York University

Heart Failure

Survival Analysis

April 5, 2023

SURVIVAL ANALYSIS MATH 4130K

Department of Mathematics

Dao Feng, Jinqi Guo, Shahzab Hussain

Professor Kevin McGregor

What we'll discuss this afternoon

Presentation Overview

What is heart failure and causes it?

A medical condition in which the heart is unable to pump enough blood to meet the body's demands

The main reasons for heart failure include coronary heart disease, diabetes, high blood pressure, and other diseases like HIV

This can occur when the heart muscle is weakened or stiffened, which makes it harder for the heart to fill with blood or pump blood out to the body.

DATA OVERVIEW

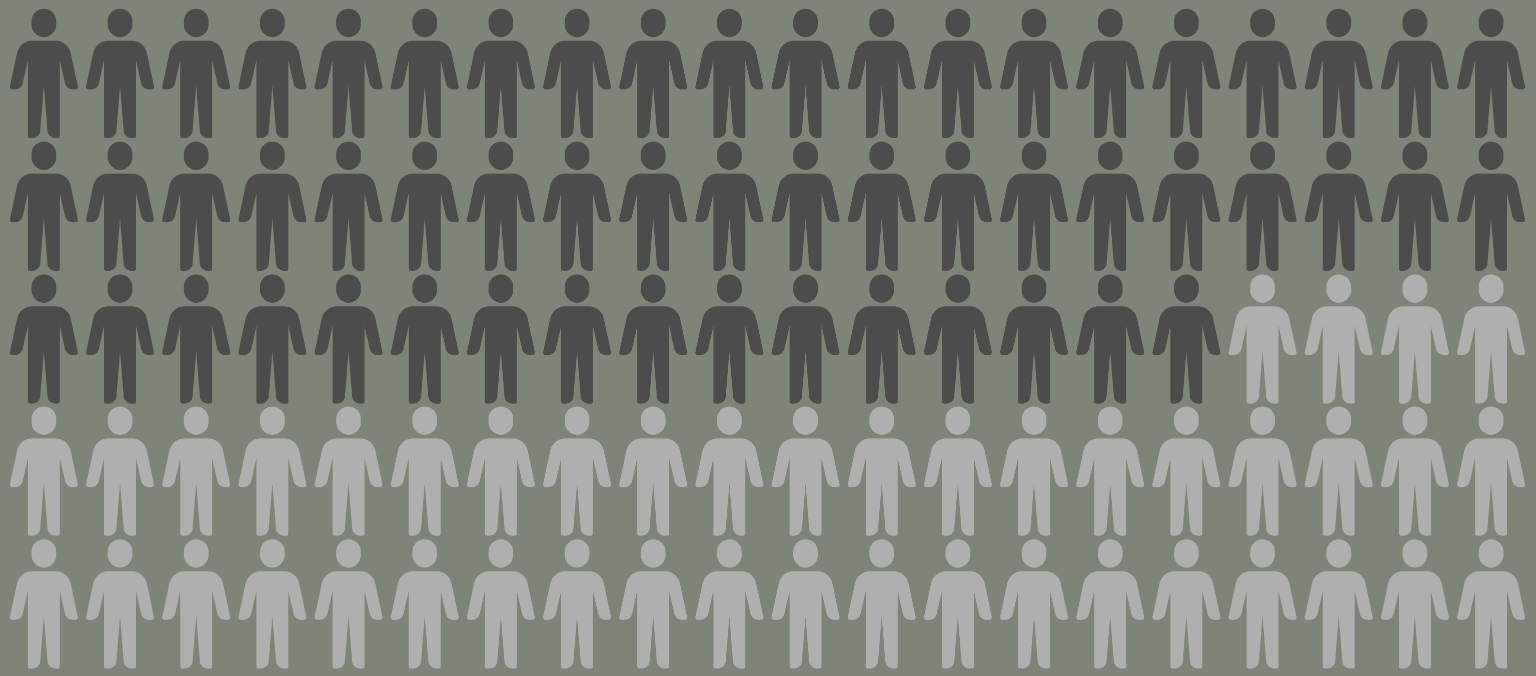
[Source: UCI LEARNING REPOSITORY]

DATA COLLECTION :

INSTITUTE OF CARDIOLOGY AND ALLIED HOSPITAL IN FAISALABAD

COLLECTION TIME :

APRIL-DECEMBER 2015



194 M 105 F

299 Subjects

203 Censored 96 Observed

13 Clinical Features

13 Clinical Features

AGE

SEX

DIABETES

SMOKING

TIME

ANAEMIA

CREATININE PHOSPHOKINASE

EJECTION FRACTION

HIGH BLOOD PRESSURE

DEATH [TARGET EVENT]

PLATELETS

SERUM CREATININE

SERUM SODIUM

13 Clinical Features

AGE
(numerical)

SEX
(binary)

DIABETES
(boolean)

SMOKING
(boolean)

TIME
follow-up period (days)

ANAEMIA
decrease of red blood cells or hemoglobin (boolean)

CREATININE PHOSPHOKINASE
CPK enzyme level in blood msg/L

EJECTION FRACTION
percentage of blood leaving the heart at each contraction

HIGH BLOOD PRESSURE
(boolean)

DEATH [TARGET EVENT]
if the patient deceased during the follow-up period (boolean)

PLATELETS
platelets in the blood (kiloplatelets/mL)

SERUM CREATININE
level of serum creatinine in the blood (mg/dL)

SERUM SODIUM
level of serum sodium in the blood (mEq/L)

13 Clinical Features

AGE
(numerical)

SEX
(binary)

DIABETES
(boolean)

SMOKING
(boolean)

TIME
follow-up period (days)

ANAEMIA
decrease of red blood cells or hemoglobin (boolean)

CREATININE PHOSPHOKINASE
CPK enzyme level in blood msg/L

EJECTION FRACTION
percentage of blood leaving the heart at each contraction

HIGH BLOOD PRESSURE
(boolean)

DEATH [TARGET EVENT]
if the patient deceased during the follow-up period (boolean)

PLATELETS
platelets in the blood (kiloplatelets/mL)

SERUM CREATININE
level of serum creatinine in the blood (mg/dL)

SERUM SODIUM
level of serum sodium in the blood (mEq/L)

Features of Interest

[TARGET EVENT]
DEATH

AGE	SERUM CREATININE (SCR)
SEX	EJECTION FRACTION (EF)
TIME	HIGH BLOOD PRESSURE (HBP)

Methodology

Kaplan-Meier PL Estimators

$$\hat{S}(t) = \prod_{j:t_j \le t} (1 - d_j/n_j) \text{ where } d_j/n_j \text{ is the estimated hazard for } j$$

01

Non-parametric Method

02

Utilized in comparing the survival rates between different levels within categorical covariates

03

Variance of the KM curve approximated by Greenwood's formula with confidence interval at each value of t

Cox Proportional Hazard Model

$$\lambda_i(t, x_i) = \lambda_0(t) exp\{x_i^T \beta\}$$

where $\lambda_0(t)$ is the baseline hazard

Proportional Hazard Assumption

Hazard ratio between two individuals with different predictors does not depend on time.

$$\frac{\lambda_2(t, x_2)}{\lambda_1(t, x_1)} = \frac{exp\{x_2^T \beta\}}{exp\{x_1^T \beta\}}$$

Model Assumptions

Methods of Verification

01 Log-Log Plots

Using the log-log plot on the survival functions of two individuals to check for parallel trends

The difference between the individuals are independent of time

02 Schoenfeld Residuals

Using the zph() and cox.zph() functions to check for proportionality assumption in R

Having a small p-values would indicate undesired time dependent coefficients

Analysis

First Steps

- 55 to 70% – Normal heart function.
- 40 to 55% – Below normal heart function. Can indicate previous heart damage from heart attack or cardiomyopathy.
- >75% – Can indicate a heart condition like hypertrophic cardiomyopathy, a common cause of sudden cardiac arrest.
- <=40% – May confirm the diagnosis of heart failure.

Penn Medicine

Prepare and clean data through:

01 Data Wrangling

02 Variable Manipulation

Ejection Fraction (EF) is converted into a categorical variable with 4 factor levels, according to Penn Medicine.

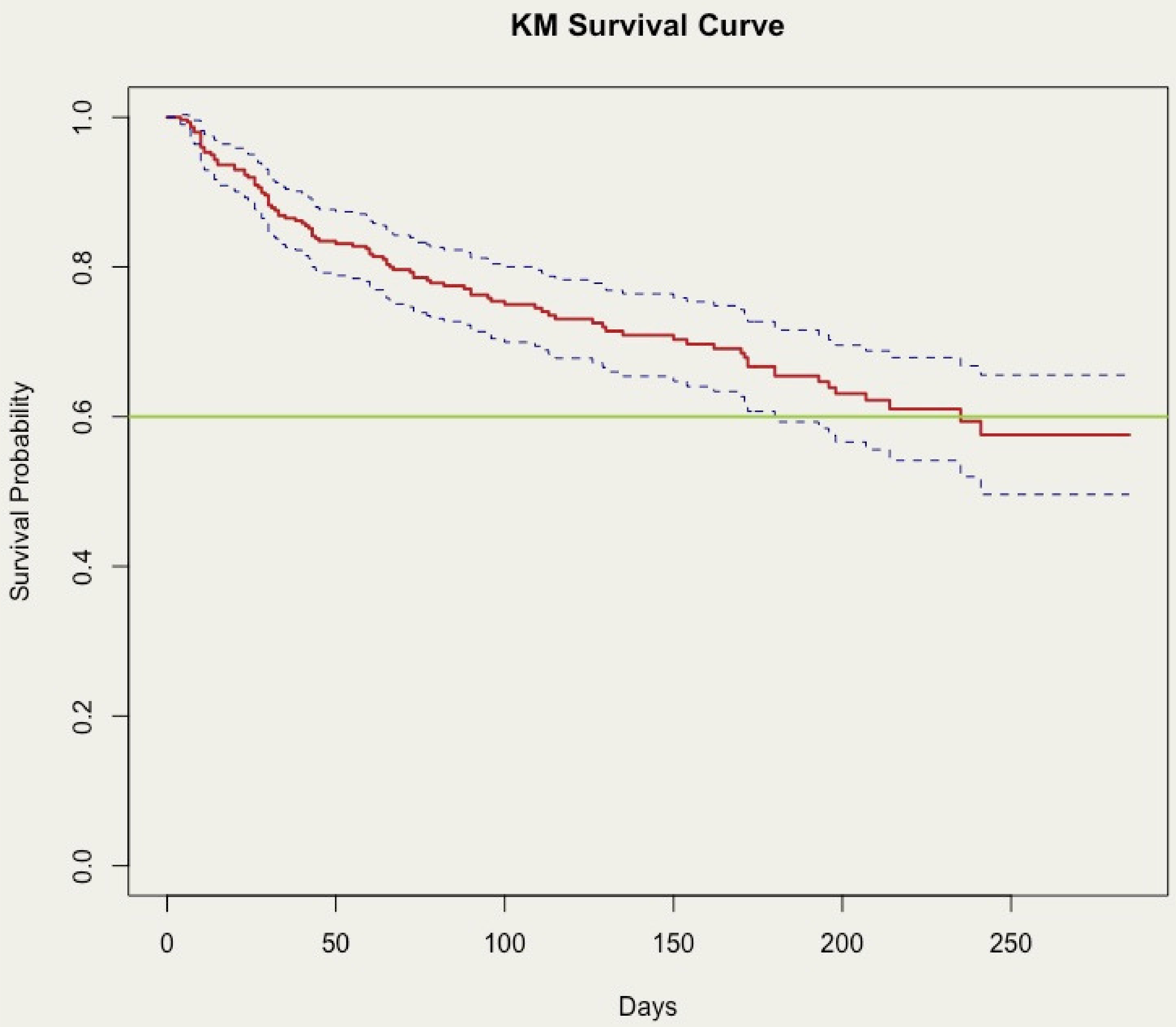
03 Standardize Continuous Variables

Analysis

Kaplan-Meier Survival Curve

The survival curve before modelling

Shows a gradual trend with a relatively high survival probability at just below 60% over long survival times



Cox Proportional Hazard Model

Regression Output from R

	term	estimate	std.error	z statistic	p.value
1	age	0.5096	0.1086	4.6930	0.0000
2	CPK	0.1897	0.0992	1.9128	0.0558
3	EF1	-0.2167	0.3194	-0.6784	0.4975
4	EF2	-0.8666	0.4369	-1.9837	0.0473
5	EF3	-12.5179	2168.8692	-0.0058	0.9954
6	platelets	-0.0360	0.1101	-0.3271	0.7436
7	SCR	0.3090	0.0689	4.4838	0.0000
8	SSO	-0.2560	0.0963	-2.6570	0.0079
9	anaemia	0.4226	0.2170	1.9471	0.0515
10	diabetes	0.1271	0.2231	0.5696	0.5689
11	HBP	0.4978	0.2172	2.2923	0.0219
12	sex	-0.1207	0.2529	-0.4773	0.6331
13	smoking	0.1641	0.2536	0.6472	0.5175

Full Model

$$\hat{\lambda}(t, x_i) = \hat{\lambda_0}(t) exp\{x_{i1}\hat{\beta}_{age} + x_{i2}\hat{\beta}_{CPK} + x_{i3}\hat{\beta}_{EF1} + x_{i4}\hat{\beta}_{EF2} + x_{i5}\hat{\beta}_{EF3} + x_{i6}\hat{\beta}_{platelets}$$
$$+ x_{i7}\hat{\beta}_{SCR} + x_{i8}\hat{\beta}_{SSO} + x_{i9}\hat{\beta}_{anaemia} + x_{i10}\hat{\beta}_{diabetes} + x_{i11}\hat{\beta}_{HBP} + x_{i12}\hat{\beta}_{sex} + x_{i13}\hat{\beta}_{smoking}\}$$

Reduced Model

Four significant covariates are selected for further analysis:

Age

EF (ejection fraction)

SCR (serum creatinine)

HBP (high blood pressure)

EF and HBP are categorical indicator variables*

The **two-way interaction effects** between them must be considered

$$\hat{\lambda}_0(t)exp\{x_{i1}\hat{\beta}_{age}+x_{i2}\hat{\beta}_{EF1}+x_{i3}\hat{\beta}_{EF2}+x_{i4}\hat{\beta}_{EF3}+x_{i5}\hat{\beta}_{SCR}+x_{i6}\hat{\beta}_{HBP}$$

$$+x_{i7}\hat{\beta}_{EF1*HBP}+x_{i8}\hat{\beta}_{EF2*HBP}+x_{i9}\hat{\beta}_{EF3*HBP}\}$$

Reduced Model

Regression Output

term	coef	exp(coef)	std.error	z statistic	p-value
age	0.45	1.57	0.11	4.21	0.00
EF1	-0.28	0.75	0.44	-0.65	0.52
EF2	-0.05	0.95	0.44	-0.12	0.90
EF3	-12.76	0.00	2094.77	-0.01	1.00
HBP	0.64	1.89	0.24	2.72	0.01
SCR	0.40	1.49	0.07	5.72	0.00
EF1:HBP	-0.15	0.86	0.63	-0.24	0.81
EF2:HBP	-2.73	0.07	1.14	-2.39	0.02
EF3:HBP	NA	NA	0.00	NA	NA

Age, **HBP**, and **SCR** are important predictors of dying of heart failure

The relationship between **EF** and the death is complex and depends on the presence of **HBP**

NA Values in the third interaction term due to the fact that there is only **one observation** in the third level of the variable **EF**

EF Frequency Table

	$\leq$ 40%	40 - 55%	55 - 70%	>70%
EF Level	0	1	2	3
Counts	219	44	35	1

Interpretations

Hazard Ratio

EF Frequency Table

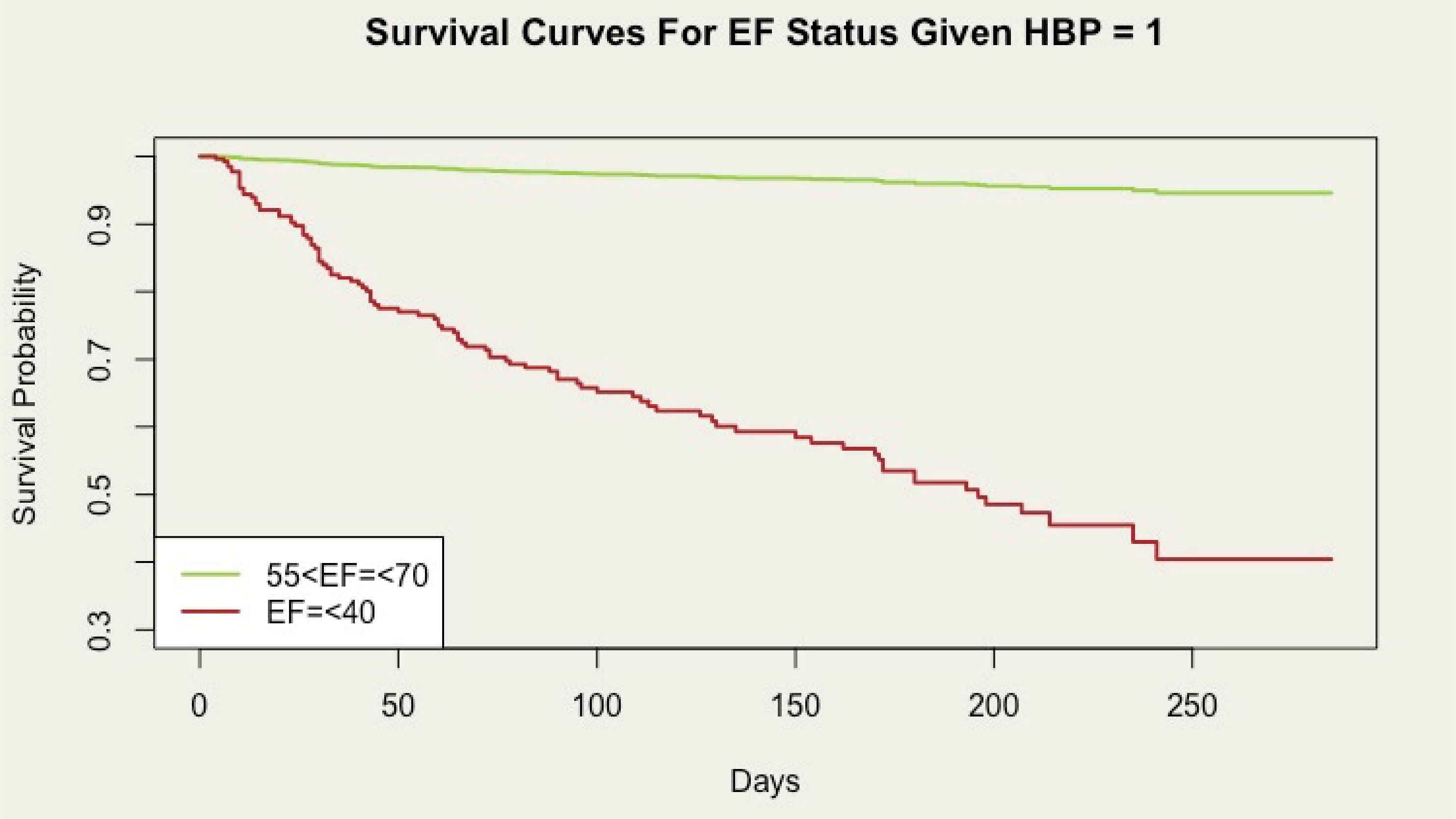
	$\leq$ 40%	40 - 55%	55 - 70%	>70%
EF Level	0	1	2	3
Counts	219	44	35	1

Term	$exp\{\hat{\beta}_j\}$	HR	Implication
age	1.5659208	$HR > 1$	Hazard increases by a multiple of $\hat{\beta}_j$ per unit increase in age
EF1	0.7541553	$HR < 1$	Hazard decreases by a multiple of $\hat{\beta}_j$ per unit increase in EF1
EF2	0.9476188	$HR < 1$	Hazard decreases by a multiple of $\hat{\beta}_j$ per unit increase in EF2
EF3	0.0000029	$HR < 1$	Hazard decreases by a multiple of $\hat{\beta}_j$ per unit increase in EF3
HBP	1.8939248	$HR > 1$	Hazard increases by a multiple of $\hat{\beta}_j$ per unit increase in HBP
SCR	1.4947745	$HR > 1$	Hazard increases by a multiple of $\hat{\beta}_j$ per unit increase in SCR
EF1:HBP	0.8617743	$HR < 1$	Hazard decreases by a multiple of $\hat{\beta}_j$ per unit increase in EF1:HBP
EF2:HBP	0.0650996	$HR < 1$	Hazard decreases by a multiple of $\hat{\beta}_j$ per unit increase in EF2:HBP

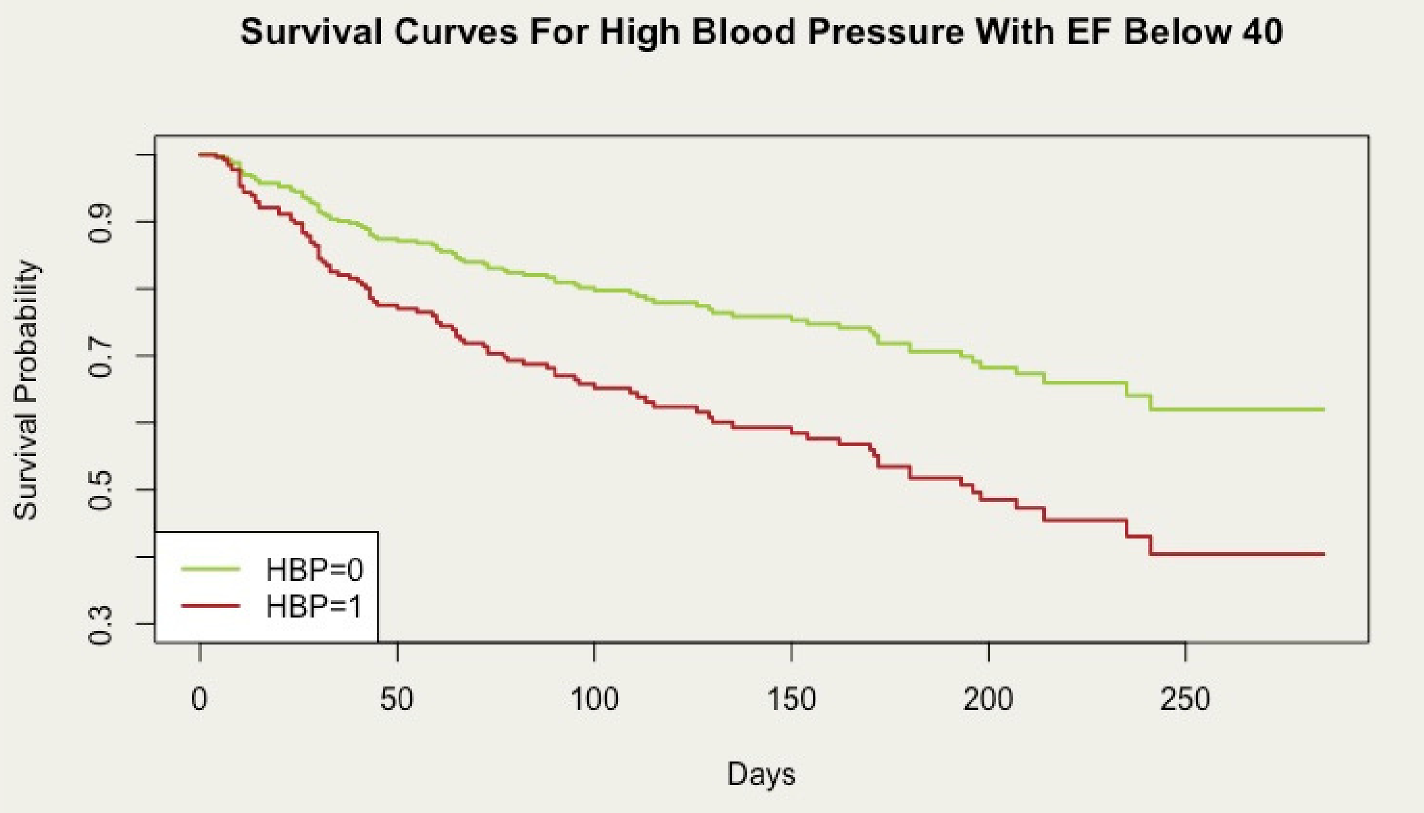
Conditional Survival Curves

Holding Age and SCR constant...

The presence of **HBP**

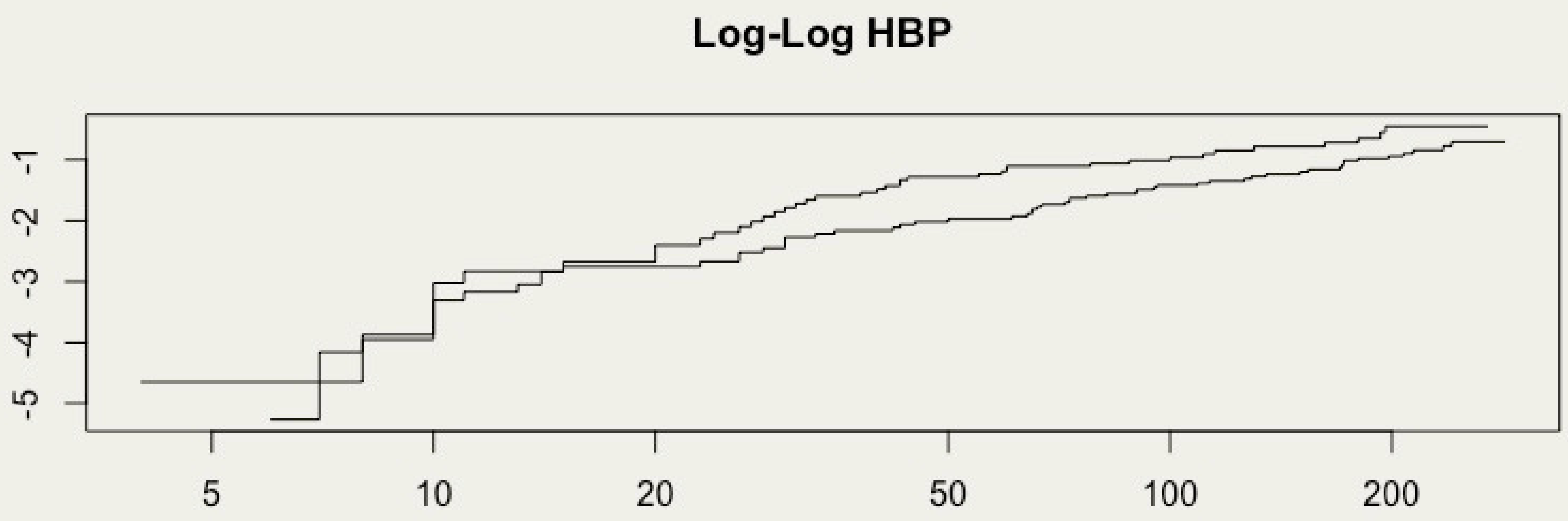


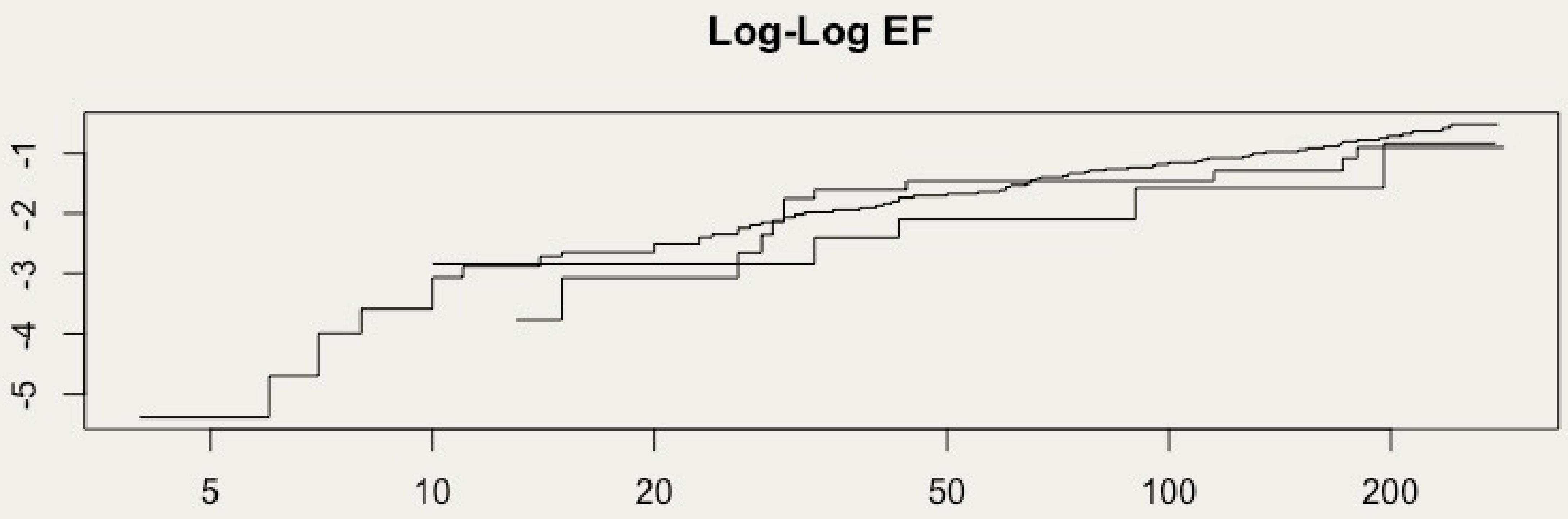
The presence of **EF** below of 40%



Verifying Assumptions

01 Log-Log Plots



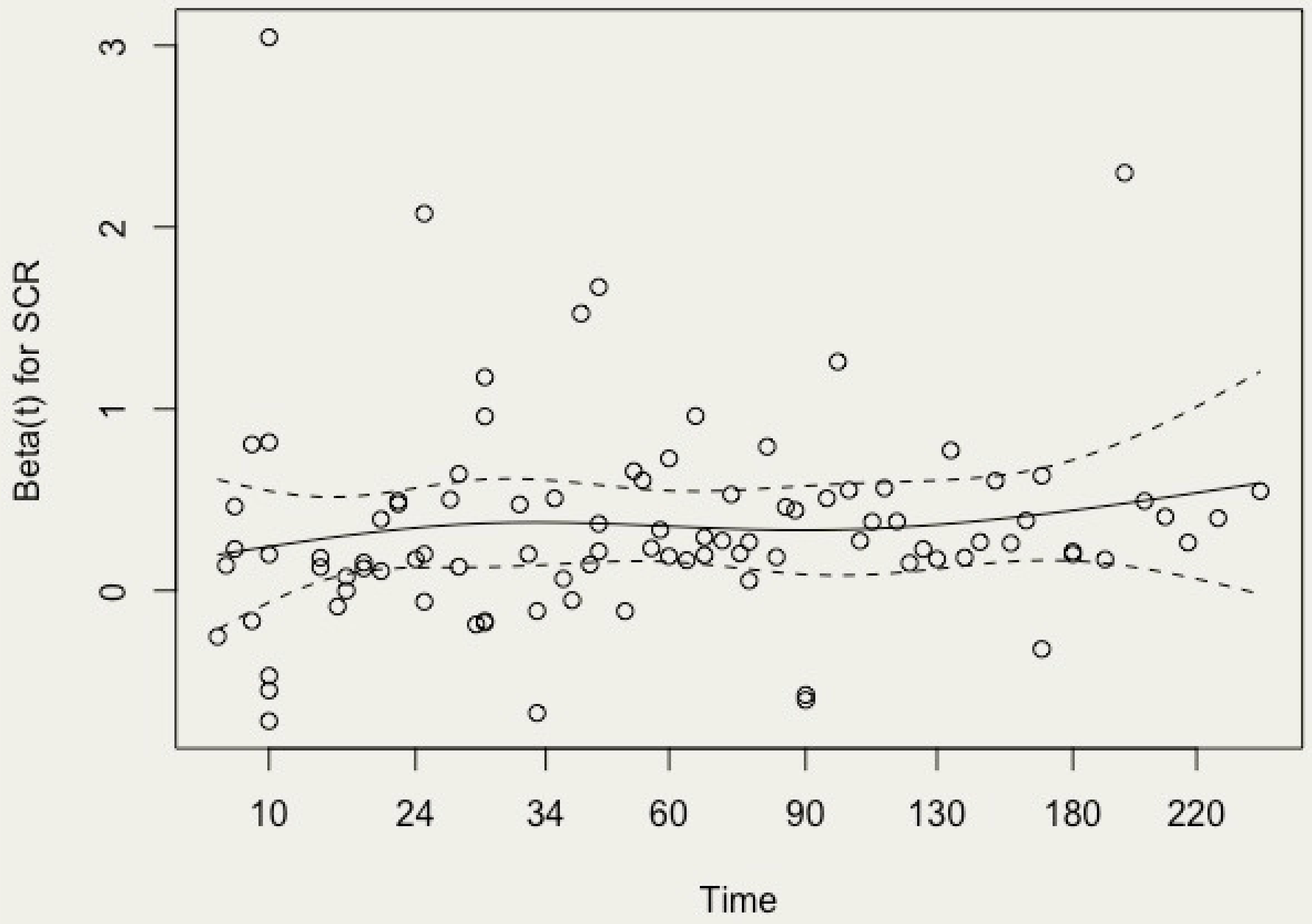


We proceed by checking the assumptions of the two categorical variables **HBP** and **EF**

	$\leq$ 40%	40 - 55%	55 - 70%	>70%
EF Level	0	1	2	3
Counts	219	44	35	1

Chi-Squared Table

	χ^2	df	p
age	0.3067	1	0.58
EF	0.0637	3	1.00
HBP	0.2038	1	0.65
SCR	0.8187	1	0.37
GLOBAL	1.7389	6	0.94



	χ^2	df	p
age	0.2175	1.0000	0.6409
EF	0.1685	3.0000	0.9825
HBP	0.1488	1.0000	0.6997
SCR	3.8766	1.0000	0.0490
EF:HBP	3.0418	2.0000	0.2185
GLOBAL	5.9843	8.0000	0.6490

LIMITATIONS

01

Stratified Model Without Interaction & Assumption Inconsistencies

02

Experimental Design Limitations & Sample Bias

03

Lead Time Bias

Stratified Model

Assumptions regarding Schoenfeld Residuals are all met

Two-way interaction effects are no longer present

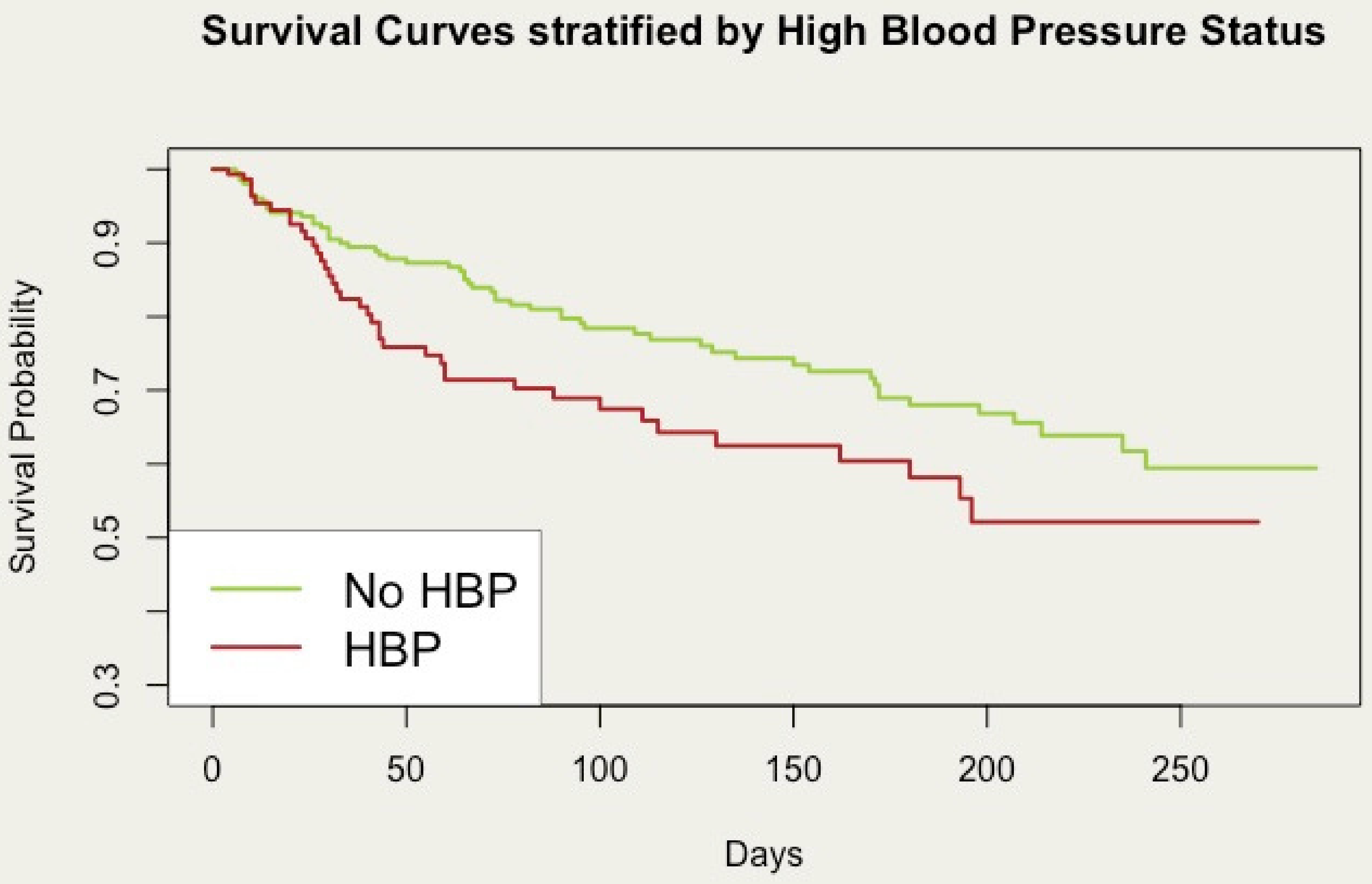
Stratifying by HBP eliminates it from from parametric portion of our model

$$\hat{\lambda}_0(t)exp\{x_{i1}\hat{\beta}_{age}+x_{i2}\hat{\beta}_{EF1}+x_{i3}\hat{\beta}_{EF2}+x_{i4}\hat{\beta}_{EF3}+x_{i5}\hat{\beta}_{SCR}$$

Stratified Model

$$\hat{\lambda}_0(t)exp\{x_{i1}\hat{\beta}_{age}+x_{i2}\hat{\beta}_{EF1}+x_{i3}\hat{\beta}_{EF2}+x_{i4}\hat{\beta}_{EF3}+x_{i5}\hat{\beta}_{SCR}$$

	χ^2	df	p
age	0.486	1	0.49
EF	0.066	3	1.00
SCR	0.480	1	0.49
GLOBAL	1.7389	5	0.94



	term	estimate	std.error	statistic	p.value
1	age	0.49	0.11	4.67	0.00
2	EF1	-0.37	0.31	-1.18	0.24
3	EF2	-0.97	0.43	-2.23	0.03
4	EF3	-12.64	2122.89	-0.01	1.00
5	SCR	0.35	0.07	5.17	0.00

Experimental Design

Inclusion Criteria

All the patients were aged 40 years or above, having left ventricular systolic dysfunction, belonging to NYHA class III and IV.

Small Sample Size

Observational Study

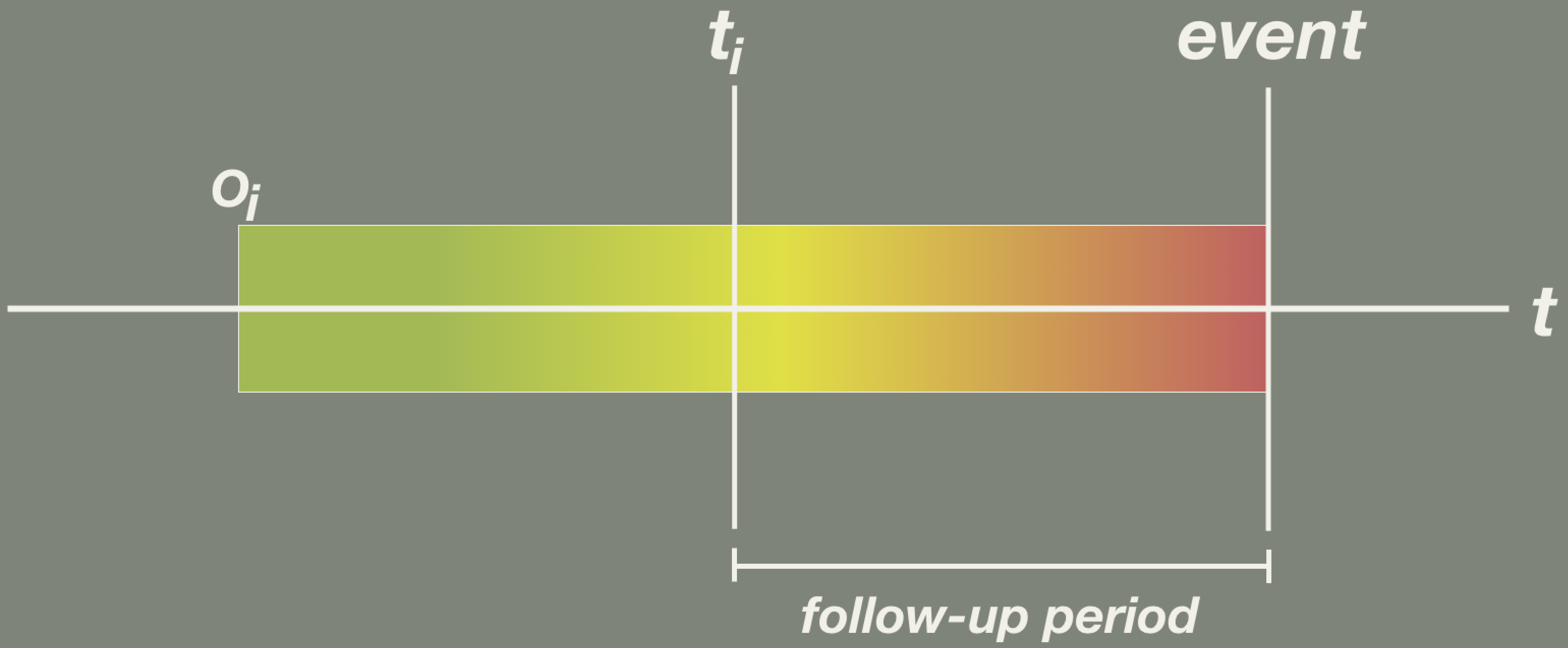
No Randomization

Selection Bias

All the patients were selected in 1 hospital, of 1 geographical region.

LEAD-TIME BIAS

Chronic vs. Infectious Disease

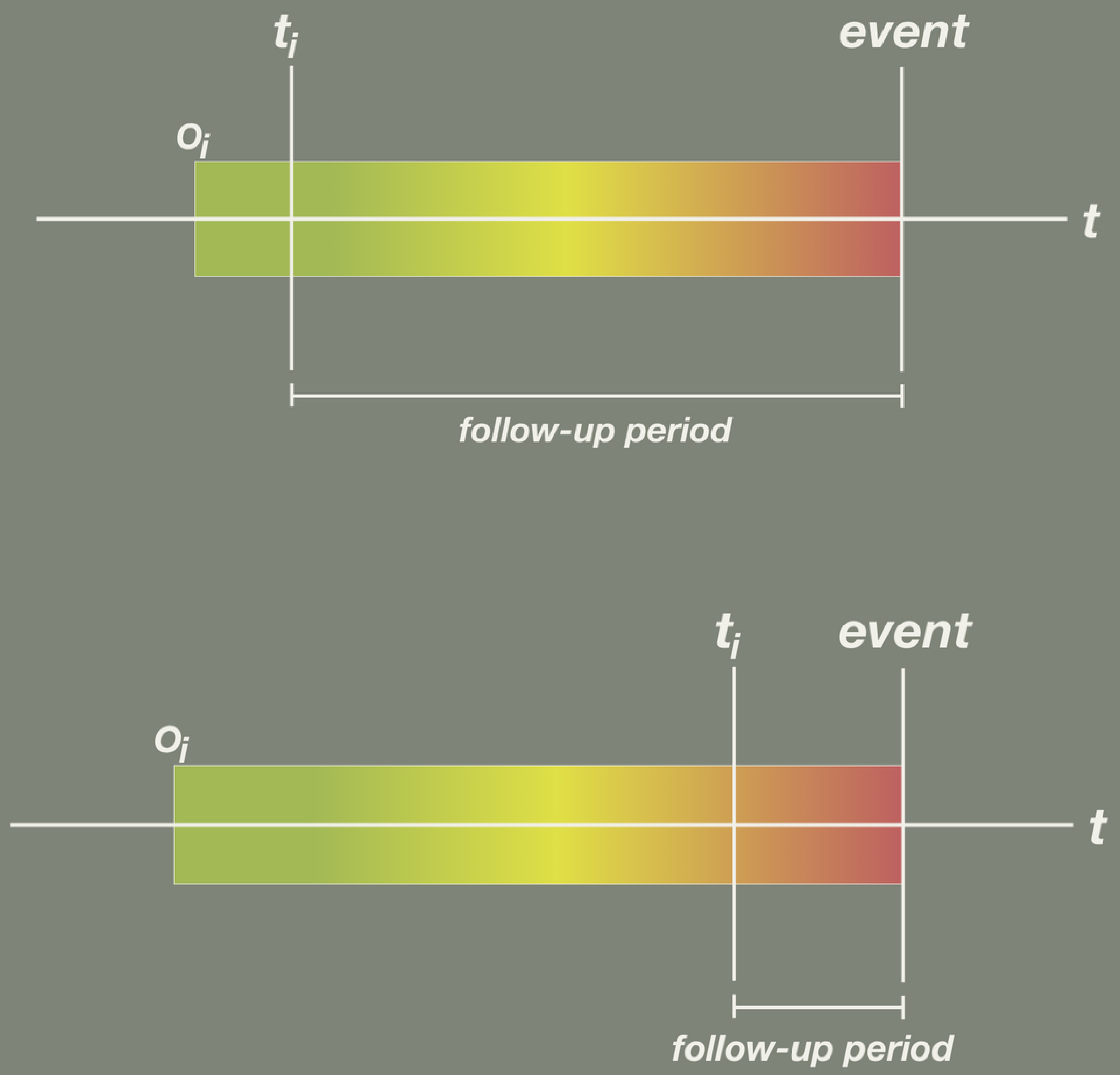


LEAD-TIME BIAS

Unable to determine exact onset period

Variability in diagnoses time and seriousness of disease

Variations can influences survival and hazard accuracy



CONCLUSION


Thank You!