Time Varying Covariates in a Cox Model

Steve Simon

Abstract

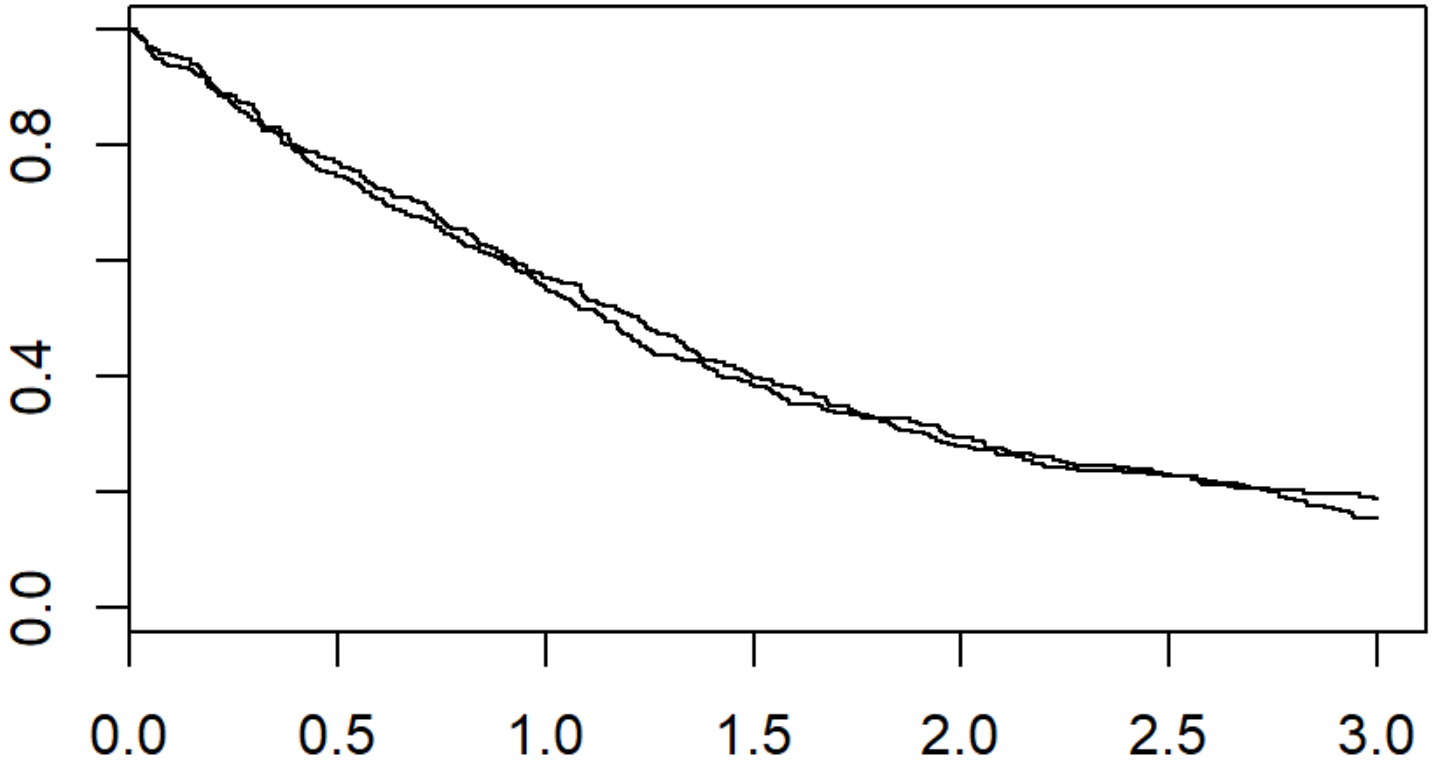
Time varying covariates allow you to account for non-proportional hazards and can model settings where patients switch from one therapy to another. You will code data for time-varying covariates, fit time-varying models, and interpret the results.

Testing the Proportional Hazards Assumption

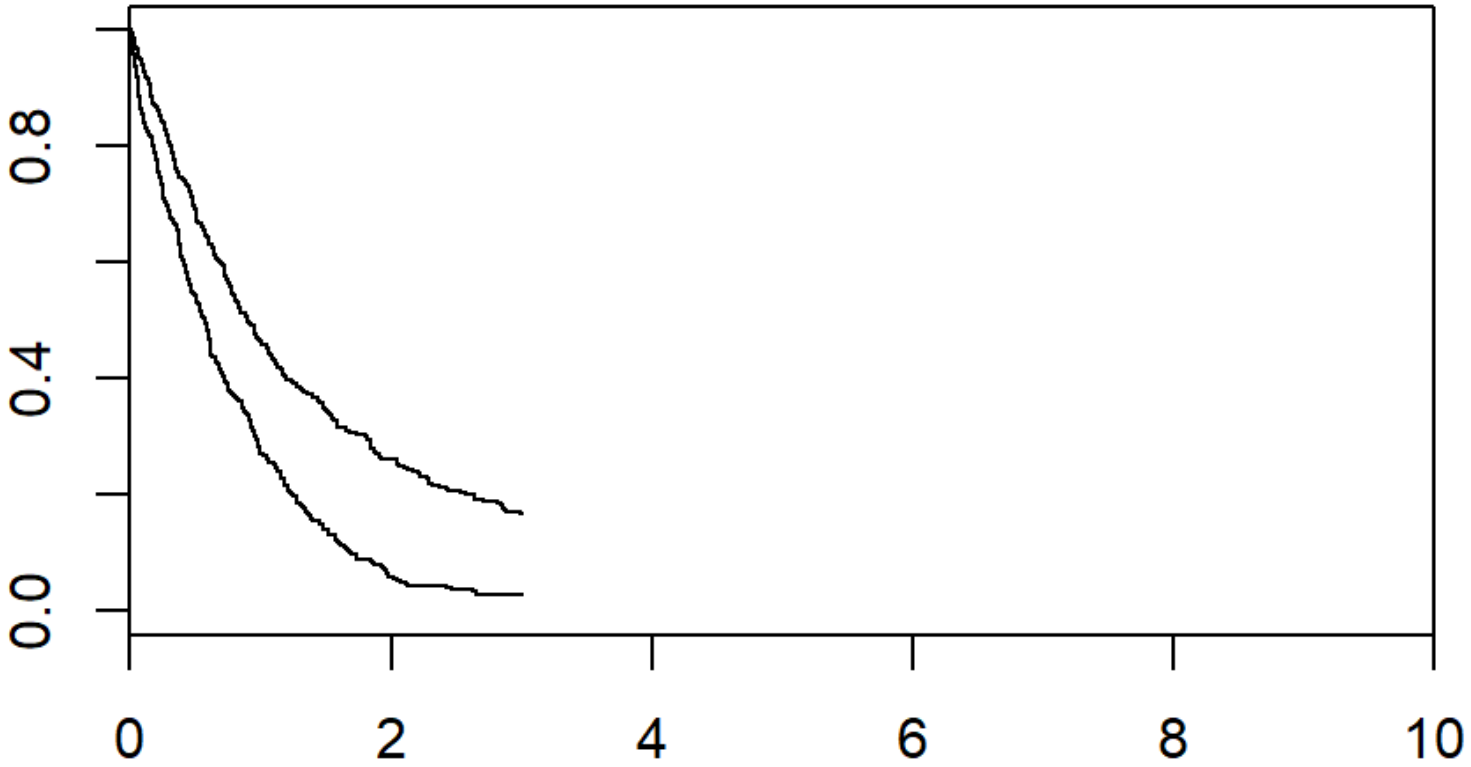
Several approaches

1. Patterns in Kaplan-Meier curves
2. Complementary log-log plot
3. Schoenfeld Residuals
4. Stratified analysis
5. Fit time varying covariates

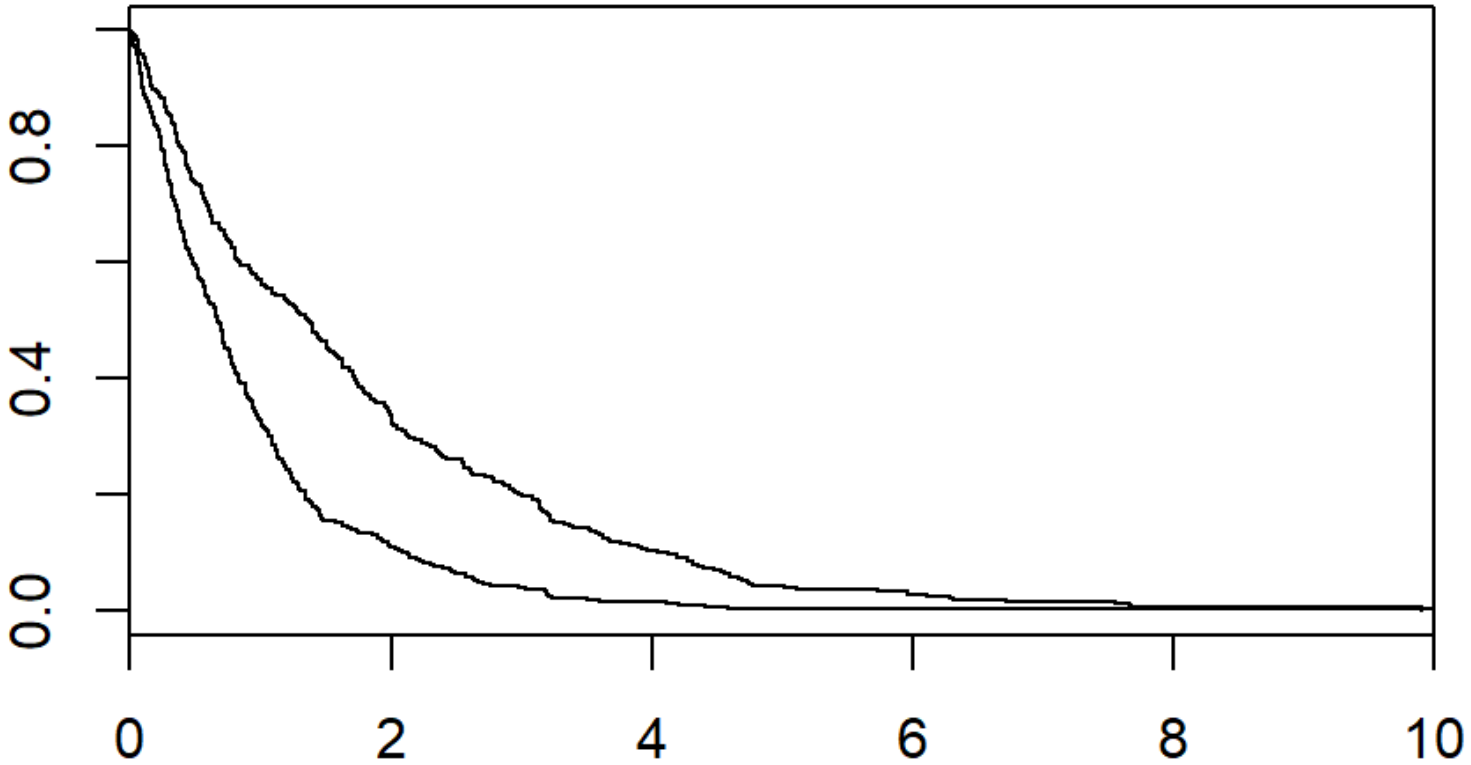
Patterns in the Kaplan-Meier Curves



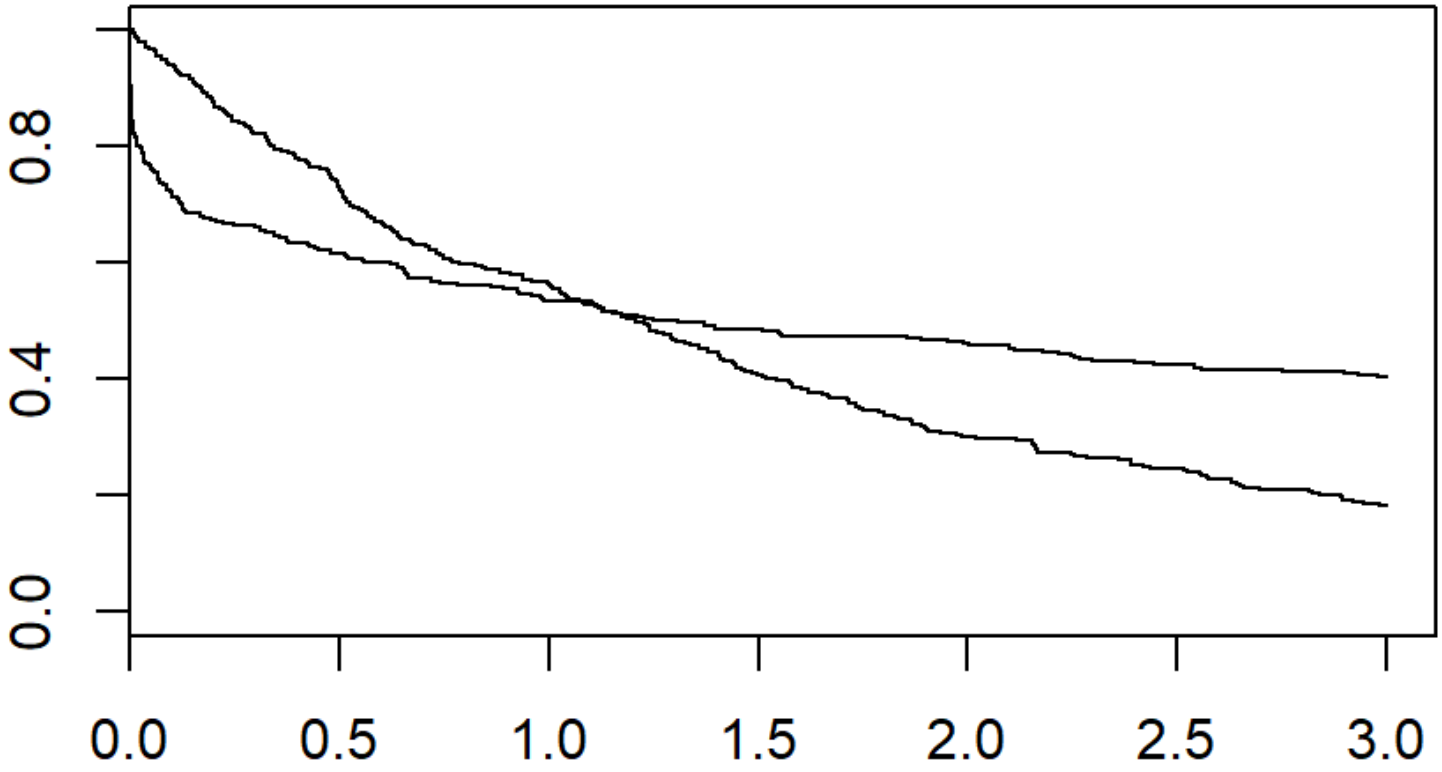
Patterns in the Kaplan-Meier Curves



Patterns in the Kaplan-Meier Curves



Patterns in the Kaplan-Meier Curves



Example: Heroin Clinic Discharge Times

StatSci.org / Home

OzDASL

Methadone Treatment of Heroin Addicts

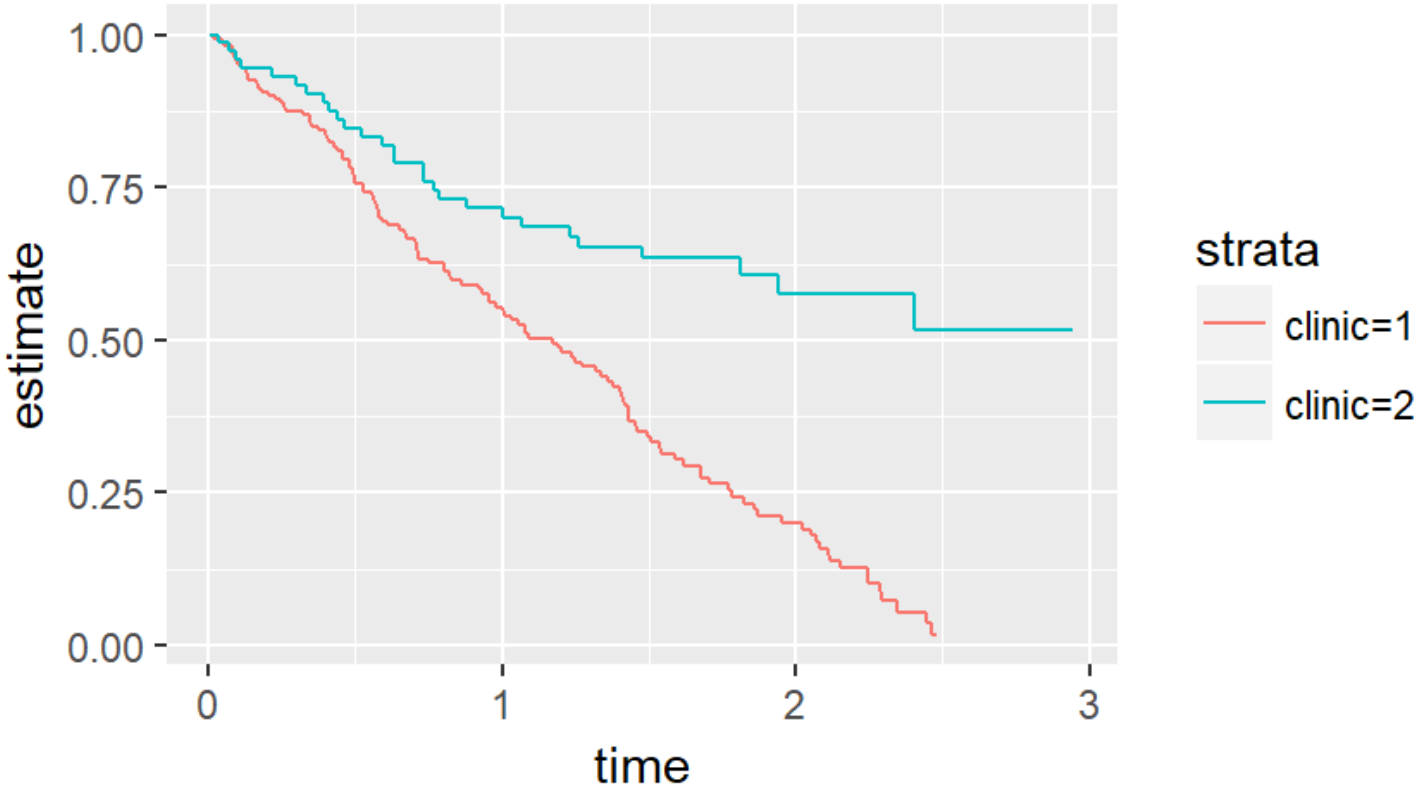
Keywords: censored data, Cox regression.

Description

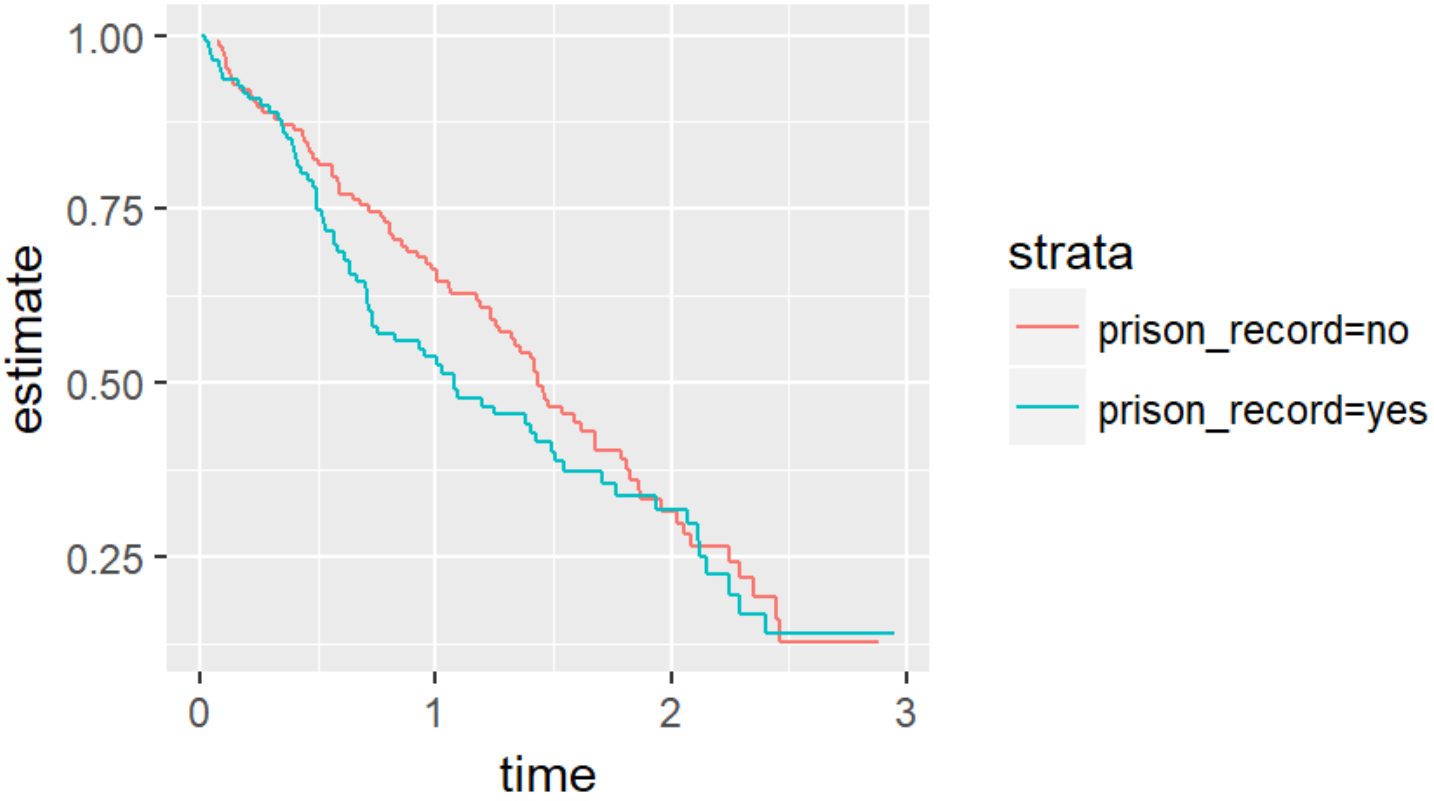
The data are the times, in days, that heroin addicts spend in a clinic. There are two clinics and the covariates are believed to affect the times spent in the clinic by addicts.

Variable	Description
Clinic	1 or 2
Status	0 = still in clinic at end of study (censored) or 1 = departed from clinic
Time	days spent in clinic
Prison	1 = prison record or 0 = no record
Dose	methadone dosage (mg/day)

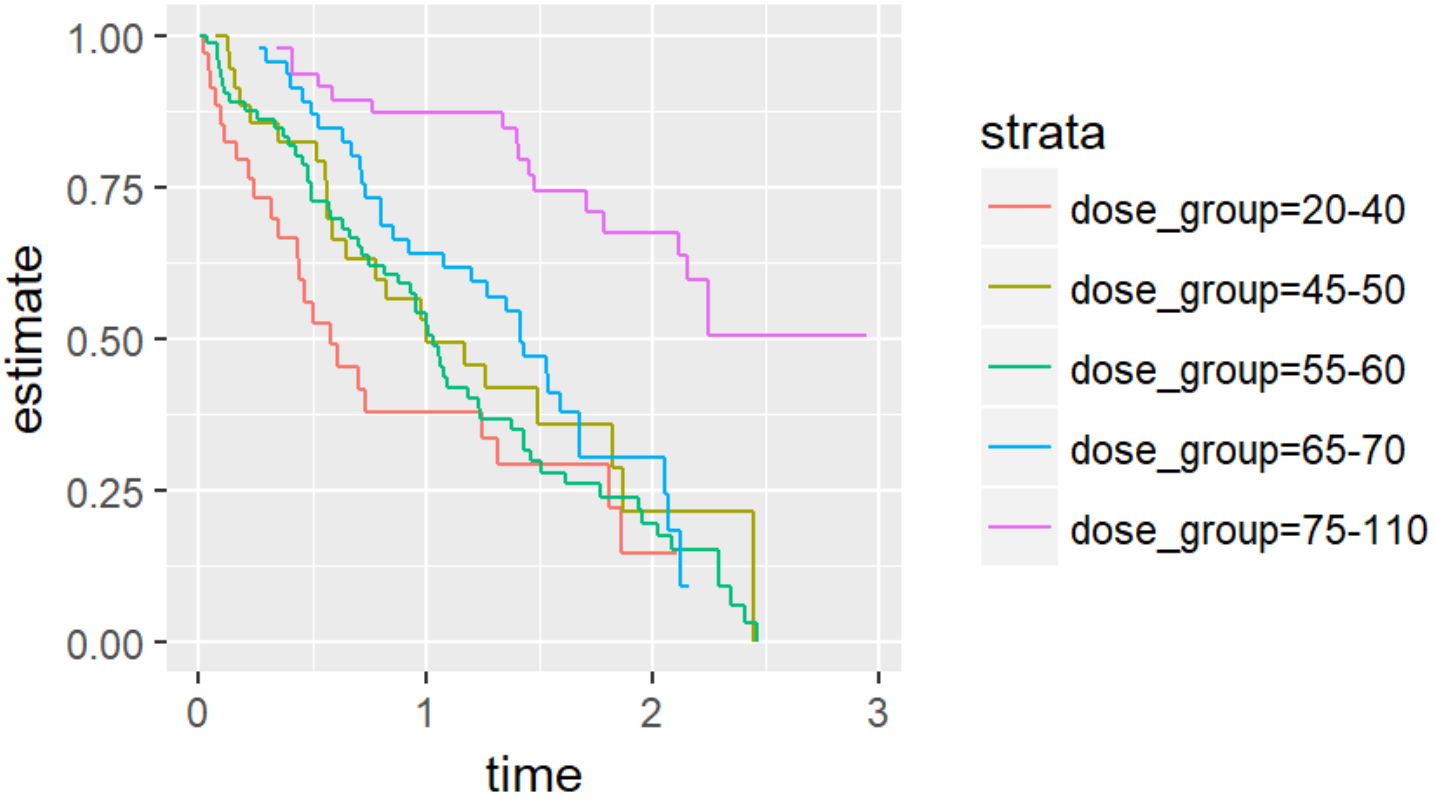
Example: Heroin Data Set



Example: Heroin Data Set



Example: Heroin Data Set



Complementary Log-Log

If the proportional hazards assumption holds then the survival curves for two different groups should be related by

$$S_2(t) = S_1(t)^{hr}$$

Take the logarithm of both sides to get

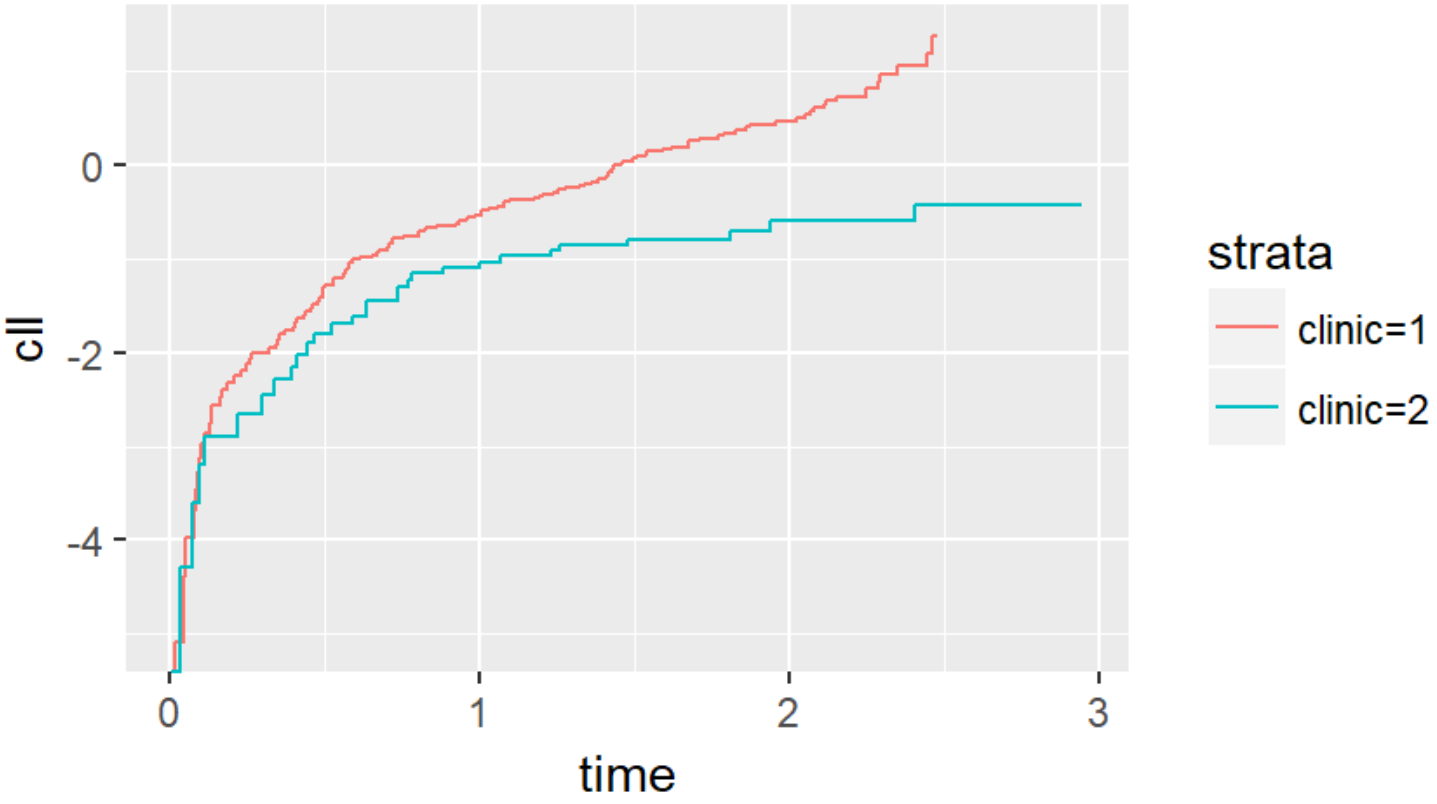
$$log(S_2(t)) = log(S_1(t))hr$$

We'd like to take a second log here, but since S_2 and S_1 are always between 0 and 1, their logarithms would be negative. You have to flip this to a positive value and then take a second logarithm.

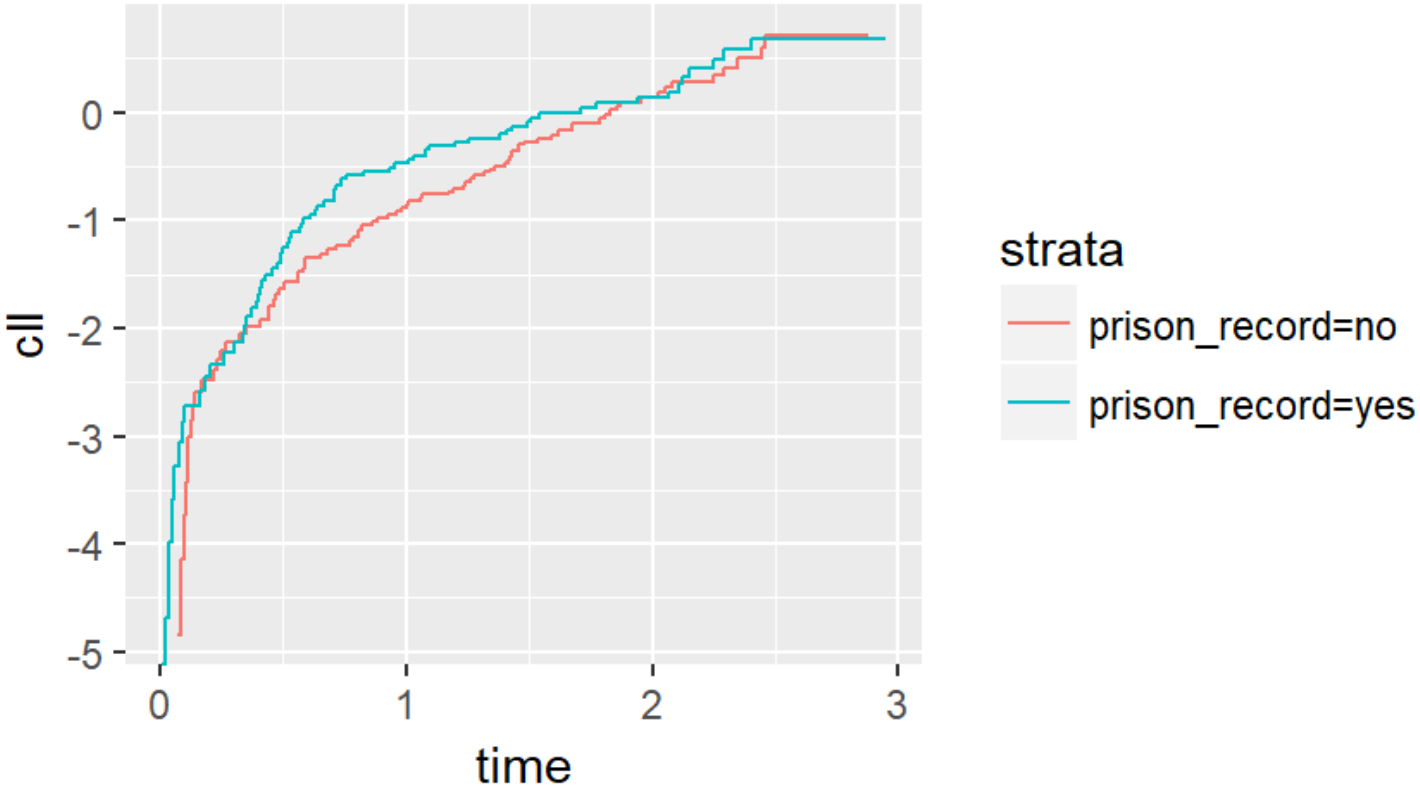
$$log(-log(S_2(t))) = log(-log(S_1(t))) + log(hr)$$

If this transformation, the complementary log-log transformation, produces two curves that are separated by a constant for all values of t, then you have evidence to support the proportional hazards assumption.

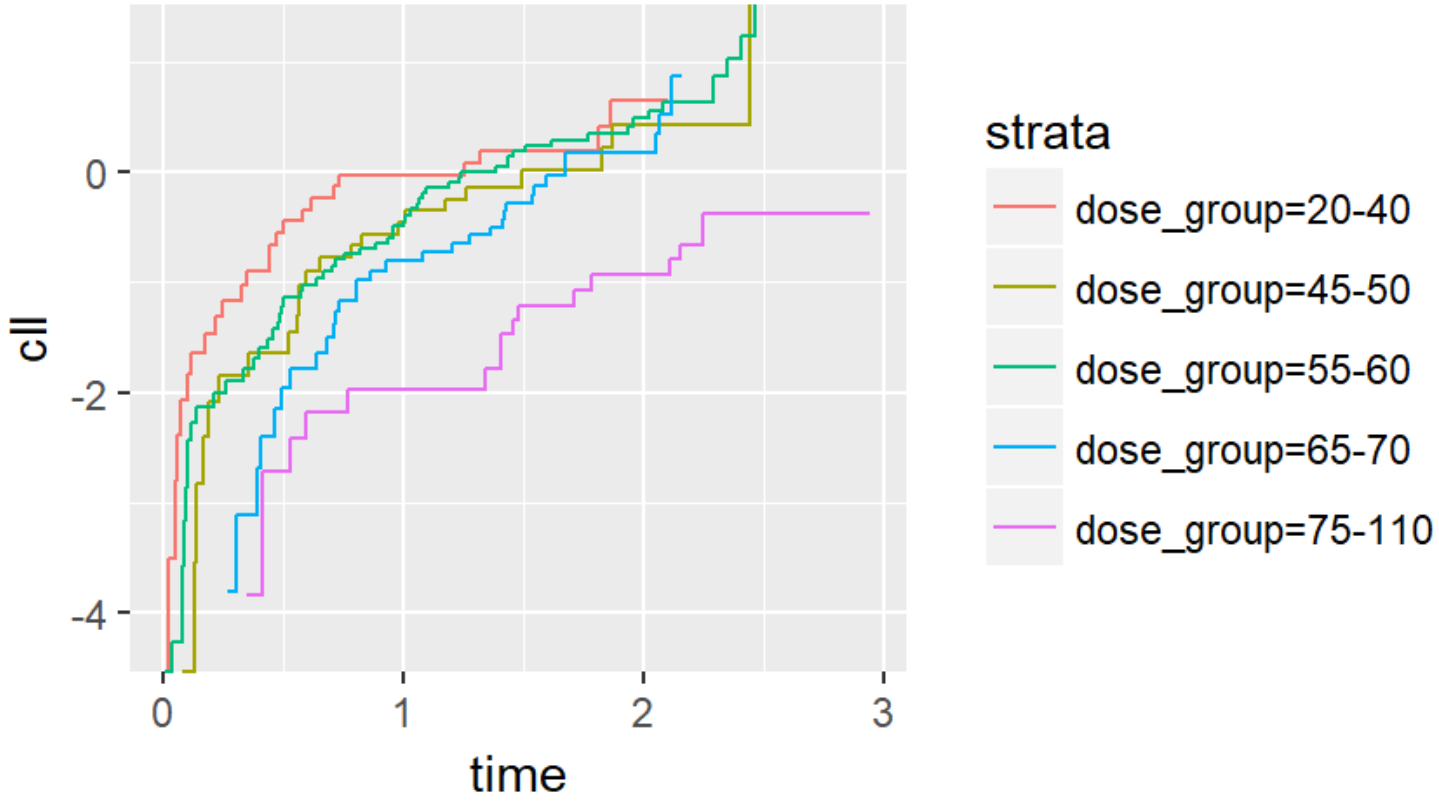
Complementary Log-Log



Complementary Log-Log



Complementary Log-Log



Review of Likelihood

The partial likelihood is

$$l_p = \prod_i \frac{e^{X_{(i)}\beta}}{\sum_{j \in R_{(i)}} e^{X_j \beta}}$$

The log partial likelihood is

$$L_p = \sum_i \left(X_{(i)}\beta - log\left(\sum_{j \in R_{(i)}} e^{X_j \beta} \right) \right)$$

Review of Likelihood

The derivative of the log partial likelihood is

$$\frac{\partial L_p}{\partial \beta} = \sum_{i=1}^{m} \left(X_{(i)} - \bar{X}_i(\beta) \right)$$

where $\bar{X}_i(\beta)$ is a weighted average of all the X's remaining in the risk set and with weights equal to

$$w_{ij} = \frac{e^{X_j \beta}}{\sum_{l \in R_i} e^{X_l \beta}}.$$

A positive derivative implies that we could maximize the log partial likelihood by increasing from the current value of β and a negative derivative implies the opposite.

Schoenfeld Residuals

The slope of the likelihood function is zero at its maximum, which implies that

$$\sum_{i=1}^{m} \left(X_{(i)} - \bar{X}_i(\hat{\beta}) \right) = 0.$$

The individual terms in this sum,

$$X_{(i)} - \hat{X}_{w(i)},$$

are called the Schoenfeld residuals.

Schoenfeld Residuals

Interpretation

Normally, the Schoenfeld residuals are standardized.

A time trend is evidence of a violation of the proportional hazards assumption.

Consider this as evidence of an interaction between time and your independent variable.

Schoenfeld Residuals

Interpretation

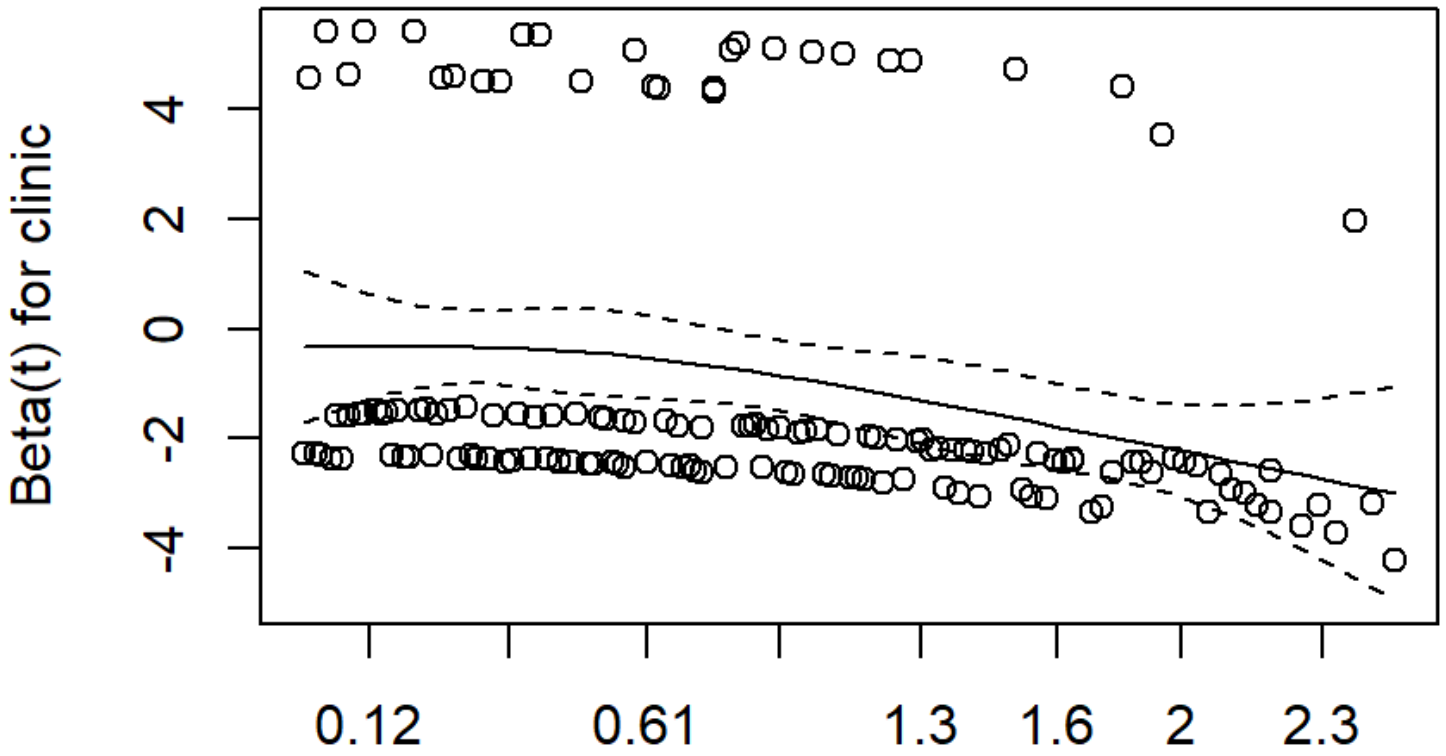
Regions of time where residuals are negative

- the hazard ratio is a bit too large.

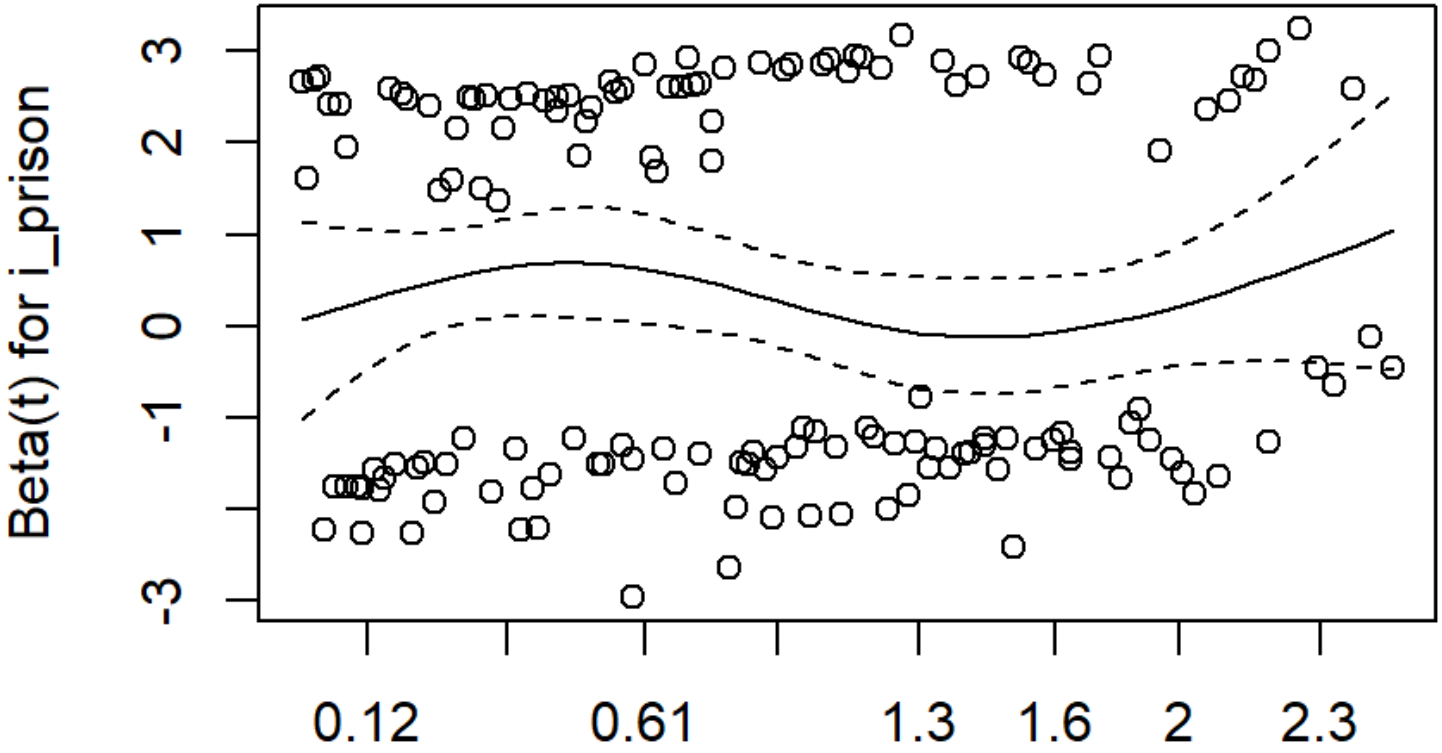
Regions of time where residuals are positive

- the hazard ratio is a bit too small.

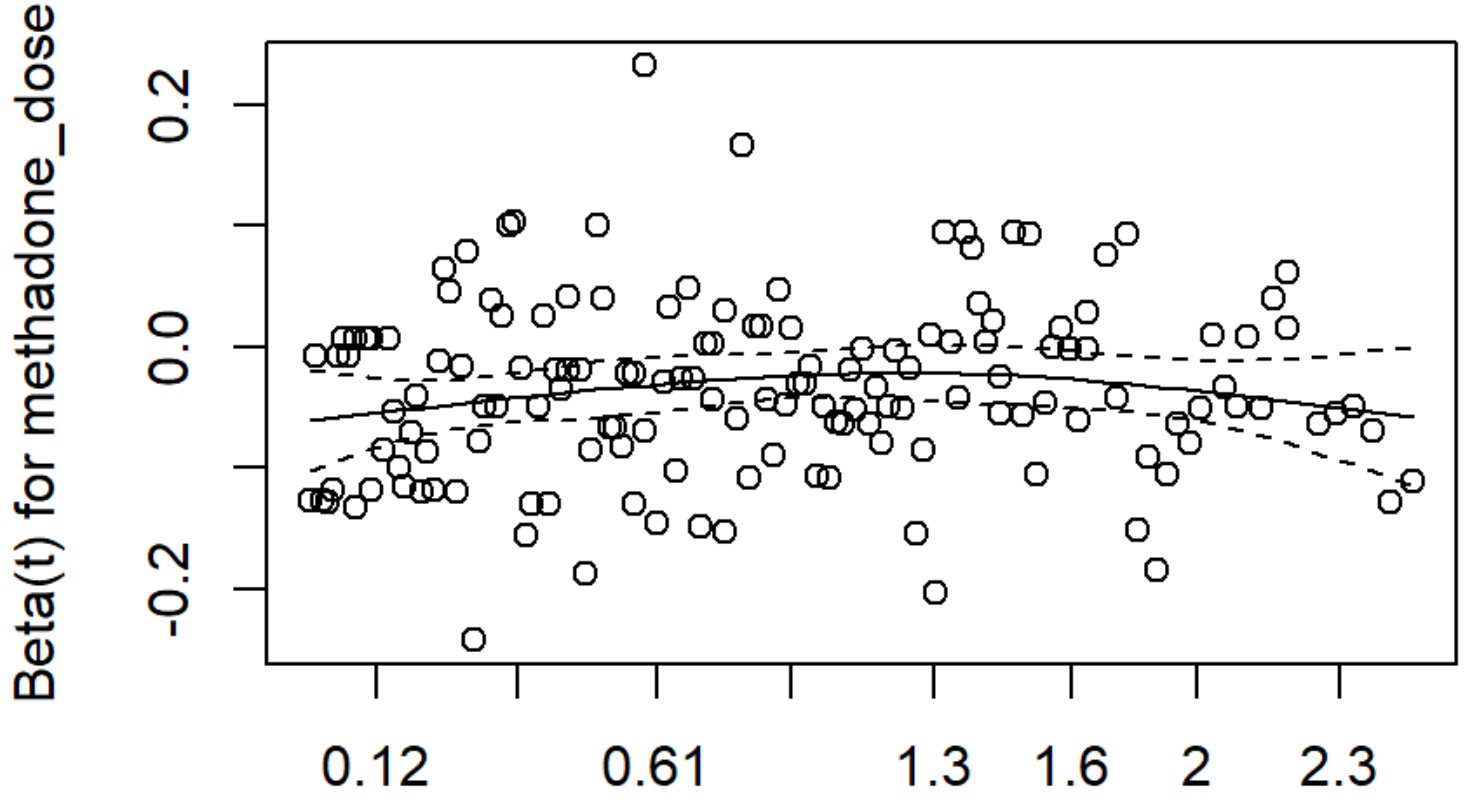
Schoenfeld Residuals



Schoenfeld Residuals



Schoenfeld Residuals



Stratified Models

The easiest solution for time-varying covariates

Stratification creates a separate baseline hazard for each level of your strata.

Only helpful when the time-varying covariate is a nuisance parameter.

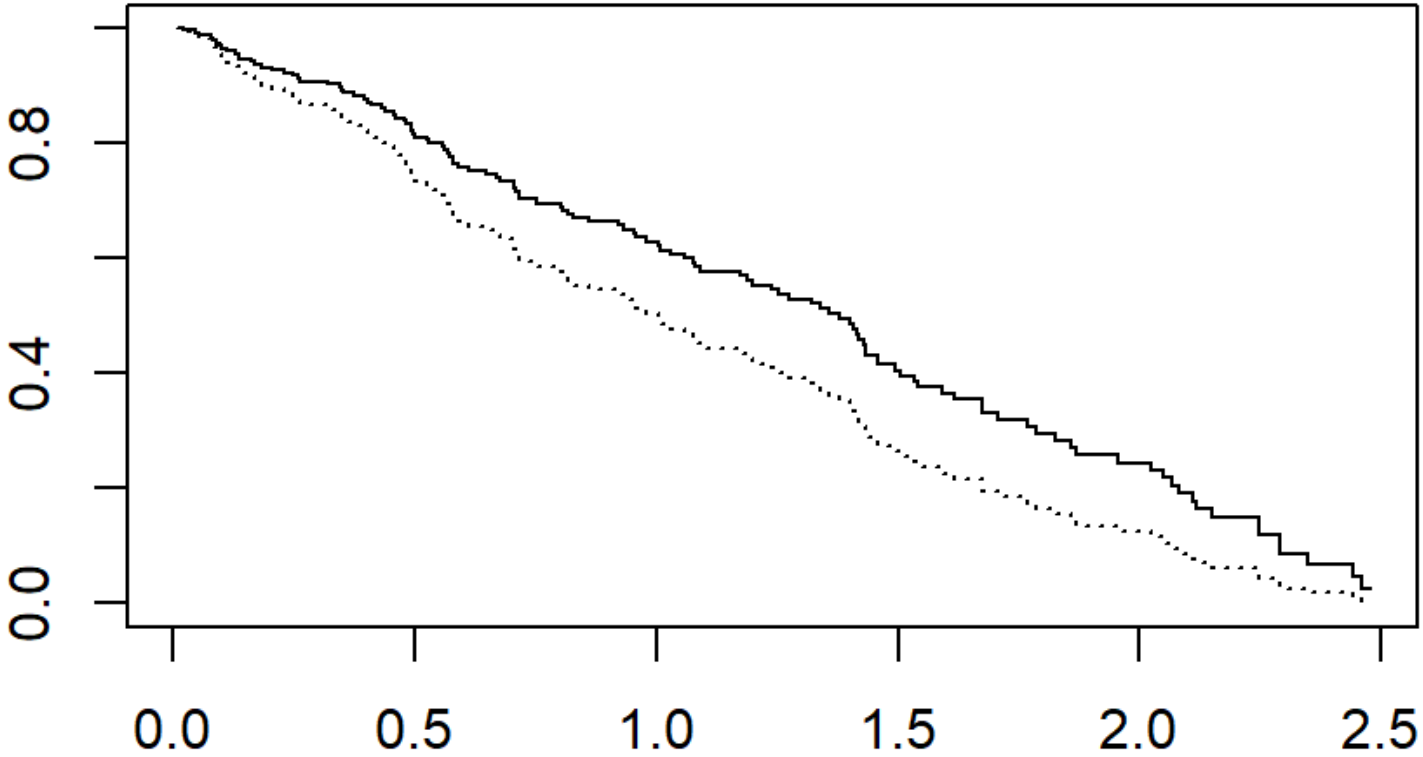
Stratified Models

	coef	exp(coef)	se(coef)	z	p
i_prison	0.38960	1.47640	0.16893	2.31	0.021
methadone_dose	-0.03511	0.96549	0.00646	-5.43	5.6e-08

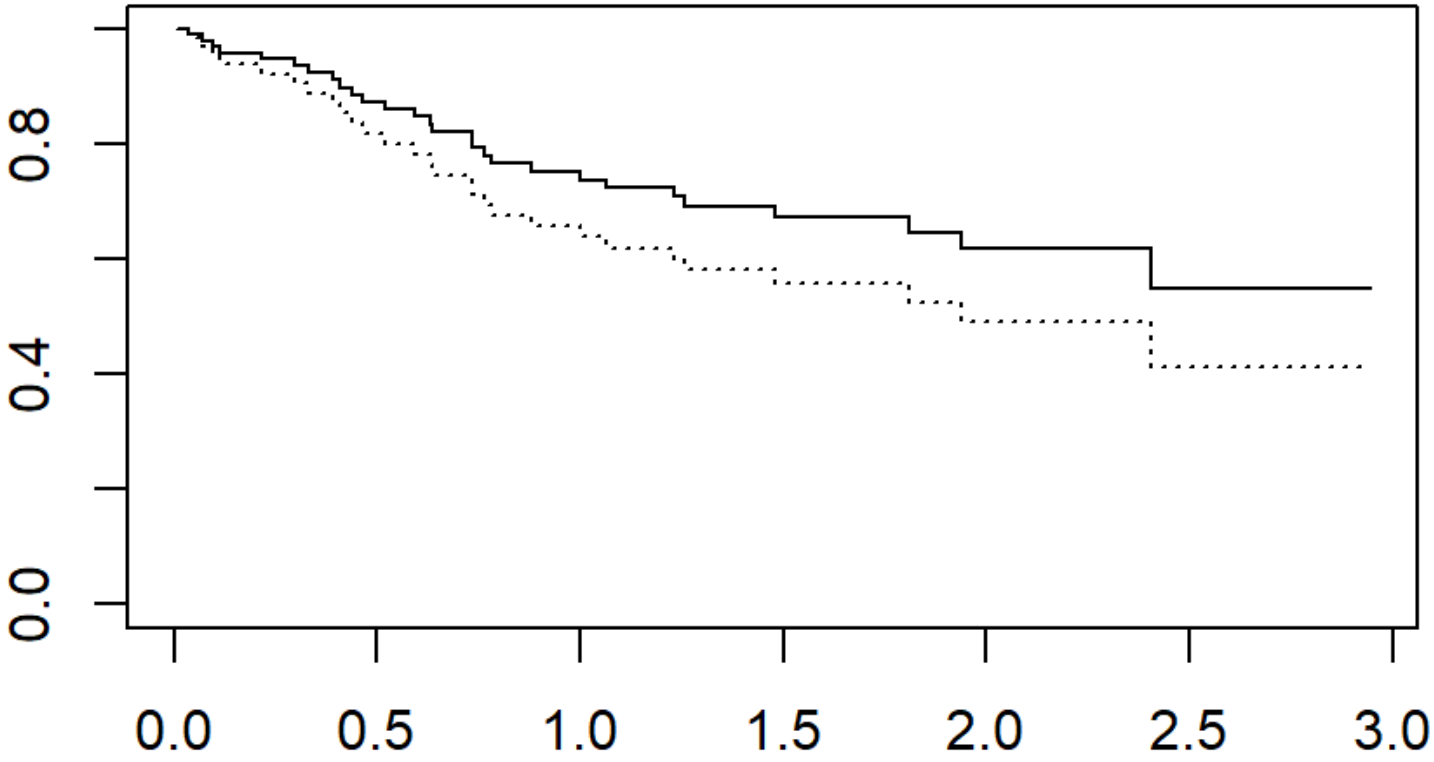
Likelihood ratio test=33.9 on 2 df, p=4.32e-08

n= 238, number of events= 150

Clinic 1 Survivals: Solid Line is prison_record=no



Clinic 2 Survivals: Solid Line is prison_record=no



Time-Varying Covariates

Interaction with time

You can address problems with non-proportional hazards by creating an interaction involving time.

- Also consider interactions involving log(time).

Time-Varying Covariates

Interaction with time, heroin dataset

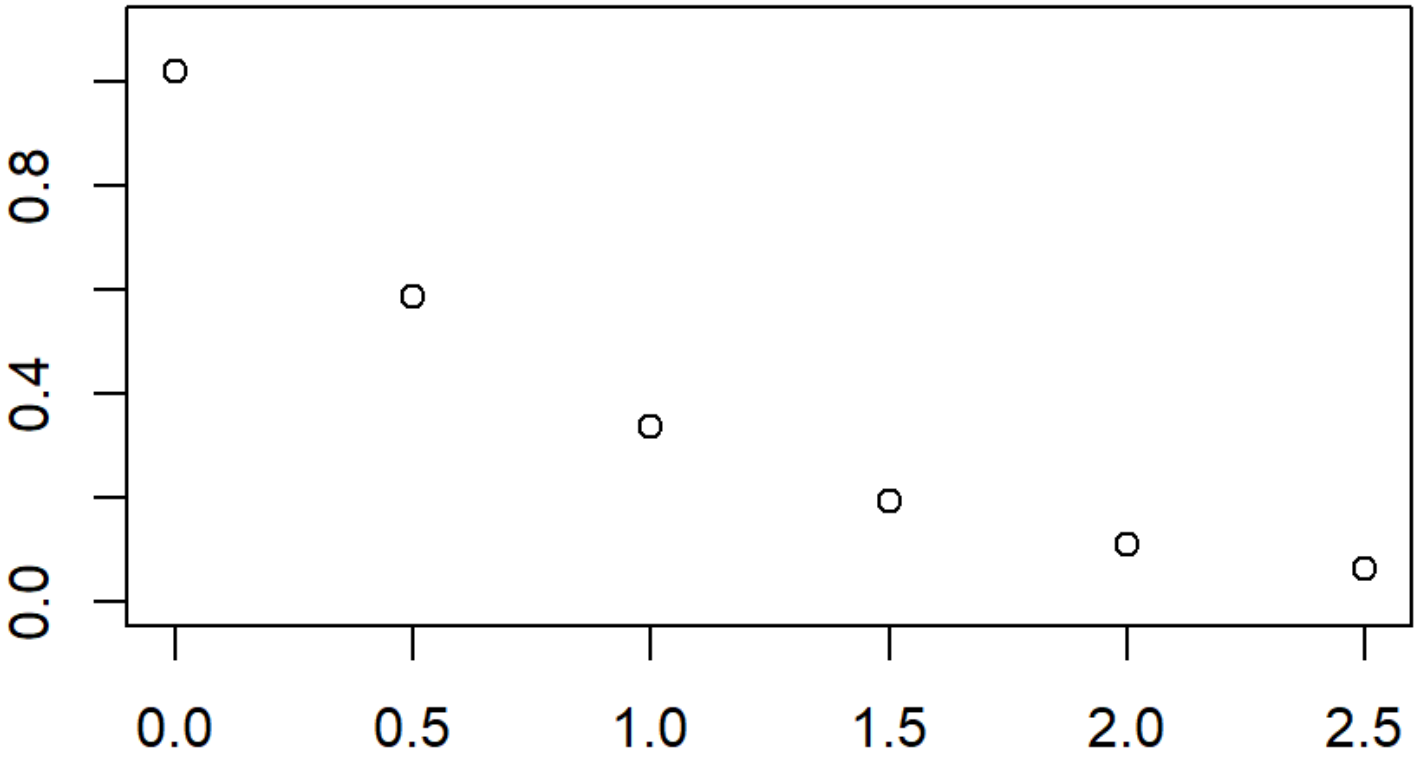
```
                    coef exp(coef) se(coef)     z        p
clinic           0.01940   1.01958  0.34717  0.06   0.9554
tt(clinic)      -1.10331   0.33177  0.34528 -3.20   0.0014
i_prison         0.38997   1.47693  0.16889  2.31   0.0209
methadone_dose -0.03519   0.96543  0.00644 -5.46 4.7e-08

Likelihood ratio test=76.1  on 4 df, p=1.11e-15
n= 238, number of events= 150
```

Time-Varying Covariates

Interaction with time, heroin dataset

	beta_clinic	beta_interaction	t	log_hr	hr
1	0.019	-1.103	0.0	0.019	1.020
2	0.019	-1.103	0.5	-0.532	0.587
3	0.019	-1.103	1.0	-1.084	0.338
4	0.019	-1.103	1.5	-1.636	0.195
5	0.019	-1.103	2.0	-2.187	0.112
6	0.019	-1.103	2.5	-2.739	0.065

Time-Varying Covariates



Stanford Transplant Data

Data pp 45-50 | Cite as

Stanford Heart Transplant Data

Authors | Authors and affiliations

D. F. Andrews, A. M. Herzberg

Chapter

1.6k
Downloads

Part of the Springer Series in Statistics book series (SSS)

Abstract

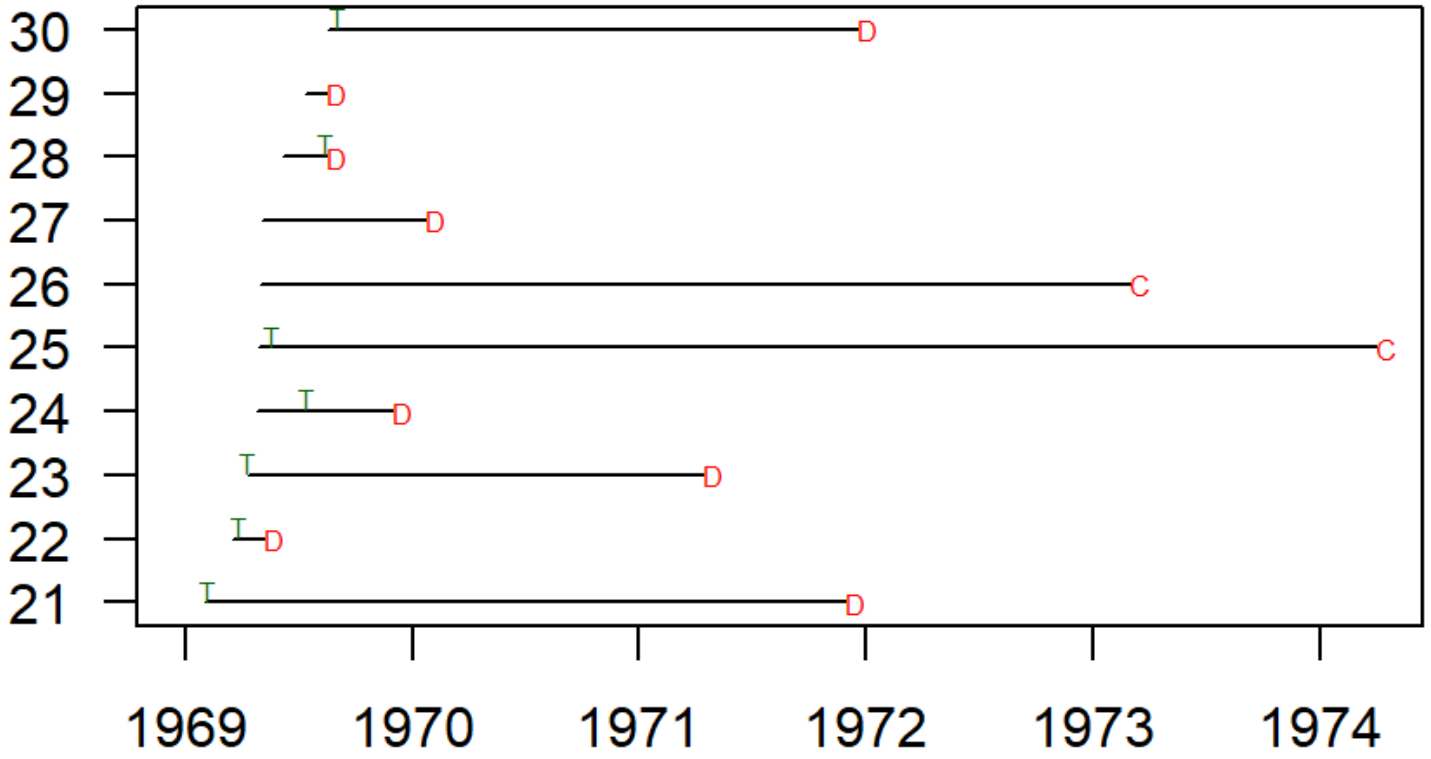
The Stanford Heart Transplantation Program began in October 1967. Patients are admitted to the program after review by a committee, and then they wait for donor hearts to become available. While waiting, some may die or be transferred out of the program, but most receive a transplant. The cut-off date for the data presented in Table 7.1 was in February 1980, and by that time 184 patients had received a transplant.

Stanford Transplant Data

jasa: original data

birth.dt:	birth date
accept.dt:	acceptance into program
tx.date:	transplant date
fu.date:	end of followup
fustat:	dead or alive
surgery:	prior bypass surgery
age:	age (in years)
futime:	followup time
wait.time:	time before transplant
transplant:	transplant indicator
mismatch:	mismatch score
hla.a2:	particular type of mismatch
mscore:	another mismatch score
reject:	rejection occurred

Stanford Transplant Data, Patients 21-30



Stanford Transplant Data, Naive Analysis

	coef	exp(coef)	se(coef)	z	p
transplant	-1.7171	0.1796	0.2785	-6.16	7.1e-10
age	0.0589	1.0607	0.0150	3.91	9.1e-05
surgery	-0.4190	0.6577	0.3712	-1.13	0.26

Likelihood ratio test=45.9 on 3 df, p=6.11e-10

n= 103, number of events= 75

Stanford Transplant Data

Patient #21

```
id  accept.dt    tx.date    fu.date fustat transplant
21 1969-02-01 1969-02-08 1971-11-29      1           1

id   time1     time2 event transplant
21  7 days 1031 days     D           1

id start stop event transplant
21     0    7     0           0
21     7 1031     1           1
```

Stanford Transplant Data

Patient #22

```
id  accept.dt    tx.date    fu.date fustat transplant
22 1969-03-18 1969-03-29 1969-05-07      1           1

id   time1      time2 event transplant
22 11 days   50 days     D           1

id start stop event transplant
22     0   11     0           0
22    11   50     1           1
```

Stanford Transplant Data

Patient #23

```
id  accept.dt    tx.date    fu.date fustat transplant
23 1969-04-11 1969-04-13 1971-04-13      1           1

id    time1      time2 event transplant
23   2 days   732 days     D           1

id start stop event transplant
23     0    2     0           0
23     2  732     1           1
```

Stanford Transplant Data

Patient #24

```
id  accept.dt    tx.date    fu.date fustat transplant
24 1969-04-25 1969-07-16 1969-11-29      1           1

id   time1     time2 event transplant
24 82 days  218 days     D           1

id start stop event transplant
24     0   82     0           0
24    82  218     1           1
```

Stanford Transplant Data

Patient #25

```
id  accept.dt    tx.date    fu.date fustat transplant
25 1969-04-28 1969-05-22 1974-04-01      0          1

id   time1     time2 event transplant
25 24 days 1799 days     C          1

id start stop event transplant
25     0   24     0          0
25    24 1799     0          1
```

Stanford Transplant Data

Patient #26

```
id  accept.dt    tx.date    fu.date fustat transplant
26 1969-05-01       <NA> 1973-03-01      0           0

id   time1     time2 event transplant
26 NA days 1400 days     C           0

id start stop event transplant
26     0 1400     0           0
```

Stanford Transplant Data

Patient #27

```
id  accept.dt    tx.date    fu.date fustat transplant
27 1969-05-04       <NA> 1970-01-21      1           0

id   time1     time2 event transplant
27 NA days  262 days     D           0

id start stop event transplant
27     0  262     1           0
```

Stanford Transplant Data

Patient #28

```
id  accept.dt    tx.date    fu.date fustat transplant
28 1969-06-07 1969-08-16 1969-08-17      1           1

id   time1     time2 event transplant
28 70 days   71 days     D           1

id start stop event transplant
28     0   70     0           0
28    70   71     1           1
```

Stanford Transplant Data

Patient #29

```
id  accept.dt    tx.date    fu.date fustat transplant
29 1969-07-14       <NA> 1969-08-17      1           0

id   time1     time2 event transplant
29 NA days   34 days     D           0

id start stop event transplant
29     0   34     1           0
```

Stanford Transplant Data

Patient #30

```
id  accept.dt    tx.date    fu.date fustat transplant
30 1969-08-19 1969-09-03 1971-12-18      1           1

id   time1      time2 event transplant
30 15 days  851 days     D           1

id start stop event transplant
30     0   15     0           0
30    15  851     1           1
```

Stanford Transplant Data, Time-varying Model

	coef	exp(coef)	se(coef)	z	p
transplant	0.0141	1.0142	0.3082	0.05	0.964
age	0.0306	1.0310	0.0139	2.20	0.028
surgery	-0.7733	0.4615	0.3597	-2.15	0.032

Likelihood ratio test=10.7 on 3 df, p=0.0134

n= 170, number of events= 75

Conclusion

What have you learned today?

1. There are several ways to check for/control for time varying covariates: patterns in Kaplan-Meier curves, the complementary log-log plot, Schoenfeld Residuals, stratified analysis, and fitting time varying covariates.
2. You can fit time-varying covariates using start/stop coding or a time transfer function.