Reghan Lanning
BS852 Statistical Methods in Epidemiology
9 December 2024

Systolic Blood Pressure and Mortality in Framingham Heart Study Cohort

Introduction

The Framingham Heart Study is a longitudinal study collecting information on common factors contributing to cardiovascular disease. This analysis uses data taken at exam four such as behavioral factors and measures of coronary health, as well as incidence of coronary heart disease, type 2 diabetes, and death in the 22-year follow-up period. The primary goal is to investigate the association between systolic blood pressure and overall survival of the cohort. Table 1 provides more information about the variables within this analysis, and their distributions within the study group.

Methods

To test if systolic blood pressure (SPF4) is significantly associated with death, a univariate Cox Proportional Hazards (PH) model was used with death as the censoring variable and survival as the time until the censored event from exam 4. Schoenfeld residuals determined if covariates failed the PH assumption at the alpha = 0.05 level of significance and time-varying coefficient plots confirmed these results. To determine influence of covariates in the data, a multivariable model was built using a stepwise automatic variable selection technique. Menopause was not included in inputted covariates for the stepwise selection since it is not a measure for all patients within the study. Coronary heart disease and type 2 diabetes were not included in the inputted covariates due to competing risks. Predictors that were significant at the alpha = 0.15 level of significance and lowered model AIC were incorporated into the model. Missing values were assumed random and removed from the data, ensuring consistent data usage for this model. The significant covariates identified through stepwise selection had Schoenfeld residuals tested to meet the PH assumption. AICs for the crude and adjusted model were compared to determine which model fit mortality the best. Potential confounding was tested through the ratio of the crude Hazards Ratio (HR) to the adjusted HR for SPF4. To determine if the association between SPF4 and mortality is the same in males and females, an interaction term was added to the multivariate model, and the term was evaluated at the alpha = 0.05 level of significance with the null hypothesis that the association between systolic blood pressure at visit 4 and mortality does not differ between males and females after adjusting for other covariates. The data was stratified by sex (male =1, female = 2) to further examine HR estimates between men and women. Each sex strata underwent its own stepwise variable selection to discover more about the covariates that may be more significant without the variable sex influencing AIC, and a significance level of alpha = 0.15 was used to select covariates. PH assumptions were tested for both models, and models were adjusted accordingly through stratification.

Results

In the crude analysis of systolic blood pressure and mortality, we test under the assumption that the coefficient for SPF4 is equal to one, and that there is no association between SPF4 and mortality. SPF4 does not fail the PH assumption (χ^2 = 0.4718, df = 1, p = 0.4399), so the effect does not appear to change with time. Using a Cox PH model, the chi-square statistic for SPF4 is 246.17 (df = 1, p < 0.0001), so we reject the null hypothesis. There is a significant relationship between systolic blood pressure measured at exam 4 and mortality. At any given time over follow-up a one-unit increase in SPF4, the risk of mortality increases by 2.2% (HR = 1.022, 95% CI 1.019, 1.024). The confidence interval does not contain 1, so this finding is significant.

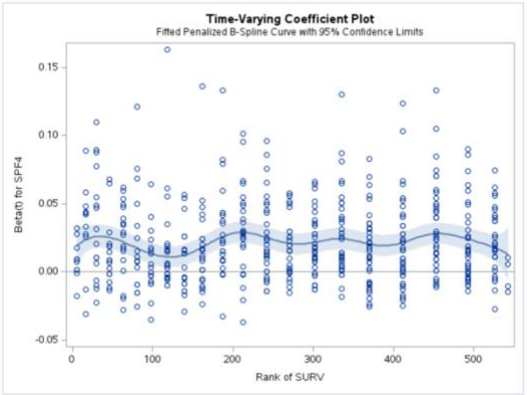

Figure 1: Systolic BP Modeled Over Time

Variables found significant in the data were systolic blood pressure, sex, age, cigarettes smoked daily (CIGS4), and pulmonary function (FVC4). Age violated the null hypothesis that PH assumptions are satisfied (χ^2 = 4.16, df = 1, p = 0.0467), so age groups were built into SAS, and the model was stratified by four age groups: 34-43 years, 44-53 years, 54-63 years, 64-73 years. Since age was incorporated into the baseline, hazard tested covariates were SPF4, sex, CIGS4, and FVC4. For the adjusted analysis, the null hypothesis predicted that SPF4 is equal to 1, after adjusting for sex, CIGS4, and FVC4, while stratifying by age group. At the alpha = 0.05 level of significance, the null was rejected, and SPF4 was significantly associated with mortality after adjusting for significant covariates and stratifying by age group (χ^2 = 76.1194, df = 1, p < 0.0001). At any given time over follow-up a one-unit increase in SPF4, the risk of mortality increases by 1.4% (HR = 1.014, 95% CI 1.011, 1.018), when adjusting for covariates and stratifying by age group. The confidence interval does not contain 1, so this finding is significant. The ratio of the crude hazard ratio (cHR = 1.022) to the adjusted hazard ratio (aHR = 1.014) is above 0.9 and below 1.1 (cHR:aHR = 1.008), so we do not find evidence of confounding when adjusting for covariates and stratifying by age group (Table 2). The multivariate model best fit the data (adjusted AIC with covariates = 6086.9, crude AIC with covariates = 7882.85).

To determine if the effect of SPF4 and mortality is the same between men and women we can insert an interaction term into the model. The interaction term between sex and systolic blood pressure was not found to be significant at the alpha = 0.05 level of significance (χ^2 = 0.0773, df = 1, p = 0.7810). There is not enough evidence to conclude that the association between SPF4 and mortality is different between males and females when adjusting for covariates and stratifying by age. Significant covariates in the association between systolic blood pressure and mortality in women are age, cigarettes smoked per day, pulmonary function, weight, BMI, menopause, and hypertension using a significance level of alpha = 0.15. Cigarettes smoked per day violated the PH assumption when using Schoenfeld residuals. Five groups were created to adjust for the violation: 0-9 cigarettes, 10-19 cigarettes, 20-29 cigarettes, 30-39 cigarettes, and 40-49 cigarettes. All covariates meet PH assumptions after model stratification. Significant covariates in the association between systolic blood pressure and mortality in men are age, cigarettes smoked per day, pulmonary function, and hypertension using a significance level of alpha = 0.15. All covariates pass the proportional hazards assumption using a significance level of alpha = 0.05. Estimates found in Table 3 through the female strata, male strata, and all-data Cox models confirm that there is not enough evidence to suggest the association between systolic blood pressure and mortality is different between females and males.

Conclusions

Systolic blood pressure is significantly associated with mortality when accounting for longitudinal study design and various covariates. Along with SPF4, sex, age, cigarettes smoked daily, and pulmonary function best fit the model to predict death. These covariates do not appear to confound the relationship found between systolic blood pressure and death. There is not enough evidence to support that the relationship between systolic blood pressure measured at exam four and mortality behaves differently in men than it does for women, which can be observed through a non-significant interaction term and similar hazard ratio estimates.

Discussion

This analysis does not account for the information provided with coronary heart disease, type 2 diabetes, and the time to those events. These would have to be accounted for through a competing risk analysis, such as a Fine and Grays model, or discovery through a cumulative incidence function. This could be done through building a column that categorizes mortality within the dataset; 1 for coronary heart disease death, 2 for unrelated death, and 0 for censored observation. Using this column as the new censoring column could provide information on the association between CHD and total mortality, when accounting for systolic blood pressure. In my analysis, I was unable to produce a HR that accurately captured the data (HRs discovered were highly inflated) when accounting for competing risks, so this is a future direction I would like to explore. This could be a potential limitation to this analysis – since all the data provided was not utilized, some findings may have been missed. Nevertheless, discovery from this analysis can provide insight as to what lifestyle factors and health measurements most predict the variability in death and can be potential targets for intervention.

Appendix A: Tables

Table 1: Summary of Patient Demographics, Characteristic Association with Death (DTH)

	Overall (N=2000)	**DTH (N=544)**	**No DTH (N=1456)**	**p**
Male (N, %)	871 (43.55)	291 (33.41)	580 (66.59)	<0.001
Female (N, %)	1129 (56.45)	253 (22.41)	876 (77.59)	
Cholesterol (mean, 95% CI)	230.03 (228.08, 231.99)	235.53 (231.43, 239.63)	227.96 (225.77, 230.16)	<0.001
Systolic BP (mean, 95% CI)	134.13 (133.06, 135.19)	146.88 (144.41, 149.35)	129.37 (128.33, 130.30)	<0.001
Diastolic BP (mean, 95% CI)	82.93 (82.37, 83.48)	87.19 (85.95, 88.42)	81.34 (80.74, 81.93)	<0.001
Cigarettes Smoked (mean, 95% CI)	8.31 (7.80, 8.82)	9.15 (8.10, 10.21)	8 (7.43, 8.58)	0.049
Smoking Status (N, %)				
No (0)	1075 (54.38)	288 (26.79)	787 (73.21)	0.67
Yes (1)	902 (45.62)	247 (27.38)	655 (72.62)	
Missing	23			
Hypertension (N, %)				
No (0)	1227 (62.22)	239 (19.48)	988 (80.52)	<0.001
Yes (1)	745 (37.78)	299 (40.13)	446 (59.87)	
Missing	28			
BMI (mean, 95% CI)	25.98 (25.80, 26.16)	26.67 (26.30, 27.03)	25.73 (25.52, 25.93)	<0.001
Weight in lbs (mean, 95% CI)	154.79 (153.55, 156.02)	159.61 (157.31, 161.91)	153 (151.54, 154.45)	<.0001
Pulmonary Function (mean, 95% CI)	469.66 (464.82, 474.49)	438.89 (428.88, 448.90)	481.05 (475.67, 486.43)	<.0001
CHD (N, %)				
No (0)	1532 (80.72)	292 (19.06)	1240 (80.94)	<.0001
Yes (1)	366 (19.28)	188 (51.37)	178 (48.63)	
Missing	102			
T2 Diabetes (N, %)				
No (0)	1587 (91.63)	429 (27.03)	1158 (72.97)	0.8223
Yes (1)	145 (8.37)	43 (29.66)	102 (70.34)	
Missing	268			

Table 2: Analysis on Crude and Adjusted Cox PH Models

	Parameter Estimate	P-Value	HR (95% CI)
Crude	0.02172	<.0001	1.022 (1.019, 1.025)
Adjusted	0.01421	<0.001	1.014 (1.011, 1.018)

Table 3: Analysis of MLE Across Sex-Specific Adjusted Models

	Parameter Estimate	P-Value	HR (95% CI)
Female	0.01610	<.0001	1.016 (1.010, 1.022)
Male	0.01142	0.0008	1.011 (1.005, 1.018)
All Gender	0.01421	<.0001	1.014 (1.011, 1.018)

Appendix B: SAS Code from Analysis

```
proc import datafile='/home/u63575405/BS852/FINAL/framdat4(3) (1).csv'
out=fram dbms=csv replace; getnames=yes; datarow=2;
run;
proc print data=fram (obs=20);run;

* part 1, assembling demographics table ;

                                * categorical / binary data ;
proc freq data=fram;
        tables dth*(sex smoke htn4 chd t2d);
run;

                                * continuous data ;
proc means data = fram mean clm;
        var chol4 bmi4 wgt4 fvc4 spf4 dpf4 cigs4;
run;

proc means data = fram mean clm;
        class dth;
        var chol4 bmi4 wgt4 fvc4 spf4 dpf4 cigs4;
run;

proc ttest data=fram;
        class dth;
        var chol4 bmi4 wgt4 fvc4 spf4 dpf4 cigs4;
run;

proc logistic data=fram descending;
        class dth(ref='0') sex(ref='1') smoke(ref='0') htn4(ref='0')/param=ref;
        model dth=sex htn4 smoke chd t2d;
run;


* univariate analysis ;

*Run Univariate, check time-varying coefficient to see if PH assumption met;

proc phreg data=fram zph;
        model SURV*DTH(0) = spf4 / rl ties = efron;
run;

* passes assumption ;
* investigating multivariate model ;
proc phreg data=fram;
  model SURV*DTH(0) = spf4 SEX age4 chol4 cigs4 smoke wgt4 fvc4 bmi4 htn4 t2d /
  selection=stepwise slentry=0.15 slstay=0.15;
run; * age, cigs, sex, spf4 ;
```

```
* after variable selection of significant covariates,
  remove missing values from final model ;

data fram2; set fram;
        keep age4 spf4 cigs4 sex spf4 fvc4 surv dth;
        if cmiss(of age4 spf4 cigs4 sex spf4 fvc4 surv dth) then delete;
run;

* multivariate model with significant covariates ;
proc phreg data=fram2 zph;
        class sex(ref=first) / order=internal;
        model surv*dth(0) = spf4 sex age4 cigs4 fvc4  /rl ties=efron;
run;

* see distribution of age to make appropriate groups ;
proc univariate data=fram; var age4; run;

proc format;
  value agef 1='34-43' 2='44-53' 3='54-63' 4='64-73';
run;

data fram2;
   set fram2;
        agegrp = floor((age4-34)/10 + 1); * create age categories ;
        format agegrp agef.;
run;

*age status violates, test with strata;
proc phreg data=fram2;
        class sex(ref=first) / order=internal;
        model SURV*DTH(0) = spf4 sex cigs4 fvc4 / rl ties=efron ;
        strata agegrp;
run;

* does sex modify effect? Testing interaction ;
proc phreg data=fram2;
        class sex(ref=first) / order=internal;
        model SURV*DTH(0) = spf4 sex spf4*sex cigs4 fvc4 / rl ties=efron ;
        strata agegrp;
run;

* stratify data by gender to get gender-specific estimates ;
data female;
        set fram;
        if SEX = 2 ;
run;

* female model selection ;
proc phreg data=female;
```

```
  model SURV*DTH(0) = spf4 SEX age4 chol4 cigs4 meno4 smoke wgt4 fvc4 bmi4 htn4 /
  selection=stepwise slentry=0.15 slstay=0.15;
run;

* testing PH assumption ;
proc phreg data=female zph;
        model surv*dth(0) = age4 spf4 cigs4 fvc4 wgt4 bmi4 meno4 htn4 /rl ties=efron;
run;

* cig smoking violates, see distribution of cigs/day to build groups ;
proc univariate data=female; var cigs4; run;

proc format;
  value cigsf 1='0-9' 2='10-19' 3='20-29' 4='30-39' 5='40-49';
run;

data female2;
        set female;
        cigsgrp = floor((cigs4)/10 + 1);
        format cigsgrp cigsf.;
run;

* stratify by cigsgrp, check ph and estimates ;
proc phreg data=female2 zph;
        model surv*dth(0) = age4 spf4 cigs4 fvc4 wgt4 bmi4 meno4 htn4 /rl ties=efron;
        strata cigsgrp;
run;

* data stratification by males ;
data male;
        set fram;
        if SEX = 1 ;
run;
* automatic variable selection ;
proc phreg data=male;
  model SURV*DTH(0) = spf4 SEX age4 chol4 cigs4 smoke wgt4 fvc4 bmi4 htn4 /
  selection=stepwise slentry=0.15 slstay=0.15;
run;
* model, check PH assumptions ;
proc phreg data=male zph;
        model surv*dth(0) = spf4 age4 cigs4 fvc4 htn4 / rl ties = efron;
run; * all variables pass ;

* is CHD associated with mortality? exploratory ;

data fram_nna; set fram; drop meno4; run;
data fram_nna; set fram_nna; if cmiss(of _all_) then delete; run;

data fram_cr;
```

```
        set fram_nna;
        if (chd=1) then disease=1;
        else if (dth=1 & chd=0) then disease=2;
        else disease=0;
run;

proc phreg data=fram_cr zph;
   class CHD(ref=first)/order=internal;
   model SURV*disease(0) = CHD SPF4 spf4 sex age4 cigs4 fvc4 / rl ties=efron;
        strata CHD;
run;

proc phreg data=fram_cr zph plots(overlay=stratum)=cif;
   class chd(order=internal ref=first);
   model SURV*dth(0) = chd/ eventcode =1 rl ties=efron;
run;
```