2015

AN ASSOCIATION STUDY BETWEEN ADULT BLOOD PRESSURE AND TIME TO FIRST CARDIOVASCULAR DISEASE

Yongjia Pu

AN ASSOCIATION STUDY BETWEEN ADULT BLOOD PRESSURE AND TIME TO FIRST CARDIOVASCULAR DISEASE

By

Yongjia Pu

A Thesis

Submitted to the Faculty of
Virginia Commonwealth University
in Partial Fulfillment of the Requirements
for the Master of Science Degree
in Biostatistics
in the Department of Biostatistics

Richmond, Virginia

May 2015

Name: Yongjia Pu

Date of Degree: May 1, 2015

Institution: Virginia Commonwealth University

Major Field: Biostatistics

Major Professor: Dr. Le Kang

Title of Study: AN ASSOCIATION STUDY BETWEEN ADULT BLOOD PRESSURE AND TIME TO FIRST CARDIOVASCULAR DISEASE

Pages in Study: 29

Candidate for Master of Science Degree

BACKGROUND: Several studies have demonstrated the association between the time to hypertension event and multiple baseline measurements for adults, yet other survival cardiovascular disease (CVD) outcomes such as high cholesterol and heart attack have been somewhat less considered. The Fels Longitudinal Study (FLS) provides us an opportunity to connect adult blood pressure (BP) at certain ages to the time to first CVD outcomes. The availability of long-term serial BP measurements from FLS also potentially allows us to evaluate if the trend of the measured BP biomarkers over time predicts survival outcomes in adulthood through statistical modeling.

METHODS: When the reference standard is right-censored time-to-event (survival) outcome, the *C* index or concordance *C*, is commonly used as a summary measure of discrimination between a survival outcome that is possibly right censored and a predictive-score variable, say, a measured biomarker or a composite-score output from a statistical model that combines multiple biomarkers. When we have subjects longitudinally followed up, it is of primary interest to assess if some baseline measurements predict the time-to-event outcome. Specifically, in this study, systolic blood pressure, diastolic blood pressure, as well as their variations over time, are

considered predictive biomarkers, and we assess their predictive ability for certain time-to-event outcomes in terms of the *C* index.

RESULTS: There are a few summary *C* index differences that are statistically significant in predicting and discriminating certain CVD outcomes at some age stage, though some of these differences are altered in the presence of medicine treatment and lifestyle characteristics. The variation of systolic BP measures over time has a significantly different predicting ability comparing with systolic BP measures at certain given time points, for predicting survival outcome such as high cholesterol.

CONCLUSIONS: Adult systolic and diastolic BP measurements may have significantly different ability in predicting time to first CVD events. The fluctuation of BP measurements over time may have significant association with the time to first CVD events at a single baseline time point.

Key words: adulthood, systolic, diastolic, high blood pressure, high cholesterol, heart attack, heart failure, FELS Longitudinal Study, *C* index, survival analysis, time to event, cardiovascular disease

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to Dr. Le Kang. He has been a tremendous mentor to me. His advice on both research and career has been priceless. Without his supervision and constant help this dissertation would not have been possible.

I would also like to thank Dr. Shumei Sun and Dr. Qiqi Lu, for serving as my committee members, and for the constructive comments and suggestions they provided.

In addition, I take this opportunity to thank all of the Department faulty members and my colleagues, Mrs. Xiaoyan Deng, Mr. Kingston Kang, their brilliant suggestions help me a lot in my scientific research. A special thanks goes to Mr. Yangyang Deng, who provides insights and expertise that greatly assist.

I place on record, my deeply appreciation to my husband, Dr. Liang Zhao and my lovely son Eric Zhao, my brother Yongjie Pu and my mom Professor Youwei Gu for their continued support and unconditional love.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1. Background

Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels.[1] They are the leading cause of death globally: more people die annually from CVDs than from any other disease[2]. An estimated 17.5 million people died from CVDs in 2012, representing 31% of all global deaths[2]. There are a lot of risk factors involved for developing CVDs, among which, blood pressure (BP) is a common one that acts on the arterial wall and is responsible largely for various CVD events[3], such as heart attack, congestion heart failure and other serious heart diseases. Also, blood pressure is one of major risk factors that affect LDL (bad) cholesterol--the major source of cholesterol buildup and blockage in the arteries [4]. Because of the development of several national database and scientific discoveries on HBP throughout adulthood, its association with high cholesterol has been established. It is evident that the long-term health risks for hypertensive adults could be substantial. However, not too much is known about the blood pressure measurements during adulthood for predicting time to hypertension and other types of CVDs. For this purpose, it is of interest to investigate the association between adult BP at some baseline time point, as well as the pattern of these measured biomarker over time, and time to CVDs.

The 1997 United State Joint National Committee on Prevention, Detection Evaluation, and Treatment of High blood Pressure (JNC) report (JNC-VI) [5] and the 1999 World Health Organization (WHO) guidelines [6], which try to classify hypertension into stage, based on levels

of systolic blood pressure (SBP) and diastolic blood pressure (DBP), determine when an individual's SBP and DBP fall into as 'upstaging'[7]. It is well accepted that a subject with SBP ≥ 140 mm Hg, or DBP ≥ 90 mm Hg, is considered as having hypertension. From previous studies, controversial conclusions have been made about the difference of predictive ability between some baseline SBP and DBP for predicting certain CVDs. In this project, we are able to reexamine the predictive ability between SBP and DBP, using the Fels Longitudinal Study (FLS) data. Adult participants in the FLS were followed up periodically from birth into adulthood [8]. The long-term serial data on these measurements from childhood into adulthood could also provide us possible information to make statistical inference about fluctuation of adulthood SBP/DBP and CVDs events.

To measure discrimination in survival analysis (i.e. time to event analysis), it is more difficult and ambiguous than that in logistic regression [9]. Instead of having two possible outcomes into which each subject falls, we have continuous survival times which are subject to censoring[9]. We consider using specific methodology[10], i.e., an one-shot nonparametric approach to compare discrimination performance in terms of *C* index. An assumption to be made in this project is that subjects with lower value of predict variables usually survive longer without experiencing event of interest[9] . To be more specific, we consider baseline SBP and DBP measurements at some time point as predictive variable for event-free survival time. The high blood pressure (HBP), high cholesterol, heart attack and congestion heart failure (CHF) are the events of interest in this project. Besides the estimation the *C* index value, we are able to assess if there is significant difference between SBP and DBP in predicting certain right-censored survival time.

We consider several paired predictive variables from longitudinal followed measurements of SBP and DBP to evaluate the difference of ability between the paired variables. Figure 1.1 below shows the design of the analysis.

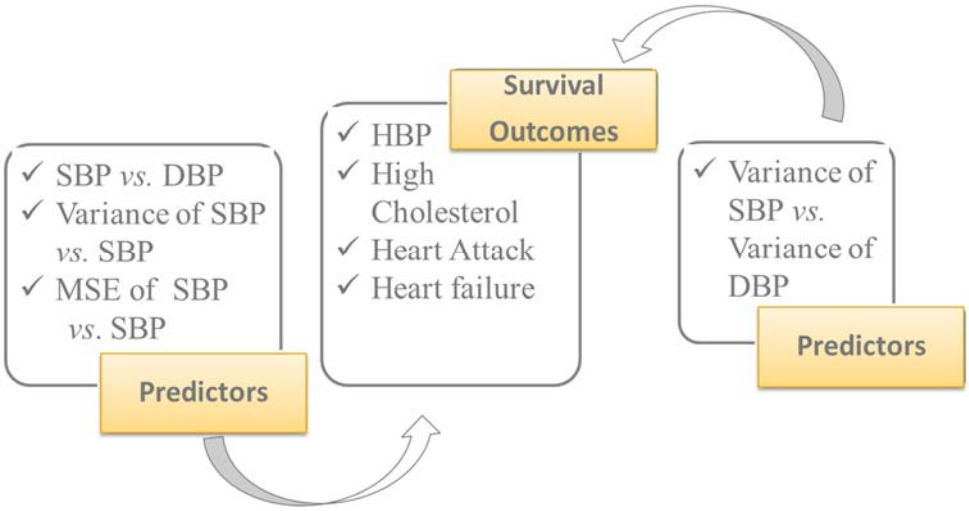


Fig.1.1 Analysis Approach and Design

This project is organized in the following three steps: study population selecting, statistic testing, results analysis and summarization. In the first step, based on selection criteria, a study sample were selected from FLS longitudinal dataset. In the second step, we presented individual estimated *C* indices, and test statistic for the difference between two correlated *C* indices that arise from the paired predictive variables. In the third step, statistical analysis and results interpretation are performed.

1.2. Objective of Study

The overall objective of this study was to evaluate the relationship between adult blood pressure (including HBP and DBP) and first time to CVDs. Two specific objectives of this study were:

1. To assess if some baseline SBP and DBP measurements at certain time points, predict the time-to-event outcome statistically differently.
2. To evaluate if the fluctuation for those measured biomarkers over time predicts the right censored survival outcomes better or worse.

1.3. Organization of Thesis

This thesis includes four chapters and an appendix. Chapter I is about brief introduction and objective for this study. Chapter II is designed as the description for study sample and discussion for the *C* index and approach in detail. Chapter III is interpretation and summarization for test statistic and statistical comparison resulting from the utilization of the one-shot nonparametric approach. Finally, in Chapter IV, a broader discussion as well as the future work on assessing and comparing prognostic accuracy is presented. In addition, the appendix includes the SAS programming and references.

CHAPTER II

METHODS

2.1. Study Sample

2.1.1 Fels Longitudinal Study

The Fels Longitudinal Study is the world's largest and longest running study of human development, growth, body composition and aging [11].Since 1929, the Fels Longitudinal Study (FLS) has recruited children primarily from three counties in the Dayton, Ohio metropolitan area[12]. At the beginning, FLS was designed to study child growth and development. Today, FLS focuses on physical growth, body composition, risk factor for cardiovascular disease and obesity, longitudinal biostatistical analyses and aging[12,13]. It was initiated at Fels Institute in Yellow Springs, OH, by Samuel Fels and Arthur Morgan, President of Antioch College. In 1977, the Fels Research Institute and the FLS became part of the Wright State University's Boonshoft School of Medicine[12,14].

FLS participants are generally enrolled at birth and are not selected in regard to factor known to be associated with disease, body composition, or other clinical conditions [15]. A total of 2,567 infants have been enrolled at birth in annual cohorts of 25-35 up to the present time. The oldest participants are now 82 years old [16,17]. This project addresses the question of the predictive ability of adulthood blood pressure for survival time of CVDs, using blood pressure measurements from a selected sample of 1791 participants in the Fels Longitudinal Study. All BP observations for a single participant collected at 18 years old and thereafter were selected. Other related information about medical history and physical activity were included as well.

2.1.2 Measurement Protocols

2.1.2.1 Measurements and covariates

The predictive variables SBP, DBP and right censored survival outcomes are obtained from the FLS. The following Table.*2.1.* presents a summarization of the variables and outcomes.

Table.2.1 Measurements pertinent to the proposed study sample

Predict Variables (Blood Pressure)	SBP, DBP
Survival Type Outcomes	HBP, CHF, Heart Attack, High Cholesterol level

Systolic blood pressures (SBP) and diastolic blood pressures (DBP) were measured in adult FLS participants in a standardized manner using a sphygmomanometer as recommended by the Second NHLBI Task Force on blood pressure (BP) Control [15]. Well trained technician recorded the SBP and fourth and fifth phase DBP from the subject seated in upright position by inflating the arm cuff to the maximum level and deflating at a rate of 2 mm Hg per second, with 30 s rest between each determination [15,18], SBP is determined by the onset of "tapping" Korotkoff sound (K1), the disappearance of fifth Korotkoff sound (K5) be recorded as the DBP[19]. Both SBP and DBP measurements are recorded as the average of three readings during the single visit.

High blood pressure is defined as a systolic blood pressure (SBP) of 140 mm Hg or more, or a diastolic blood pressure (DBP) of 90 mm Hg or more, or if you have diabetes or chronic kidney disease, a blood pressure of 130/80 mmHg or higher is considered HBP[18] as clinical diagnoses standard. In this study, we compare two time points between the age when first time participant has been diagnosed as HBP and age when first time SBP measurement is greater than 140 mm Hg or DBP is greater than 90 mm Hg, then the earlier age is defined as event time.

Cholesterol level is screened by performing blood test called lipid profile for men at least ages 35 and older and women ages 45 and older. The content of lipid profile include : Total cholesterol , low-density lipoprotein cholesterol, also called 'bad' cholesterol (LDL), high-density lipoprotein cholesterol, also called 'good' cholesterol (HDL), and Triglycerides which is fats exist in blood from the excess calories we consume. For adult, if triglycerides level is greater than 200 mg/dL; HDL smaller than 40 mg/dL for men or small than 50 mg/mL for women; LDL is greater than 160 mg/dL and Total cholesterol level greater than 240 mg/dL , means they are under high risk of developing CVDs and treatment needed [18,20]. Similarly, participants were required to write down the date when they been diagnosed by doctor. This date will be transferred to related age metric which is another survival outcome we considered in this study.

About 720,000 people in the U.S. suffer heart attacks each year. Of these, 515,000 are the first heart attack and 205,000 happen in people who have already had a heart attack. [21] The diagnosis of the heart attack is based on test results as following, electrocardiogram (EKG) which will translate heart's electrical activity into line tracings on paper; cardiac enzymes and troponin level are important indicators of heart attack from blood test, they are solid evidence to determine the heart attack and its size, together with approximately when the heart attack started; Echocardiography is an imaging test which is used for identify which part of heart is dysfunction after heart attack; Cardia catheterization is another visualize method to let doctor find blockage from artery by an inserted tube through blood vessels during the first hour of heart attack[22-24]. Based on above one or more clinical examinations, patients could be diagnosed heart attack, the date they reported is time to event outcome we are interested in.

Congestion heart failure (CHF) is one of the other heart failures. Cardiologist might run several tests to diagnose it. The diagnostic tests include but not limited to B-type Natriuretic

Peptide (BNP) blood test, EKG, the ejection Fraction (EF), Chest X-ray, stress test etc., Once participant were confirmed with CHF disease, dates for diagnosis were recorded as well.

2.1.2.2 Selection Criteria

All subjects are longitudinally followed from birth in FLS, serial clinical examination data has included SBP and DBP, so well as the biological age between 18 years old to age at the last follow up. In this analysis, the event time of interest is the first time that the subjects had a cardiovascular disease event (heart attack, congestion heart failure, high blood pressure and high cholesterol level). Since age variable in this study is continuous, we could not find exact time point for the baseline measurement at specific ages. We determine an age which has minimum distance to 30 years old, 40 years old and 50 years old for each individual and select the BP measurements at that age as baseline measurement at desired time points. The selection criteria for the proposed study sample were stated as: i) participants age at least 18 years old ii) participants with birthday records and reasonable visit age records available iii) participants with at least one pair of measurements for SBP and DBP. iv) date of diagnoses (survival outcome) available. The data flow diagram of progress through the study was presented as in Figure 2.1 below.

Excluding participants without SBP and DBP measurements, we have a subset with 2149 unique subjects. Furthermore, after excluding subjects who did not have visit after 18 years old, we have 1791 individuals eligible for the analysis.

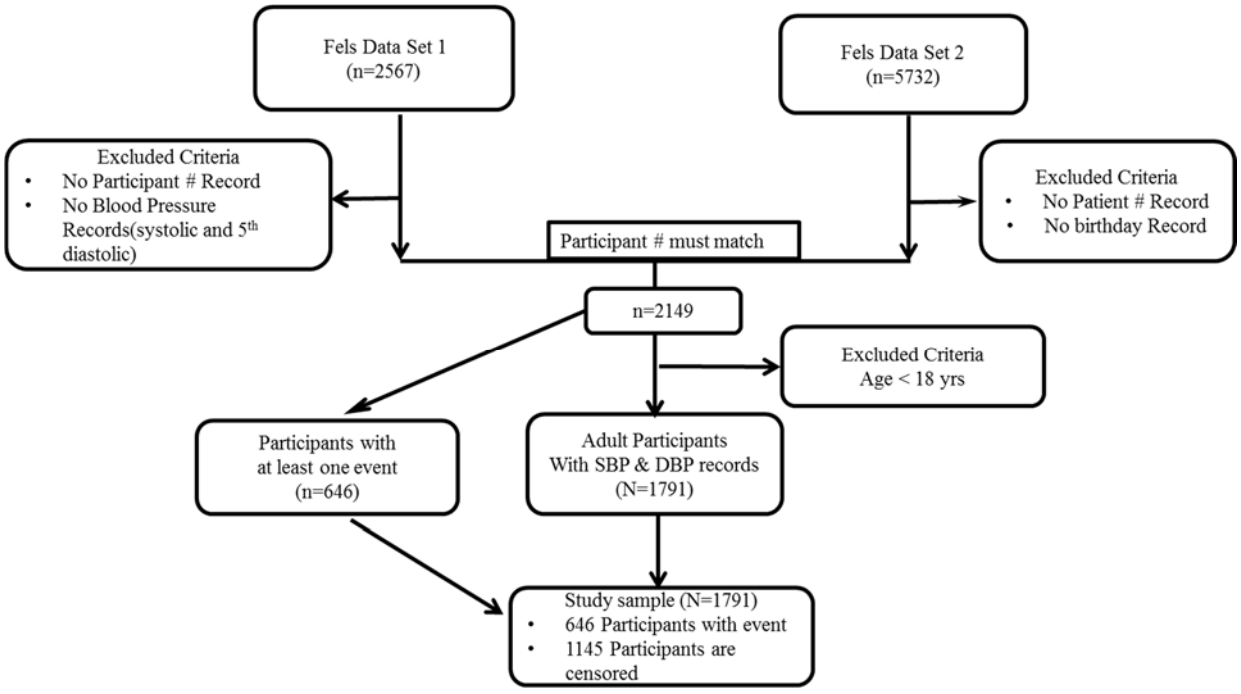


Fig.2.1 Data Flow Diagram of Progress through the Study

2.2 Statistical Modeling and Testing

Many methods have been proposed for evaluating and comparing diagnostic performance. Here we are interested in statistical comparison of two predictive variables, e.g. SBP and DBP, in term of *C* index [10].

An ideal prognostic biomarker should allow the early and accurate identification of individuals at risk for a given outcomes, and should be relatively easy to measure with acceptable costs [25]. SBP and DBP are common risk factors and popular indicators for CVDs and high cholesterol level. In this study, SBP and DBP are used as our two correlated predictive variables to event-free survival outcomes and a *z*-score test is conducted to compare the two *C* indices.

2.2.1 The Overall *C* Index

To characterize the performance of a statistical model predicting a dichotomous outcome, two type of measures are commonly considered: discrimination and calibration [26]. One of the most popular measures of discrimination used in the context of dichotomous outcomes is the area under the receiver operating characteristics (ROC) curve, i.e., AUC[9,27]. Comparing with the logistic regression, it is more complex and difficult measuring discrimination in scenario of time to event analysis.

An assumption we have for our predictive variables is that subjects with lower predictive score actually survive longer without experiencing the event of interest [9,10] . Harrell [28] proposed an overall *C* index, by generalizing and extending the concept of area under the ROC curve (AUC) with binary outcomes to survival outcome. The overall *C* index is a conditional concordance probability measure between a survival outcome that is possibly right censored and a predictive score variable, which can represent a measured biomarker or a composite-score output from an algorithm that combines multiple biomarkers [10].

2.2.1.1 Outcomes with no censoring

Denote X as actually observed survival time without experiencing event. All different pairs of subjects, say (i,j) are considered, where $i \neq j$. To express five types of pairs between the survival time X and the predictive score Y, we introduce the follow notation [10]:

1. a *concordance* with probability $\Pi_c = P(X_i < X_j \text{ and } Y_i < Y_j \text{ or } X_i > X_j \text{ and } Y_i > Y_j)$;
2. a *discordance* with probability $\Pi_d = P(X_i < X_j \text{ and } Y_i > Y_j \text{ or } X_i > X_j \text{ and } Y_i < Y_j)$;
3. an X-only tie with probability $\Pi_{tX} = P(X_i = X_j \text{ and } Y_i > Y_j \text{ or } X_i = X_j \text{ and } Y_i < Y_j)$;
4. a Y-only tie with probability $\Pi_{tY} = P(X_i < X_j \text{ and } Y_i = Y_j \text{ or } X_i > X_j \text{ and } Y_i = Y_j)$;
5. a joint tie in both X and Y with probability $\Pi_{tXY} = P(X_i = X_j \text{ and } Y_i = Y_j)$.

These five kind of possibilities are mutually exclusive, then we have [10] :

$$\Pi_c + \Pi_d + \Pi_{tX} + \Pi_{tY} + \Pi_{tXY} = 1.$$

$$P_K = \frac{1}{2}(d_{X\cdot Y} + 1) = \frac{\Pi_c + \frac{1}{2}\Pi_{tY}}{\Pi_c + \Pi_d + \Pi_{tY}},$$

The above is the probability of a concordance plus one-half the probability of a predictive-score-only (*Y*-only) tie, both conditioned on distinct values or outcomes *X*. [10]

2.2.1.2 Outcomes with Random Right-censoring Assuming Continuous Predictive score

In this study, predictive variable is continuous random variable, say, $P(Y_i = Y_j) = 0$. [10] Under this assumption, Harrell's definition of the *C* index can be expressed in terms of functions of the probability of a concordance and the probability of a discordance,

$$C_{XY} = P(X_i < X_j \text{ and } Y_i < Y_j \text{ or } X_i > X_j \text{ and } Y_i > Y_j | X_i \neq X_j) = \frac{\Pi_c}{\Pi_c + \Pi_d}.$$

Thus a pair is said to be concordant if $X_i < X_j$ and $Y_i < Y_j$ or $X_i > X_j$ and $Y_i > Y_j$, i.e., inequalities go in the same direction, otherwise that pair is to be considered as discordant.

2.2.1.3 Method for Testing An Individual C index

Since we know if value of *C* index near to 0.5, that indicates the performance of predictive score is no better than tossing a coin in determining which subject will survive longer, thus after reporting individual C index for each of each of predictive score variable, we would like to know if the individual C indices have significant difference from 0.5.

The null hypothesis is that $H_0 : C = 0.5$, The test statistic is z= $\frac{\hat{c}-0.5}{\sqrt{\widehat{var}(\hat{c})}}$ we will reject H_0, if $|z| > z_{1-\alpha/2}$, and claim there is statistical significant difference between individual C indices and 0.5, the random guess.

2.2.2 Method for Comparing Two Correlated *C* indices

Denote by X_i a random variable describing the a right-censored survival time, δ_i as censoring indicator where (i =1, 2, …, n), if $\delta_i = 1$ if subject experienced an event, and δ_i equal 0 otherwise. Denote a paired of predictive variables by *Y* and *Z*[10].

Given a general *C* index estimator $\hat{C}^g_{XY}$ for predictive score *Y* and $\hat{C}^g_{XZ}$ for predictive score Z, both conditional on the right-censored survival outcome *X*,

$$\begin{aligned} var(\hat{C}^g_{XY} - \hat{C}^g_{XZ}) &= var(\hat{C}^g_{XY}) + var(\hat{C}^g_{XZ}) - 2cov(\hat{C}^g_{XY}, \hat{C}^g_{XZ}) \\ &= \frac{1}{4}\left[var\left(\frac{t_{XY}}{t^*_{XX}}\right) + var\left(\frac{t_{XZ}}{t^*_{XX}}\right) - 2cov\left(\frac{t_{XY}}{t^*_{XX}}, \frac{t_{XZ}}{t^*_{XX}}\right)\right]. \end{aligned}$$

Using multivariate Delta Method in conjunction with the U-statistics [10,29], variance could be estimated. Recall our original interests is to compare two biomarkers in terms of C index, the z score test is conducted to test null hypothesis $H_0 : C^g_{XY} = C^g_{XZ}$ The test statistic is $z = \frac{\hat{C}^g_{XY} - \hat{C}^g_{XZ}}{\sqrt{\hat{var}(\hat{C}^g_{XY} - \hat{C}^g_{XZ})}}$, we will reject the H_0 if $|z| > z_{1-\alpha/2}$, and claim there is statistical significant difference between two *C* indices [10].

2.2.3 Predictive Score Variables Selection

Based on our approach setting, four pairs of predictors were collected for single subject at 18 years old and thereafter. The first pair is SBP and DBP measurements at three specific time points age 30, age 40 and age 50 respectively, as variables *Y* and *Z*; The second pair is the variance of SBP and SBP itself as *Y* and *Z*, respectively. To be more specific, *i*th subject may have several observations during a given age range, where *i*= 1, 2, 3,…, N represents unique

subject, j represent a specific age. For example, for first subject, we may have the variance formula as following, Var ($SBP_{1j} | \ 25 \leq Age_{1j} < 35$) = $\frac{\sum(SBP_{1j} - \overline{SBP_{1j}})^2}{n1-1}$, in same way, variance of SBP_1 at age interval $35 \leq Age_{1j} < 45$ and $45 \leq Age_{1j} < 55$ were calculated. The third pair of predictive score variables are the mean square error (MSE) of SBP and SBP, denoted as *Y* and *Z*, respectively. Given an example as following, MSE ($SBP_{1j} | \ 25 \leq Age_{1j} < 35$) = $\frac{\sum(SBP_{1j} - \widehat{SBP_{1j}})^2}{n1-1}$, where $\widehat{SBP}$ is prediction value based on linear regression model. The last pair of predictor of interest is the variance of SBP versus the variance of DBP. All of predictors are conditional on right censored survival outcomes (denoted as *X)*, in this study, we also created four indicator variables as our censoring indicator δ for each CVD event.

In general, we could always rearrange data in matrix format with each row representing a unique participant [10].

$$\begin{pmatrix} X & \delta & Y & Z \\ \hline X_1 & \delta_1 & Y_1 & Z_1 \\ X_2 & \delta_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots & \vdots \\ X_n & \delta_n & Y_n & Z_n \end{pmatrix}.$$

2.3 Statistical Analyses

Data preparation step was performed using SAS 9.4. To obtain right-censored survival time *X*, the time to event outcome is calculated as the time from birth to the time of first CVD event diagnosed or reported, otherwise the time from birth to the last known doctor visit time is recorded to be the censored event time. The indicator variables would be 1 if the corresponding CVD event occurred or 0 otherwise.

SBP and DBP are extracted from original datasheet given that the associated age variable is the nearest to 30 years old, 40 years old and 50 years old. Some other statistical summarizations are produced by SAS.

All the statistical analysis in terms of C index was performed using R package named 'compare*C*' [10] under R x64 3.1.1. The package aims to statistically compare two C indices with right censored survival outcome, which commonly arise from paired design [10] .

CHAPTER III

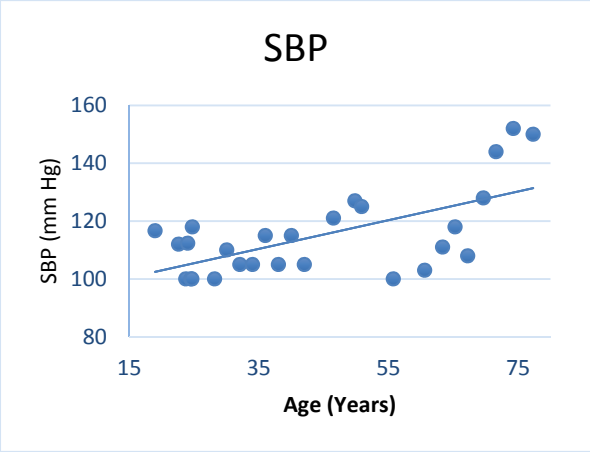
RESULTS

3.1. Summary of Statistics

3.1.1. Predictive Score – SBP and DBP

3.1.1.1.Patterns of Change in SBP and DBP

Several clinical trials already showed, SBP continues to rise due to the fact that loss of elasticity of the major arteries was an unavoidable consequence of aging and augmented by hypertension, in contrast to DBP which was thought to be a function of peripheral resistance [7]. Also, another landmark observational Framingham Heart Study found 94% of the population had an elevated SBP [7]. Here, we choose a random participant whose SBP and DBP measurements are longitudinally followed up over time from age range 18 years old to end of follow up for an illustration. Figure 3.1 illustrated the observed SBP and DBP elevated pattern for this participant, which has been followed from 18.98 to 77.33 years old. Maximum value of SBP (152 mm Hg) was detected at age of 74.28 years old, maximum value of DBP (92 mm Hg) was detected at age 77.33 years old. We also notice that SBP measurements may have more fluctuation around linear trend line comparing with DBP measurements around its trend line after age 55 years old. DBP value goes more steadily beyond 55 years old.



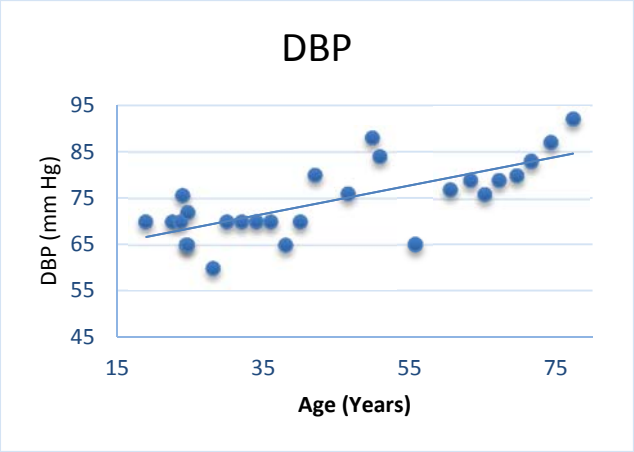


Fig.3.1 Observed Subject and linear pattern of SBP *Fig.3.2 Observed Subject and linear pattern of DBP*

Base on the table 3.1, we could tell mean of SBP (111.05±12.74 mm Hg) in age interval 18 to 34 years old is less than SBP (113.77±13.17 mm Hg) in the age interval 35 to 44 years old, similarly, SBP (132.51±20.94 mm Hg) in age interval 55 years old to the end of follow up is bigger than SBP (119.76±17.61 mm Hg) in the age interval 45 to 55 years old. From the second and the third column of the following table, mean of SBP and DBP rise along with the rising of age, and standard deviation of SBP has the same elevated changing pattern. SBP value has biggest variation with standard deviation (20.93) during the age range (>55). In contrast, DBP value has biggest variation with SD (11.99) at age range (45 -55), then SD decrease to 11.20 at age range (>55). To this end, we observe similar pattern for mean and standard deviation changing from sample population to that of the randomly selected participant previously.

Table.3.1 SBP changing pattern

SBP (mm Hg)	**Obs.**	**Mean**	**SD**	**Min**	**Max**
Age Range(18 -34)	1251	111.0546219	12.2746256	80	170
Age Range(35 -44)	750	113.7713333	13.17407	80	164
Age Range(45 -55)	582	119.7551546	17.6150684	75	210
Age Range > 55	587	132.5085172	20.9372328	80	232.5

Table.3.2 DBP changing pattern

DBP (mm Hg)	Obs.	Mean	SD	Min	Max
Age Range(18 -34)	1251	69.4281887	9.711961	40	118
Age Range(35 -44)	750	73.0053333	10.8057542	37	115
Age Range(45 -55)	582	75.9415808	11.9957332	35	125
Age Range > 55	587	77.130725	11.2035562	48	162

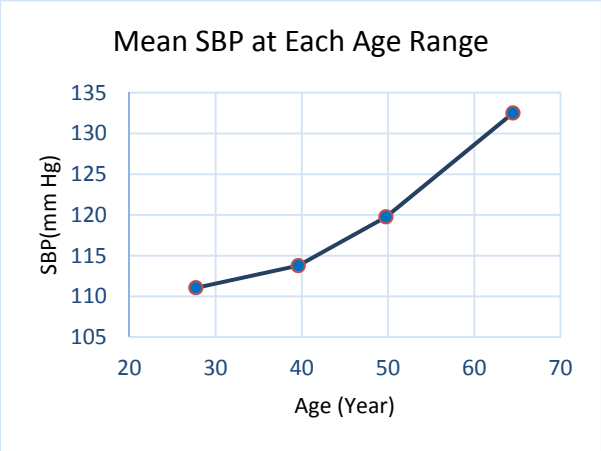


Fig.3.3 Mean SBP at different age groups

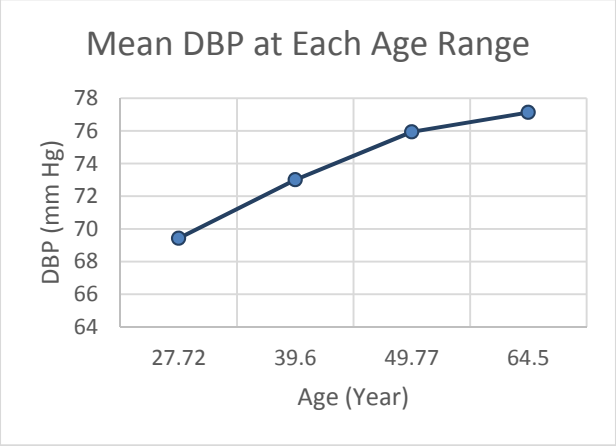


Fig.3.4 Mean DBP at different age groups

The patterns of mean changes in SBP and DBP were summarized based on our study sample which was divided into four different age groups. Visually, compared with trend of the mean SBP elevating model, the mean of DBP keeps rising, but the curve goes smoothly over time. We also noticed the rate of change of mean DBP at age interval (49.77 – 64.5 years old) is smaller than the rate of change of mean SBP at same age interval.

3.1.1.2. The patterns of Variance for SBP and DBP

Variance changing patterns of SBP and DBP were summarized based on the results from 1791 subjects. Mean ages are 27.72, 39.6, 49.77 and 64.5 years old with SBP variance

being 150.67, 173.56, 310.29 and 438.37, respectively, in four distinct age groups, while the DBP variance is 94.32, 116.764, 143.898 and 125.52 respectively.

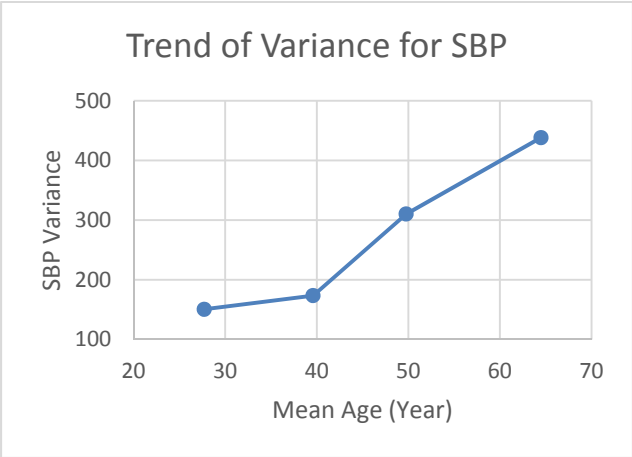


Fig.3.5 SBP Variance at different age groups

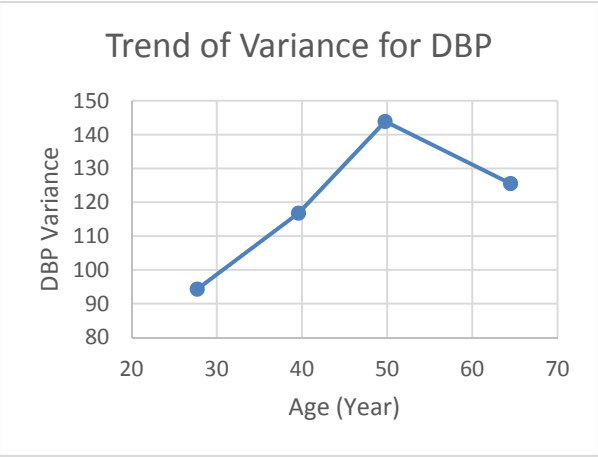


Fig.3.6 DBP Variance at different age groups

Variance plot of SBP indicate the variation grows up while the biological parameter-age rises, however it's still not clear if the aging affects variability of SBP. Meanwhile, the variation for DBP decrease with elder age group. These result shows aging not merely has influence on the magnitude of blood pressure including SBP and DBP, but causes a complex variability pattern [30] .

3.1.2. Survival outcomes

3.1.2.1.Summarization for survival outcomes

We define high cholesterol, high blood pressure, heart attack and heart failure as events of interest. Subjects may carry one or multiple events. The time to event is the age of participant

at which he was diagnosed of having certain events by doctor. The following table is the summarization for our time to event outcomes.

Table.3.3 Summarization for survival outcomes-Age at diagnosed

Outcomes	No.	Mean	Std Dev	Min	Max
High Cholesterol	215	50.875302	13.277934	15.31	87.33
High BP	570	49.261143	16.2761952	8.65	93.00
Heart Attack	11	62.512727	8.8608883	56.07	87.17
Heart Failure	11	63.067273	15.314839	36.44	80.7

3.2. Comparison of C indices between predict variables

3.2.1. Individual *C* indices and related P value

In our scenario, if event indicator is 1, then observed right-censored survival time is true survival time (age), otherwise, it recorded the censoring time (age). Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are common risk factors for cardiovascular disease [10]. We calculate overall *C* index of individual biomarker for each event at three different age group. Under this setting, the C_{xy} is a conditional probability between the survival time *X* for a certain disease event and the predictive score *Y,* SBP, C_{xz} is a conditional probability between the same survival time *X* and the predictive score *Z,* DBP.

If the estimate value of C index is less 0.5, this indicates that lower predictive scores or lower observed values link to longer event-free survival time. In contrast, if *C* index value is greater than 0.5, this suggests that subject has higher observed value relates to longer time to developing certain event/disease. The *Z* test reports the p value for each of individual *C* index, we denote P_{xy} is the p value for C_{xy}, while P_{xz} is the p value for C_{xz}, and if p value is smaller

than 0.05, that suggests us *C* value is statistically different from 0.5. The following table is a summation for predictive score variable SBP and DBP.

Table.3.4 Summarization for individual C indices and P value -for SBP ~ DBP

Events	***Predictive Score***	***C indices & Pvalue***	25-35 (Years)	35-45 (Years)	45-55 (Years)
High BP	**SBP**	***Cxy***	0.253	0.26	0.25
		Pxy	0	0	0
	DBP	***Cxz***	0.276	0.25	0.25
		Pxz	0	0	0
		ΔC	-0.022	0.01	0
High Cholesterol	**SBP**	***Cxy***	0.455	0.46	0.529
		Pxy	0.13	0.18	0.32
	DBP	***Cxz***	0.442	0.43	0.475
		Pxz	0.07	0.02	0.4
		ΔC	0.013	0.03	0.054
Heart Attack	**SBP**	***Cxy***	0.459	0.421	0.396
	DBP	***Cxz***	0.416	0.409	0.447
		ΔC	0.043	0.012	-0.051
Heart Failure	**SBP**	***Cxy***	0.48	0.552	0.512
	DBP	***Cxz***	0.714	0.609	0.646
		ΔC	-0.233	-0.057	-0.134

Since participants are longitudinally followed up, intra-subject variances of observed SBP value for each subjects are also considered as predictive scores, variance is 0 if a subject only has one observation for SBP in this study sample. Here, we want to investigate if lower observed variation of a certain biomarker, e.g., SBP, has longer survival time. C_{xy} and C_{xz} have almost same definition as before, however, the predictive score *Y* is variance of SBP, but *Z* is SBP itself. Meanwhile, *P* value is reported for each of *C* index value. For instance, *P* value for individual C indices of SBP in predicting high cholesterol event is significant at either mid-age

and elder-age groups, say, SBP always has better prognostic performance in determining which subject has later onset of high cholesterol. As illustrated in the following table, all the C indices and related *P* values are reported.

Table.3.5 Summarization for individual C indices-for Variance of SBP ~ SBP

Events	***Predictive Score***	***C indices & Pvalue***	25-35 (Years)	35-45 (Years)	45-55 (Years)
High BP	**Var of SBP**	***Cxy***	0.45	0.38	0.51
		Pxy	0.13	<0.001	0.87
	SBP	***Cxz***	0.32	0.3	0.42
		Pxz	<0.001	<0.001	0.28
		Δ*C*	0.13	0.07	0.09
High Cholesterol	**Var of SBP**	***Cxy***	0.505	0.42	0.416
		Pxy	0.86	0.003	0.002
	SBP	***Cxz***	0.455	0.463	0.529
		Pxz	0.13	0.18	0.32
		Δ*C*	0.05	-0.038	-0.112
Heart Attack	**Var of SBP SBP**	***Cxy***	0.463	0.346	0.521
		Cxz	0.459	0.421	0.396
		Δ*C*	0.003	-0.074	0.126
Heart Failure	**Var of SBP SBP**	***Cxy***	0.576	0.343	0.477
		Cxz	0.48	0.552	0.512
		ΔC	0.094	-0.209	-0.034

Previous table illustrates association between intra-subject variance of observed SBP value and event free survival time. Also, we consider model based mean squared error (MSE) of SBP as our predictive scores *Y*, linear regression model is applied for prediction of SBP value. According the prior formula, we are able to calculate MSE value for each subject at specific age interval, MSE value is considered as zero if there is only one observation in specific age group for single subject.

Table.3.6 Summarization for individual C indices-for MSE of SBP ~ SBP

Events	***Predictive Score***	***C indices & Pvalue***	25-35 (Years)	35-45 (Years)	45-55 (Years)
High BP	**MSE of SBP**	***Cxy***	0.47	0.467	0.52
		Pxy	0.11	0.07	0.45
	SBP	***Cxz***	0.25	0.26	0.25
		Pxz	0	0	<0.001
		ΔC	0.216	0.209	0.264
High Cholesterol	**MSE of SBP**	***Cxy***	0.53	0.486	0.454
		Pxy	0.27	0.53	0.05
	SBP	***Cxz***	0.455	0.463	0.529
		Pxz	0.13	0.18	0.31
		ΔC	0.075	0.023	-0.075
Heart Attack	**MSE of SBP**	***Cxy***	0.531	0.518	0.511
	SBP	***Cxz***	0.459	0.421	0.396
		ΔC	0.072	0.097	0.115
Heart Failure	**MSE of SBP**	***Cxy***	0.592	0.435	0.447
	SBP	***Cxz***	0.48	0.552	0.512
		ΔC	0.111	-0.117	-0.065

Variance of SBP and variance of DBP are estimated for each subjects within study sample, which are considered as target predictive scores. In this proposed setting, *Y* is variance of SBP and *Z* is variance of DBP. We notice most of *C* indices value in the following table are all less than 0.5 in this scenario, which could explain that lower variation of blood pressure over time relates to later onset of cardiovascular disease.

Table.3.7 Summarization for individual C indices-for Variance of SBP ~ Variance of DBP

High BP			**High Cholesterol**			**Heart Attack**			**Heart Failure**		
Cxy	***Cxz***	***ΔC***	***Cxy***	***Cxz***	***ΔC***	***Cxy***	***Cxz***	***ΔC***	***Cxy***	***Cxz***	**ΔC**
0.342	0.316	0.03	0.423	0.385	0.038	0.345	0.31	0.035	0.344	0.353	0.009

3.2.2. Comparing two correlated *C* indices

In the view of above results, we have individual *C* index value and difference between two correlated *C* indices. The null hypothesis is that there is no difference between two *C* indices, $H_0 : C_{XY}^{g} = C_{XZ}^{g}$. The test to be used is a z score test. We would reject null hypothesis if the absolute value of z score is great than 0.975 quantile from a standard normal distribution (two-sided test).

Table.3.8 Summarization of P-values for ΔC: SBP ~ DBP

Predict Variable: SBP vs. DBP				
	High BP	**High Cholesterol**	**Heart Attack**	**Heart Failure**
Age30	0.12	0.61	N/A	N/A
Age40	0.19	0.05	N/A	N/A
Age50	0.98	0.004	N/A	N/A

No strong evidence shows two *C* index values for predictive variables SBP and DBP are different for the high BP, the high cholesterol, heart attack and heart failure event free survival time at age group 30, 40 and 50. Since there are only 11 heart attack and heart failure disease events out of 1791 subjects, with this high rate of censoring, the p-value estimates are not reported here.

For testing for another pair of predictive variable, i.e., variance of SBP and SBP, predictive ability at different age interval are compared under similar setting.

Table.3.9 Summarization of P-values for ΔC: variance of SBP ~ SBP

Predict Variable: Variance of SBP vs. SBP				
	High BP	**High Cholesterol**	**Heart Attack**	**Heart Failure**
Age30	0.007	0.244	N/A	N/A
Age40	0.077	0.337	N/A	N/A
Age50	0.372	0.002	N/A	N/A

In this application, we conclude that two *C* indices difference between variance of SBP and SBP is significant for high cholesterol event survival time at age50 group, also, difference is significant for high blood pressure event free survival time at age group30. That indicates variance of SBP is significantly different than SBP in predicting certain event-free survival time in certain age interval.

Performing the third statistical test, the difference is significant between two corrected *C* indices at specific age group, and we could conclude MSE of SBP has inferior performance in predict high blood pressure at age30 and age40.

Table.4.0 Summarization of P-values for ΔC: MSE of SBP ~ SBP

Predict Variable: Variance of SBP vs. SBP				
	High BP	**High Cholesterol**	**Heart Attack**	**Heart Failure**
Age30	<0.001	0.08	N/A	N/A
Age40	0	0.51	N/A	N/A
Age50	<0.01	0.04	N/A	N/A

We also are interesting in comparing the predictive ability between the predictive scores, variance of SBP and variance of DBP.

Table.4.1 Summarization of P-values for ΔC: variance of SBP ~variance of DBP

Predict Varible : Variance of SBP vs. Variance of DBP				
	High BP	**High Cholesterol**	**Heart attack**	**Heart failure**
p-value	0.01	0.022	N/A	N/A

Z test reports the *p*-value of 0.01 and 0.022, i.e., variance of DBP is a better prognostic biomarker in predicting high blood pressure and high blood cholesterol event free survival time when comparing with the variance of SBP. It is reasonable to estimate the certain event-free survival time through looking at variation of time-dependent covariate.

CHAPTER IV

DISCUSSION

4.1. Conclusion and limitation

In general, *C* index estimates the probability of concordance between the biomarker and survival outcome. The higher than 0.5 *C* statistic indicates that right-censored survival time and predict variable SBP or DBP in adulthood go in the same direction, however, in this study, most of *C* statistic are less than 0.5, this can be interpreted as lower numeric value of predictors related to longer observed event-free survival time. Then we applied *z* score test to statistically compare two *C* indices, with some of results showing significance.

Our finding does suggest that there is no significant difference between the systolic blood pressure (SBP) and diastolic blood pressure (DBP) in predicting time to high blood pressure event for adulthood in survival setting. These results are in line to those found in other literature reviews, usually, SBP elevation comes together with the elevation of DBP. In the National Health Examination Survey (NHES) and National Health and Nutrition Examination Survey (NHANES) studies, few of subjects with elevated BP readings had (isolated) elevation of SBP with a normal DBP [31] in young adults. Also from figure 3.1.3 and figure 3.1.4, mean of SBP and DBP rise simultaneously with age rise. Therefore, this might be reason that predictive ability of SBP and DBP have no significant difference in determining which subject has longer HBP free survival time. While, we also notice p value indicate that individual *C* indices of SBP and DBP have significant difference from 0.5 when predicting HBP event, thus we could claim, SBP and DBP are important individual indicators for HBP event.

P value for comparing two correlated C index in predicting both heart attack and heart failure is not reported, since there are only 11 events either for heart attack event and heart failure event out of 1791 subjects.

From Table 3.8, we could not find strong evidence shows that either SBP or DBP has better prognostic performance in predicting high cholesterol event, while the difference found in this study between two correlated *C* indices is significant. It stands to reason that difference may play an albeit non-direct role. A closer inspection of comparison between SBP and DBP may shed light on the interconnections between blood pressure and cholesterol level by providing a more biological parameters, like sex, medicine history etc.

When comparing two correlated *C* index between MSE of SBP and SBP at three different age intervals, we found the SBP itself has significantly different predictive ability in predicting HBP event. Also, the *C* index for SBP has lower value than *C* index for MSE of SBP, it suggests that MSE of SBP has inferior performance in predicting HBP and high cholesterol event.

Significant difference is observed between the variance of SBP and variance of DBP in predicting HBP and high cholesterol.

We realize that the method we applied here works well across a large sample size, and the *z*-score test is also based on a known population variance and large sample size, however, for the small sample size, t statistics might be appropriate to be estimated based on a known sample standard error. Meanwhile, the date variable which has been recorded as format like 'Month/Day/Year', '99' represent missing value if no month record, also '99' represent missing value if no day record. In this situation, we apply numeric value 12 instead of '99' for month, 28 for day value. The accuracy between the true date and date we applied might cause certain error in this study.

4.2. Future Work

Our present study and analyses shows some biomarkers have better performance in predicting the certain right-censored survival time. Linking to survival probabilities associated with HBP and CVDs may improve the focus of future research.

In our furture work, firstly, we would like to consider Cox proportional hazards regression model combining multiple predictors, $h(t)/h_0(t) = \exp(\beta_1 X_1 + \beta_2 X_2)$, here hazard ratio is exponential function of the linear combination of predictors, say, the predictors have proportional effect on the predicted hazard ratio. Through this model, we are able to estimate parameters β_1 and β_2, and then let $z = \widehat{\beta_1} X_1 + \widehat{\beta_2} X_2$, be the linear function of two predictors and adjusted through estimated parameters. We are interesting in comparing the difference between the z we propose and individual predictor X_1 or X_2 like SBP or DBP in term of C index. If the test statistics with the p-value is significant at 0.05, we could claim that the combination of risk factors have better predictive ability for the certain right-censored event free survival time than the single one.

Second, considering variation calculation based on the scaled sum of change between an individual profile and the population profile. We calculate distance Δd_j between individual blood pressure measures and U.S population blood pressure measure at same age, and Δd_{j-1} is the distance between previous measures and population mean for same subject. The distance could be positive number or negative numeric value. Since our interest is measuring the fluctuation of individual blood pressure, we will sum all the differences between two adjacent distance together of same individual, say $\sum |\Delta d_j - \Delta d_{j-1}|$, furthermore, considering times of visit for individual are varying, some individuals might have several serial data records, others

might has only one visit, thus we divided summation by times of visit. We estimate the fluctuation for individual through formula as following,

$$\frac{1}{n_{ij}} \sum_{j=1}^{n_j} |\Delta d_{ij} - \Delta d_{ij-1}|$$

Where i =1, 2,…N subject, n_{ij} are times of measurement for ***i***th individual, Δd_{ij} is distance between the *j*th blood pressure and population mean for ***i***th subject at a given age, Δd_{ij-1} is distance between the *(j*-1*)*th blood pressure and population mean for same subject at a given age. Here the population mean is mean blood pressure of U.S. population.

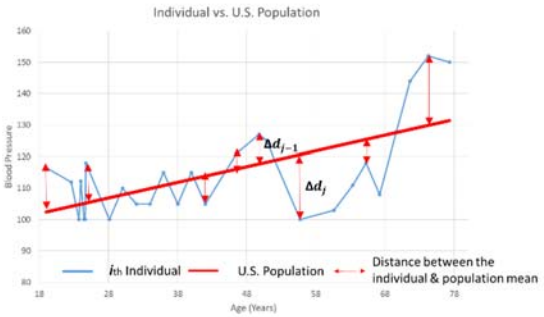


Fig.4.1 Individual BP has higher variation

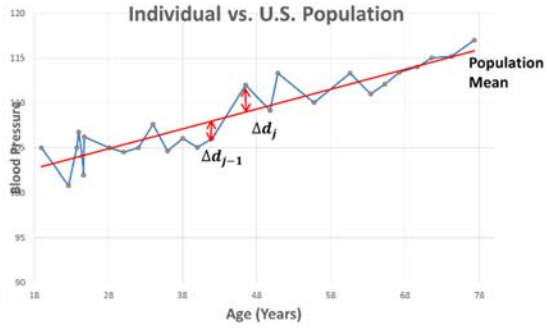


Fig.4.2 Individual BP lower variation

Fig 4.1 and *Fig 4.2* are two simulated illustrations for individual *vs*. population mean, obviously, *Fig 4.1* individual plot shows more fluctuation of blood pressure comparing with population mean. Though all the blood pressure measures in *Fig 4.2* are not exactly fall in the population mean curve, but the difference between two adjacent distance is small, also the summation of absolute difference is smaller than previous one, thus we could come up with the conclusion individual in *Fig. 4.2* has less fluctuation of blood pressure over time comparing with individual in *Fig. 4.1*.

Fig 4.3 and Fig 4.4 illustrate both individuals have higher baseline measures over lifetime, comparing with the population mean. Though subject in *Fig.4.4* has relatively larger distance in term of his measurements over time, however, we notice individual profile curve almost paralleled with the population profile curve, say, the individual blood pressure elevates smoothly when age rises, in other word, that subject has less variation of blood pressure over time. In opposite, individual profile curve in *Fig 4.3* has more narrow spikes, say, this subject we are interested in would have more variation of blood pressure over time.

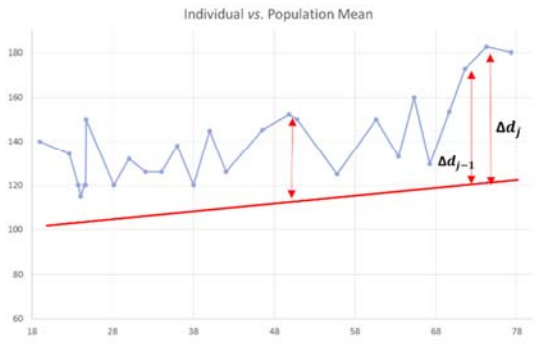


Fig.4.3 Individuals BP with higher baseline measures and higher variation

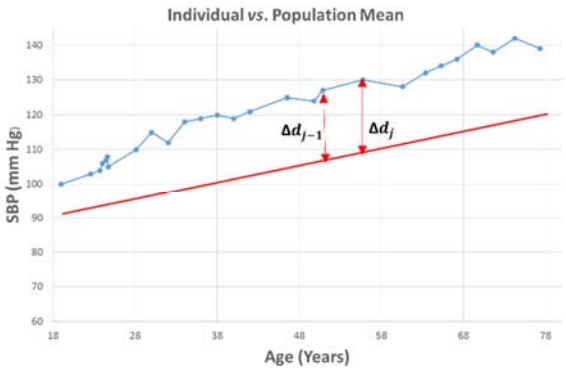


Fig.4.4 Individuals BP with higher baseline measures and lower variation

Last, based on modeling the population trend and giving each individual a time-windows based variation measure, we are interested in if the variation in this specific time-window predict the survival outcomes. We are also interested in comparing variation from different time-windows to find which has better predictive ability for time of event outcomes in survival setting.

REFERENCE

1. World Health Organization. Cardiovascular diseases (CVDs). fact sheet N°317. http://www.who.int/mediacentre/factsheets/fs317/en/. Accessed 04/06, 2015.

2. World Health Organization. Global status report on noncommunicable diseases 2014. charpter 1. globle target 1. In: Vol 1. ; 2014:9.

3. Safar ME, Levy BI, Struijker-Boudier H. Current perspectives on arterial stiffness and pulse pressure in hypertension and cardiovascular diseases. *Circulation*. 2003;107(22):2864-2869. doi: 10.1161/01.CIR.0000069826.36125.B4.

4. National Institutes of Health National Heart, Lung, and Blood Institute. High blood cholesterol: What you need to know. . 2001;No. 05-3290.

5. The sixth report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Archives of internal medicine*. 1997;157(21):2413-2446.

6. Chalmers J, MacMahon S, Mancia G, et al. 1999 world health organization-international society of hypertension guidelines for the management of hypertension. guidelines sub-committee of the world health organization. *Clinical and experimental hypertension*. ;21(5-6):1009-1060. doi: 10.3109/10641969909061028.

7. Tin LL, Beevers DG, Lip GYH. Systolic vs diastolic blood pressure and the burden of hypertension. *Journal of human hypertension*. 2002;16(3):147-150. doi: 10.1038/sj.jhh.1001373.

8. Guo SS, Zeller C, Chumlea WC, Siervogel RM. Aging, body composition, and lifestyle: The fels longitudinal study. *The American journal of clinical nutrition :*. 1999;70(3):405-411.

9. Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Statistics in medicine*. 2004;23(13):2109-2123.

10. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: A one-shot nonparametric approach. *Stat Med*. 2015;34(4):685-703. doi: 10.1002/sim.6370.

11. Din-Dzietham R, Liu Y, Bielo M, Shamsa F. High blood pressure trends in children and adolescents in national surveys, 1963 to 2002. *Circulation*. 2007;116(13):1488-1496. doi: 10.1161/CIRCULATIONAHA.106.683243.

12. Von Hippel PT, Nahhas RW. Extending the history of child obesity in the united states: The fels longitudinal study, birth years 1930-1993. *Obesity*. 2013;21(10):2153-2156. doi: 10.1002/oby.20395.

13. Alex F. Roche. *Growth, maturation, and body composition
The fels longitudinal study 1929–1991*. Cambridge University Press; 1992.

14. Lifespan Health Research Center, Wright State University. Fels longitudinal study history. https://medicine.wright.edu/lifespan-health-research-center/fels-longitudinal-study/history. Updated 2015. Accessed 04/07, 2015.

15. Sabo RT, Lu Z, Daniels S, Sun SS. Relationships between serial childhood adiposity measures and adult blood pressure: The fels longitudinal study. *Am J Hum Biol*. 2010;22(6):830-835. doi: 10.1002/ajhb.21091.

16. Guo SS, Chumlea WC, Cockram DB. Use of statistical methods to estimate body composition. *Am J Clin Nutr*. 1996;64(3):428S.

17. Guo SS, Chumlea WC, Roche AF, Siervogel RM. Age- and maturity-related changes in body composition during adolescence into adulthood: The fels longitudinal study. *International journal of obesity*. 1997;21(12):1167-1175.

18. Sun SS, Grave GD, Siervogel RM, Pickoff AA, Arslanian SS, Daniels SR. Systolic blood pressure in childhood predicts hypertension and metabolic syndrome later in life. *Pediatrics*. 2007;119(2):237.

19. National high blood pressure education program working group report on hypertension in the elderly. national high blood pressure education program working group. *Hypertension*. 1994;23(3):275-285.

20. Go AS, Mozaffarian D, Roger V, et al. Heart disease and stroke statistics--2013 update: A report from the american heart association. *Circulation*. 2013;127(1):e6-e245. doi: 10.1161/CIR.0b013e31828124ad.

21. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics--2015 update: A report from the american heart association. *Circulation*. 2015;131(4):e29-322. doi: 10.1161/CIR.0000000000000152.

22. National Institutes of Health National Heart, Lung, and Blood Institute. How is a heart attack diagnosed? http://www.nhlbi.nih.gov/health/health-topics/topics/heartattack/diagnosis. Updated 2013. Accessed 04/08, 2015.

23. Centers of Disease Control and Prevention. Heart disease facts. http://www.cdc.gov/heartdisease/facts.htm. Updated 2015. Accessed 04/08, 2015.

24. WebMD. Heart attacks and heart disease. http://www.webmd.com/heart-disease/guide/heart-disease-heart-attacks?page=4. Updated 2013. Accessed 04/08, 2015.

25. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Statistical methods for the assessment of prognostic biomarkers (part I): Discrimination. *Nephrology dialysis transplantation*. 2010;25(5):1399-1401. doi: 10.1093/ndt/gfq018.

26. D'Agostino RB, Griffith JL, Schmidt CH, Terrin N. Measures for evaluating model performance. . 1997(American Statistical Association):253-258.

27. Bamber D.
The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*. 1975:387-415.

28. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. 1996;15(4):361-387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

29. Casella G BR. *Statistical inference.* CA: Ducbury Press; 2001.

30. Veerman DP, Imholz BP, Wieling W, Karemaker JM, van Montfrans,G.A. Effects of aging on blood pressure variability in resting conditions. *Hypertension*. 1994;24(1):120-130.

31. Blank SG, Mann SJ, James GD, West JE, Pickering TG. Isolated elevation of diastolic blood pressure. real or artifactual? *Hypertension*. 1995;26(3):383-389.

APPENDIX A

SAS CODE FOR PREPARATION OF DATA

```
/*************************************************************************
                    Data set with all the cardiovascular events and
                      baseline measurements at some time points
*************************************************************************/

/*************************************************************************
data set with all the cardiovascule events and part of part
of baseline measurements
*************************************************************************/

libname fels "E:\Projects\VCU\fels";
options fmtsearch=(fels.fmt64);
proc contents data=fels.vcu0610;run;
ODS HTML CLOSE;       ODS HTML;


/* joint two data sets by ptno*/


data sample_4; set fels.vcu0610;
keep ptno visit age sex  ANbpsys /*ANbpd4*/ ANbpd5 ANstature ANweight BCbmi  HQcardt4 HQcardt5
HQcardt6 HQcardt7;
if ptno=. then delete;
run;

data birthday; set fels.vid0609;  /*there are 4115 patients*/
keep ptno sex IDbirthdate;
if ptno=. then delete;
if IDbirthdate=. then delete;
run;

/* since ANbpd4 has much more missing value than ANbpd5, thus i would like exclude the ANbpd4 from dataset
and i also exclude missing value from ANbpd5 and ANbpsys as joint condition */

proc sql;
create table sample_study as select
a.*, b.* from birthday as a join sample_4 as b
on
a.ptno = b.ptno and
b.ANbpsys^=. and
b.ANbpd5^=. ;
quit;


/*proc sort data=sample_study nodupkey ;
by ptno;
run;*/

/* delete missing value from data set sample_study*/


data one; set sample_study;
if HQcardt4^=. or HQcardt5^=. or HQcardt6^=. or HQcardt7^=.;
run;
```

```
data one; set one;
if  HQcardt4^="" or HQcardt5^="" or HQcardt6^="" or HQcardt7^="";
drop sex visit ANstature ANweight BCbmi ;
run;

data one_update; set one;
run;

data one_update; set one_update;
if HQcardt6="99999999" or HQcardt7="99999999" or HQcardt4="99999999" or HQcardt5="99999999" then
delete;
run;

proc sql;  /*there are 303 unique ptno*/
select count(*) as total, count(distinct ptno)as number
from one_update;
quit;



data two; set one_update;
if substr(HQcardt4,1,2)=99 then substr(HQcardt4,1,2)=12;
if substr(HQcardt4,3,2)=99 then substr(HQcardt4,3,2)=28;
if substr(HQcardt5,1,2)=99 then substr(HQcardt5,1,2)=12;
if substr(HQcardt5,3,2)=99 then substr(HQcardt5,3,2)=28;
if substr(HQcardt6,1,2)=99 then substr(HQcardt6,1,2)=12;
if substr(HQcardt6,3,2)=99 then substr(HQcardt6,3,2)=28;
if substr(HQcardt7,1,2)=99 then substr(HQcardt7,1,2)=12;
if substr(HQcardt7,3,2)=99 then substr(HQcardt7,3,2)=28;
run;


data three;
set two;
format HQcardt4_date HQcardt5_date HQcardt6_date HQcardt7_date mmddyy10.;
HQcardt4_date=input(trim(HQcardt4),MMDDYY10.);  /*convert charactor to numeric-> input(souce, informat)*/
HQcardt5_date=input(trim(HQcardt5),MMDDYY10.);
HQcardt6_date=input(trim(HQcardt6),MMDDYY10.);
HQcardt7_date=input(trim(HQcardt7),MMDDYY10.);
datest = put(IDbirthdate,8.0);                /*convert numeric to charactor*/
yr = input(substr(datest,1,4),4.0);           /*break the string, convert substring to numeric*/
mon= input(substr(datest,5,2),2.0);
dat = input(substr(datest,7,2),2.0);
birth = mdy(mon,dat,yr);                    /*return to numeric value*/
testdate4 = HQcardt4_date;
testdate5 = HQcardt5_date;
testdate6 = HQcardt6_date;
testdate7 = HQcardt7_date;
format birth mmddyy10. testdate4 mmddyy10. testdate5 mmddyy10. testdate6 mmddyy10. testdate7 mmddyy10. ;
/*birthyr = year(birth);*/
/*agenew = (testdate6 - birth)/365.25;*/
agenew4 = yrdif(birth,testdate4,'AGE of heart failure');
agenew5 = yrdif(birth,testdate5,'AGE of heart attack');
agenew6 = yrdif(birth,testdate6,'AGE at high bp');
agenew7 = yrdif(birth,testdate7,'AGE of high chol');
run;
```

```
data four;
set three;
age_heart_failure=round(agenew4,.01);
age_heart_attack=round(agenew5,.01);
age_hp=round(agenew6,.01);
age_hc=round(agenew7,.01);
drop testdate4 testdate5 testdate6 testdate7
    HQcardt4_date HQcardt5_date HQcardt6_date HQcardt7_date
    HQcardt4 HQcardt5 HQcardt6 HQcardt7
            yr mon dat datest birth
    agenew4 agenew5 agenew6 agenew7;
run;


/*************************************************************************************/
/*check ptno from questionnaire, data set four from sas code fels0204 */

data ck_hp; set four; /* dataset four from fels0204 */
keep ptno age_hp;
if age_hp^=. or age_hp^="";
run;

data ck_hp; set ck_hp; /*190 ptno*/
proc sort data=ck_hp nodupkey;
by ptno;
run;

data ck_hc; set four; /* dataset four from fels0204 */
keep ptno age_hc;
if age_hc^=. or age_hc^="";
run;

data ck_hc; set ck_hc; /*216 ptno*/
proc sort data=ck_hc nodupkey;
by ptno;
run;

data ck_ha; set four; /* dataset four from fels0204 */
keep ptno age_heart_attack;
if age_heart_attack^=. or age_heart_attack^="";
run;

data ck_ha; set ck_ha; /*11 ptno*/
proc sort data=ck_ha nodupkey;
by ptno;
run;

data ck_hf; set four; /* dataset four from fels0204 */
keep ptno age_heart_failure;
if age_heart_failure^=. or age_heart_failure^="";
run;

data ck_hf; set ck_hf; /*11 ptno*/
proc sort data=ck_hf nodupkey;
by ptno;
```

```
run;

data five;
merge ck_hc ck_hp ck_ha ck_hf;
by ptno;
run;

proc means data=five; /*total 303 ptno*/
var age_hp age_hc age_heart_attack age_heart_failure;
run;

proc options option=work;run; /* export data set*/


libname lib "C:\Users\Yongjia\Desktop";
data five; set lib.five;
run;
proc means data=five;
run;

data all; set lib.all; /*from sas code 'fels0124_all'*/
run;

data all_five; set lib.all_five;
run;

proc sql;
create table all_five as select    /*all_five have all the patients with event*/
a.*, b.* from all as a left join five as b on
a.ptno=b.ptno ;
quit;

proc means data=all_five;  /*after left joint, lost 1 event from age_hc*/
var age_hp age_hc age_heart_attack age_heart_failure;
run;

/*hp=190, hc=215, hta=11, htf=11*/

/**********THIS dataset include all the event defined as diagnosis date event***************/

proc export data=all_five
  outfile='C:\Users\Yongjia\Desktop\all_five.csv';
  run;


/*****************************************************************
  add measurments for no event part at age 30 40 and 50
 *****************************************************************/
libname FELS "C:\Users\Yongjia\Desktop";
data test; set FELS.new;      /*dataset "new" contain all the obs where age >= 18*/
run;

data test; set test;
drop age_event age_30;
run;
proc sql;
```

```
create table test as
select *, max(ind) as IND_MAX
from test
group by ptno;
quit;

data no_event; set test;  /*subset for no event patient*/
if IND_MAX=0;
run;

data no_event; set no_event;
drop IND_MAX;
run;

 proc sort data=no_event;
 by ptno age;
 run;

proc sql;                /*creat subset at age 30*/
create table test1 as
select *, abs(age-30) as age_30, min(abs(age-30)) as age_30_min
from no_event
group by ptno;
quit;

data test1_30; set test1;  /*there are 2 obs for ptno 2769*/
if age_30=age_30_min;
run;
data test1_30; set test1_30;
age_30=1;
drop age_30_min;
run;

data test1_30; set test1_30;
if first.ptno;
by ptno;
run;

proc sql;                /*creat subset at age 40*/
create table test2 as
select *, abs(age-40) as age_40, min(abs(age-40)) as age_40_min
from no_event
group by ptno;
quit;
data test2_40; set test2;
if age_40=age_40_min;
run;
data test2_40; set test2_40;
age_40=1;
drop age_40_min;
run;

proc sql;                /*creat subset at age 50*/
create table test3 as
select *, abs(age-50) as age_50, min(abs(age-50)) as age_50_min
from no_event
```

```
group by ptno;
quit;
data test3_50; set test3;
if age_50=age_50_min;
run;
data test3_50; set test3_50;
age_50=1;
drop age_50_min;
run;

data noevent_sample;          /*append subsets together*/
set test1_30 test2_40 test3_50;
run;


proc sort data=noevent_sample; by ptno;
run;

data noevent_sample; set noevent_sample ;  /*blood pressure for all patient without event at age30, 40,50 */
if age_30=1 then age_ind=30;
if age_40=1 then age_ind=40;
if age_50=1 then age_ind=50;
drop age_30 age_40 age_50;
run;


proc transpose data=noevent_sample out=trans_noevent prefix=age_ind;
  by ptno;
  id age_ind;
  var ANbpsys ANbpd5;
run;

data trans_1; set trans_noevent; /* subset with sysbolic at age 30,40,50*/
if first.ptno ;
by ptno;
run;

data trans_2; set trans_noevent;  /* subset with diabolic at age 30,40,50*/
if last.ptno;
by ptno;
run;

data trans_sys30;
set trans_1(rename=(age_ind30=sys_age30 age_ind40=sys_age40 age_ind50=sys_age50));
keep ptno sys_age30;
run;

data trans_dia30;
set trans_2(rename=(age_ind30=dia_age30 age_ind40=dia_age40 age_ind50=dia_age50));
keep ptno dia_age30;
run;

data trans_30;          /*subset at age 30*/
merge trans_sys30 trans_dia30;
by ptno;
run;
```

```
data trans_sys40;
set trans_1(rename=(age_ind30=sys_age30 age_ind40=sys_age40 age_ind50=sys_age50));
keep ptno sys_age40;
run;

data trans_dia40;
set trans_2(rename=(age_ind30=dia_age30 age_ind40=dia_age40 age_ind50=dia_age50));
keep ptno dia_age40;
run;

data trans_40;   /*subset at age 40*/
merge trans_sys40 trans_dia40;
by ptno;
run;

data trans_sys50;
set trans_1(rename=(age_ind30=sys_age30 age_ind40=sys_age40 age_ind50=sys_age50));
keep ptno sys_age50;
run;

data trans_dia50;
set trans_2(rename=(age_ind30=dia_age30 age_ind40=dia_age40 age_ind50=dia_age50));
keep ptno dia_age50;
run;

data trans_50;   /*subset at age 50*/
merge trans_sys50 trans_dia50;
by ptno;
run;

data trans_all;  /*1791 ptno with sys and bio at 30, 40, 50*/
merge trans_30 trans_40 trans_50;
by ptno;
run;


data age_30;
set noevent_sample;
if first.ptno;
by ptno;
age_30=age;
keep age_30 ptno;
run;

data new30; /* subset with real age 30 and blood pressure*/
merge age_30 trans_30;
by ptno;
run;


data age_40;
set noevent_sample;
if age_ind=40;
age_40=age;
keep age_40 ptno;
```

```
run;

data new40; /* subset with real age 40 and blood pressure*/
merge age_40 trans_40;
by ptno;
run;

data age_50;
set noevent_sample;
if last.ptno ;
by ptno;
age_50=age;
keep age_50 ptno;
run;


proc sql; /*specified how many distinct ptno in data set five comparing with "all"*/
select count(ptno) from age_50 where ptno not in (select ptno from trans_50);
quit;

data new50; /* subset with real age 50 and blood pressure*/
merge age_50 trans_50;
by ptno;
run;



data trans_all_update ;  /* wide format with age */
merge new30 new40 new50;
by ptno;
run;


data final;
merge all_five trans_all_update;
by ptno;
run;


proc means data=final;
var age_hp age_hc age_heart_attack age_heart_failure;
run;

/*************************************************************************************************
**************
   add sex in final data set

*************************************************************************************************
***************/


data newSample_0; set fels.vcu0610;
run;

data newsample_1; set newsample_0; /*21823 obs with systolic and diastolic measurements*/
keep ptno visit age sex  ANbpsys  ANbpd5 ANstature ANweight BCbmi ;
```

```
if ptno=. then delete;
if ANbpsys^=. and  ANbpd5^=.;
run;



data birthday; set fels.vid0609;  /*there are 4115 patients*/
keep ptno sex IDbirthdate;
if ptno=. then delete;
if IDbirthdate=. then delete;
run;

proc sql;   /*joint two sample sets by  patient number*/
create table newsample_study as select
a.*, b.* from birthday as a join newsample_1 as b
on
a.ptno = b.ptno ; /*and
b.ANbpsys^=. and
b.ANbpd5^=. ; */
quit;


proc sort data=newsample_study; /*21823 obs */
by ptno;
run;


data newsample_2; set newsample_study;  /*creat event indicator*/
if ANbpsys ge 140 or ANbpd5 ge 90 then ind=1 ;
 else ind=0;
run;

 data new_sex; set newsample_2;  /*create subset, age >= 18*/
 if age ge 18;
 drop  visit IDbirthdate ANstature ANweight BCbmi ind age ANbpsys ANbpd5;
 run;

 proc sort data=new_sex nodupkey;
 by ptno;
 run;

 data final_update;
 merge new_sex final;
 by ptno;
 run;

/**********  this dataset with event defined as SBP >140 AND DBP >90  ***************/

proc export data=final_update
   outfile='C:\Users\Yongjia\Desktop\final_update.csv';
   run;

proc options option=work; run;
```

```
/*************  var_sys30 between age 25-30**********************/

libname lib "C:\Users\Yongjia\Desktop";

data V1 ; set lib.new;
drop age_event age_30 ind;
if  age <=35 and age >= 25;
run;

proc means var noprint;
output out=Variance1 var=var_sys30;
var ANbpsys ;
by ptno;
run;

data V1 ; set lib.new;
drop age_event age_30 ind;
if  age <=35 and age >= 18;
run;
proc means var noprint;
output out=Variance2 var=var_dia;
var ANbpd5 ;
by ptno;
run;

data V2;
merge variance1 variance2;
by ptno;
run;

data v3; set v2;
if var_sys30=. then var_sys30=0;
if var_dia=. then var_dia=0;
drop _type_ _freq_;
run;

data Q0; set lib.test3;
run;


proc sql;
create table t1 as select    /*merge two data set */
a.*, b.* from v3 as a left join q0 as b on
a.ptno=b.ptno ;
quit;

data t1; set t1;
keep ptno var_sys30;
run;

/*proc export data=t1
   outfile='C:\Users\Yongjia\Desktop\t1.csv';
   run; */


/****************** var_sys40 between 35-45**********************/
```

```
libname lib "C:\Users\Yongjia\Desktop";

data V1 ; set lib.new;
drop age_event age_30 ind;
if  age <=45 and age > 35;
run;

proc means var noprint;
output out=Variance1 var=var_sys40;
var ANbpsys ;
by ptno;
run;

data V1 ; set lib.new;
drop age_event age_30 ind;
if  age <=45 and age > 35;
run;
proc means var noprint;
output out=Variance2 var=var_dia;
var ANbpd5 ;
by ptno;
run;

data V2;
merge variance1 variance2;
by ptno;
run;

data v3; set v2;
if var_sys40=. then var_sys40=0;
if var_dia=. then var_dia=0;
drop _type_ _freq_;
run;

data Q0; set lib.test3;
run;


proc sql;
create table t2 as select    /*merge two data set */
a.*, b.* from v3 as a left join q0 as b on
a.ptno=b.ptno ;
quit;

data t2; set t2;
keep ptno var_sys40;
run;

/*proc export data=t2
   outfile='C:\Users\Yongjia\Desktop\t2.csv';

   run;*/

/************ var_sys50 between age 45-55**********/
```

```
libname lib "C:\Users\Yongjia\Desktop";

data V1 ; set lib.new;
drop age_event age_30 ind;
if  age <=55 and age > 45;
run;

proc means var noprint;
output out=Variance1 var=var_sys50;
var ANbpsys ;
by ptno;
run;

data V1 ; set lib.new;
drop age_event age_30 ind;
if  age <=55 and age > 45;
run;
proc means var noprint;
output out=Variance2 var=var_dia;
var ANbpd5 ;
by ptno;
run;

data V2;
merge variance1 variance2;
by ptno;
run;

data v3; set v2;
if var_sys50=. then var_sys50=0;
if var_dia=. then var_dia=0;
drop _type_ _freq_;
run;

data Q0; set lib.test3;
run;


proc sql;
create table t3 as select    /*merge two data set */
a.*, b.* from v3 as a left join q0 as b on
a.ptno=b.ptno ;
quit;

data t3; set t3;
keep ptno var_sys50;
run;

/*proc export data=t3
   outfile='C:\Users\Yongjia\Desktop\t3.csv';
run;*/



/******************joint var_sys30, var_sys40, var_sys50 together **********************/
```

```
data var_sys;
merge t1 t2 t3;
by ptno;
run;


data new0415; set lib.test3;
run;

proc sql;
create table new0415_1 as select    /*merge two data set */
a.*, b.* from new0415 as a left join var_sys as b on
a.ptno=b.ptno ;
quit;

proc export data=new0415_1
  outfile='C:\Users\Yongjia\Desktop\new0415_1.csv';
  run;

/******************joint mse_sys30, mse_sys40, mse_sys50 together **********************/


libname lib "C:\Users\Yongjia\Desktop";
/********** MSE at age 30******************/

data M1 ; set lib.new;
drop age_event age_30 ind;
if  age <=35 and age >= 25;
run;


PROC reg data=M1 OUTEST=est30  noprint;
model ANbpsys=age;
by ptno;
run;

proc print data=est30;
run;

data est30; set est30;
keep ptno _RMSE_;
RUN;

data mse30; set est30;
mse_30=_RMSE_**2;
if mse_30=. then mse_30=0;
RUN;

/********** MSE at age 40******************/

data M2 ; set lib.new;
drop age_event age_30 ind;
if  age <=45 and age > 35;
run;
```

```
PROC reg data=M2 OUTEST=est40  noprint;
model ANbpsys=age;
by ptno;
run;

proc print data=est40;
run;

data est40; set est40;
keep ptno _RMSE_;
RUN;

data mse40; set est40;
mse_40=_RMSE_**2;
if mse_40=. then mse_40=0;
RUN;

/********** MSE at age 50*****************/

data M3 ; set lib.new;
drop age_event age_30 ind;
if  age <=55 and age > 45;
run;


PROC reg data=M3 OUTEST=est50  noprint;
model ANbpsys=age;
by ptno;
run;

proc print data=est50;
run;

data est50; set est50;
keep ptno _RMSE_;
RUN;

data mse50; set est50;
mse_50=_RMSE_**2;
if mse_50=. then mse_50=0;
RUN;

/************ joint mse30 mse40 mse 50 together and merge into 1791 data set*********/
data mse;
merge mse30 mse40 mse50;
by ptno;
drop _RMSE_;
run;

data new0415_1; set lib.new0415_1;
run;

proc sql;
create table new0415_2 as select    /*merge two data set */
a.*, b.* from new0415_1 as a left join mse as b on
a.ptno=b.ptno ;
```

```
quit;

data new0415_3; set new0415_2;
label var_sys30 = "var_sys30"
      var_sys40 = "var_sys40"
      var_sys50 = "var_sys50";

 run;

proc export data=new0415_3
   outfile='C:\Users\Yongjia\Desktop\new0415_3.csv';
   run;


/******* consider HBP event is SBP >140 or DBP >90, also from questionnaire*********/
libname lib "C:\Users\Yongjia\Desktop";

data t1_0428; set new0415_3;         /******this data set include all events from questionnaires ***/
keep ptno new_censor age_hp ind_hp;
run;

data t2_0428; set lib.final_update;  /*******this data set include all events from SBP >140 OR DBP>90 ***/
keep ptno age_event ind;
run;

data t3_0428;
merge t1_0428 t2_0428;
by ptno;
run;

data t4_0428; set t3_0428;
if age_hp =. then age_hp=0;
if age_event =. then age_event=0;
run;

data t5_0428; set t4_0428;   /*****comparing two events age  and find ealiest age *******/
if age_hp =0 and age_event^=0 then new_age_hp= age_event;
if age_event=0 and age_hp^=0 then new_age_hp= age_hp;
if  age_hp < age_event  and age_hp^=0 then new_age_hp= age_hp;
if  age_event < age_hp  and age_event^=0 then new_age_hp= age_event;
if age_hp = 0 and age_event=0 then new_age_hp =age_event;
run;

data t6_0428; set t5_0428;        /*********define new indicator ********/
if new_age_hp=0 then new_hp_ind=0 ;
  else new_hp_ind=1;
 run;

data t7_0428; set t6_0428;
keep new_age_hp new_hp_ind ptno;
run;


data t8_0428;
merge t7_0428 new0415_3;
by ptno;
```

```
if new_age_hp=0 then new_age_hp=.;
run;

proc export data=t8_0428
  outfile='C:\Users\Yongjia\Desktop\t8_0428.csv';
  run;
```

R CODE FOR ANALYSIS OF DATA

```
################  For Blood Pressure####################
For SBP vs. DBP
########  @Age30, p value=0.1195808############

rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
library(compareC)
compareC(fels_var30,fels_var$new_hp_ind,
       fels_var$sys_age30,fels_var$dia_age30)

######## at age 40 : pval=0.1930183
fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
compareC(fels_var40,fels_var$new_hp_ind,
       fels_var$sys_age40,fels_var$dia_age40)

######## at age 50 : pval= 0.9785448
fels_var<-fels[(fels$age_40<=55)&(fels$age_40>45),]
fels_var50<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
compareC(fels_var50,fels_var$new_hp_ind,
       fels_var$sys_age50,fels_var$dia_age50)

############  For high chol #################
SBP vs. DBP

######## at age 30 :
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/t8_0428.csv")
```

```
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var30,fels_var$ind_hc,
       fels_var$sys_age30,fels_var$dia_age30)


####### at age 40 :

fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var40,fels_var$ind_hc,
       fels_var$sys_age40,fels_var$dia_age40)

####### at age 50 :

fels_var<-fels[(fels$age_50<=55)&(fels$age_50>45),]
fels_var50<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var50,fels_var$ind_hc,
       fels_var$sys_age50,fels_var$dia_age50)

############## Test for heart attack ################
##### at age 30,
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
analysis_30 = fels[(fels$age_30<35)&(fels$age_30>=25),]
head(analysis_30)
analysis_30$event.time<- ifelse(analysis_30$ind_hta==0,
                          analysis_30$new_censor,
                          analysis_30$age_heart_attack)
compareC(analysis_30$event.time,analysis_30$ind_hta,
       analysis_30$sys_age30,analysis_30$dia_age30)

##### At age 40,
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
analysis_40 = fels[(fels$age_40<45)&(fels$age_40>=35),]
head(analysis_40)
```

```
analysis_40$event.time<- ifelse(analysis_40$ind_hta==0,
                        analysis_40$new_censor,
                        analysis_40$age_heart_attack)
compareC(analysis_40$event.time,analysis_40$ind_hta,
       analysis_40$sys_age40,analysis_40$dia_age40)

##### at age 50,
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
analysis_50 = fels[(fels$age_50<55)&(fels$age_50>=45),]
head(analysis_50)
analysis_50$event.time<- ifelse(analysis_50$ind_hta==0,
                        analysis_50$new_censor,
                        analysis_50$age_heart_attack)
compareC(analysis_50$event.time,analysis_50$ind_hta,
       analysis_50$sys_age50,analysis_50$dia_age50)

############## Test for heart failure ################
##### at age 30,
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
analysis_30 = fels[(fels$age_30<35)&(fels$age_30>=25),]
head(analysis_30)
analysis_30$event.time<- ifelse(analysis_30$ind_htf==0,
                        analysis_30$new_censor,
                        analysis_30$age_heart_failure)
compareC(analysis_30$event.time,analysis_30$ind_htf,
       analysis_30$sys_age30,analysis_30$dia_age30)

##### At age 40,
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
analysis_40 = fels[(fels$age_40<45)&(fels$age_40>=35),]
head(analysis_40)
analysis_40$event.time<- ifelse(analysis_40$ind_htf==0,
                        analysis_40$new_censor,
                        analysis_40$age_heart_failure)
compareC(analysis_40$event.time,analysis_40$ind_htf,
       analysis_40$sys_age40,analysis_40$dia_age40)

##### at age 50,
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
```

```
head(fels)
analysis_50 = fels[(fels$age_50<55)&(fels$age_50>=45),]
head(analysis_50)
analysis_50$event.time<- ifelse(analysis_50$ind_htf==0,
                        analysis_50$new_censor,
                        analysis_50$age_heart_failure)
compareC(analysis_50$event.time,analysis_50$ind_htf,
       analysis_50$sys_age50,analysis_50$dia_age50)


################  For Blood Pressure####################
##For Variance SBP vs. SBP
########### at age 30, ####################

rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
library(compareC)
compareC(fels_var30,fels_var$ind_hp,
       fels_var$var_sys30,fels_var$sys_age30)

####### at age 40 :
fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
compareC(fels_var40,fels_var$ind_hp,
       fels_var$var_sys40,fels_var$sys_age40)

####### at age 50 :
fels_var<-fels[(fels$age_40<=55)&(fels$age_40>45),]
fels_var50<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
compareC(fels_var50,fels_var$ind_hp,
       fels_var$var_sys50,fels_var$sys_age50)

################  For High chol####################
##For Variance SBP vs. SBP
########### at age 30, ####################
####### at age 30 :
rm(list=ls())
```

```
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var30,fels_var$ind_hc,
       fels_var$var_sys,fels_var$sys_age30)


####### at age 40 :

fels_var<-fels[(fels$age_40<45)&(fels$age_40>=35),]
fels_var40<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var40,fels_var$ind_hc,
       fels_var$var_sys,fels_var$sys_age40)

####### at age 50 :

fels_var<-fels[(fels$age_50<55)&(fels$age_50>=45),]
fels_var50<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var50,fels_var$ind_hc,
       fels_var$var_sys,fels_var$sys_age50)

###########  Heart attack : var~ sys ################
#### at age 30:
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$ind_hta==0,
                fels_var$new_censor,
                fels_var$age_heart_attack)
compareC(fels_var30,fels_var$ind_hta,
       fels_var$var_sys,fels_var$sys_age30)

##### at age 40:
fels_var<-fels[(fels$age_40<45)&(fels$age_40>=35),]
fels_var40<- ifelse(fels_var$ind_hta==0,
                fels_var$new_censor,
                fels_var$age_heart_attack)
compareC(fels_var40,fels_var$ind_hta,
```

```
        fels_var$var_sys,fels_var$sys_age40)

##### at age 50:
fels_var<-fels[(fels$age_50<55)&(fels$age_50>=45),]
fels_var50<- ifelse(fels_var$ind_hta==0,
                fels_var$new_censor,
                fels_var$age_heart_attack)
compareC(fels_var50,fels_var$ind_hta,
        fels_var$var_sys,fels_var$sys_age50)

############  Heart failure : var~ sys ################
#### at age 30:
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$ind_htf==0,
                fels_var$new_censor,
                fels_var$age_heart_failure)
compareC(fels_var30,fels_var$ind_htf,
        fels_var$var_sys,fels_var$sys_age30)

#### at age 40:
fels_var<-fels[(fels$age_40<45)&(fels$age_40>=35),]
fels_var40<- ifelse(fels_var$ind_htf==0,
                fels_var$new_censor,
                fels_var$age_heart_failure)
compareC(fels_var40,fels_var$ind_htf,
        fels_var$var_sys,fels_var$sys_age40)

#### at age 50:
fels_var<-fels[(fels$age_50<55)&(fels$age_50>=45),]
fels_var50<- ifelse(fels_var$ind_htf==0,
                fels_var$new_censor,
                fels_var$age_heart_failure)
compareC(fels_var50,fels_var$ind_htf,
        fels_var$var_sys,fels_var$sys_age50)

################  For Blood Pressure####################
For MSE of SBP vs. SBP

############## Age 30
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
```

```
fels_var30<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
library(compareC)
compareC(fels_var30,fels_var$new_hp_ind,
       fels_var$mse_30,fels_var$sys_age30)

####### at age 40 :
fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
compareC(fels_var40,fels_var$new_hp_ind,
       fels_var$mse_40,fels_var$sys_age40)

####### at age 50 :
fels_var<-fels[(fels$age_40<=55)&(fels$age_40>45),]
fels_var50<- ifelse(fels_var$new_hp_ind==0,
                fels_var$new_censor,
                fels_var$new_age_hp)
compareC(fels_var50,fels_var$new_hp_ind,
       fels_var$mse_50,fels_var$sys_age50)

#################  For High Chol ####################
For MSE of SBP vs. SBP

############### Age 30

rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var30,fels_var$ind_hc,
       fels_var$mse_30,fels_var$sys_age30)

####### at age 40 :

fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var40,fels_var$ind_hc,
       fels_var$mse_40,fels_var$sys_age40)
```

```
####### at age 50 :
fels_var<-fels[(fels$age_50<=55)&(fels$age_50>45),]
fels_var50<- ifelse(fels_var$ind_hc==0,
                fels_var$new_censor,
                fels_var$age_hc)
compareC(fels_var50,fels_var$ind_hc,
       fels_var$mse_50,fels_var$sys_age50)

###########  Heart attack : var~ sys ################

#### at age 30:
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
fels_var30<- ifelse(fels_var$ind_hta==0,
                fels_var$new_censor,
                fels_var$age_heart_attack)
compareC(fels_var30,fels_var$ind_hta,
       fels_var$mse_30,fels_var$sys_age30)


##### at age 40:
fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$ind_hta==0,
                fels_var$new_censor,
                fels_var$age_heart_attack)
compareC(fels_var40,fels_var$ind_hta,
       fels_var$mse_40,fels_var$sys_age40)


##### at age 50:
fels_var<-fels[(fels$age_50<55)&(fels$age_50>45),]
fels_var50<- ifelse(fels_var$ind_hta==0,
                fels_var$new_censor,
                fels_var$age_heart_attack)
compareC(fels_var50,fels_var$ind_hta,
       fels_var$mse_50,fels_var$sys_age50)

###########  Heart failure : var~ sys ################
#### at age 30:
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<-fels[(fels$age_30<=35)&(fels$age_30>=25),]
```

```
fels_var30<- ifelse(fels_var$ind_htf==0,
                fels_var$new_censor,
                fels_var$age_heart_failure)
compareC(fels_var30,fels_var$ind_htf,
       fels_var$mse_30,fels_var$sys_age30)

#### at age 40:

fels_var<-fels[(fels$age_40<=45)&(fels$age_40>35),]
fels_var40<- ifelse(fels_var$ind_htf==0,
                fels_var$new_censor,
                fels_var$age_heart_failure)
compareC(fels_var40,fels_var$ind_htf,
       fels_var$mse_40,fels_var$sys_age40)

#### at age 50:

fels_var<-fels[(fels$age_50<=55)&(fels$age_50>45),]
fels_var50<- ifelse(fels_var$ind_htf==0,
                fels_var$new_censor,
                fels_var$age_heart_failure)
compareC(fels_var50,fels_var$ind_htf,
       fels_var$mse_50,fels_var$sys_age50)


######### Compairing Var_sys  and Var_dia###################
##########compare C for blood pressure##############

rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<- ifelse(fels$ind_hp==0,
fels$new_censor,
fels$age_hp)
library(compareC)
compareC(fels_var,fels$ind_hp,
fels$var_sys,fels$var_dia)

##########compare C for high chol##############
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<- ifelse(fels$ind_hc==0,
               fels$new_censor,
               fels$age_hc)
compareC(fels_var,fels$ind_hc,
```

```
        fels$var_sys,fels$var_dia)

##########compare C for heart attack ###############
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<- ifelse(fels$ind_hta==0,
                  fels$new_censor,
                  fels$age_heart_attack)
compareC(fels_var,fels$ind_hta,
        fels$var_sys,fels$var_dia)

##########compare C for heart failure ###############
rm(list=ls())
fels<-read.csv("C:/Users/Yongjia/Desktop/ t8_0428.csv")
head(fels)
fels_var<- ifelse(fels$ind_htf==0,
                  fels$new_censor,
                  fels$age_heart_failure)
compareC(fels_var,fels$ind_htf,
        fels$var_sys,fels$var_dia)
```