional hazard models with logarithmically transformed continuous variables

Model	Covariates	SBC Values		
		CVD	CHD	Stroke
C	Log Total Cholesterol, Cigarette Category, Sex, Log Age, Log SBP, BP_Meds, Diabetes, Log BMI	17222.656	18624.123	6127.548

Table 8: SBC values for Cox proportional hazard models with interaction terms

Model	Covariates	SBC Values		
		CVD	CHD	Stroke
D1	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, total_chol*age	17218.851	18623.479	6132.931
D2	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, total_chol*sex	17227.769	18630.692	6131.973
D3	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, cigarette_cat*age	17234.640	18641.426	6140.327
D4	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, cigarette_cat*sex	17239.428	18644.196	6145.981
D5	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, sbp*age	17227.997	18630.819	6134.558
D6	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, sbp*sex	17228.365	18630.168	6135.025
D7	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, diabetes*age	17225.991	18627.899	6132.961
D8	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, diabetes*sex	17227.538	18630.565	6134.421
D9	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, bmi*age	17227.970	18628.161	6135.194
D10	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, bmi*sex	17228.303	18630.830	6135.090
D11	Total Cholesterol, Cigarette Category, Sex, Age, SBP, BP_Meds, Diabetes, BMI, total_chol*cigarette_cat	17240.466	18644.435	6150.721

Table 9: Cox proportional hazard model regression coefficients for CVD among males

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	p-value
Total Cholesterol	0.00339	1.003	1.002-1.005	<0.0001
1-10 Cigarettes	0.16939	1.185	0.908-1.545	0.2112
11-20 Cigarettes	0.31857	1.375	1.137-1.663	0.0010
21+ Cigarettes	0.30807	1.361	1.090-1.698	0.0064
Age	0.05698	1.059	1.048-1.069	<0.0001
Systolic Blood Pressure	0.01447	1.015	1.011-1.019	<0.0001
BP Meds	0.54776	1.729	1.133-2.639	0.0111
Diabetes	0.70138	2.017	1.429-2.639	<0.0001
BMI	0.02884	1.029	1.005-1.054	0.0173

Table 10: Cox proportional hazard model regression coefficients for CHD among males

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	p-value
Total Cholesterol	0.00407	1.004	1.003-1.006	<0.0001
1-10 Cigarettes	0.10019	1.105	0.849-1.438	0.4558
11-20 Cigarettes	0.21978	1.246	1.032-1.504	0.0220
21+ Cigarettes	0.30208	1.353	1.094-1.673	0.0053
Age	0.04598	1.047	1.037-1.057	<0.0001
Systolic Blood Pressure	0.01055	1.011	1.007-1.015	<0.0001
BP Meds	0.31260	1.367	0.876-2.133	0.1684
Diabetes	0.57042	1.769	1.223-2.539	0.0020
BMI	0.03732	1.038	1.014-1.062	0.0014

Table 11: Cox proportional hazard model regression coefficients for stroke among males

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	p-value
Total Cholesterol	0.0003558	1.001	0.997-1.004	0.7554
1-10 Cigarettes	0.16839	1.144	0.697-1.877	0.5949
11-20 Cigarettes	0.53980	1.725	1.226-2.426	0.0018
21+ Cigarettes	0.17683	1.150	0.728-1.815	0.5493
Age	0.09237	1.097	1.076-1.118	<0.0001
Systolic Blood Pressure	0.01865	1.019	1.012-1.026	<0.0001
BP Meds	1.04148	2.833	1.502-5.344	0.0013
Diabetes	1.09311	2.984	1.700-5.236	0.0004
BMI	0.01682	1.017	0.972-1.064	0.4687

Table 12: Cox proportional hazard model regression coefficients for CVD among females

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	p-value
Total Cholesterol	0.00246	1.002	1.000-1.005	0.0213
1-10 Cigarettes	0.22444	1.252	0.966-1.621	0.0890
11-20 Cigarettes	0.60593	1.833	1.422-2.363	<0.0001
21+ Cigarettes	0.32654	1.386	0.789-2.436	0.2565
Age	0.06240	1.064	1.050-1.079	<0.0001
Systolic Blood Pressure	0.01617	1.016	1.012-1.020	<0.0001
BP Meds	0.24215	1.274	0.909-1.785	0.1591
Diabetes	0.94093	2.562	1.794-3.659	<0.0001
BMI	0.02593	1.026	1.007-1.046	0.0084

Table 13: Cox proportional hazard model regression coefficients for CHD among females

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	p-value
Total Cholesterol	0.00454	1.005	1.003-1.007	<0.0001
1-10 Cigarettes	0.04746	1.049	0.818-1.344	0.7076
11-20 Cigarettes	0.18591	1.204	0.933-1.555	0.1535
21+ Cigarettes	0.04278	1.044	0.607-1.794	0.8770
Age	0.04184	1.043	1.029-1.056	<0.0001
Systolic Blood Pressure	0.01314	1.014	1.009-1.017	<0.0001
BP Meds	0.37257	1.451	1.056-1.995	0.0217
Diabetes	0.71362	2.041	1.405-2.967	0.0002
BMI	0.03435	1.035	1.016-1.054	0.0003

Table 14: Cox proportional hazard model regression coefficients for stroke among females

Variable	Parameter Estimate	Hazard Ratio	95% Hazard Ratio Confidence Interval	p-value
Total Cholesterol	-0.00358	0.996	0.996-1.000	0.0273
1-10 Cigarettes	0.24007	1.271	0.863-1.873	0.2249
11-20 Cigarettes	0.74361	2.104	1.463-3.024	<0.0001
21+ Cigarettes	0.18368	1.202	0.488-2.962	0.6898
Age	0.08277	1.086	1.065-1.108	<0.0001
Systolic Blood Pressure	0.01920	1.019	1.014-1.025	<0.0001
BP Meds	0.62296	1.864	1.219-2.852	0.0041
Diabetes	0.69638	2.007	1.150-3.505	0.0143
BMI	0.01759	1.018	0.990-1.046	0.2111

Irene Hsueh

<table>
<tr><td colspan="9">**Pearson Correlation Coefficients**
Prob > |r| under H0: Rho=0
Number of Observations</td></tr>
<tr><td></td><td>total_chol</td><td>cigarette_cat</td><td>sex</td><td>age</td><td>sbp</td><td>bp_meds</td><td>diabetes</td><td>bmi</td></tr>
<tr><td>**total_chol**
Serum Total Cholesterol</td><td>1.00000

4382</td><td>-0.03891
0.0103
4350</td><td>0.06787
<.0001
4382</td><td>0.24931
<.0001
4382</td><td>0.19996
<.0001
4382</td><td>0.08158
<.0001
4322</td><td>0.04165
0.0058
4382</td><td>0.12268
<.0001
4364</td></tr>
<tr><td>**cigarette_cat**
Cigarettes Per Day Category</td><td>-0.03891
0.0103
4350</td><td>1.00000

4402</td><td>-0.29556
<.0001
4402</td><td>-0.20976
<.0001
4402</td><td>-0.11421
<.0001
4402</td><td>-0.05608
0.0002
4342</td><td>-0.04269
0.0046
4402</td><td>-0.13231
<.0001
4383</td></tr>
<tr><td>**sex**
Sex</td><td>0.06787
<.0001
4382</td><td>-0.29556
<.0001
4402</td><td>1.00000

4434</td><td>0.01418
0.3450
4434</td><td>0.04615
0.0021
4434</td><td>0.05497
0.0003
4373</td><td>-0.01660
0.2691
4434</td><td>-0.06978
<.0001
4415</td></tr>
<tr><td>**age**
Age</td><td>0.24931
<.0001
4382</td><td>-0.20976
<.0001
4402</td><td>0.01418
0.3450
4434</td><td>1.00000

4434</td><td>0.39849
<.0001
4434</td><td>0.13506
<.0001
4373</td><td>0.10626
<.0001
4434</td><td>0.13421
<.0001
4415</td></tr>
<tr><td>**sbp**
Systolic Blood Pressure</td><td>0.19996
<.0001
4382</td><td>-0.11421
<.0001
4402</td><td>0.04615
0.0021
4434</td><td>0.39849
<.0001
4434</td><td>1.00000

4434</td><td>0.26607
<.0001
4373</td><td>0.11519
<.0001
4434</td><td>0.32809
<.0001
4415</td></tr>
<tr><td>**bp_meds**
Use of BP Meds</td><td>0.08158
<.0001
4322</td><td>-0.05608
0.0002
4342</td><td>0.05497
0.0003
4373</td><td>0.13506
<.0001
4373</td><td>0.26607
<.0001
4373</td><td>1.00000

4373</td><td>0.04045
0.0075
4373</td><td>0.09848
<.0001
4354</td></tr>
<tr><td>**diabetes**
Diabetes</td><td>0.04165
0.0058
4382</td><td>-0.04269
0.0046
4402</td><td>-0.01660
0.2691
4434</td><td>0.10626
<.0001
4434</td><td>0.11519
<.0001
4434</td><td>0.04045
0.0075
4373</td><td>1.00000

4434</td><td>0.08778
<.0001
4415</td></tr>
<tr><td>**bmi**
BMI</td><td>0.12268
<.0001
4364</td><td>-0.13231
<.0001
4383</td><td>-0.06978
<.0001
4415</td><td>0.13421
<.0001
4415</td><td>0.32809
<.0001
4415</td><td>0.09848
<.0001
4354</td><td>0.08778
<.0001
4415</td><td>1.00000

4415</td></tr>
</table>

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations								
	total_chol	cigarette_cat	sex	age	sbp	bp_meds	diabetes	bmi
total_chol **Serum Total Cholesterol**	1.00000 4382	-0.04571 0.0026 4350	0.06327 <.0001 4382	0.27395 <.0001 4382	0.21732 <.0001 4382	0.07970 <.0001 4322	0.03208 0.0337 4382	0.15082 <.0001 4364
cigarette_cat **Cigarettes Per Day Category**	-0.04571 0.0026 4350	1.00000 4402	-0.27513 <.0001 4402	-0.21428 <.0001 4402	-0.11879 <.0001 4402	-0.05606 0.0002 4342	-0.04487 0.0029 4402	-0.14807 <.0001 4383
sex **Sex**	0.06327 <.0001 4382	-0.27513 <.0001 4402	1.00000 4434	0.01605 0.2852 4434	0.01195 0.4263 4434	0.05497 0.0003 4373	-0.01660 0.2691 4434	-0.13049 <.0001 4415
age **Age**	0.27395 <.0001 4382	-0.21428 <.0001 4402	0.01605 0.2852 4434	1.00000 4434	0.39597 <.0001 4434	0.13084 <.0001 4373	0.10594 <.0001 4434	0.14478 <.0001 4415
sbp **Systolic Blood Pressure**	0.21732 <.0001 4382	-0.11879 <.0001 4402	0.01195 0.4263 4434	0.39597 <.0001 4434	1.00000 4434	0.21180 <.0001 4373	0.09447 <.0001 4434	0.32403 <.0001 4415
bp_meds **Use of BP Meds**	0.07970 <.0001 4322	-0.05606 0.0002 4342	0.05497 0.0003 4373	0.13084 <.0001 4373	0.21180 <.0001 4373	1.00000 4373	0.04045 0.0075 4373	0.08660 <.0001 4354
diabetes **Diabetes**	0.03208 0.0337 4382	-0.04487 0.0029 4402	-0.01660 0.2691 4434	0.10594 <.0001 4434	0.09447 <.0001 4434	0.04045 0.0075 4373	1.00000 4434	0.07501 <.0001 4415
bmi **BMI**	0.15082 <.0001 4364	-0.14807 <.0001 4383	-0.13049 <.0001 4415	0.14478 <.0001 4415	0.32403 <.0001 4415	0.08660 <.0001 4354	0.07501 <.0001 4415	1.00000 4415

18