ACM 661 Harsha Kankanamge PROJECT 3

The data set WHAS500.SAV represents a sample of 500 subjects admitted to hospitals with acute myocardial infarction (MI) in Worcester, Massachusetts. The goal of the study is to describe factors associated with trends over time in the survival following hospital admission for acute myocardial infarction (MI).

Use length of follow-up (LENFOL) as the survival time variable, status at last follow-up (FSTAT) as the censoring variable, and 5 percent as the level of significance.
In the data set WHAS500.SAV, consider the variables AGE, 'HR' (Initial heart rate in beats per minute-V4), SYSBP (Initial systolic blood pressure in mm/Hg-V5), DIASBP (Initial Diastolic blood pressure in mm/Hg-V6), BMI (Body mass index in kg/meter squared), GENDER, CVD (History of cardiovascular disease-V8, 0 = No, 1 = Yes), AFB (Atrial Fibrillation-V9, 0 = No, 1 = Yes), CHF (Congestive heart complications-V11, 0 = No, 1 = Yes), MIORD (MI Order-V13, 0 = First, 1 = Recurrent)), and MITYPE (MI Type-V14, 0 = Non Q-wave, 1 = Q-wave).

1)Conduct univariate analysis of variables in analysis.

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	300	60.0	60.0	60.0
	female	200	40.0	40.0	100.0
	Total	500	100.0	100.0	

In the sample of 500 subjects, 60% of them are males and 40% of them are females. So in the sample male percentage is higher than female percentage.

history of cardiovascular disease

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	125	25.0	25.0	25.0
	Yes	375	75.0	75.0	100.0
	Total	500	100.0	100.0	

In the sample of 500 subjects, 75% of them have history of cardiovascular disease and 25% of them do not have history of cardiovascular disease. So in the sample most of the people have history of cardiovascular disease.

atrial fibrillation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	422	84.4	84.4	84.4
	1	78	15.6	15.6	100.0
	Total	500	100.0	100.0	

In the sample of 500 subjects, 15.6% of them have atrial fibrillation and 84.4% of them do not have atrial fibrillation. So in the sample most of the people do not have atrial fibrillation.

congestive heart complications

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	345	69.0	69.0	69.0
	Yes	155	31.0	31.0	100.0
	Total	500	100.0	100.0	

In the sample of 500 subjects, 31% of them have congestive heart complications and 69% of them do not have congestive heart complications. So in the sample most of the people do not have congestive heart complications.

MI order

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	First	329	65.8	65.8	65.8
	Recurrent	171	34.2	34.2	100.0
	Total	500	100.0	100.0	

In the sample of 500 subjects, 65.8% of them have MI order first and 34.2% of them have MI order recurrent. So in the sample most of people have MI order first.

MITYPE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Non-Q-wave	347	69.4	69.4	69.4
	Q-wave	153	30.6	30.6	100.0
	Total	500	100.0	100.0	

In the sample of 500 subjects, 69.4% of them have MITYPE non-q-wave and 30.6% of them have MITYPE q-wave. So in the sample most of people have MITYPE non-q-wave.

vital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Alive	285	57.0	57.0	57.0
	Dead	215	43.0	43.0	100.0
	Total	500	100.0	100.0	

Duration of the study 57% of people were alive and 43% were died. Higher percentage of people were alive at the end of the study.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
age at hospital admission	500	30	104	69.85	14.491
initial heart rate in beats per minute	500	35	186	87.02	23.586
initial systolic blood pressure in mm/Hg	500	57	244	144.70	32.295
initrial diastolic blood pressure in mm/Hg	500	6	198	78.27	21.545
body mass index	500	13.05	44.84	26.6138	5.40566
total length of follow-up	500	1	2358	882.44	705.665
Valid N (listwise)	500				

Mean age at hospital admission is 69.85 and std. Deviation is 14.491.

Mean initial heart rate in beats per minute is 87.02 and std. Deviation is 23.586.

Mean initial systolic blood pressure in mm/Hg is 144.70 and std. Deviation is 32.295.

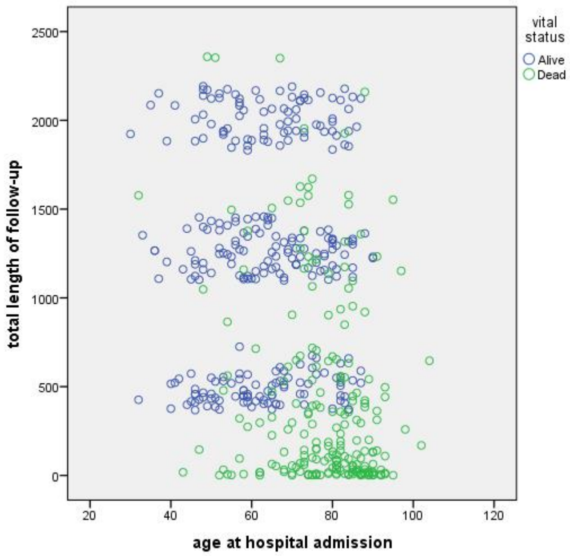
Mean initial diastolic blood pressure in mm/Hg is 78.27 and std. Deviation is 21.545.

Mean total length of the follow-up time in days is 882.44 and std. Deviation is 705.665.

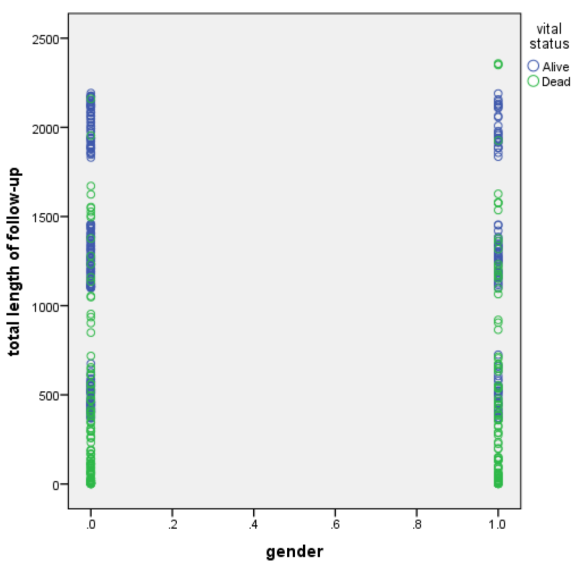
Mean body mass index is 26.6138 and std. Deviation is 5.40566.

2.Initial exploratory analysis: Plot lenfol versus each of the selected covariates, setting markers by censoring status to see if there are any visual differences between censoring status and the covariate. Summarize your findings.

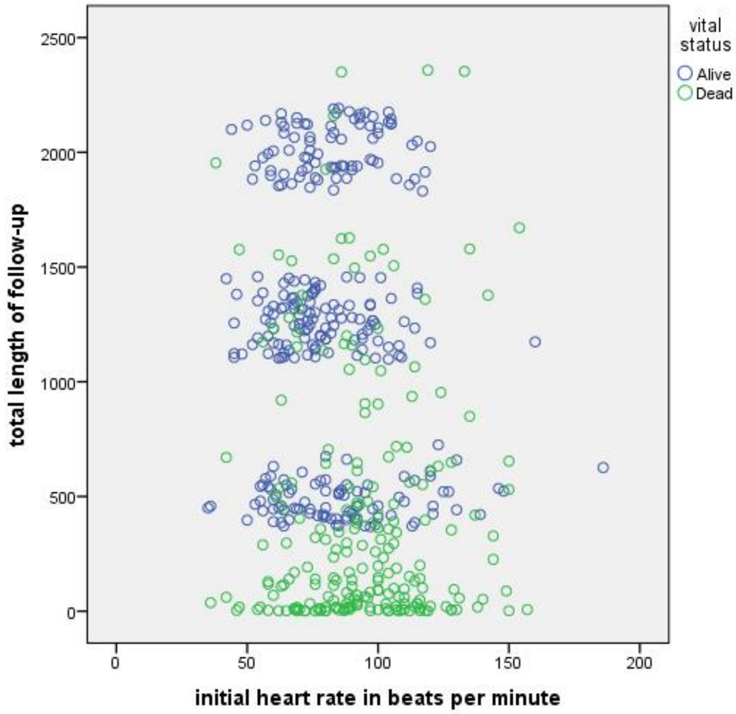
lenfol versus age



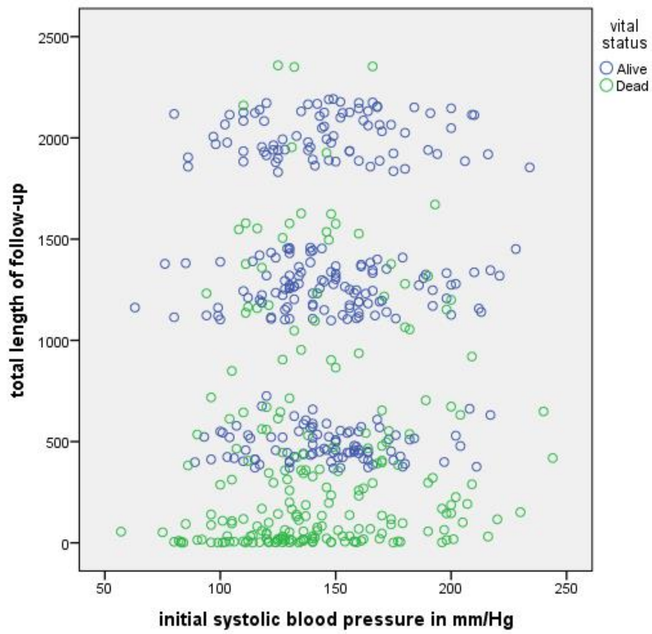
Most of elderly people died at the beginning of the study. Young people survived than elderly people. Lot of people were withdrawal from the study around 500 days, between 1200 to 1400 days and around 2000 days.



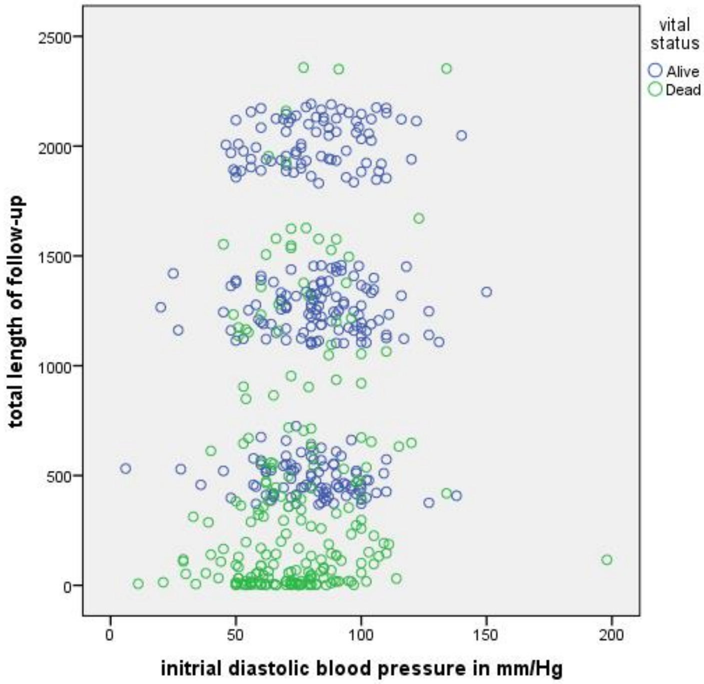
Most of the peoples had been died during first 600 days of the study in both gender groups. Males subjects were survived more days than female subjects. Most of male subjects and female subject were withdrawal from the study around 500 days, between 1200 to 1400 days and around 2000 days. At the end female subject was died. Closer to end of the study period a lots of male subjects were removed from the study than the female subjects.



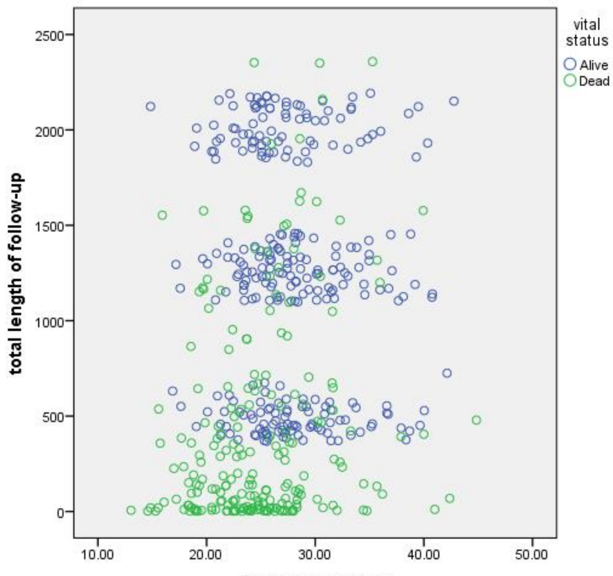
Lot of people have been died at the beginning of the study. Most of the people who have been died had higher ‘initial heart in beats per minute’ compare with people who are alive. Most subjects were withdrawal from the study around 500 days, between 1200 to 1400 days and around 2000 days.



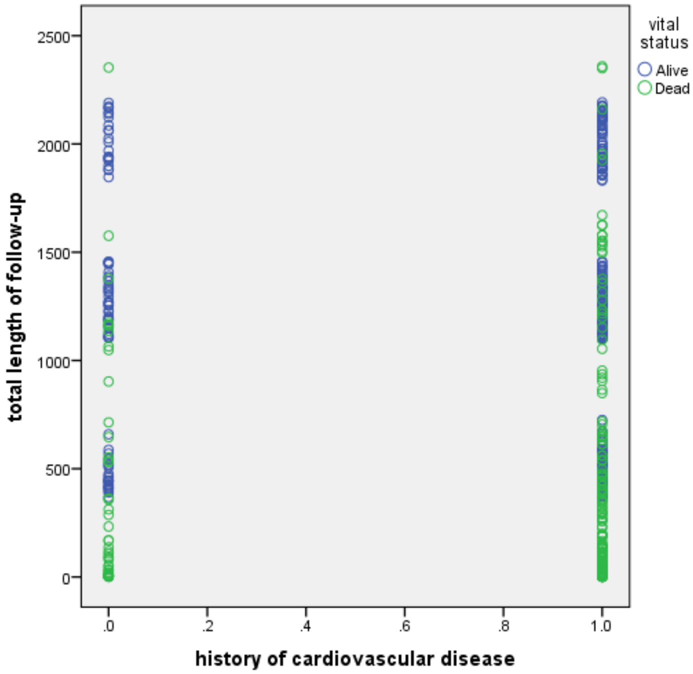
Lot of people have been died at the beginning of the study. Most of the people have higher ‘’initial systolic blood pressure in mm/Hg’’ either alive or dead. Most subjects were withdrawal from the study around 500 days, between 1200 to 1400 days and around 2000 days.



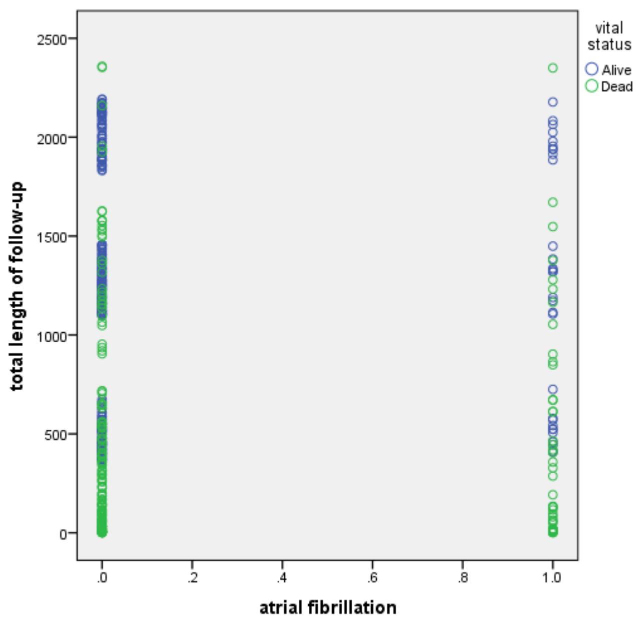
Lot of people have been died at the beginning of the study. On average ''initial diastolic blood pressure in mm/Hg'' is very closer for both groups (alive or dead). Most subjects were withdrawal from the study around 500 days, between 1200 to 1400 days and around 2000 days.



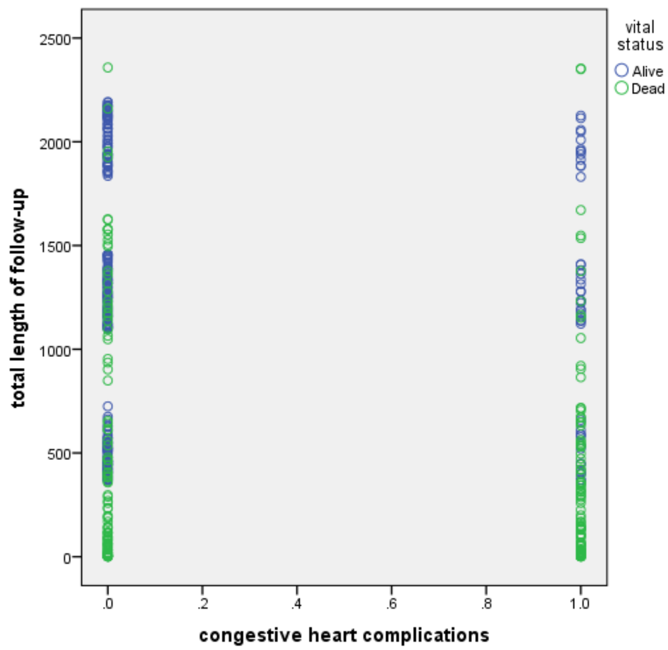
Most of the subject ware died at the beginning of the study. As well as most of people who died had less BMI value. So people how have less BMI have high risk of having the event. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.



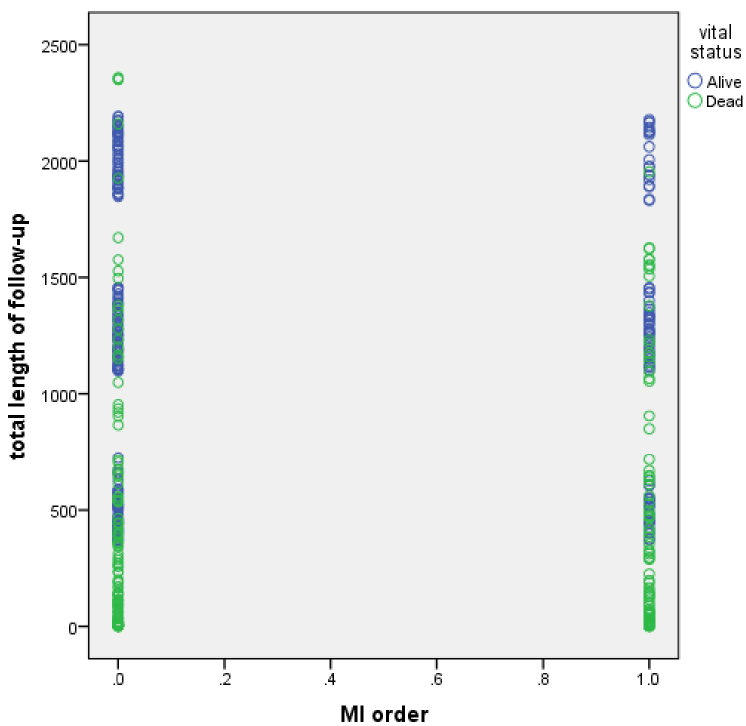
Most of the people who have been died had history of cardiovascular disease. Also lot of people died during first 500 days of the study had the disease. As well as lot of people have 'history of cardiovascular disease'. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.



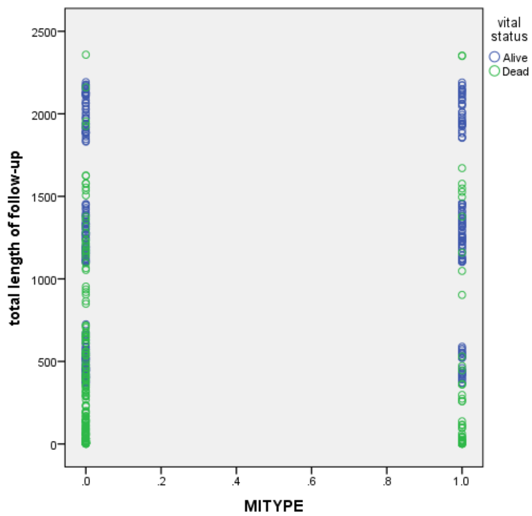
Most of the people do not have 'atrial fibrillation'. As well as most of the people who have been died have not had 'atrial fibrillation'. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.



Most of the people do not have ‘congestive heart complications’. As well as most of the people who have been died have had ‘congestive heart complications’ on proportion. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.



Most of the people have ‘MI order first’. As well as dead proportion seems to be same for both groups. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.



Most of the people have ‘MITYPE non-q-wave’. Higher number of people have been died in ‘MITYPE non-q-wave’ group. Which happened within first 600 days of study. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.

3. Conduct bivariate analysis of each covariate with outcome.

gender * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
gender	male	Count	189	111	300
		% within gender	63.0%	37.0%	100.0%
	female	Count	96	104	200
		% within gender	48.0%	52.0%	100.0%
Total		Count	285	215	500
		% within gender	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11.016[a]	1	.001		
Continuity Correction[b]	10.412	1	.001		
Likelihood Ratio	11.003	1	.001		
Fisher's Exact Test				.001	.001
Linear-by-Linear Association	10.994	1	.001		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 86.00.

b. Computed only for a 2x2 table

37.0% of males have died during the study in the male group. As well as 52% females have died during the study in the female group. So higher percentage of females have been died compare with males during the study. So this deference is significant. Which is conform by Chi-square test. Chi-square statistic is 11.016 and p-value = 0.001 < 0.05. So there is a significant association between gender and vital status.

history of cardiovascular disease * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
history of cardiovascular disease	No	Count	80	45	125
		% within history of cardiovascular disease	64.0%	36.0%	100.0%
	Yes	Count	205	170	375
		% within history of cardiovascular disease	54.7%	45.3%	100.0%
Total		Count	285	215	500
		% within history of cardiovascular disease	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.332[a]	1	.068		
Continuity Correction[b]	2.962	1	.085		
Likelihood Ratio	3.371	1	.066		
Fisher's Exact Test				.076	.042
Linear-by-Linear Association	3.325	1	.068		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 53.75.

b. Computed only for a 2x2 table

45.3% of people have died during the study in the group of people who have had ‘history of cardiovascular disease’. As well as 36% of people who haven’t had ‘history of cardiovascular disease’ have died during the study. So higher percentage of people died in the group of people who had ‘history of cardiovascular disease’ during the study. But this deference is not significant (border line significant). Which is conform by Chi-square test. Chi-square statistic is 3.332 and p-value = 0.068 >0.05. So there is a border line significant association between ‘history of cardiovascular disease’ and vital status.

atrial fibrillation * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
atrial fibrillation	0	Count	254	168	422
		% within atrial fibrillation	60.2%	39.8%	100.0%
	1	Count	31	47	78
		% within atrial fibrillation	39.7%	60.3%	100.0%
Total		Count	285	215	500
		% within atrial fibrillation	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11.228[a]	1	.001		
Continuity Correction[b]	10.409	1	.001		
Likelihood Ratio	11.123	1	.001		
Fisher's Exact Test				.001	.001
Linear-by-Linear Association	11.206	1	.001		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 33.54.

b. Computed only for a 2x2 table

60.3% of people have died during the study in the group of people who have had ‘Atrial Fibrillation’. As well as 39.8% of people who haven’t had ‘Atrial Fibrillation’ have died during the study. So higher percentage of people died in the group of people who had ‘Atrial Fibrillation’ during the study. This deference is significant. Which is conform by Chi-square test. Chi-square statistic is 11.228 and p-value = 0.001 <0.05. So there is a significant association between ‘Atrial Fibrillation’ and vital status.

MI order * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
MI order	First	Count	204	125	329
		% within MI order	62.0%	38.0%	100.0%
	Recurrent	Count	81	90	171
		% within MI order	47.4%	52.6%	100.0%
Total		Count	285	215	500
		% within MI order	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.836[a]	1	.002		
Continuity Correction[b]	9.248	1	.002		
Likelihood Ratio	9.798	1	.002		
Fisher's Exact Test				.002	.001
Linear-by-Linear Association	9.816	1	.002		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 73.53.

b. Computed only for a 2x2 table

38.0% of people have died during the study in the group of people who have had 'MI order first'. As well as 52.6% of people have died during the study in the group of people who have had 'MI order recurrent'. So higher percentage of people died in the group of people who had 'MI order recurrent' during the study. This deference is significant. Which is conform by Chi-square test. Chi-square statistic is 9.836 and p-value = 0.002 <0.05. So there is a significant association between 'MI order' and vital status.

congestive heart complications * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
congestive heart complications	No	Count	240	105	345
		% within congestive heart complications	69.6%	30.4%	100.0%
	Yes	Count	45	110	155
		% within congestive heart complications	29.0%	71.0%	100.0%
Total		Count	285	215	500
		% within congestive heart complications	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	71.689[a]	1	.000		
Continuity Correction[b]	70.045	1	.000		
Likelihood Ratio	72.551	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	71.546	1	.000		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 66.65.

b. Computed only for a 2x2 table

71% of people have died during the study in the group of people who have had ‘Congestive heart complications’. As well as 30.4% of people who haven’t had ‘Congestive heart complications’ have died during the study. So higher percentage of people died in the group of people who had ‘Congestive heart complications’ during the study. This deference is significant. Which is conform by Chi-square test. Chi-square statistic is 71.689 and p-value = 0.000 <0.05. So there is a significant association between ‘Congestive heart complications’ and vital status.

MITYPE * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
MITYPE	Non-Q-wave	Count	178	169	347
		% within MITYPE	51.3%	48.7%	100.0%
	Q-wave	Count	107	46	153
		% within MITYPE	69.9%	30.1%	100.0%
Total		Count	285	215	500
		% within MITYPE	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	15.049[a]	1	.000		
Continuity Correction[b]	14.298	1	.000		
Likelihood Ratio	15.411	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	15.019	1	.000		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 65.79.

b. Computed only for a 2x2 table

48.7% of people have died during the study in the group of people who have had 'MITYPE Non-Q-wave'. As well as 30.1% of people have died during the study in the group of people who have had 'MITYPE Q-wave'. So higher percentage of people died in the group of people who had 'MITYPE Non-Q-wave' during the study. This deference is significant. Which is conform by Chi-square test. Chi-square statistic is 15.049 and p-value = 0.000 <0.05. So there is a significant association between 'MITYPE' and vital status.

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
body mass index	Alive	285	27.9153	5.08444	.30118
	Dead	215	24.8885	5.34633	.36462

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
body mass index	Equal variances assumed	.109	.741	6.445	498	.000	3.02674	.46960	2.10410	3.94939
	Equal variances not assumed			6.400	448.364	.000	3.02674	.47292	2.09733	3.95616

Mean 'body mass index' for people who have died is 24.8885 and mean 'body mass index' for people who are alive is 27.9153. People who have died has less mean body mass index compare with mean body mass index of alive people. Different is significant. Which is confirm by independent sample t test. T-statistic 6.445 and p-value = 0.000 <0.05. which means mean 'body mass index' is significantly different in the two groups (alive and dead). So there is a significant association between "body mass index" and "vital status"

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
age at hospital admission	Alive	285	63.73	13.445	.796
	Dead	215	77.95	11.579	.790

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
age at hospital admission	Equal variances assumed	11.191	.001	-12.414	498	.000	-14.216	1.145	-16.465	-11.966
	Equal variances not assumed			-12.675	489.293	.000	-14.216	1.122	-16.419	-12.012

Mean ‘age’ for people who have died is 77.95 and mean ‘age’ for people who are alive is 63.73. People who have died have higher mean age compare with mean age of alive people. Difference is significant. Which is confirm by independent sample t test. T-statistic -12.675 and p-value = 0.000 <0.05. which means mean ‘age’ is significantly different in the two groups (alive and dead). So there is a significant association between “age” and “vital status”

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
initial heart rate in beats per minute	Alive	285	81.71	21.757	1.289
	Dead	215	94.05	24.120	1.645

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
initial heart rate in beats per minute	Equal variances assumed	2.409	.121	-5.990	498	.000	-12.339	2.060	-16.386	-8.292
	Equal variances not assumed			-5.905	434.107	.000	-12.339	2.090	-16.446	-8.232

Mean ‘initial heart rate in beats per minute’ for people who have died is 94.05 and mean ‘initial heart rate in beats per minute’ for people who are alive is 81.71. People who have died have higher mean ‘initial heart rate in beats per minute’ compare with mean ‘initial heart rate in beats per minute’ of alive people. Difference is significant. Which is confirm by independent sample t test. T-statistic -5.990 and p-value = 0.000 <0.05. which means mean ‘initial heart rate in beats per minute’ is significantly different in the two groups (alive and dead). So there is a significant association between “initial heart rate in beats per minute” and “vital status”

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
initial systolic blood pressure in mm/Hg	Alive	285	146.80	30.585	1.812
	Dead	215	141.93	34.307	2.340

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
initial systolic blood pressure in mm/Hg	Equal variances assumed	3.925	.048	1.671	498	.095	4.866	2.912	-.855	10.588
	Equal variances not assumed			1.644	430.846	.101	4.866	2.959	-.950	10.682

Mean 'Initial Diastolic blood pressure in mm/Hg' for people who have died is 141.93 and mean 'Initial Diastolic blood pressure in mm/Hg' for people who are alive is 146.80. People who have died have less mean 'Initial Diastolic blood pressure in mm/Hg' compare with mean 'Initial Diastolic blood pressure in mm/Hg' of alive people. Difference is not significant. Which is confirm by independent sample t test. T-statistic 1.644 and p-value = 0.101 <0.05. which means mean 'Initial Diastolic blood pressure in mm/Hg' is not significantly different in the two groups (alive and dead). So there is no significant association between 'Initial Diastolic blood pressure in mm/Hg' and "vital status"

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
initrial diastolic blood pressure in mm/Hg	Alive	285	81.79	20.472	1.213
	Dead	215	73.60	22.088	1.506

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
initrial diastolic blood pressure in mm/Hg	Equal variances assumed	.634	.426	4.278	498	.000	8.186	1.913	4.427	11.945
	Equal variances not assumed			4.233	441.518	.000	8.186	1.934	4.385	11.987

Mean 'initial diastolic blood pressure in mm/Hg' for people who have died is 73.60 and mean 'initial diastolic blood pressure in mm/Hg' for people who are alive is 81.79. People who have died have lower mean 'initial diastolic blood pressure in mm/Hg' compare with mean 'initial diastolic blood pressure in mm/Hg' of alive people. Difference is significant. Which is confirm by independent sample t test. T-statistic 4.278 and p-value = 0.000 <0.05. which means mean 'initial diastolic blood pressure in mm/Hg' is significantly different in the two groups (alive and dead). So there is a significant association between 'initial diastolic blood pressure in mm/Hg' and "vital status"

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
total length of follow-up	Alive	285	1227.79	607.687	35.996
	Dead	215	424.64	549.235	37.458

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
total length of follow-up	Equal variances assumed	6.492	.011	15.243	498	.000	803.156	52.690	699.634	906.677
	Equal variances not assumed			15.460	482.009	.000	803.156	51.950	701.079	905.232

Mean ‘total length of follow-up’ in day for people who have died is 424.64 and mean ‘total length of follow-up’ in days for people who are alive is 1227.79. People who have died have lower mean ‘total length of follow-up’ in days compare with mean ‘total length of follow-up’ of alive people. Difference is significant. Which is confirm by independent sample t test. T-statistic 15.460 and p-value = 0.000 <0.05. which means mean ‘total length of follow-up’ is significantly different in the two groups (alive and dead). So there is a significant association between ‘total length of follow-up’ and “vital status”

4. Find a best model: conduct multivariate analysis (Forward LR) using the Cox Proportional Hazards model.

Categorical Variable Codings[a,c,d,e,f,g]

		Frequency	(1)
gender[b]	0=male	300	0
	1=female	200	1
history of cardiovascular disease[b]	0=No	125	0
	1=Yes	375	1
atrial fibrillation[b]	0	422	0
	1	78	1
congestive heart complications[b]	0=No	345	0
	1=Yes	155	1
MI order[b]	0=First	329	0
	1=Recurrent	171	1
MITYPE[b]	0=Non-Q-wave	347	0
	1=Q-wave	153	1

a. Category variable: gender (gender)

b. Indicator Parameter Coding

c. Category variable: history of cardiovascular disease (CVD)

d. Category variable: atrial fibrillation (AFB)

e. Category variable: congestive heart complications (CHF)

f. Category variable: MI order (MIORD)

g. Category variable: MITYPE

Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	age at hospital admission	.066	.006	118.799	1	.000	1.068
Step 2	age at hospital admission	.059	.006	92.407	1	.000	1.060
	congestive heart complications	.861	.142	36.570	1	.000	2.365
Step 3	age at hospital admission	.059	.006	91.645	1	.000	1.061
	initial heart rate in beats per minute	.009	.003	9.119	1	.003	1.009
	congestive heart complications	.757	.146	26.899	1	.000	2.131
Step 4	age at hospital admission	.054	.006	72.053	1	.000	1.055
	initial heart rate in beats per minute	.011	.003	13.411	1	.000	1.011
	initrial diastolic blood pressure in mm/Hg	-.011	.004	9.358	1	.002	.989
	congestive heart complications	.741	.146	25.653	1	.000	2.097
Step 5	age at hospital admission	.048	.007	53.655	1	.000	1.049
	initial heart rate in beats per minute	.011	.003	13.305	1	.000	1.011
	initrial diastolic blood pressure in mm/Hg	-.010	.004	8.237	1	.004	.990
	body mass index	-.043	.016	7.172	1	.007	.958
	congestive heart complications	.745	.146	26.113	1	.000	2.107

$$h(t)\hat{}=h_0(t)\exp(0.048*age+0.011*HR-0.010*DIASBP-0.043*BMI+0.745*CHF)$$

Omnibus Tests of Model Coefficients[f]

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	2313.358	126.260	1	.000	141.800	1	.000	141.800	1	.000
2[b]	2277.467	173.939	2	.000	35.891	1	.000	177.692	2	.000
3[c]	2268.602	181.548	3	.000	8.865	1	.003	186.556	3	.000
4[d]	2258.900	196.228	4	.000	9.701	1	.002	196.258	4	.000
5[e]	2251.536	202.675	5	.000	7.364	1	.007	203.622	5	.000

a. Variable(s) Entered at Step Number 1: age at hospital admission

b. Variable(s) Entered at Step Number 2: congestive heart complications

c. Variable(s) Entered at Step Number 3: initial heart rate in beats per minute

d. Variable(s) Entered at Step Number 4: initrial diastolic blood pressure in mm/Hg

e. Variable(s) Entered at Step Number 5: body mass index

f. Beginning Block Number 1. Method = Forward Stepwise (Likelihood Ratio)

5. Check interactions between covariates selected in the stepwise model in step 4 and make a table. Ignore significant interactions between the covariates and instead, Add GENDER and GENDER*AGE to the model in step 4 which will be the **final model**. Write the fitted equation.

	Age	HR	DIASBP	BMI	Gender	CHF
Age						
HR	p-va=0.065 not Sig (bor)					
DIASBP	p-va=0.117 not signific	p-va=0.787 not signific				
BMI	p-va=0.621 not signific	p-va=0.856 not signific	p-va=0.950 not signific			
Gender	p-va=0.008 significant	p-va=0.152 Not signific	p-va=0.044 significant	pval=0.918 not sig		
CHF	p-va=0.017 Significant	p-va=0.115 not Signific	p-va=0.877 not signific	p-va=0.701 not signific	p-va=0.341 not signi	

Interactions between Age and 'Congestive heart complications' is significant.
Interactions between Age and 'Gender' is significant.
Interactions between Gender and 'initial diastolic blood pressure in mm/Hg' is significant.
Interaction between age and HR is border line significant.
All other interactions are (age* DIASBP, age*BMI, HR* DIASBP, HR*BMI, HR*gender, HR*CHF, DIASBP*BMI, DIASBP*CHF, BMI*gender, BMI*CHF, Gender*CHF) are not significant.

Categorical Variable Codings[a,c]

		Frequency	(1)
gender[b]	0=male	300	0
	1=female	200	1
congestive heart complications[b]	0=No	345	0
	1=Yes	155	1

a. Category variable: gender (gender)

b. Indicator Parameter Coding

c. Category variable: congestive heart complications (CHF)

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2241.106	207.855	7	.000	214.052	7	.000	214.052	7	.000

a. Beginning Block Number 1. Method = Enter

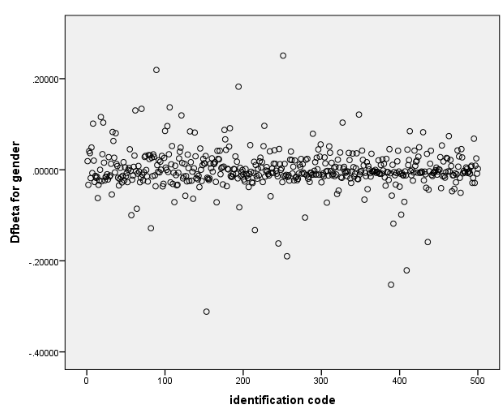
Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
gender	2.249	.960	5.487	1	.019	9.480
age at hospital admission	.062	.008	55.424	1	.000	1.064
initial heart rate in beats per minute	.011	.003	14.925	1	.000	1.011
initrial diastolic blood pressure in mm/Hg	-.011	.003	9.684	1	.002	.989
body mass index	-.044	.016	7.478	1	.006	.957
congestive heart complications	.776	.146	28.362	1	.000	2.172
age at hospital admission*gender	-.032	.012	6.979	1	.008	.968

h(t)^=h_0(t)exp(2.249*gender+0.062*age+0.011*HR-0.011* DIASBP-0.044*BMI+0.776*CHF-0.032*age*gender)

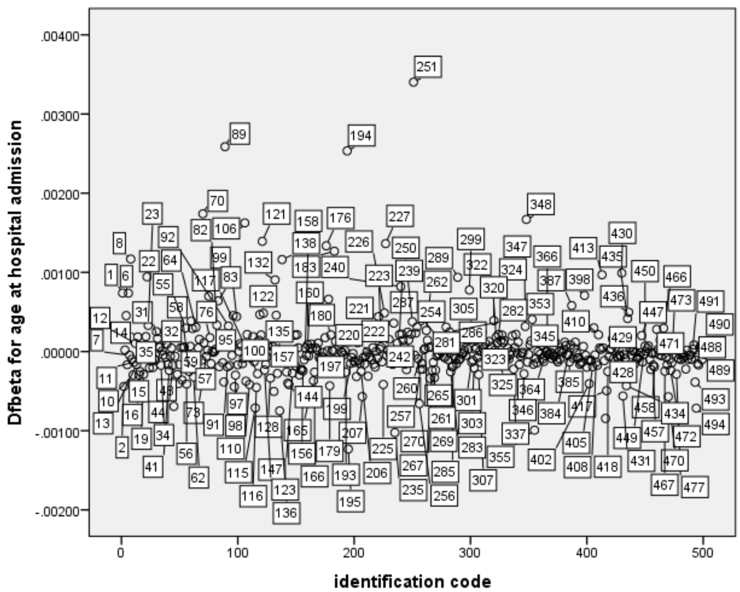
Log hazard estimated functions lnH(t)= ln[H_0(t)]+ 2.249*gender+ 0.062*age+0.011*HR-0.011* DIASBP-0.044*BMI+0.776*CHF-0.032*age*gender

6. Check for influential observations (Plot dfbeta versus ID and partial residual versus lenfol for variables in **final model**). Interpret plots.



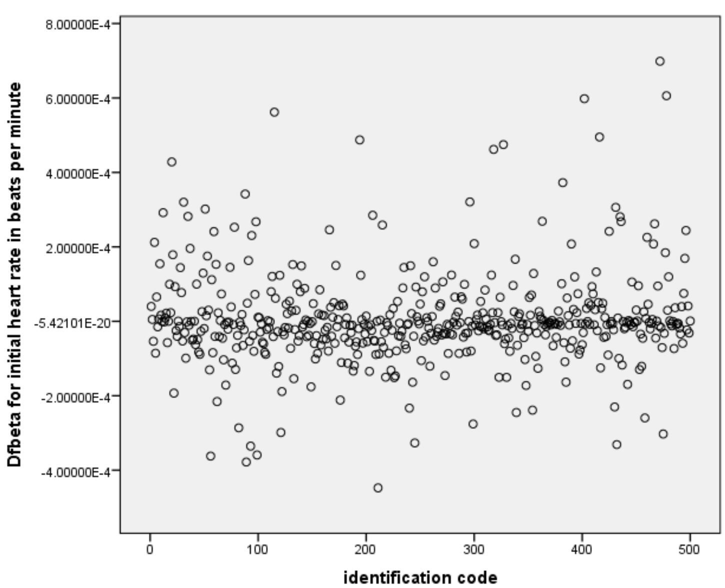
Dfbeta (B1i) = B1 - B1i

Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of -0.2 to 0.2). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of gender.



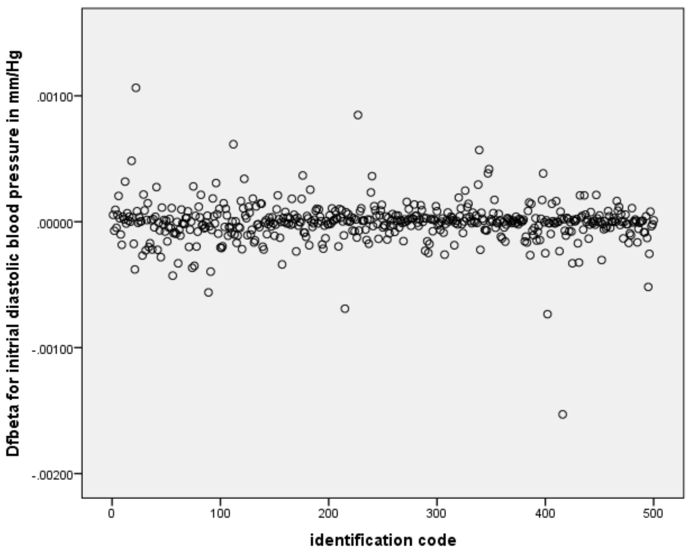
Dfbeta (B1i) = B1 - B1i

Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of 0.001 to -0.001). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of age. Case 89, 194 and 251 stay little away from zero

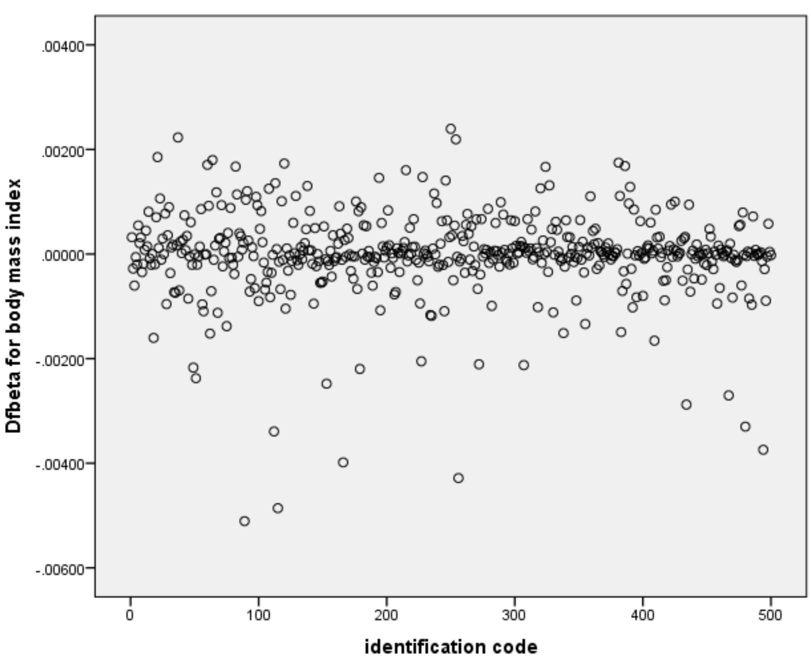


Dfbeta (B1i) = B1 - B1i

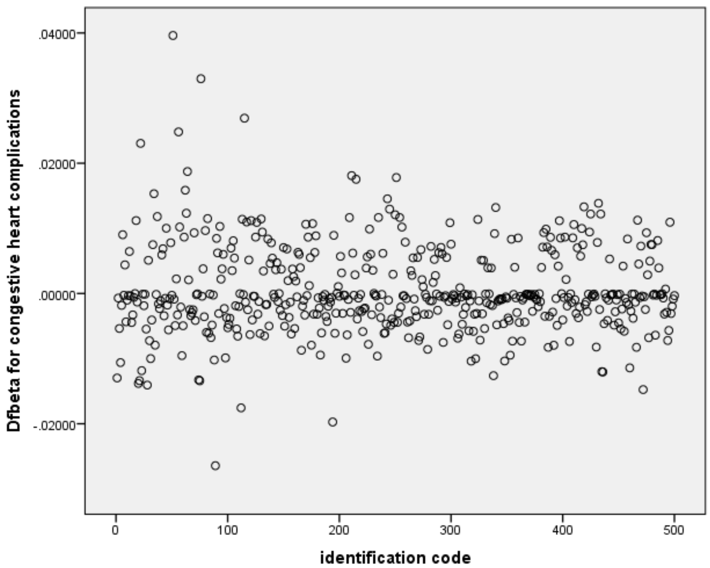
Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of 0.0002 to -0.0002). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of HR.



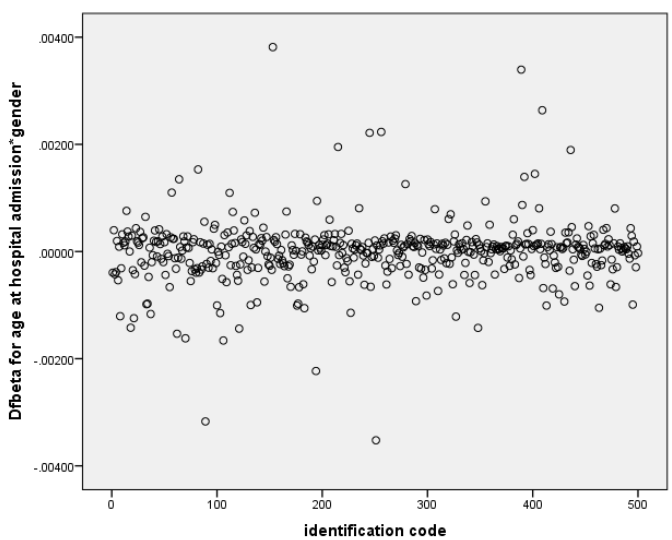
Dfbeta (B1i) = B1 - B1i
Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of 0.001 to -0.001). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of DIASBP.



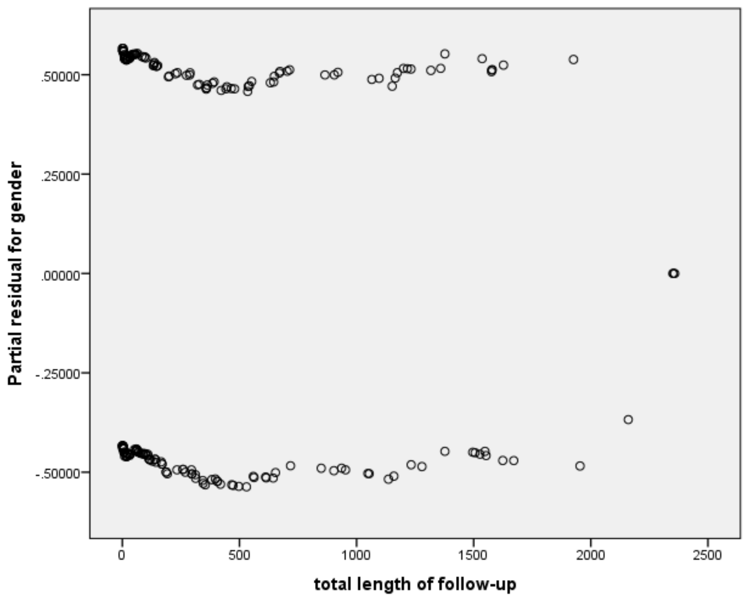
Dfbeta (B1i) = B1 - B1i
Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of 0.002 to -0.002). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of BMI.



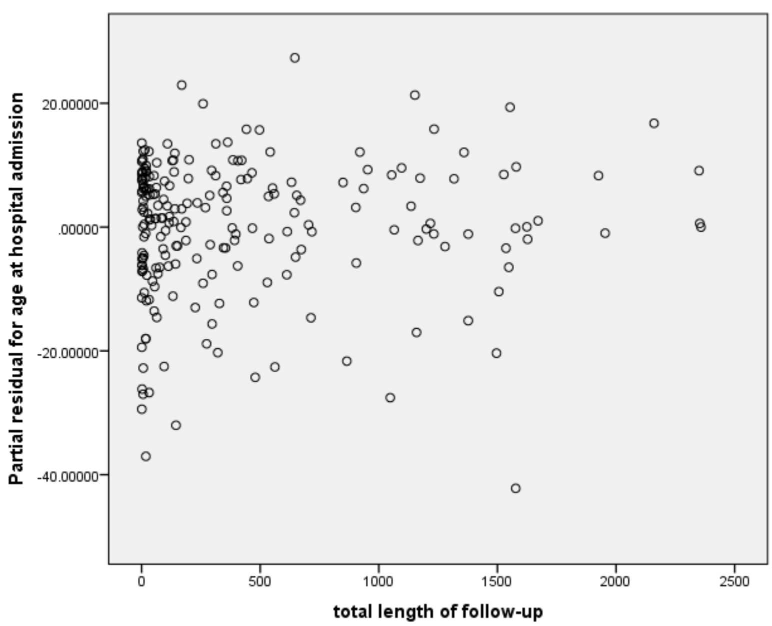
Dfbeta (B1i) = B1 - B1i
Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of 0.02 to -0.02). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of ‘congestive heart complications’.



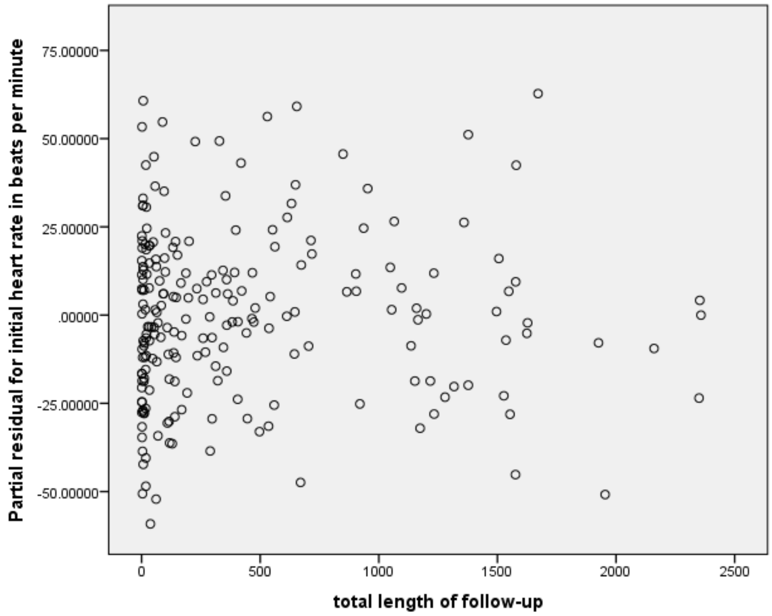
Dfbeta (B1i) = B1 - B1i
Where B1 is the value of the coefficient when all cases are included and B1i is the value of the coefficient when the ith case is excluded. Most of the data points stay very closer to zero (within the band of 0.002 to -0.002). Which mean when most of the case excluding from the model, it does not make big difference on coefficient of age*gender.



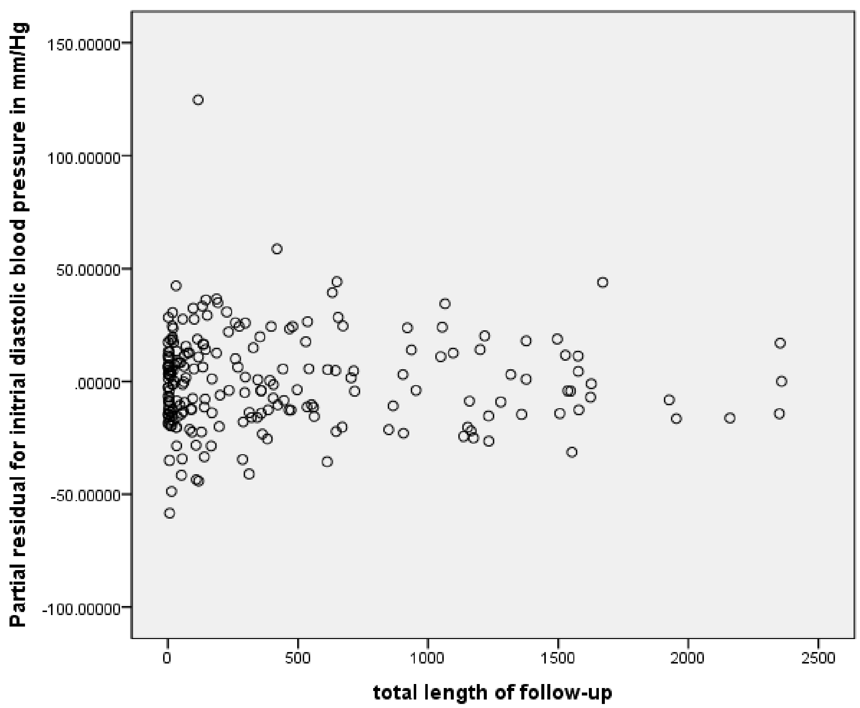
Gender is a categorical variable. So there are two lines. One line is very closer to 0.5 and other is very closer to -0.5. But its seems like residual for gender is changing with time. Which means assumption not met.



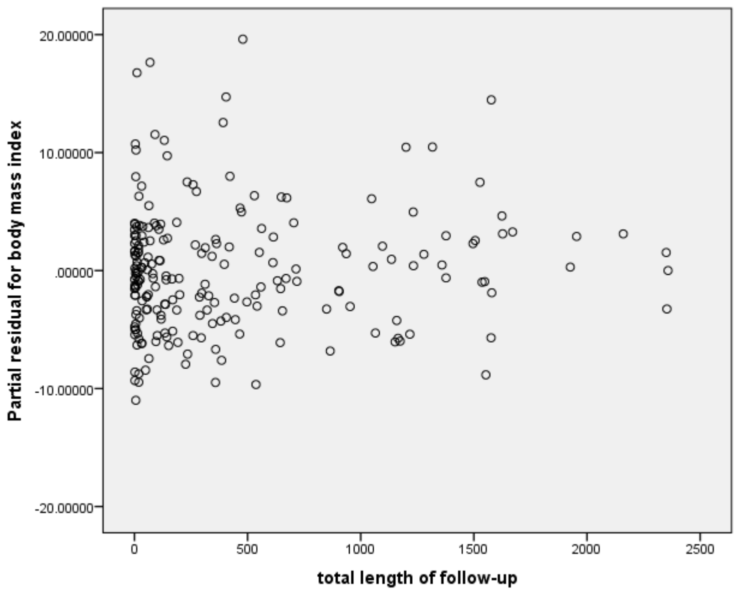
Most of the data points are within the band of 20 to -20. So Its seems to be age is not changing with time. Which mean proportional assumption appear to be satisfy.



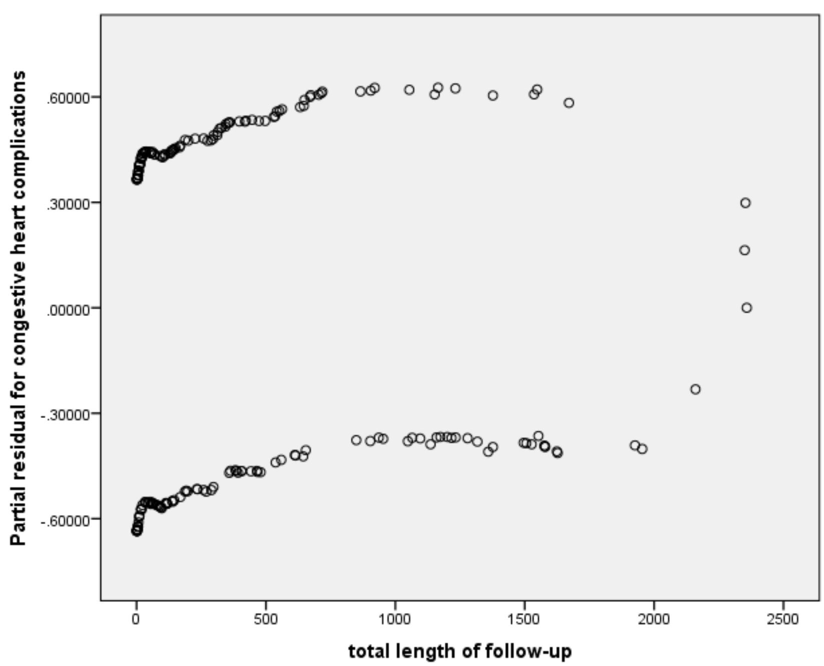
Most of the data points are within the band of -50 to 50. So Its seems to be Initial heart rate is not changing with time. Which means proportional assumption appear to be satisfy.



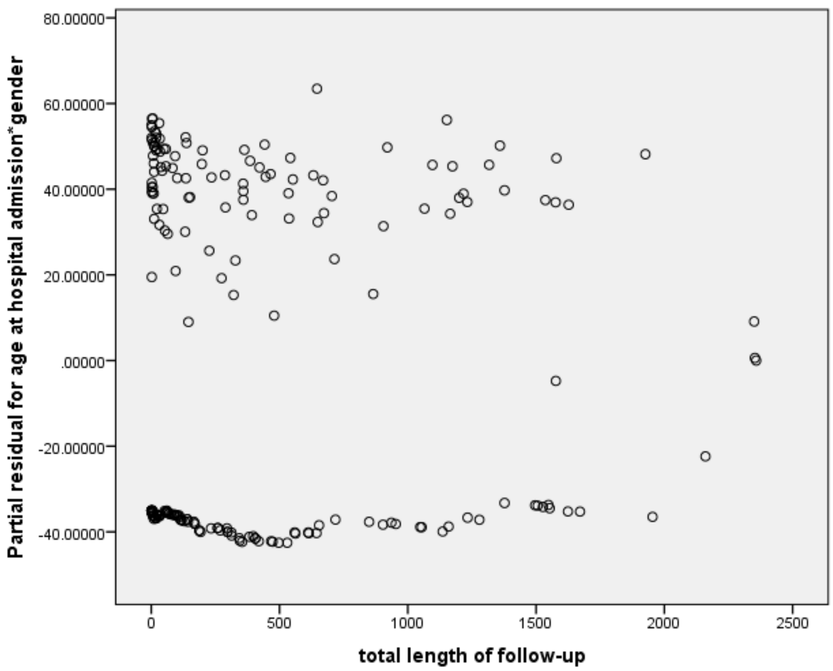
Most of the data points are within the band of -50 to 50. So Its seems to be initrial diastolic blood pressure is not changing with time. Which means proportional assumption appear to be satisfy.



Most of the data points are within the band of -10 to 10. So Its seems to be BMI is not changing with time Which means proportional assumption appear to be satisfy.



Congestive heart complications is a categorical variable. So there are two lines. But its seems like residual for Congestive heart complications is changing with time. Which means assumption not met.



It seems to be there are two lines. But its seems like residuals are changing with time. Which means assumption not met.

7. Do the data support the assumption of linearity in the log hazard of all selected continuous covariates in the **final model**? (Use the quartile design variable method)

Age

Statistics

		age at hospital admission	initial heart rate in beats per minute	initrial diastolic blood pressure in mm/Hg	body mass index
N	Valid	500	500	500	500
	Missing	0	0	0	0
Percentiles	25	59.00	69.00	63.00	23.1898
	50	72.00	85.00	79.00	25.9459
	75	82.00	100.75	91.75	29.4012

Age group	Mid poins
30-59	44.5
60-72	66
73-82	77.5
83-104	93.5

Age_group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	138	27.6	27.6	27.6
	2	119	23.8	23.8	51.4
	3	130	26.0	26.0	77.4
	4	113	22.6	22.6	100.0
	Total	500	100.0	100.0	

Categorical Variable Codings[a,c,d]

		Frequency	(1)	(2)	(3)
gender[b]	0=male	300	0		
	1=female	200	1		
congestive heart complications[b]	0=No	345	0		
	1=Yes	155	1		
Age_group[b]	1	138	0	0	0
	2	119	1	0	0
	3	130	0	1	0
	4	113	0	0	1

a. Category variable: gender (gender)

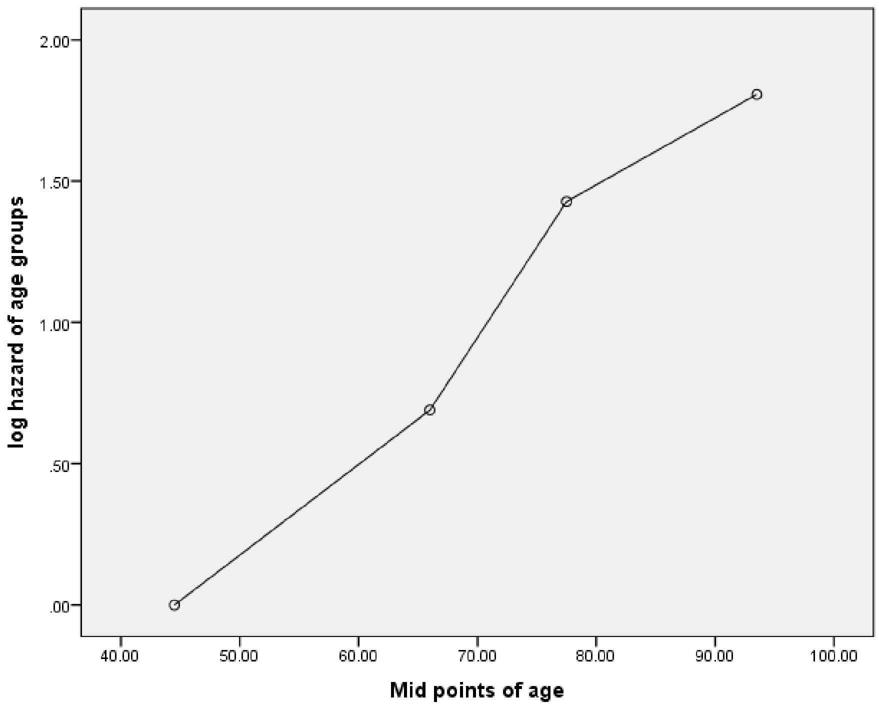
b. Indicator Parameter Coding

c. Category variable: congestive heart complications (CHF)

d. Category variable: Age_group

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
initial heart rate in beats per minute	.010	.003	11.411	1	.001	1.010
congestive heart complications	.834	.146	32.593	1	.000	2.302
initrial diastolic blood pressure in mm/Hg	-.011	.003	10.438	1	.001	.989
body mass index	-.049	.016	9.204	1	.002	.952
gender	-.299	.145	4.270	1	.039	.741
Age_group			53.176	3	.000	
Age_group(1)	.690	.294	5.487	1	.019	1.993
Age_group(2)	1.428	.273	27.317	1	.000	4.172
Age_group(3)	1.807	.278	42.206	1	.000	6.093



There is no departure from the linear sight. Keep increasing. Log Hazard keep changing linearly. So data support the assumption of linearity in the log hazard of age

Initial heart rate in beats per minute

Statistics

		age at hospital admission	initial heart rate in beats per minute	initrial diastolic blood pressure in mm/Hg	body mass index
N	Valid	500	500	500	500
	Missing	0	0	0	0
Percentiles	25	59.00	69.00	63.00	23.1898
	50	72.00	85.00	79.00	25.9459
	75	82.00	100.75	91.75	29.4012

Descriptive Statistics

	N	Minimum	Maximum
initial heart rate in beats per minute	500	35	186
Valid N (listwise)	500		

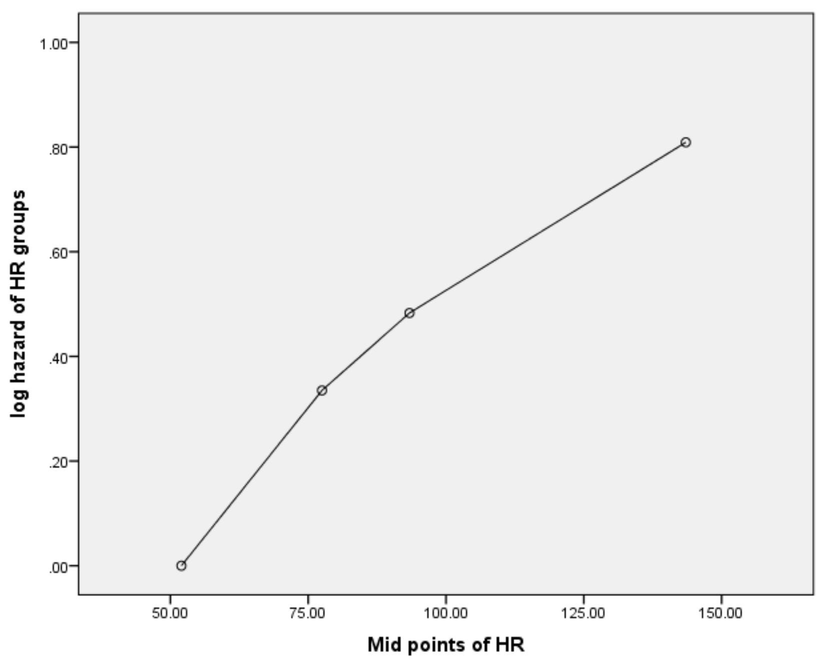
HR group	Mid poins
35-69	52
70-85	77.5
86-100.75	93.375
101-186	143.5

HR_group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	129	25.8	25.8	25.8
	2	123	24.6	24.6	50.4
	3	123	24.6	24.6	75.0
	4	125	25.0	25.0	100.0
	Total	500	100.0	100.0	

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
age at hospital admission	.050	.007	57.160	1	.000	1.052
gender	-.269	.144	3.494	1	.062	.764
initrial diastolic blood pressure in mm/Hg	-.010	.004	8.154	1	.004	.990
body mass index	-.046	.016	7.921	1	.005	.955
congestive heart complications	.750	.149	25.429	1	.000	2.117
HR_group			15.965	3	.001	
HR_group(1)	.335	.237	1.995	1	.158	1.398
HR_group(2)	.483	.212	5.174	1	.023	1.621
HR_group(3)	.809	.208	15.154	1	.000	2.246



There is no departure from the linear sight. Keep increasing. Log Hazard keep changing linearly. So data support the assumption of linearity in the log hazard of HR.

Initial diastolic blood pressure in mm/HG

Statistics

		age at hospital admission	initial heart rate in beats per minute	initrial diastolic blood pressure in mm/Hg	body mass index
N	Valid	500	500	500	500
	Missing	0	0	0	0
Percentiles	25	59.00	69.00	63.00	23.1898
	50	72.00	85.00	79.00	25.9459
	75	82.00	100.75	91.75	29.4012

Descriptive Statistics

	N	Minimum	Maximum
initrial diastolic blood pressure in mm/Hg	500	6	198
Valid N (listwise)	500		

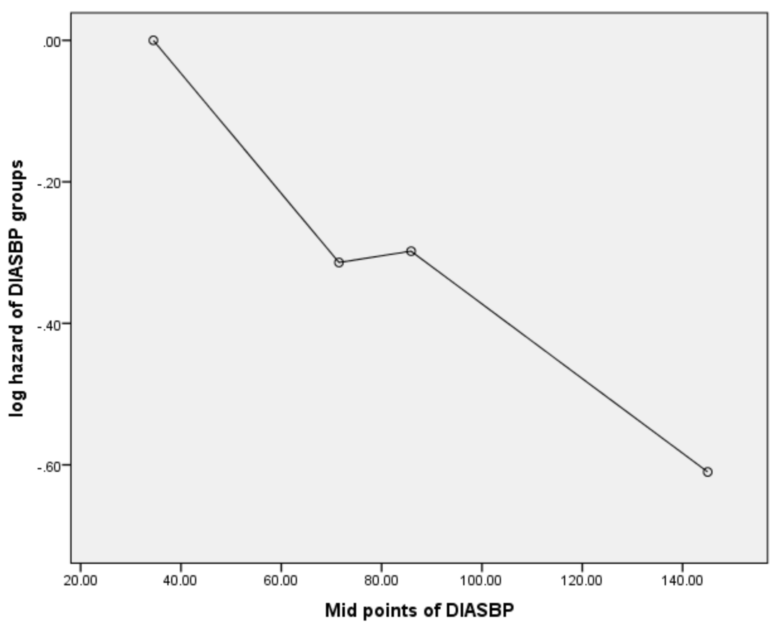
DIASBP group	Mid poins
6-63	34.5
64-79	71.5
80-91.75	85.875
92-198	145

DIASBP_group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	126	25.2	25.2	25.2
	2	126	25.2	25.2	50.4
	3	123	24.6	24.6	75.0
	4	125	25.0	25.0	100.0
	Total	500	100.0	100.0	

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
age at hospital admission	.050	.007	57.262	1	.000	1.051
gender	-.253	.144	3.063	1	.080	.777
body mass index	-.046	.016	7.755	1	.005	.955
congestive heart complications	.769	.151	26.080	1	.000	2.157
initial heart rate in beats per minute	.011	.003	14.505	1	.000	1.011
DIASBP_group			8.644	3	.034	
DIASBP_group(1)	-.314	.177	3.144	1	.076	.731
DIASBP_group(2)	-.298	.211	1.990	1	.158	.742
DIASBP_group(3)	-.610	.213	8.252	1	.004	.543



Difficult to tell departure from the linear sight or just random variation.

Body mass index

Statistics

		age at hospital admission	initial heart rate in beats per minute	initrial diastolic blood pressure in mm/Hg	body mass index
N	Valid	500	500	500	500
	Missing	0	0	0	0
Percentiles	25	59.00	69.00	63.00	23.1898
	50	72.00	85.00	79.00	25.9459
	75	82.00	100.75	91.75	29.4012

Descriptive Statistics

	N	Minimum	Maximum
body mass index	500	13.05	44.84
Valid N (listwise)	500		

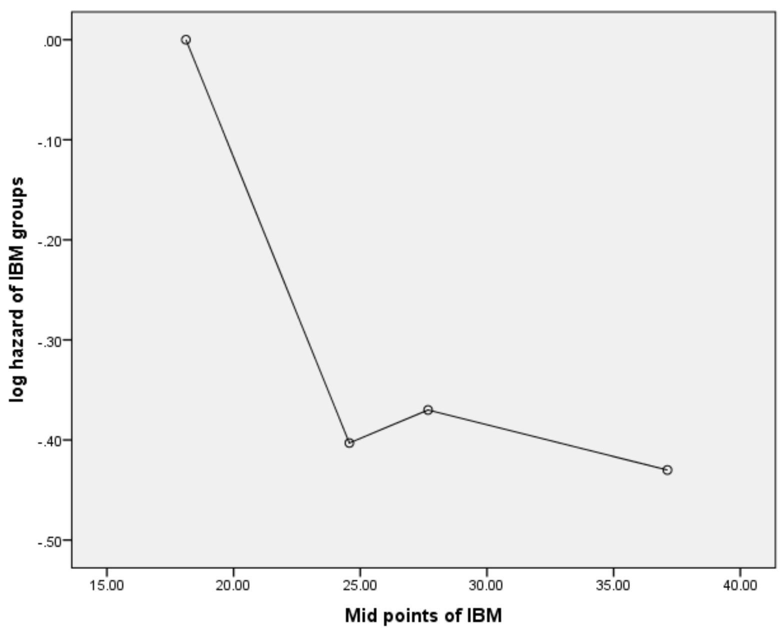
BMI group	Mid poins
13.05-23.1898	18.1199
23.1899-25.9459	24.5679
25.946-29.4012	27.6736
29.4013-44.84	37.121

BMI_group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	125	25.0	25.0	25.0
	2	125	25.0	25.0	50.0
	3	125	25.0	25.0	75.0
	4	125	25.0	25.0	100.0
	Total	500	100.0	100.0	

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
age at hospital admission	.052	.007	60.689	1	.000	1.053
gender	-.262	.143	3.357	1	.067	.769
initial heart rate in beats per minute	.011	.003	14.580	1	.000	1.011
initrial diastolic blood pressure in mm/Hg	-.011	.004	8.942	1	.003	.989
congestive heart complications	.769	.146	27.576	1	.000	2.157
BMI_group			7.268	3	.064	
BMI_group(1)	-.403	.180	5.035	1	.025	.668
BMI_group(2)	-.370	.190	3.803	1	.051	.690
BMI_group(3)	-.430	.223	3.722	1	.054	.650



Difficult to tell departure from the linear sight or just random variation.

8. From the final model,

$$\hat{h}(t) = h_0(t)\exp(2.249*gender+0.062*age+0.011*HR-0.011*DIASBP-0.044*BMI+0.776*CHF-0.032*age*gender)$$

a) Estimate and interpret the hazard ratio for a 10 beat/minute increase in heart rate.

Estimated hazard ratio for an increase of 10 beat/minute in heart rate is exp (0.011*10) = exp (0.11) = 1.116
1.116 Indicates that for every 10 beat/minute increase in heart rate, the hazard rate of death increases by 11.6%. Which is adjusted for other variables.

b) Estimate and interpret the hazard ratio for a 10 mm/Hg increase in diastolic blood pressure.

Estimated hazard ratio for an increase of 10 mm/Hg in diastolic blood pressure is exp (-0.011*10) = exp (-0.11) = 0.8958
0.8958 Indicates that for every 10 mm/Hg increase in diastolic blood pressure, the hazard rate of death decreases by 10%. Which is adjusted for other variables.

c) Estimate and interpret the hazard ratio for congestive heart complications.

Estimated hazard ratio for heart complications is exp (0.776) = 2.1727

Estimated (Adjusted for other variables) hazard ratio of heart complications is 2.1727. Which means at any time during the study, the per day rate of death among people who have congestive heart complications is 2.1727 times higher than that for people who do not have congestive heart complications when other variable remains constant.

9. Find a best model: conduct multivariate analysis (Forward LR) using Logistic Regression which is the final model. From the final model, repeat step 8 using odds ratio.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age	.088	.009	95.140	1	.000	1.092
	Constant	-6.573	.667	97.183	1	.000	.001
Step 2[b]	age	.081	.009	74.865	1	.000	1.085
	CHF(1)	1.390	.235	34.973	1	.000	4.013
	Constant	-6.533	.697	87.911	1	.000	.001
Step 3[c]	age	.080	.010	71.293	1	.000	1.084
	HR	.017	.005	11.898	1	.001	1.017
	CHF(1)	1.230	.240	26.178	1	.000	3.420
	Constant	-7.874	.828	90.372	1	.000	.000
Step 4[d]	age	.077	.010	64.739	1	.000	1.080
	HR	.019	.005	14.730	1	.000	1.019
	DIASBP	-.015	.006	7.781	1	.005	.985
	CHF(1)	1.216	.243	24.979	1	.000	3.373
	Constant	-6.646	.916	52.592	1	.000	.001
Step 5[e]	age	.071	.010	49.892	1	.000	1.073
	HR	.019	.005	14.459	1	.000	1.019
	DIASBP	-.015	.006	6.681	1	.010	.986
	bmi	-.045	.023	3.948	1	.047	.956
	CHF(1)	1.236	.245	25.444	1	.000	3.440
	Constant	-5.060	1.196	17.903	1	.000	.006

a. Variable(s) entered on step 1: age.
b. Variable(s) entered on step 2: CHF.
c. Variable(s) entered on step 3: HR.
d. Variable(s) entered on step 4: DIASBP.
e. Variable(s) entered on step 5: bmi.

Categorical Variables Codings

		Frequency	Parameter coding (1)
congestive heart complications	No	345	.000
	Yes	155	1.000
gender	male	300	.000
	female	200	1.000

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age	.103	.015	48.147	1	.000	1.108
	gender(1)	5.161	1.510	11.682	1	.001	174.404
	HR	.019	.005	14.023	1	.000	1.019
	DIASBP	-.016	.006	7.819	1	.005	.984
	bmi	-.053	.023	5.321	1	.021	.948
	CHF(1)	1.229	.249	24.404	1	.000	3.419
	age by gender(1)	-.072	.020	12.534	1	.000	.930
	Constant	-6.922	1.397	24.566	1	.000	.001

a. Variable(s) entered on step 1: age, gender, HR, DIASBP, bmi, CHF, age * gender .

Estimated logit transformation

$\ln[\frac{p^{\wedge}}{1-p^{\wedge}}]$= -6.922 +0.103*age + 0.019*HR-0.053*bmi-0.016*DIASBP +1.229*CHF +5.161*gender -0.072*age*gender

logistic probabilities

$$\mathbf{P^{\wedge} =} \frac{\mathbf{1}}{\boldsymbol{e}^{-(-6.922\ +0.103*\mathrm{age}\ +\ 0.019*\mathrm{HR}-0.053*\mathrm{bmi}-0.016*\mathrm{DIASBP}\ +1.229*\mathrm{CHF}\ +5.161*\mathrm{gender}\ -0.072*\mathrm{age}*\mathrm{gender})}}$$

a)Estimate and interpret the odds for a 10 beat/minute increase in heart rate.

Coefficient of heart rate = 0.019

OR^=exp(10*0.019) = 1.209 odds ratio for Coefficient of heart rate.

1.209 Indicates that for every 10 beat/minute' increase of heart rate, the risk of having the event (dead) increases by 20.9% while others variable remains constant.

b)Estimate and interpret the odds ratio for a 10 mm/Hg increase in diastolic blood pressure.

Coefficient of diastolic blood pressure = -0.016

OR^=exp(10*-0.016) = 0.8521 odds ratio for diastolic blood pressure
0.8521 Indicates that for every 10 mm/Hg increase in diastolic blood pressure, the risk of having the event (dead) decreases by 15% while others variable remains constant.

C)Estimate and interpret the odds ratio for congestive heart complications.

Coefficient of congestive heart complications = 1.229

OR^=exp(1.236) = 3.417 odds ratio for congestive heart complications.

Odds of having the event ("death") among those people who do have "congestive heart complications" is 3.417 times as likely than those people who do not have "congestive heart complications"