# Fixed Time Survival Analysis

Weisi Chen

2025-03-13

## Table of contents



```
# Load the needed packages
library(ggplot2)
library(dplyr)
library(lubridate)
library(survival)
library(ggsurvfit)
library(gtsummary)
library(here)
library(survminer)
library(broom)
library(forestploter)
library(tidyr)
```

```r
# Load example data
df <- colon
```

This analysis focus on survival following the chemotherapy treatment for colon cancer.

### About the sample data

The data come from the `colon` dataset, available from the *survival* package. These data include information from a clinical trial on the effectiveness of two different types of chemotherapy (levamisole and levamisole+5-fluorouracil) compared to controls (i.e. no chemotherapy treatment) on survival from stage B/C colon cancer.

There are two rows per person in the dataset, one for cancer recurrence and one for death, indicated by the event type (`etype`) variable (`etype==1` corresponds to recurrence and `etype==2` to death). In analysis below, I only focus on analysing death as an outcome.

Note: there is some incomplete values on the `differ` variable, for simplicity, in the below analysis, I drop those incomplete values.

Some important variables:

`rx`: Treatment - Obs(ervation), Lev(amisole), Lev(amisole)+5-FU `sex`: 1=male `age`: in years `obstruct`: obstruction of colon by tumour `perfor`: perforation of colon `adhere`: adherence to nearby organs `nodes`: number of lymph nodes with detectable cancer `time`: days until event or censoring `status`: censoring status `differ`: differentiation of tumour (1=well, 2=moderate, 3=poor) `extent`: Extent of local spread (1=submucosa, 2=muscle, 3=serosa, 4=contiguous structures) `surg`: time from surgery to registration (0=short, 1=long) `node4`: more than 4 positive lymph nodes `etype`: event type: 1=recurrence,2=death

### Data cleaning

- Filter records with death outcome
- Drop incomplete values on the `diff` variable
- Label the `diff` and `extent` variables
- Stratify the `age` variable

```r
df1 <- df %>%
  filter(etype == 2) %>% # Filter to deaths
  filter(!is.na(differ)) %>%
  mutate(
    differF = factor(differ, levels = 1:3, labels = c("well","moderate","poor")),
    extentF = factor(extent, levels = 1:4, labels = c("submucosa","muscle","serosa","contigu
```

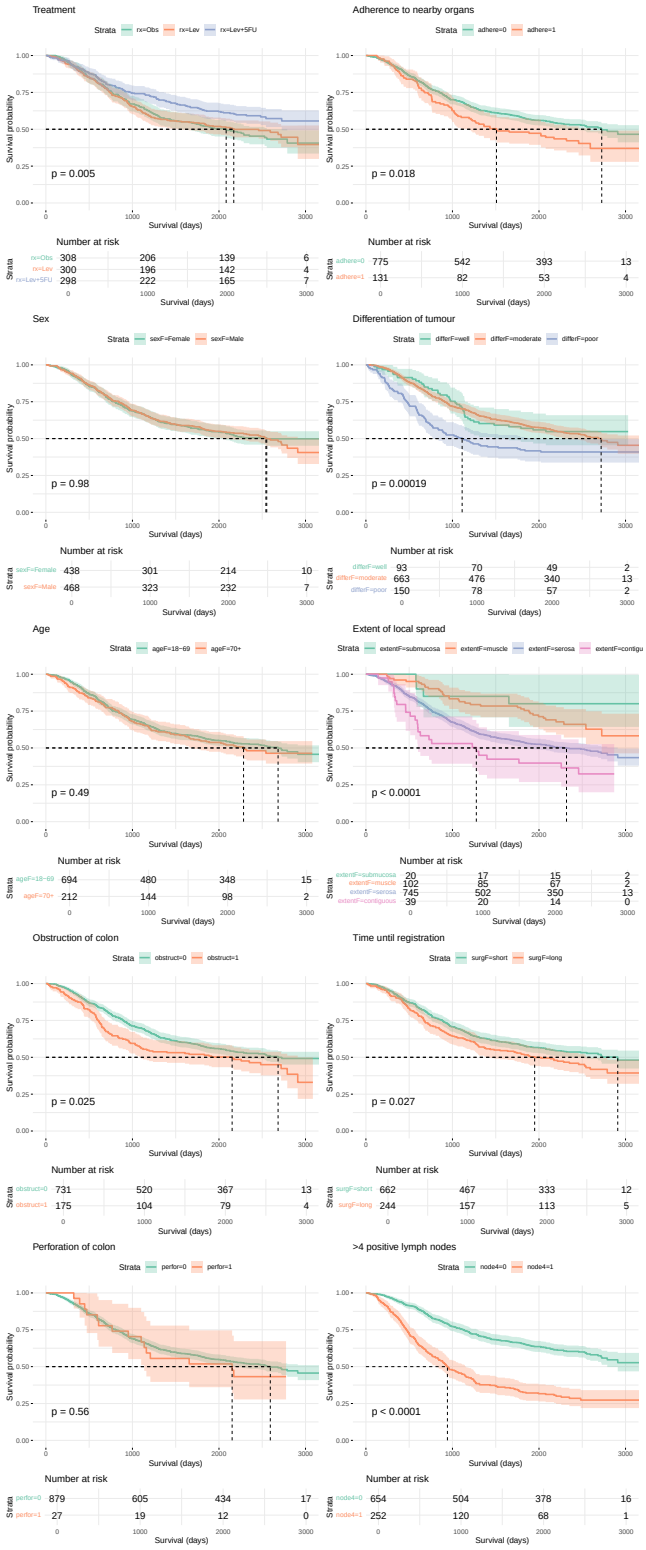
```
    ageF = factor(ifelse(age<70, 1, 2), levels = 1:2, labels = c('18-69', '70+')),
    sexF = factor(sex, levels = 0:1, labels = c("Female","Male")),
    surgF = factor(surg, levels = 0:1, labels = c("short", "long"))
  )
```

## EDA

**Table 1: Summary of demographics and disease status by treatment group**

**Figure 1: Kaplan-Meier suvival plots for key predictors**

| | Obs | Lev | Lev+5FU | |
|---|---|---|---|---|
| **Total (column denominator)** | 308 (100%) | 300 (100%) | 298 (100%) | 90 |
| **Age** | | | | |
| Mean, (SD) | 59, (12) | 60, (12) | 60, (12) | 6 |
| Median, (IQR) | 61, (53, 68) | 61, (53, 69) | 62, (52, 70) | 61 |
| Range | 18, 85 | 27, 83 | 26, 81 | |
| **Sex** | | | | |
| Female | 146 (47%) | 131 (44%) | 161 (54%) | 43 |
| Male | 162 (53%) | 169 (56%) | 137 (46%) | 46 |
| **Obstruction of colon** | 62 (20%) | 59 (20%) | 54 (18%) | 17 |
| **Perforation of colon** | 9 (3%) | 10 (3%) | 8 (3%) | 2 |
| **Adherence to nearby organs** | 45 (15%) | 47 (16%) | 39 (13%) | 13 |
| **Differentiation of tumour** | | | | |
| well | 27 (9%) | 37 (12%) | 29 (10%) | 9 |
| moderate | 229 (74%) | 219 (73%) | 215 (72%) | 66 |
| poor | 52 (17%) | 44 (15%) | 54 (18%) | 15 |
| **Extent of local spread** | | | | |
| submucosa | 7 (2%) | 3 (1%) | 10 (3%) | 2 |
| muscle | 36 (12%) | 35 (12%) | 31 (10%) | 10 |
| serosa | 248 (81%) | 251 (84%) | 246 (83%) | 74 |
| contiguous | 17 (6%) | 11 (4%) | 11 (4%) | 3 |
| **Time until registration** | | | | |
| short | 218 (71%) | 222 (74%) | 222 (74%) | 66 |
| long | 90 (29%) | 78 (26%) | 76 (26%) | 24 |
| **>4 positive lymph nodes** | 87 (28%) | 87 (29%) | 78 (26%) | 25 |
| **Days until death/censored** | | | | |
| Mean, (SD) | 1,599, (857) | 1,615, (894) | 1,796, (866) | 1,6 |
| Median, (IQR) | 1,854, (759, 2,274) | 1,910, (741, 2,383) | 2,092, (977, 2,472) | 1,978, |
| Range | 113, 3,214 | 24, 3,329 | 23, 3,309 | 2 |
| **Death** | 165 (54%) | 154 (51%) | 122 (41%) | 44 |


Treatment
Strata rx=Obs rx=Lev rx=Lev+5FU
Survival probability
p = 0.005
Survival (days)
Number at risk
rx=Obs 308 206 139 6
rx=Lev 300 196 142 4
rx=Lev+5FU 298 222 165 7
Adherence to nearby organs
Strata adhere=0 adhere=1
p = 0.018
adhere=0 775 542 393 13
adhere=1 131 82 53 4
Sex
Strata sexF=Female sexF=Male
p = 0.98
sexF=Female 438 301 214 10
sexF=Male 468 323 232 7
Differentiation of tumour
Strata diffeF=well diffeF=moderate diffeF=poor
p = 0.00019
diffeF=well 93 70 49 2
diffeF=moderate 663 476 340 13
diffeF=poor 150 78 57 2
Age
Strata ageF=18−69 ageF=70+
p = 0.49
ageF=18−69 694 480 348 15
ageF=70+ 212 144 98 2
Extent of local spread
Strata extentF=submucosa extentF=muscle extentF=serosa extentF=contiguu
p < 0.0001
extentF=submucosa 20 17 15 2
extentF=muscle 102 85 67 2
extentF=serosa 745 502 350 13
extentF=contiguous 39 20 14 0
Obstruction of colon
Strata obstruct=0 obstruct=1
p = 0.025
obstruct=0 731 520 367 13
obstruct=1 175 104 79 4
Time until registration
Strata surgF=short surgF=long
p = 0.027
surgF=short 662 467 333 12
surgF=long 244 157 113 5
Perforation of colon
Strata perfor=0 perfor=1
p = 0.56
perfor=0 879 605 434 17
perfor=1 27 19 12 0
>4 positive lymph nodes
Strata node4=0 node4=1
p < 0.0001
node4=0 654 504 378 16
node4=1 252 120 68 1
Survival (days)

| | Univariable | | | Multivariable | | |
|---|---|---|---|---|---|---|
| **Characteristic** | **HR[1]** | **95% CI[1]** | **p-value** | **HR[1]** | **95% CI[1]** | **p-value** |
| Treatment | | | | | | |
| Obs | — | — | | — | — | |
| Lev | 0.96 | 0.77, 1.19 | 0.7 | 0.98 | 0.79, 1.23 | 0.9 |
| Lev+5FU | 0.70 | 0.55, 0.88 | 0.002 | 0.70 | 0.55, 0.88 | 0.003 |
| Sex | 1.00 | 0.83, 1.21 | >0.9 | | | |
| Age (70+ years) | 1.08 | 0.87, 1.34 | 0.5 | | | |
| Obstruction of colon | 1.30 | 1.03, 1.63 | 0.025 | 1.29 | 1.03, 1.63 | 0.028 |
| Perforation of colon | 1.17 | 0.70, 1.95 | 0.6 | | | |
| Adherence to nearby organs | 1.35 | 1.05, 1.73 | 0.018 | 1.19 | 0.92, 1.53 | 0.2 |
| Differentiation of tumour | | | | | | |
| well | — | — | | — | — | |
| moderate | 1.05 | 0.76, 1.45 | 0.8 | 0.93 | 0.67, 1.29 | 0.7 |
| poor | 1.70 | 1.18, 2.46 | 0.005 | 1.36 | 0.93, 1.97 | 0.11 |
| Extent of local spread | | | | | | |
| submucosa | — | — | | — | — | |
| muscle | 1.83 | 0.65, 5.15 | 0.3 | 1.34 | 0.47, 3.79 | 0.6 |
| serosa | 3.25 | 1.21, 8.71 | 0.019 | 2.18 | 0.81, 5.87 | 0.12 |
| contiguous | 5.04 | 1.75, 14.5 | 0.003 | 3.07 | 1.06, 8.94 | 0.039 |
| Time until registration | 1.26 | 1.03, 1.54 | 0.027 | 1.27 | 1.03, 1.56 | 0.022 |
| >4 positive lymph nodes | 2.58 | 2.13, 3.12 | <0.001 | 2.49 | 2.05, 3.02 | <0.001 |

[1]HR = Hazard Ratio, CI = Confidence Interval

**Key Findings:**:

Survival following treatment for colon cancer was not differentiated by age (p=0.58), sex (p=0.49) or perforation of colon (p=0.56).

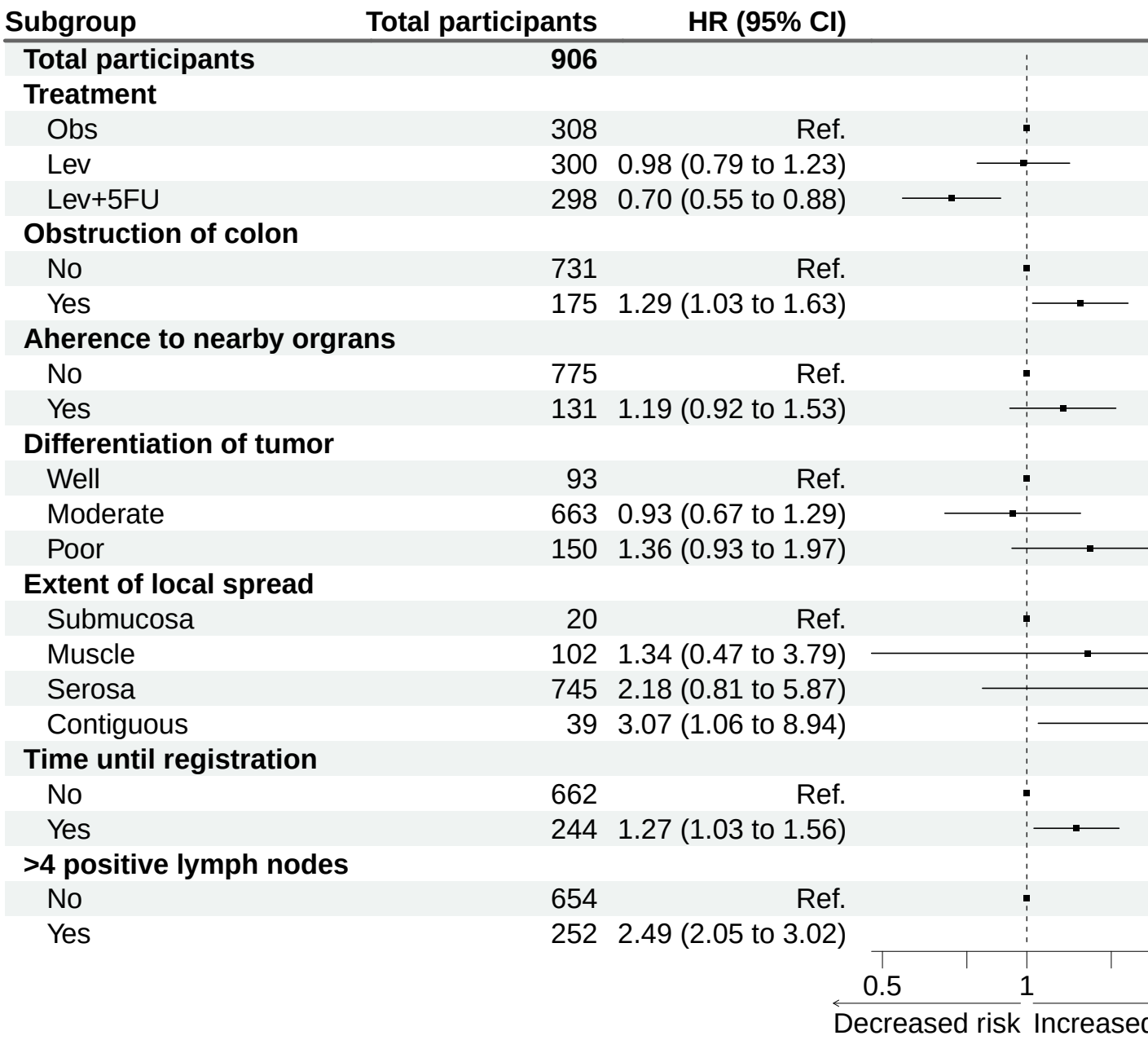
However, survival outcomes did differ across the categories of the remaining variables, with better survival rates associated with the Lev+5FU treatment, unobstructed colon, no adherence to nearby organs, well or moderately differentiated tumour, local spread limited to the submucosa or muscle, shorter time until registration and fewer positive lymph nodes.

**Table 2. Hazard ratios and 95% Confidence Intervals for univariable and multivariable Cox regression models**

**Key Findings:**

The estimates confirmed that although treatment with levamisole did not improve outcomes compared to the control group (HR = 0.98, 95% CI = 0.79-1.23), the hazard of death was 30% lower among patients treated with levamisole+5-fluorouracil (HR = 0.70, 95% CI = 0.55-0.88). Other factors significantly associated with increased hazard of death included obstruction of the colon (HR = 1.29, 95% CI = 1.03-1.63), local spread to contiguous regions (HR = 3.07, 95% CI = 1.06-8.94), longer time between surgery and registration (HR = 1.27, 95% CI = 1.03-1.56) and more than 4 positive lymph nodes (HR = 2.49, 95% CI = 2.05-3.02).

**Figure 2: Forest plot for hazard ratios and 95% confidence intervals for multivariable Cox regression models**



### Assessing the proportional harzard assumptions

One assumption of the Cox proportional hazards regression model is that the hazards are proportional at each point in time throughout follow-up. The cox.zph() function from the {survival} package allows us to check this assumption. It results in two main things:

### (1) Using a chi-squared test based on Schoenfeld residuals:

H0: Covariate effect is constant (proportional) over time HA: Covariate effect changes over time

The null hypothesis of proportional hazard is tested for each covariate individually and jointly as well.

A significant p-value indicates that the proportional hazards assumption is violated.

```
cox.zph(cox_model)
```

```
#>             chisq df       p
#> rx         2.3509  2 0.30869
#> obstruct   6.2760  1 0.01224
#> adhere     0.0775  1 0.78074
#> differF   16.0442  2 0.00033
#> extentF    7.3605  3 0.06125
#> surg       0.0247  1 0.87521
#> node4      5.8260  1 0.01579
#> GLOBAL    36.5773 11 0.00014
```

The test confirms that the proportional hazards assumption is violated for obstruction of colon (p=0.01), differentiation of tumour ($p < 0.001$) and marginally for extent of local spread (p=0.06). The test also suggests that the variable indicating more than 4 positive lymph nodes also violates the assumption (p=0.016); the global test also indicates the assumption is invalid ($p<0.001$).

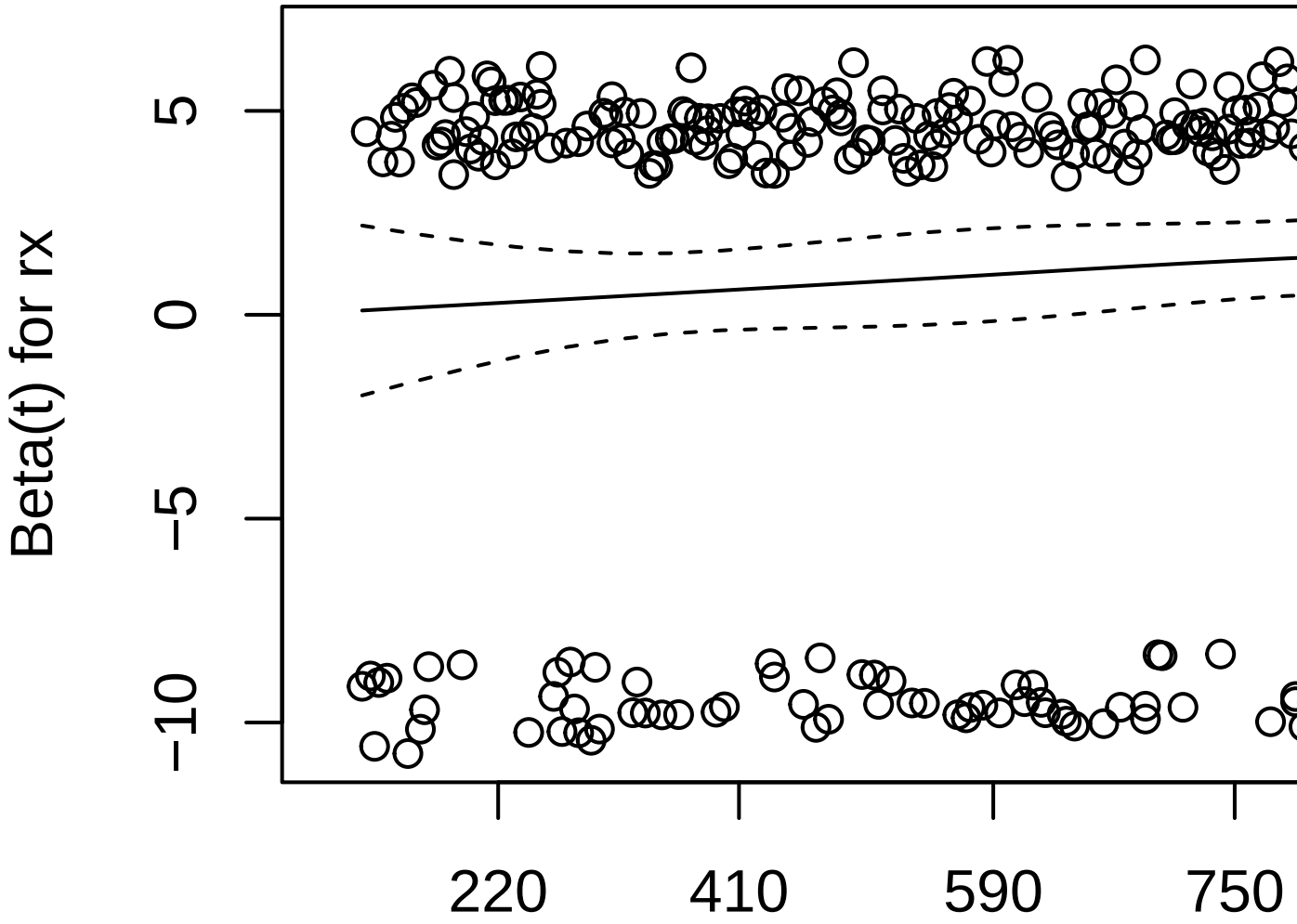
### (2) Plots of the Schoenfeld residuals

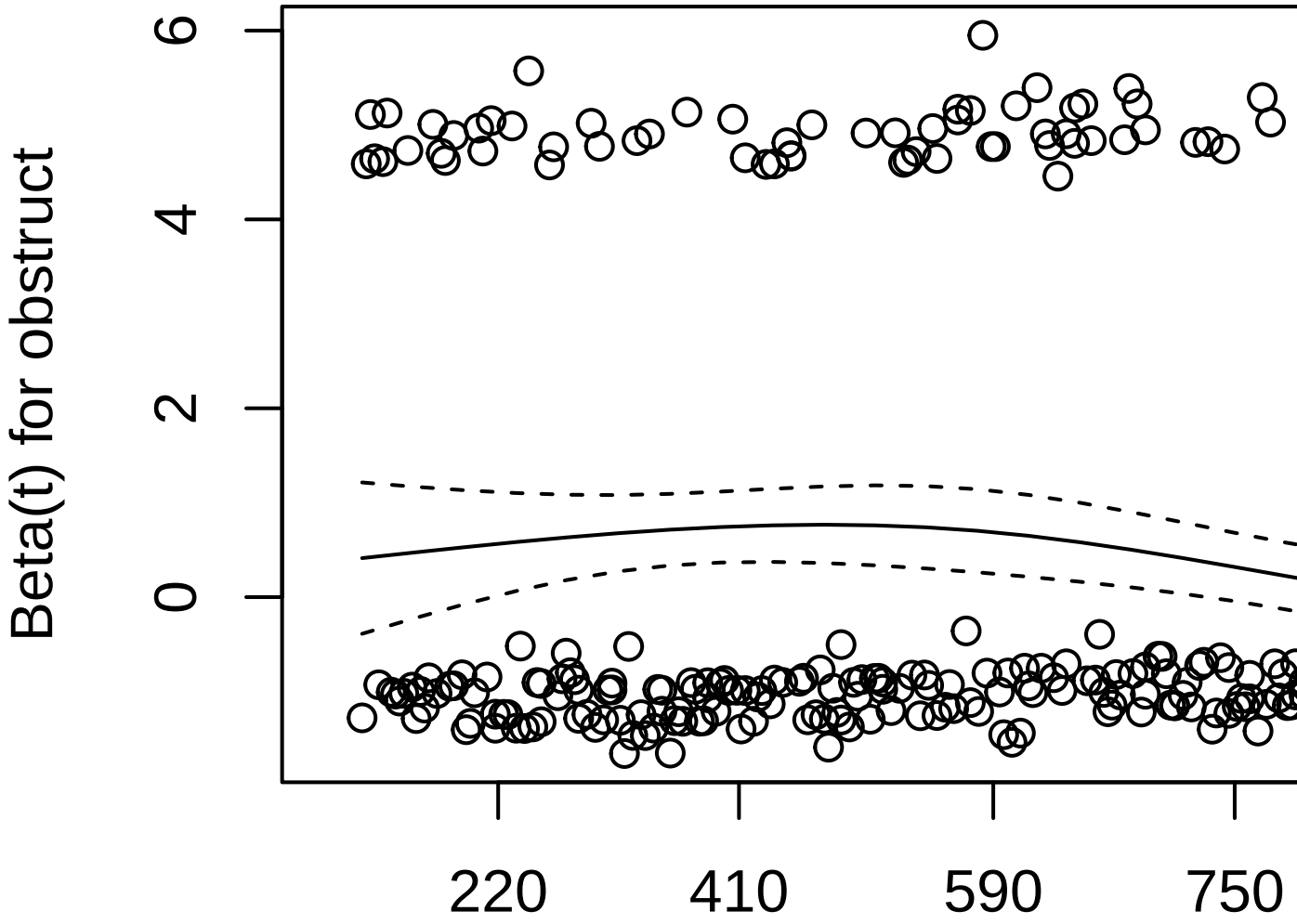
Deviation from a zero-slope (i.e., flat) line is evidence that the proportional hazards assumption is violated.

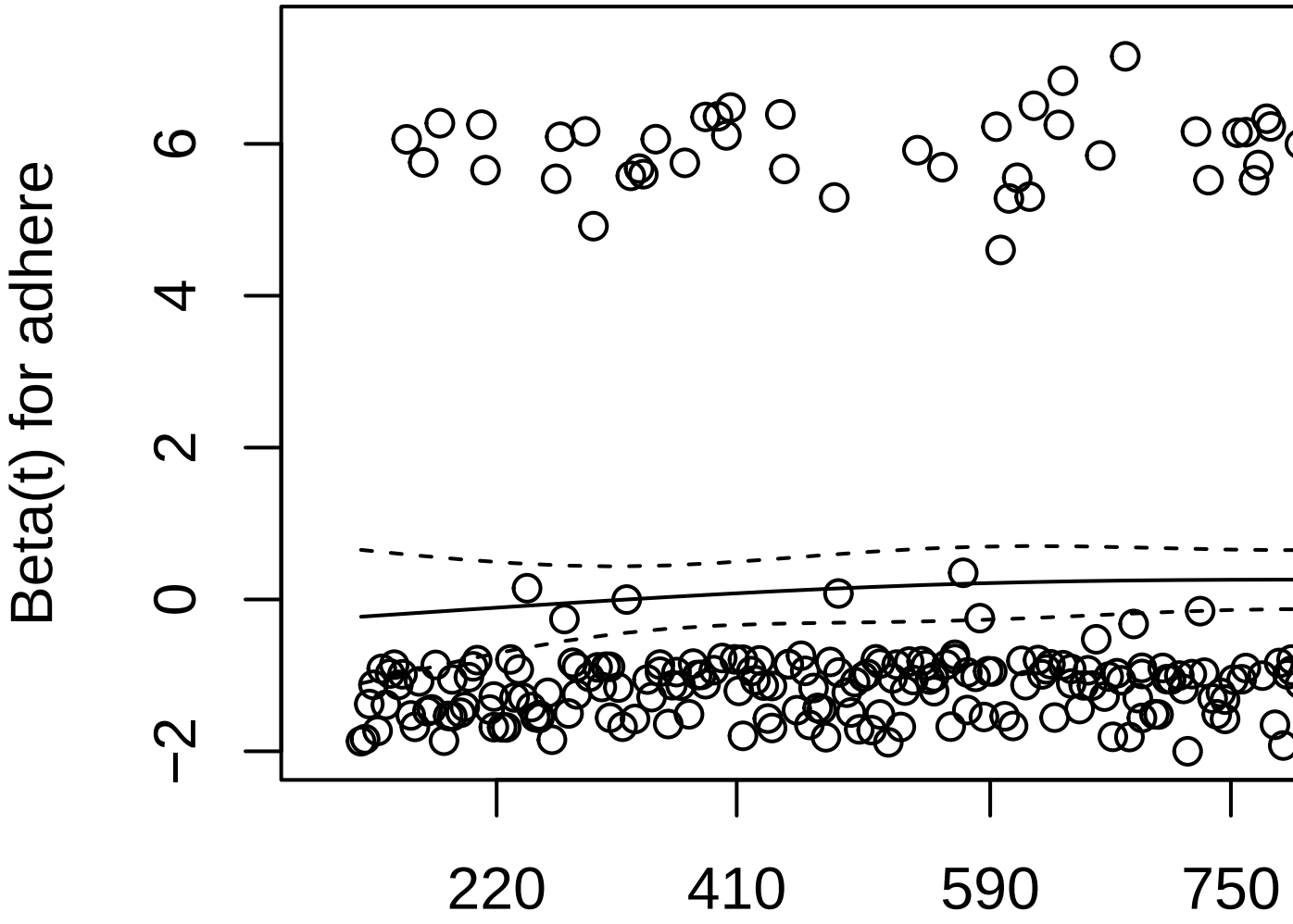
Note: It is actually plotting the coefficient for each predictor at each time point over time). We want to see a flat line over time.

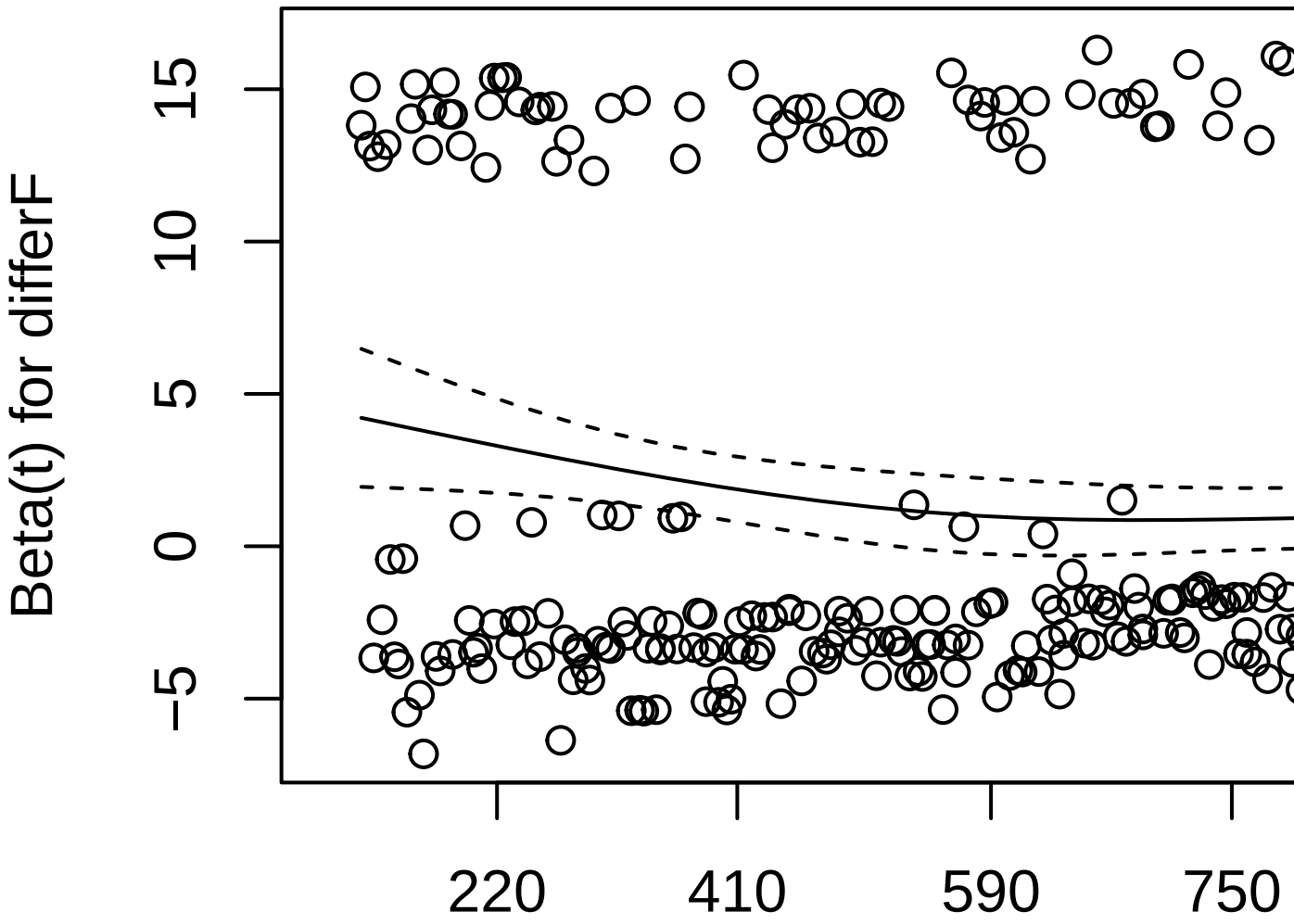
(Side note: If we have a large data, we will be able to detect very small changes of coefficients over time. So if the change in the coefficient is not large enough to be clinically meaningfully, it can perhaps be ignored as well).

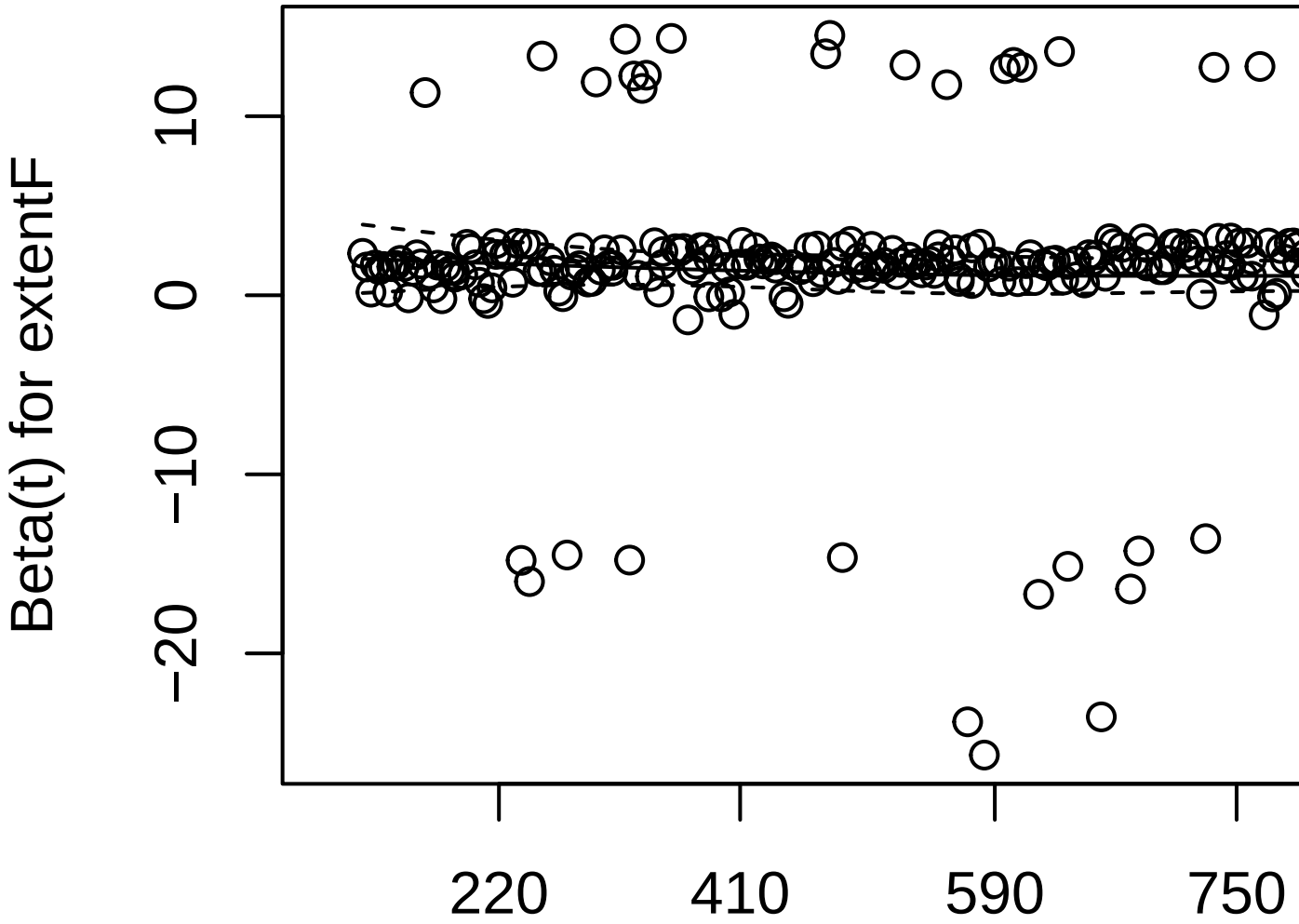
```
plot(cox.zph(cox_model))
```

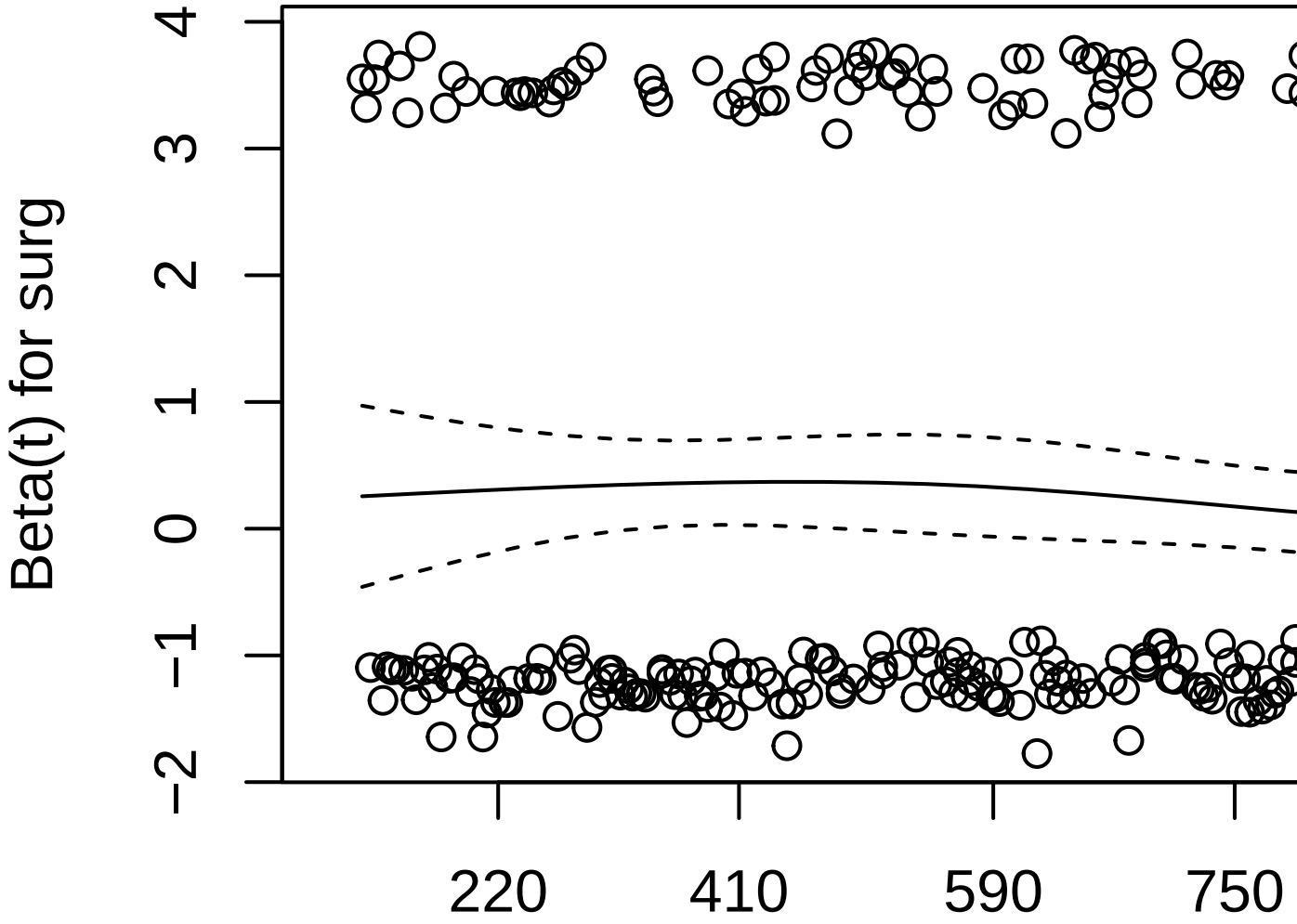

Beta(t) for rx
5
0
−5
−10
220
410
590
750

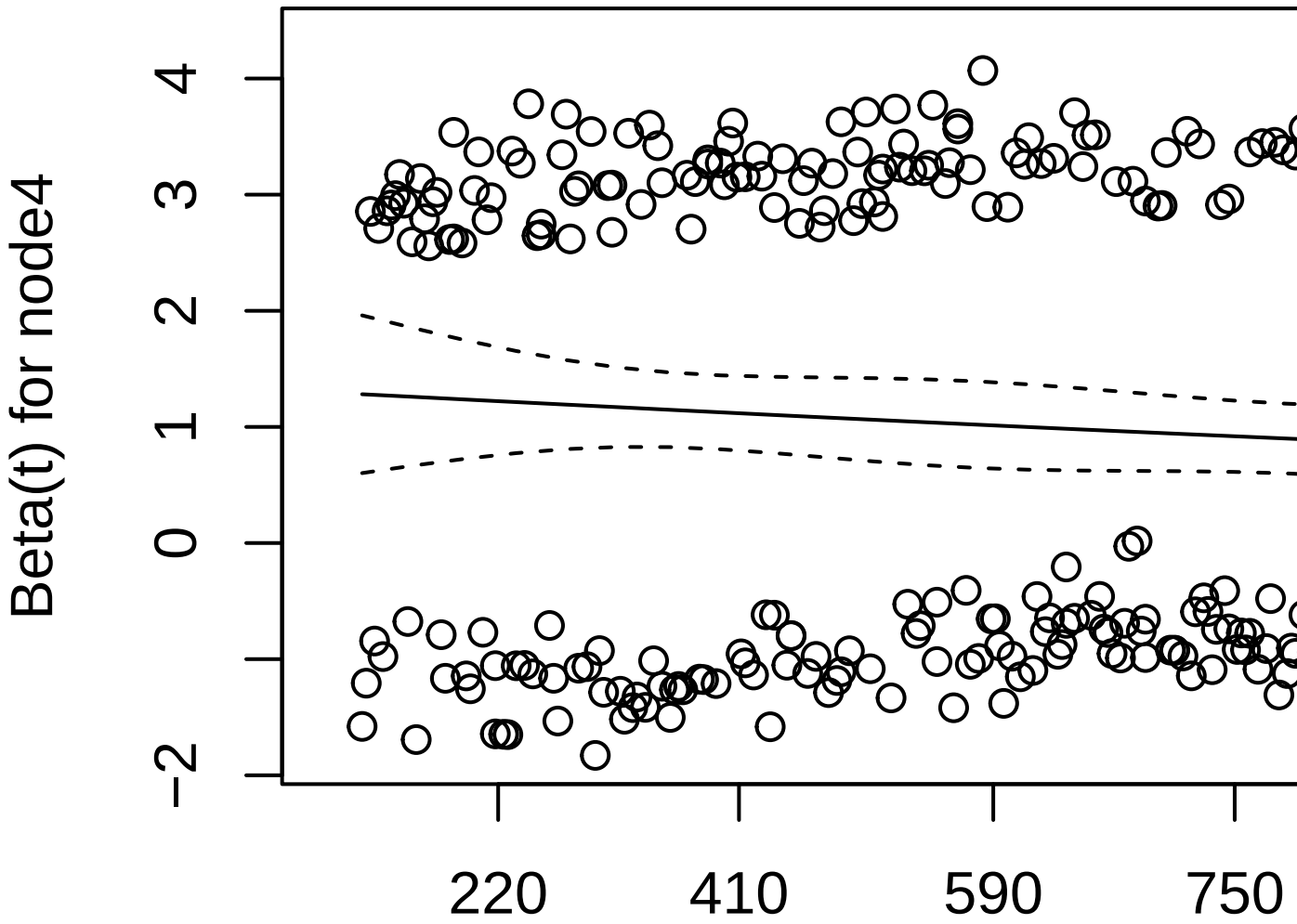

Beta(t) for obstruct
0
2
4
6
220
410
590
750


Beta(t) for adhere
−2
0
2
4
6
220
410
590
750


Beta(t) for differF
15
10
5
0
−5
220
410
590
750


Beta(t) for extentF
10
0
−10
−20
220
410
590
750


Beta(t) for surg
4
3
2
1
0
−1
−2
220
410
590
750


Beta(t) for node4
4
3
2
1
0
−2
220
410
590
750

### Dealing with proportional hazards violation as a sensitivity Analysis

#### Startify by the non-PH variable

In the stratified Cox model: - The cox model is estimated separately in each stratum - Drawback: we cannot quantify the effect of the stratification variable on survival (i.e., no coefficient will be estimated).

Because these variables are not primary factors of interest we can control for them using stratification. The resulting estimated hazard ratios and 95% confidence intervals are presented in Table 3. As can be seen, the proportional hazards assumption is met in this model.

```
mvModelStratified <- coxph(Surv(time, status) ~ rx + strata(obstruct) + adhere + strata(diffe

cox.zph(mvModelStratified)
```

```
#>           chisq df    p
#> rx      2.24767  2 0.33
#> adhere  1.50847  1 0.22
#> surg    0.00211  1 0.96
#> GLOBAL 3.87358  4 0.42
```