
Phân tích theo thời gian :

Hội quy bán tham số

Mark Stevenson

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Số dư (Residuals)
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

Tổng quan

- Phân tích hồi quy được sử dụng khi chúng ta muốn định lượng ảnh hưởng của các biến diễn giải theo thời gian đối với sự kiện
 - sự hiện diện của màng bào thai bị giữ lại làm tăng khả năng thụ thai trong chu kỳ thụ thai ở bò sữa
 - Nhưng là bao nhiêu?

Tổng quan

- Mục tiêu:
 - có được một số thước đo hiệu ứng mô tả mối quan hệ giữa một biến dự đoán và thời gian đến thất bại, điều chỉnh cho các biến khác trong mô hình
 - Hồi quy logistic: đánh giá tác động của tỷ số chênh (odds ratio)
 - Hồi quy Poisson: đánh giá tác động của tỷ số nguy cơ (risk ratio)
 - survival analysis: đánh giá tác động (trong đa số trường hợp[‡]) tỷ số mỗi nguy (hazard ratio)

[‡] Cox proportional hazards regression

BSE in British cattle holdings.

TABLE 5: Cox proportional hazards regression model showing the effect of region, holding size and holding type on the monthly hazard of experiencing a BSE index case

Explanatory variable	Number of holdings	Number BSE-positive	Regression coefficient (se)	P	Hazard ratio†	95% CI of hazard ratio
Region				<0.01*		
Eastern	5924†	1137	0.7981 (0.0359)		2.22	2.07-2.38
Mid and West	22,970	6753	0.5830 (0.0241)		1.79	1.71-1.88
Northern	16,283	4284	0.5632 (0.0254)		1.76	1.67-1.85
Scotland	16,635	2627			1.00	
South east	7702	2012	0.8865 (0.0302)		2.43	2.29-2.58
South west	24,529	8336	0.8746 (0.0234)		2.40	2.29-2.51
Wales	20,809	4315	0.5127 (0.0256)		1.67	1.59-1.76
Holding size				<0.01		
1-6	31,469	644	-0.8344 (0.0468)		0.43	0.40-0.48
7-21	25,142	2021			1.00	
22-53	29,145	8479	1.061 (0.0257)		2.89	2.75-3.04
>53	29,702	18,320	1.776 (0.0254)		5.91	5.62-6.21
Holding type				<0.01		
Dairy	38,576	21,191	1.117 (0.0168)		3.06	2.96-3.16
Mixed	9955	2221	0.5462 (0.0253)		1.73	1.64-1.81
Beef suckler	62,896	6052			1.00	

Likelihood ratio test statistic 35,570; df 11; P<0.01

* The significance of inclusion of the six region variables in the model

† Cases with missing values have been excluded, so counts vary slightly from those shown in Table 4

‡ Interpretation: compared with the reference category (holdings in Scotland), after adjusting for the effect of the size and type of holding, up to June 30, 1997, holdings in the Eastern region of England were at 2.22 (95% CI 2.07 to 2.38) times the monthly hazard of having a BSE index case

CI Confidence interval

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lockhart, J., Lin, D., Jackson, R., 2000a. Temporal aspects of the bovine spongiform encephalopathy epidemic in Great Britain: Individual animal-associated risk factors for disease. Veterinary Record 147, 349 - 354.

Tổng quan

- Dễ xây dựng mô hình mối nguy hơn mô hình sinh tồn
- Một mô hình tham số dựa trên phân bố mũ có thể được tham số hóa bằng mô hình tuyến tính cho log-hazard:

$$\log h_i(t) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}$$

Tổng quan

- Mô hình mối nguy theo tỷ lệ của Cox loại bỏ α (the intercept) và thay bằng $\alpha(t)$:

$$\log h_i(t) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}$$

$$\log h_i(t) = \alpha(t) + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi}$$

$$h_i(t) = \exp(\alpha(t) + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi})$$

- Có thể viết:

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_m x_{mi})$$



Mối nguy “Cơ bản”

Tổng quan

- Một ví dụ đơn giản:
 - bệnh nhân ung thư nhận được một trong hai liệu pháp: điều trị hoặc kiểm soát
 - nguy cơ tái phát hàng ngày đối với nhóm kiểm soát: $h_C(t) = h_0(t)$
 - nguy cơ tái phát hàng ngày đối với nhóm điều trị: $h_T(t) = h_0(t) \exp(\beta)$
 - Tỷ số nguy cơ tái phát:

$$\text{HR} = \frac{h_T(t)}{h_C(t)} = \frac{h_0(t) \exp(\beta)}{h_0(t)}$$

- so với đối chứng, nguy cơ tái phát hàng ngày đối với nhóm điều trị là $\exp(\beta)$

Tổng quan

- Một ví dụ đơn giản (tiếp):
 - tỷ lệ nguy cơ là 10 có nghĩa là nguy cơ thất bại trên một đơn vị thời gian đối với nhóm điều trị là 10 lần so với nhóm đối chứng
 - tỷ lệ nguy cơ là 0,10 có nghĩa là nguy cơ thất bại trên mỗi đơn vị thời gian đối với nhóm điều trị là 0,10 lần so với nhóm đối chứng
 - Do đó chúng ta không đánh giá mỗi nguy tuyệt đối của một sự kiện đang diễn ra mà chỉ đánh giá tỷ số của các mối nguy
 - các biến diễn giải phân loại: ‘so với nhóm tham chiếu, nguy cơ thất bại hàng ngày là ...’
 - các biến giải thích liên tục: ‘đối với sự gia tăng đơn vị trong XYZ, nguy cơ thất bại hàng năm là ...’

BSE in British cattle holdings.

TABLE 5: Cox proportional hazards regression model showing the effect of region, holding size and holding type on the monthly hazard of experiencing a BSE index case

Explanatory variable	Number of holdings	Number BSE-positive	Regression coefficient (se)	P	Hazard ratio†	95% CI of hazard ratio
Region				<0.01*		
Eastern	5924†	1137	0.7981 (0.0359)		2.22	2.07-2.38
Mid and West	22,970	6753	0.5830 (0.0241)		1.79	1.71-1.88
Northern	16,283	4284	0.5632 (0.0254)		1.76	1.67-1.85
Scotland	16,635	2627			1.00	
South east	7702	2012	0.8865 (0.0302)		2.43	2.29-2.58
South west	24,529	8336	0.8746 (0.0234)		2.40	2.29-2.51
Wales	20,809	4315	0.5127 (0.0256)		1.67	1.59-1.76
Holding size				<0.01		
1-6	31,469	644	-0.8344 (0.0468)		0.43	0.40-0.48
7-21	25,142	2021			1.00	
22-53	29,145	8479	1.061 (0.0257)		2.89	2.75-3.04
>53	29,702	18,320	1.776 (0.0254)		5.91	5.62-6.21
Holding type				<0.01		
Dairy	38,576	21,191	1.117 (0.0168)		3.06	2.96-3.16
Mixed	9955	2221	0.5462 (0.0253)		1.73	1.64-1.81
Beef suckler	62,896	6052			1.00	

Likelihood ratio test statistic 35,570; df 11; P<0.01

* The significance of inclusion of the six region variables in the model

† Cases with missing values have been excluded, so counts vary slightly from those shown in Table 4

‡ Interpretation: compared with the reference category (holdings in Scotland), after adjusting for the effect of the size and type of holding, up to June 30, 1997, holdings in the Eastern region of England were at 2.22 (95% CI 2.07 to 2.38) times the monthly hazard of having a BSE index case

CI Confidence interval

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lockhart, J., Lin, D., Jackson, R., 2000a. Temporal aspects of the bovine spongiform encephalopathy epidemic in Great Britain: Individual animal-associated risk factors for disease. Veterinary Record 147, 349 - 354.

Tổng quan

- Các giả định:
 1. Tỷ lệ của hàm số nguy cơ cho hai cá thể với các tập hợp số đồng biến khác nhau thay đổi theo thời gian
 2. Thời gian được đo trên thang đo liên tục
 3. Kiểm duyệt xảy ra ngẫu nhiên

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Residuals
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

Xây dựng mô hình

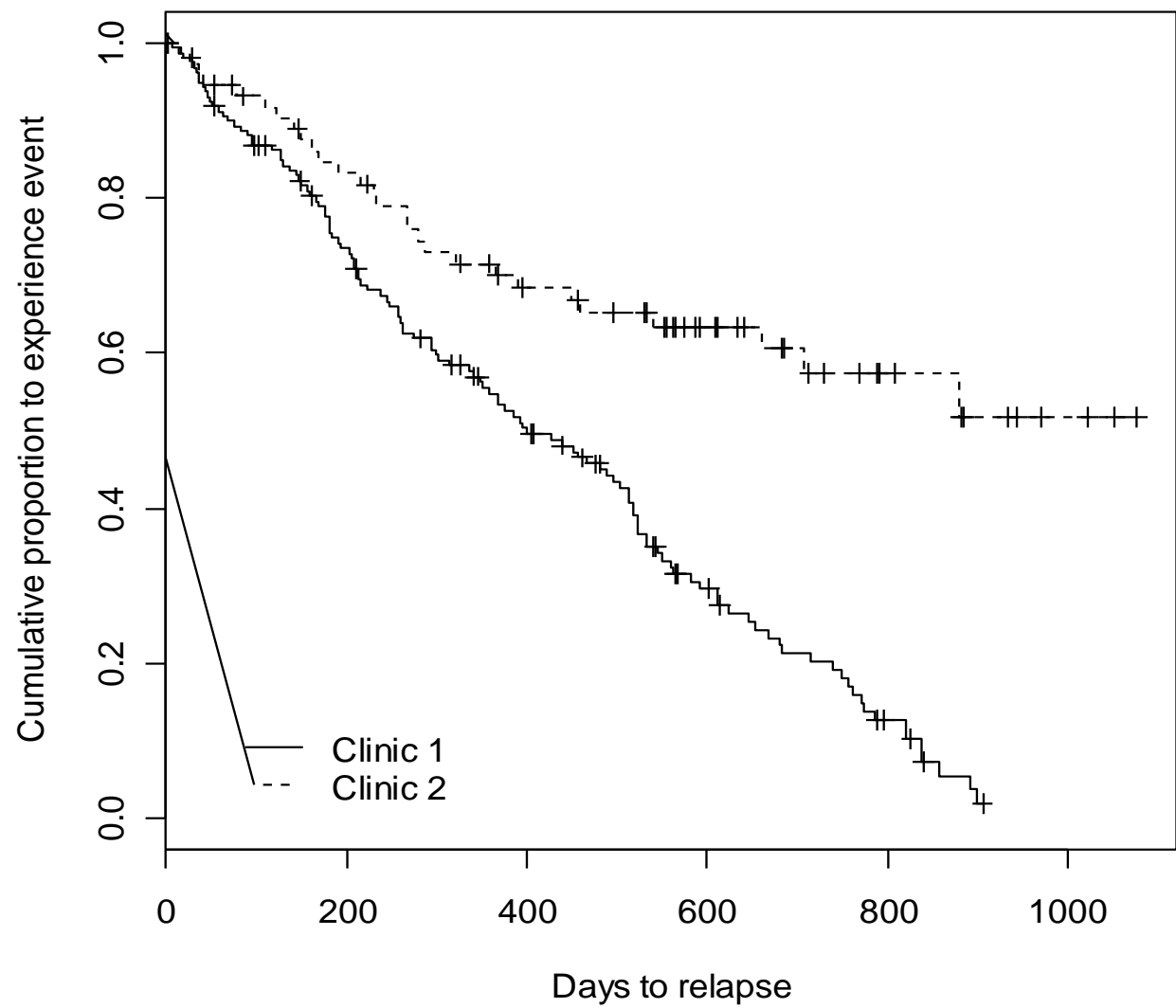
- Bước 1 lựa chọn các đồng biến:
 - phân tích kỹ lưỡng để xác định mối liên quan giữa thời gian sinh tồn và tất cả các đồng biến tiềm năng
 - Biến phân loại:
 - Bao gồm phương pháp Kaplan-Meier để ước tính các hàm số sinh tồn cụ thể của nhóm
 - Biến liên tục:
 - nên được chia thành các phần tư (hoặc các nhóm có ý nghĩa sinh học khác) và sử dụng các phương pháp tương tự nhau
 - một mô hình đa biến nên có ngay từ đầu với tất cả các biến có ý nghĩa trong các phân tích đơn biến với $P = 0.20$ đến 0.25 và với bất kỳ biến nào khác nếu nó quan trọng trong lâm sàng



```
library(survival); setwd("D:\\TEMP")
dat <- read.table("addict.csv", header = TRUE, sep = ",");

addict.km01 <- survfit(Surv(stop, status) ~ clinic, type = "kaplan-
meier", data = dat)

plot(addict.km01, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event", main = "Clinic", lty = c(1,2),
mark.time = FALSE)
legend(x = "topright", legend = c("Clinic 1", "Clinic 2"), lty =
c(1,2), bty = "n", cex = 0.80)
```



```
survdif(Surv(stop, status) ~ clinic, data = dat, na.action =  
na.omit, rho = 0)
```

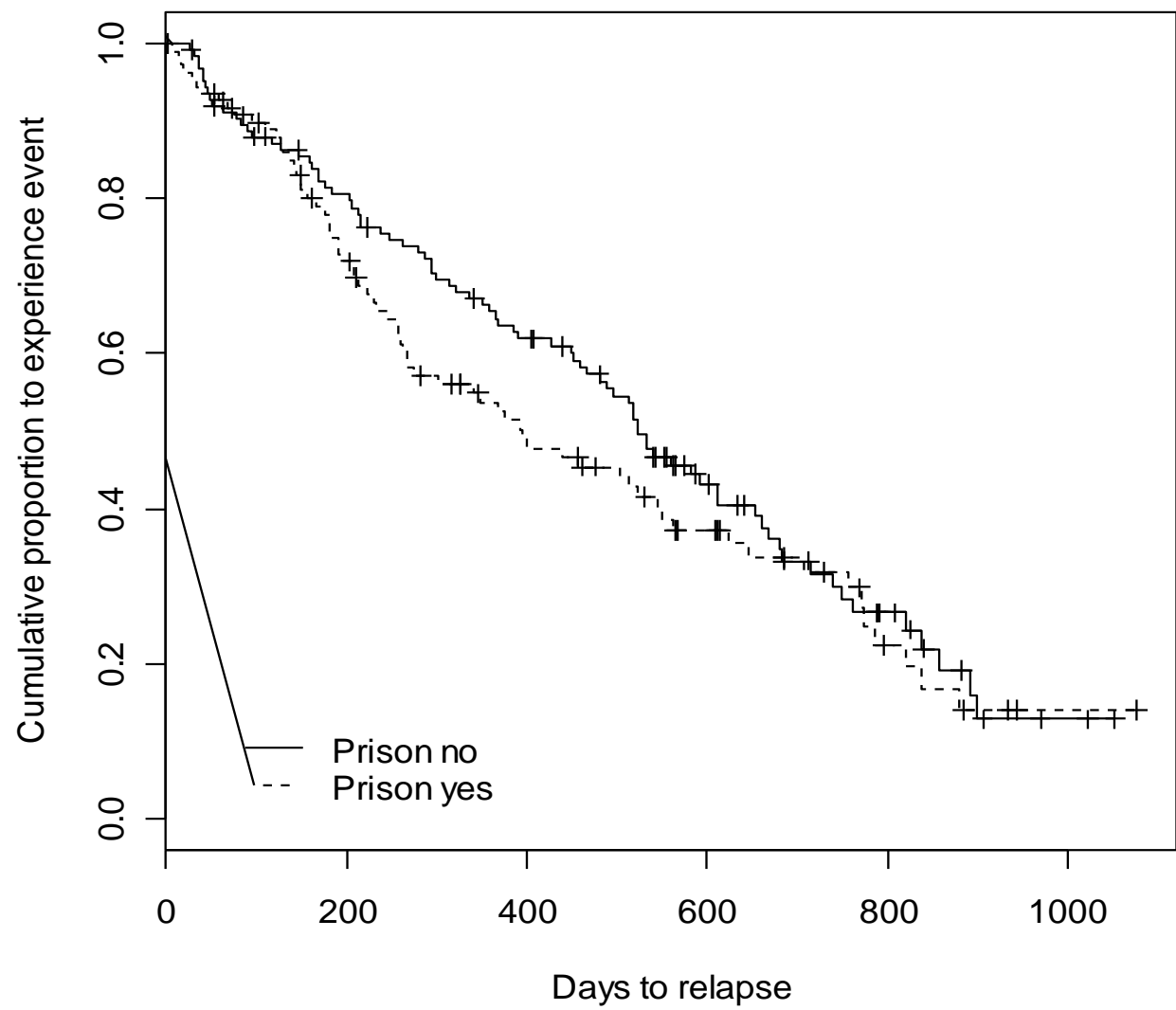
Call:

```
survdif(formula = Surv(stop, status) ~ clinic, data = dat,  
na.action = na.omit,  
rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
clinic=1	163	122	90.8	10.7	28.1
clinic=2	75	28	59.2	16.4	28.1

Chisq= 28.1 on 1 degrees of freedom, **p = 1.18e-07**

```
addict.km02 <- survfit(Surv(stop, status) ~ prison, type = "kaplan-  
meier", data = dat)  
  
plot(addict.km02, xlab = "Days to relapse", ylab = "Cumulative  
proportion to experience event", main = "Prison", lty = c(1,2),  
mark.time = FALSE)  
legend(x = "topright", legend = c("Prison absent", "Prison present"),  
lty = c(1,2), bty = "n", cex = 0.80)
```



```
survdif(Surv(stop, status) ~ prison, data = dat, na.action =  
na.omit, rho = 0)
```

Call:

```
survdif(formula = Surv(stop, status) ~ prison, data = dat,  
na.action = na.omit,  
rho = 0)
```

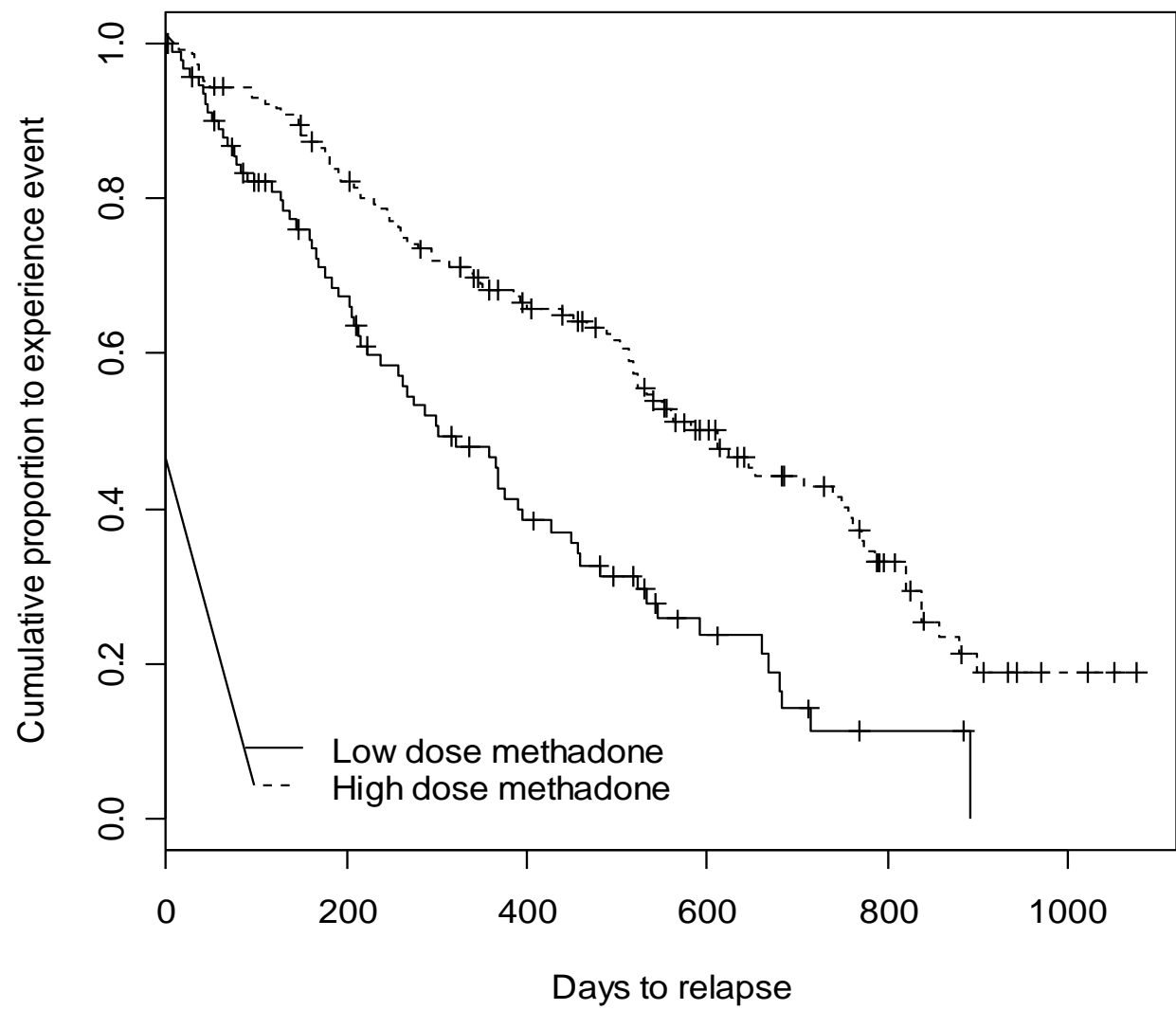
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
prison=0	127	81	87.4	0.474	1.14
prison=1	111	69	62.6	0.663	1.14

Chisq= 1.1 on 1 degrees of freedom, **p = 0.285**

```
dat$dose.cat[dat$dose < 60] <- 0;
dat$dose.cat[dat$dose >= 60] <- 1;

addict.km03 <- survfit(Surv(stop, status) ~ dose.cat, type =
"kaplan-meier", data = dat);

plot(addict.km03, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event", main = "Dose categories", lty =
c(1,2,3,4), mark.time = FALSE)
legend(x = "topright", legend = c("Low dose","High dose"), lty =
c(1,2), bty = "n", cex = 0.80)
```

```
survdiff(Surv(stop, status) ~ dose.cat, data = dat, na.action =  
na.omit, rho = 0)
```

Call:

```
survdiff(formula = Surv(stop, status) ~ dose.cat, data = dat,  
na.action = na.omit, rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
newdose=0	93	65	40.8	14.33	20.4
newdose=1	145	85	109.2	5.36	20.4

Chisq= 20.4 on 1 degrees of freedom, **p = 6.23e-06**

Xây dựng mô hình

- Bước 2 gắn với mô hình đa biến:
 - Xây dựng mô hình và sử dụng P values từ kiểm định Wald tests của các hệ số riêng lẻ để xác định các biến diễn giải có thể bị loại khỏi mô hình
 - kiểm định tỷ lệ xác suất cục bộ sẽ xác nhận biến diễn giải bị loại bỏ là không có ý nghĩa
 - tiếp tục cho đến khi không còn biến diễn giải nào nữa có thể được loại bỏ khỏi mô hình

```
addict.cph01 <- coxph(Surv(stop, status, type = "right") ~ clinic +
prison + dose, method = "breslow", data = dat)
summary(addict.cph01)
```

Call:

```
coxph(formula = Surv(stop, status, type = "right") ~ clinic +
      prison + dose, data = dat, method = "breslow")
```

	coef	exp(coef)	se(coef)	z	p
clinic	-1.0092	0.365	0.21473	-4.70	2.6e-06
prison	0.3147	1.370	0.16716	1.88	6.0e-02
dose	-0.0352	0.965	0.00637	-5.54	3.1e-08

Wald statistic: ratio of regression coefficient to its standard error. Test that β differs from zero.

	exp(coef)	exp(-coef)	lower .95	upper .95
clinic	0.365	2.74	0.239	0.555
prison	1.370	0.73	0.987	1.901
dose	0.965	1.04	0.953	0.977

```
Rsquare= 0.237 (max possible= 0.997 )
Likelihood ratio test= 64.3 on 3 df,
Wald test = 54 on 3 df,
Score (logrank) test = 56.2 on 3 df,
```

Likelihood ratio, Wald test, Score test. Test hypothesis that all β s are zero.

```
addict.cph02 <- coxph(Surv(stop, status, type = "right") ~ clinic +  
dose, method = "breslow", data = dat);  
summary(addict.cph02)
```

Call:

```
coxph(formula = Surv(stop, status, type = "right") ~ clinic +  
      dose, data = dat, method = "breslow")
```

	coef	exp(coef)	se(coef)	z	p
clinic	-0.9535	0.385	0.21204	-4.50	6.9e-06
dose	-0.0342	0.966	0.00626	-5.47	4.6e-08

	exp(coef)	exp(-coef)	lower .95	upper .95
clinic	0.385	2.59	0.254	0.584
dose	0.966	1.03	0.955	0.978

Rsquare= 0.226 (max possible= 0.997)

Likelihood ratio test= 60.8 on 2 df, p=6.14e-14

Wald test = 52.8 on 2 df, p=3.52e-12

Score (logrank) test = 54.3 on 2 df, p=1.60e-12

Xây dựng mô hình

- Bước 2 phù hợp với mô hình đa biến (tiếp.):
 - Kiểm định likelihood ratio test:
$$G = -2 [(LL_0) - (LL_1)]$$

G is compared to a χ^2 distribution with $df = (\text{number of covariates } LL_1 - \text{number of covariates } LL_0)$
 - Kiểm định univariate Wald test: được sử dụng để kiểm tra xem hệ số hồi quy có khác biệt đáng kể so với 0: có thể cung cấp manh mối về việc các biến nào có thể được loại bỏ mà không ảnh hưởng đến hiệu suất mô hình

```
x2 <- 2 * (addict.cph02$loglik[2] - addict.cph01$loglik[2]);  
P <- 1 - pchisq(x2,1);
```

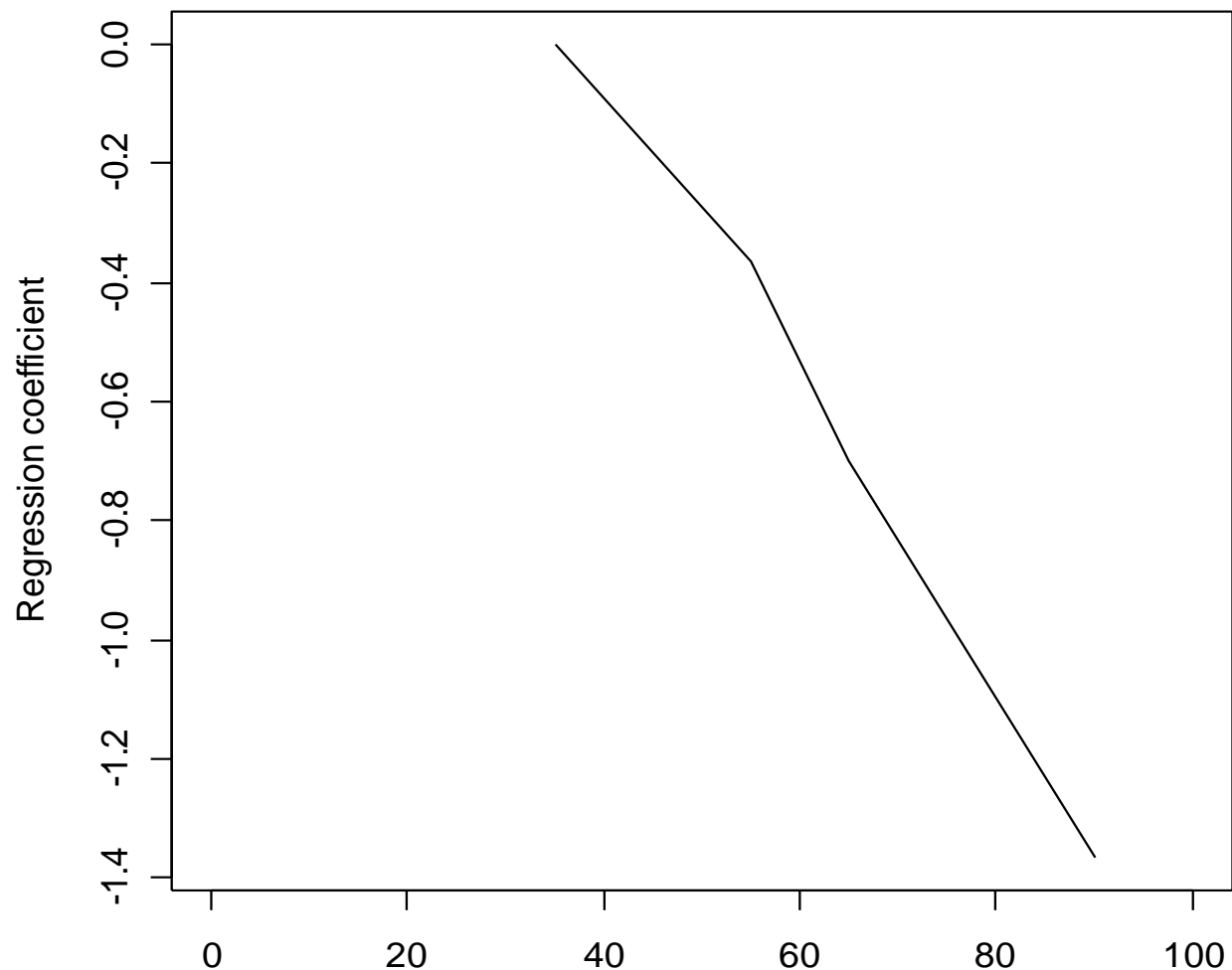
Variable	Subjects	Failed	Coefficient (SE)	P	Hazard ratio (95%)
Clinic:				< 0.01 ^a	
Clinic 1	163	122	-		1.00
Clinic 2	74	28	-1.0091 (0.2147)		0.36 (0.24 - 0.55) ^b
Prison:				0.06	
Absent	127	81	-		1.00
Present	111	69	0.3146 (0.1672)		1.37 (0.98 - 1.90)
Dose	238	150	- 0.0352 (0.0064)	< 0.01	0.96 (0.95 - 0.98)

^a Significance of the two clinic variables in the model.

^b Interpretation: compared with the reference category (patients from Clinic 1), after adjusting for the effect of methadone dose and prison status, patients from Clinic 2 had 0.36 (95% CI 0.24 - 0.55) times the daily hazard of relapse.

Xây dựng mô hình

- Bước 3 kiểm tra quy mô của đồng biến liên tục:
 - Ở đây chúng ta cần kiểm tra xem mỗi biến diễn giải liên tục là tuyến tính trong log- hazard của nó
 - Phương pháp 1
 - thay thế đồng biến liên tục bằng ba loại biến thiết kế bằng cách sử dụng Q1, Q2, và Q3 làm các điểm cắt
 - vẽ các hệ số ước tính cho các biến thiết kế so với điểm giữa của nhóm: điểm thứ tư được tính bằng 0 bằng cách sử dụng điểm giữa của nhóm thứ nhất
 - nếu quy mô chính xác là tuyến tính, thì đường nối bốn điểm sẽ xấp xỉ một đường thẳng



```
dat$dose.cat <- factor(dat$dose.cat, labels=c("1","2","3","4"))
contrasts(dat$dose.cat) <- contr.treatment(4, base = 1, contrasts =
TRUE)
addict.cph04 <- coxph(Surv(stop, status, type = "right") ~ clinic +
prison + dose.cat, method = "breslow", data = dat)
x <- c(((50 + min(dat$dose))/2), 55, 65, ((max(dat$dose) + 70)/2))
y <- c(0, addict.cph04$coefficients[3:5])
plot(x, y, xlim = c(0,100), type = "l", xlab = "Dose", ylab =
"Regression coefficient")
```

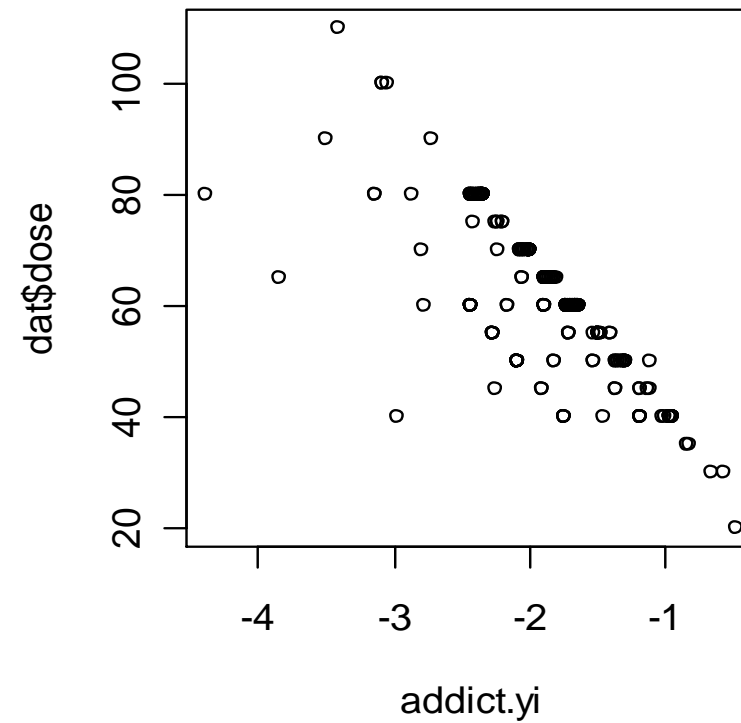
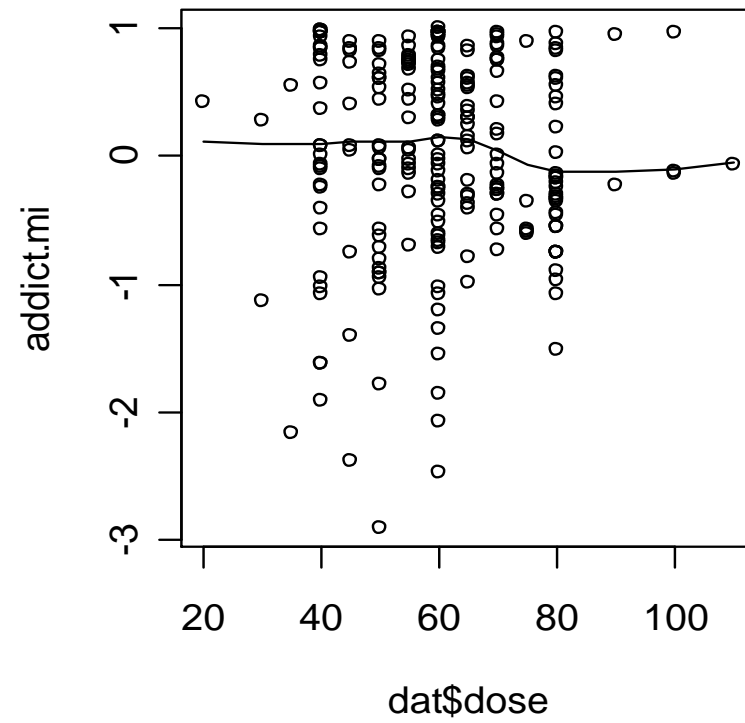
Xây dựng mô hình

- Bước 3 kiểm tra quy mô của đồng biến liên tục (tiếp.):
 - Phương pháp 2
 - Tạo sự phù hợp với mô hình hiệu ứng chính sơ bộ, bao gồm cả biến diễn giải (ví dụ: ‘tuổi’)
 - Giữ số dư Martingale residuals (M_i) từ mô hình này và tính toán $H_i = c_i - M_i$, với c_i là biến kiểm duyệt

Xây dựng mô hình

- Bước 3 kiểm tra quy mô của đồng biến liên tục (tiếp.):
 - Phương pháp 2
 - Tạo c_i như là một hàm số của biến diễn giải và tính toán giá trị lowest smooth (ký hiệu là c_{lsm})
 - Tạo H_i với một đồng biến và tính toán lowess smooth (ký hiệu là H_{LSM})
 - Tính
$$y_i = \ln \left(\frac{c_{lsm}}{H_{lsm}} \right) + \beta_{age} \times age$$
 - Tạo y_i là một hàm số về tuổi

```
addict.cph01 <- coxph(Surv(stop, status, type = "right") ~ clinic +  
prison + dose, method = "breslow", data = dat)  
  
addict.mi <- residuals(addict.cph01, type = "martingale")  
addict.hi <- dat$status - addict.mi  
addict.clsm <- lowess(dat$dose, dat$status)  
addict.hlsm <- lowess(dat$dose, addict.hi)  
addict.yi <- log(addict.clsm$y / addict.hlsm$y) +  
  (addict.cph01$coefficients[3] * dat$dose)  
  
par(pty = "s", mfrow = c(1,2))  
plot(dat$dose, addict.mg)  
lines(lowess(dat$dose, addict.mg))  
plot(addict.yi, dat$dose)
```



Nếu đồng biến là tuyến tính thì log hazard của các điểm này sẽ tạo thành một đường thẳng.

Xây dựng mô hình

- Bước 4 tương tác:
 - Xác định xem có cần một quy định về tương tác không
 - Nếu một quy định tương tác là một biến mới mà là sản phẩm của hai biến khác trong mô hình
 - Cân nhắc vấn đề sẽ quyết định một quy định tương tác cụ thể (hoặc các quy định) nên được đưa vào một mô hình, bất kể ý nghĩa thống kê như thế nào

Xây dựng mô hình

- Bước 4 tương tác (tiếp.):
 - Tác động của việc thêm một quy định tương tác nên được đánh giá bằng cách sử dụng kiểm định partial likelihood ratio test
 - Tất cả các tương tác có ý nghĩa ở $P = 0.05$ nên được duy trì trong mô hình của các tác động chủ đạo
 - Kiểm định p-values của Wald statistic có thể được sử dụng để chọn các tương tác có thể bị đưa ra khỏi mô hình
 - Tại thời điểm này, chúng ta có một “mô hình sơ bộ” và bước tiếp theo là đánh giá sự phù hợp và tuân thủ các giả định chính của nó

```
addict.cph05 <- coxph(Surv(stop, status, type="right") ~ clinic +
prison + dose + (clinic * prison), method = "breslow", data = dat);
summary(addict.cph05);
```

Call:

```
coxph(formula = Surv(stop, status, type = "right") ~ clinic +
      prison + dose + (clinic * prison), data = dat, method =
```

	coef	exp(coef)	se(coef)	z	p
clinic	-0.6641	0.515	0.2890	-2.30	2.2e-02
prison	1.1313	3.100	0.5403	2.09	3.6e-02
dose	-0.0368	0.964	0.0065	-5.66	1.5e-08
clinic:prison	-0.6819	0.506	0.4293	-1.59	1.1e-01

	exp(coef)	exp(-coef)	lower .95	upper .95
clinic	0.515	1.943	0.292	0.907
prison	3.100	0.323	1.075	8.938
dose	0.964	1.037	0.952	0.976
clinic:prison	0.506	1.978	0.218	1.173

Rsquare= 0.245 (max possible= 0.997)

Likelihood ratio test= 66.9 on 4 df, p=1.03e-13

Wald test = 57.5 on 4 df, p=9.72e-12

Score (logrank) test = 60.1 on 4 df, p=2.76e-12

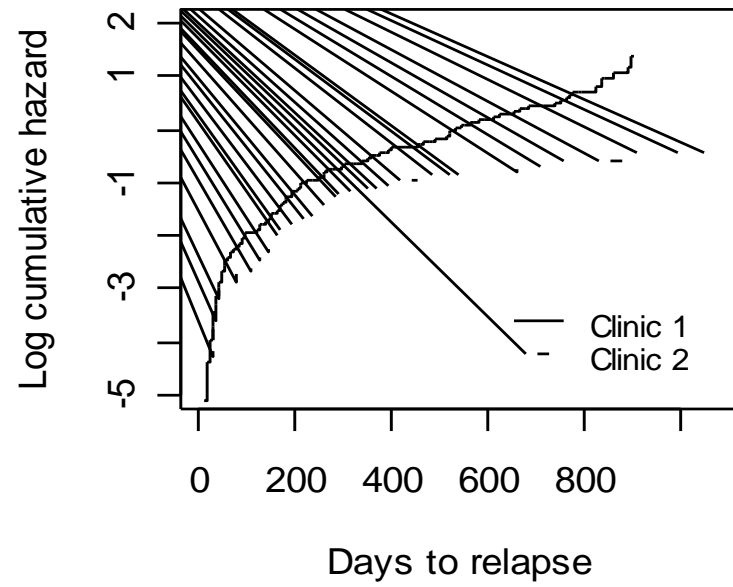
Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Residuals
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

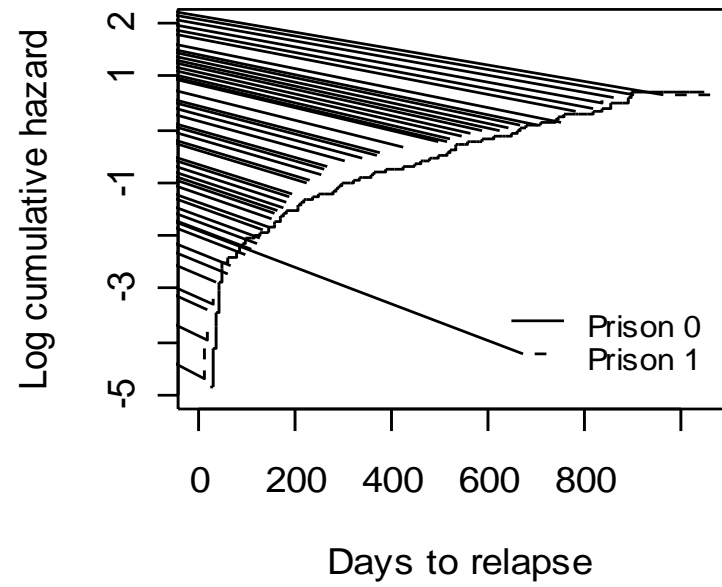
Kiểm tra giả định các mối nguy theo tỷ lệ

- Kiểm tra từng biến diễn giải để chắc chắn rằng giả thuyết CPH có hiệu lực
 1. Lập $\log[-\log S(t)]$ như một hàm thời gian cho từng phân tầng của đồng biến: Nếu giả thuyết CPH có hiệu lực thì các đường cong có thể là song song
 2. Lập số dư Schoenfeld như một hàm số về thời gian
 - Nếu giả thuyết CPH có hiệu lực với số dư Schoenfeld thì nên nằm rải rác xung quanh 0 và có độ dốc nghiêng về 0
 3. Đưa một quy định tương tác phụ thuộc thời gian đối với đồng biến: Nếu giả thuyết CPH có hiệu lực với sự tương tác thì quy định sẽ không có ý nghĩa (chi tiết ở phần sau)

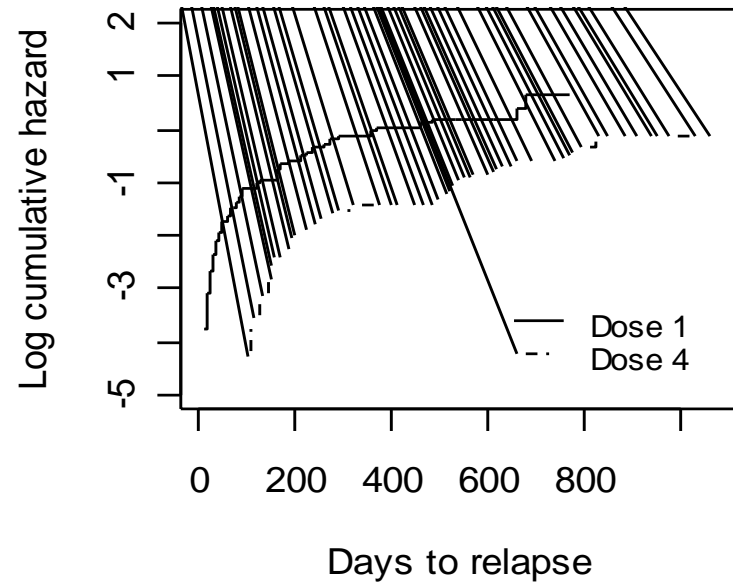
Clinic



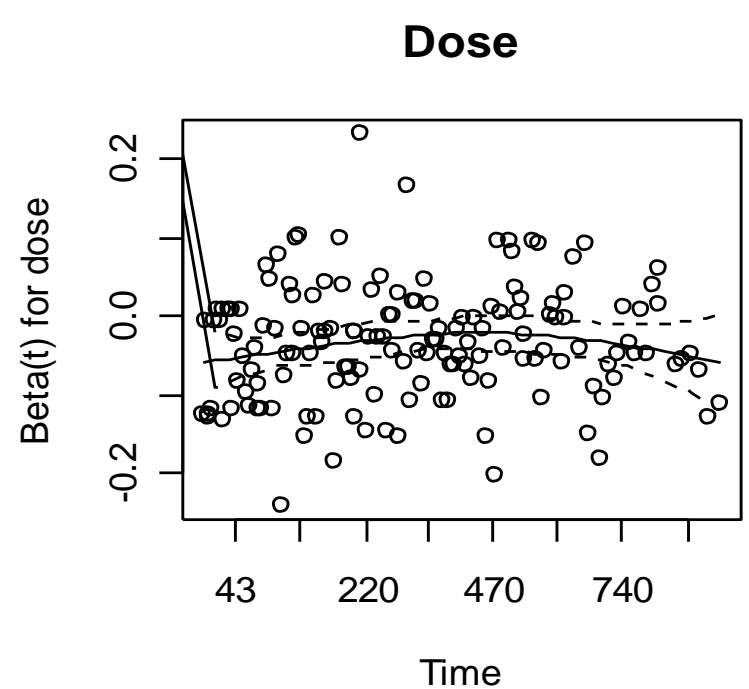
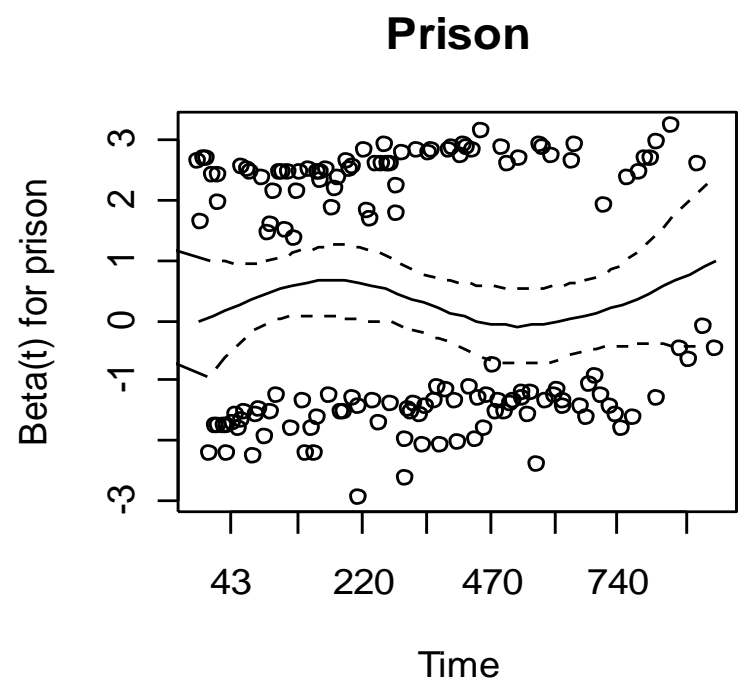
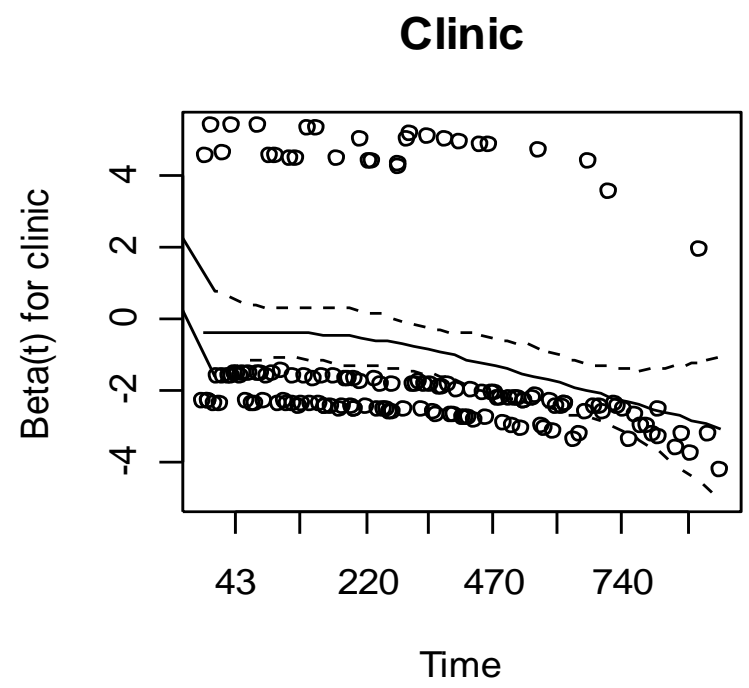
Prison



Dose



Số dư Schoenfeld và thời gian đối với từng biến diễn giải.



```
addict.cph01 <- coxph(Surv(stop, status, type = "right") ~ clinic +  
prison + dose, method = "breslow", data = dat);  
addict.zph <- cox.zph(addict.cph01);  
addict.zph
```

	rho	chisq	p
clinic	-0.2525	10.683	0.00108
prison	-0.0261	0.103	0.74867
dose	0.0703	0.657	0.41762
GLOBAL	NA	11.794	0.00812

rho là tương quan Pearson product-moment giữa số dư Schoenfeld được đánh giá và thời gian. Ở ví dụ trước, kiểm định `cox.zph` test có ý nghĩa đối với phòng khám ($P < 0.01$), điều này cho thấy giả thuyết mỗi nguy tử lệ xung đột với biến phòng khám. Lập số dư Schoenfeld sẽ hỗ trợ khái niệm này.

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Số dư (Residuals)
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

Số dư (Residuals)

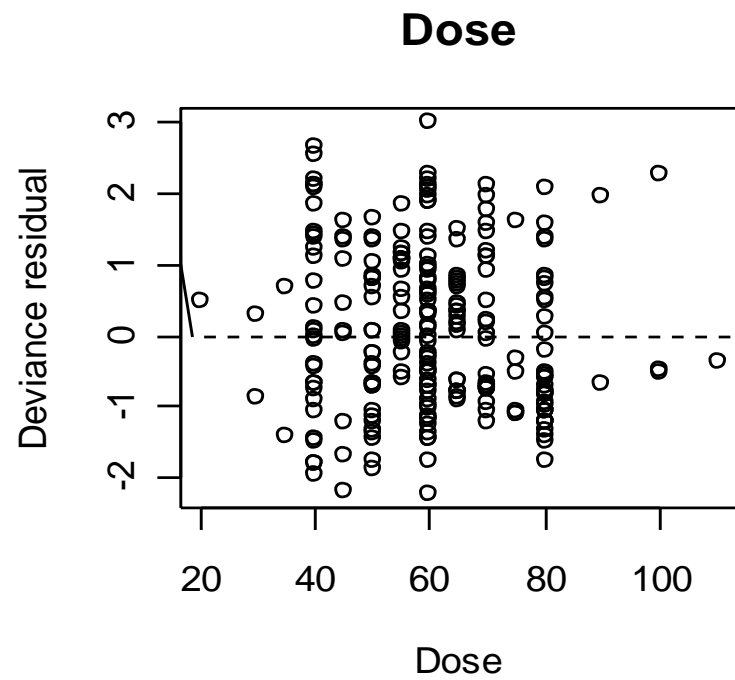
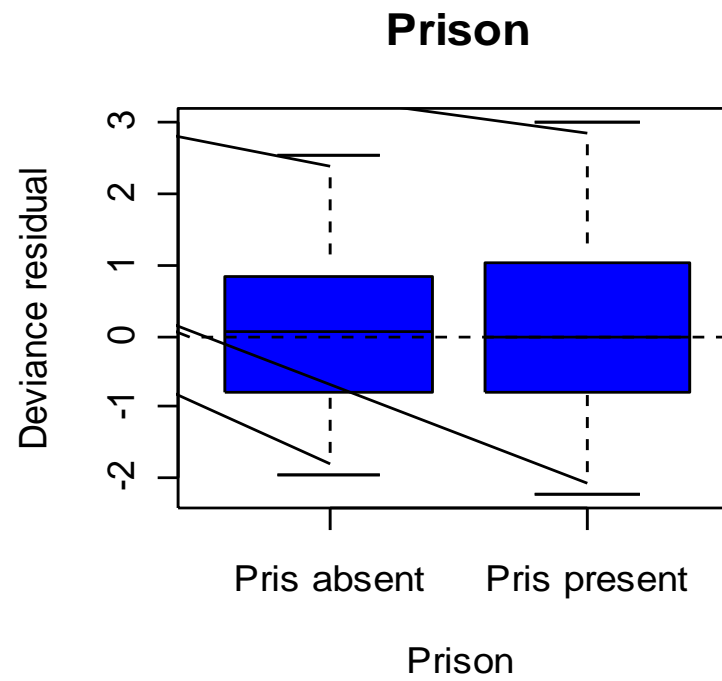
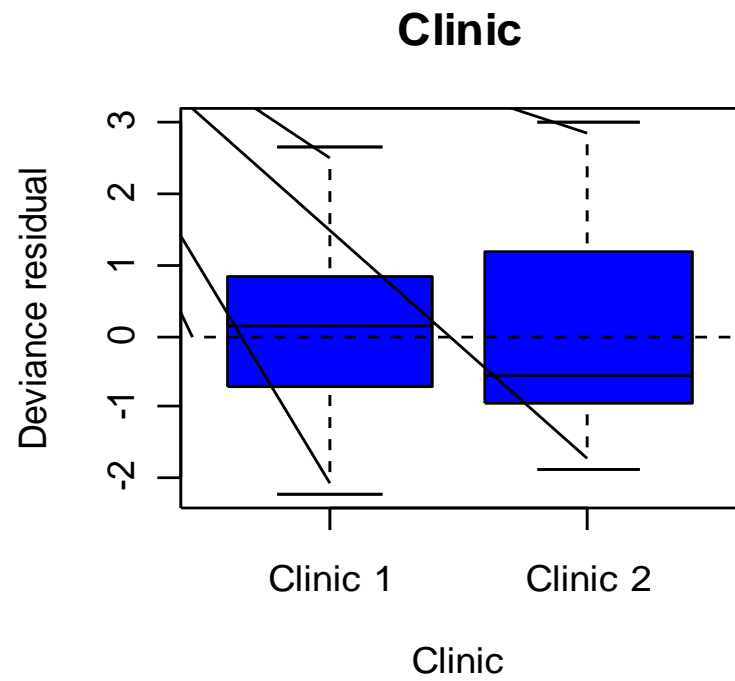
- Phân tích số dư sẽ cung cấp thông tin đánh giá sự phù hợp của mô hình mỗi nguy theo tỷ lệ
 - xác định các biện pháp đòn bẩy và biện pháp ảnh hưởng, và có thể được sử dụng để đánh giá các giả định về mỗi nguy theo tỷ lệ
 - theo định nghĩa, số dư cho các quan sát bị kiểm duyệt là âm tính và các số dư có tác dụng để có được cảm giác về số lượng kiểm duyệt trong bộ dữ liệu: một lượng lớn kiểm duyệt sẽ dẫn đến ‘dải’ của các điểm số dư

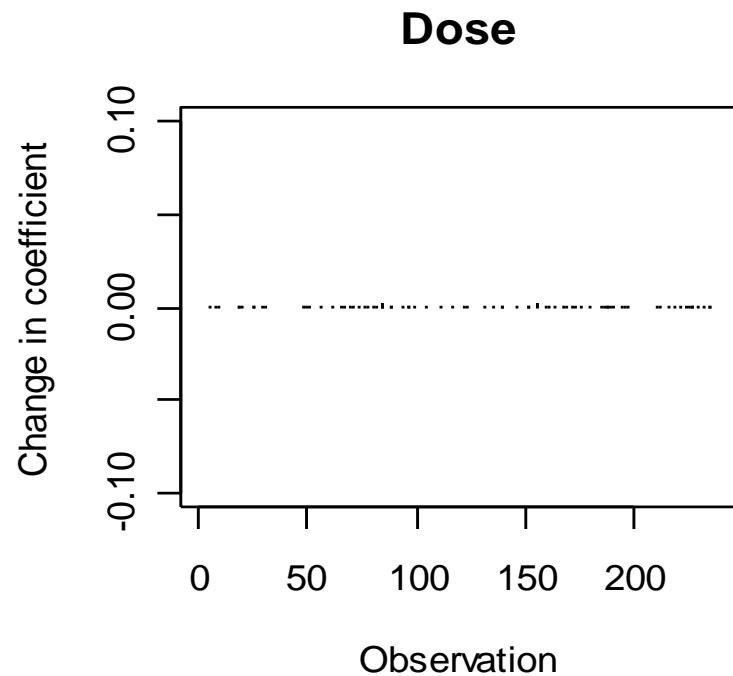
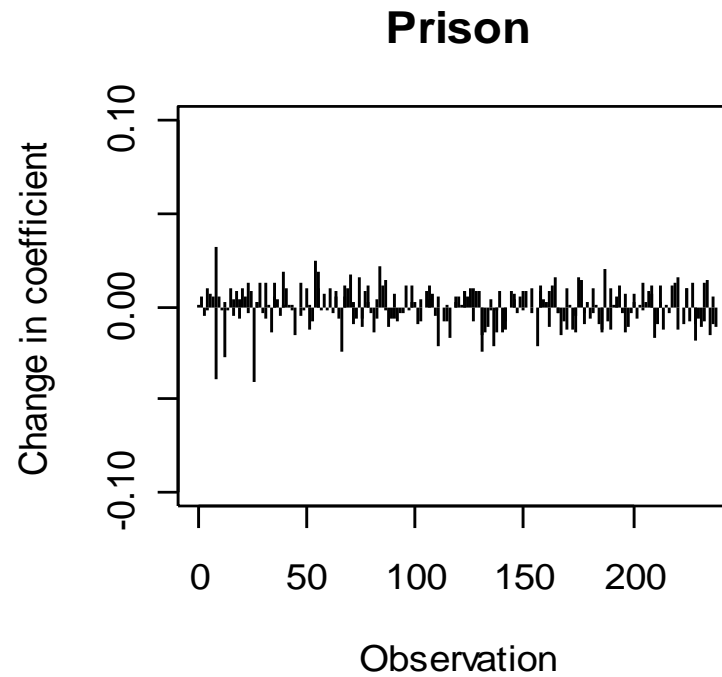
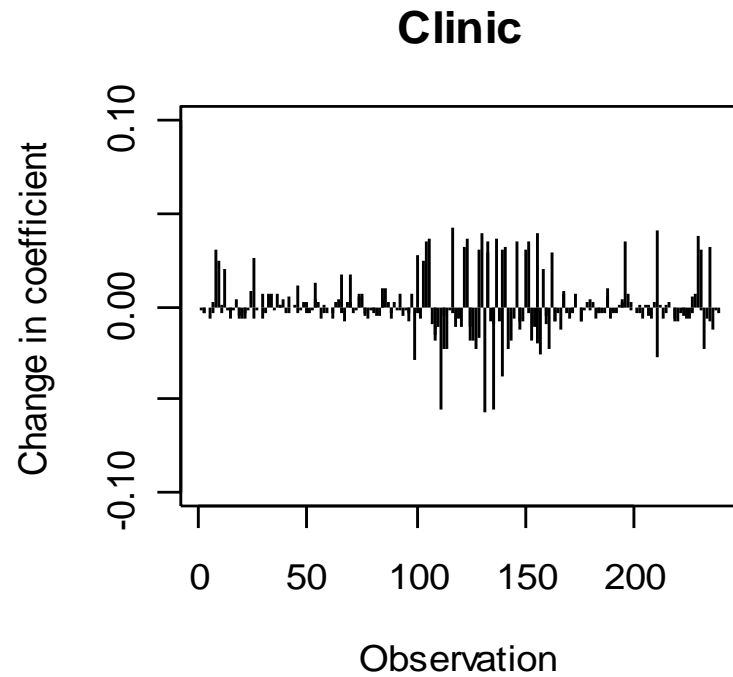
Số dư (Residuals)

- Số dư Martingale:
 - sự khác biệt giữa số lượng sự kiện được quan sát cho một cá thể và số lượng dự kiến được đưa ra cho mô hình, thời gian theo dõi và tiến trình quan sát của bất kỳ các đồng biến thay đổi theo thời gian nào
 - Lập mối nguy tích lũy và số dư Cox-Snell sẽ mang lại một đường thẳng trong một mô hình phù hợp
- Các số dư sai lệch:
 - Là một biến đổi bình của việc thường hóa số dư Martingale
 - hữu ích để xác định các ngoại lệ

Số dư (Residuals)

- Các số dư tính điểm (Score residuals):
 - một mảng ba chiều với các hướng của chủ thể, đồng biến và thời gian
 - hữu ích để đánh giá ảnh hưởng của cá thể và ước tính phương sai
- Số dư Schoenfeld :
 - Đối với chủ thể k th tác động lên biến diễn giải j th
 - hữu ích để đánh giá các mối nguy theo tỷ lệ
 - cung cấp lực chẩn đoán lớn hơn so với số dư chưa được đánh giá





Thống kê ảnh hưởng. Các sơ đồ sau đây cho thấy sự thay đổi trong từng hệ số hồi quy khi mỗi quan sát được xóa khỏi dữ liệu. Các thay đổi được chia tỷ lệ theo đơn vị của phương sai và các thay đổi nhỏ hơn 0,1 ít được quan tâm.

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Residuals
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

Goodness of fit

- Như trong tất cả các phân tích hồi quy một số loại đo lường tương tự đối với R^2 có thể được quan tâm
 - Schemper và Stare (1996) cho thấy rằng không có một biện pháp đơn giản, dễ tính toán, dễ diễn giải nào có thể đánh giá kiểm định goodness-of-fit của một mô hình hồi quy mỗi nguy theo tỷ lệ
 - thông thường, một mô hình hoàn toàn phù hợp ở giá trị mặt tiền có thể có R^2 rất thấp do tỷ lệ lưu hành của kiểm duyệt cao

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Residuals
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

TABLE 5: Cox proportional hazards regression model showing the effect of region, holding size and holding type on the monthly hazard of experiencing a BSE index case

Explanatory variable	Number of holdings	Number BSE-positive	Regression coefficient (se)	P	Hazard ratio†	95% CI of hazard ratio
Region				<0.01*		
Eastern	5924†	1137	0.7981 (0.0359)		2.22	2.07-2.38
Mid and West	22,970	6753	0.5830 (0.0241)		1.79	1.71-1.88
Northern	16,283	4284	0.5632 (0.0254)		1.76	1.67-1.85
Scotland	16,635	2627			1.00	
South east	7702	2012	0.8865 (0.0302)		2.43	2.29-2.58
South west	24,529	8336	0.8746 (0.0234)		2.40	2.29-2.51
Wales	20,809	4315	0.5127 (0.0256)		1.67	1.59-1.76
Holding size				<0.01		
1-6	31,469	644	-0.8344 (0.0468)		0.43	0.40-0.48
7-21	25,142	2021			1.00	
22-53	29,145	8479	1.061 (0.0257)		2.89	2.75-3.04
>53	29,702	18,320	1.776 (0.0254)		5.91	5.62-6.21
Holding type				<0.01		
Dairy	38,576	21,191	1.117 (0.0168)		3.06	2.96-3.16
Mixed	9955	2221	0.5462 (0.0253)		1.73	1.64-1.81
Beef suckler	62,896	6052			1.00	

Likelihood ratio test statistic 35,570; df 11; $P < 0.01$

* The significance of inclusion of the six region variables in the model

† Cases with missing values have been excluded, so counts vary slightly from those shown in Table 4

‡ Interpretation: compared with the reference category (holdings in Scotland), after adjusting for the effect of the size and type of holding, up to June 30, 1997, holdings in the Eastern region of England were at 2.22 (95% CI 2.07 to 2.38) times the monthly hazard of having a BSE index case

CI Confidence interval

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lockhart, J., Lin, D., Jackson, R., 2000a. Temporal aspects of the bovine spongiform encephalopathy epidemic in Great Britain: Individual animal-associated risk factors for disease. *Veterinary Record* 147, 349 - 354.

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Residuals
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ

Xử lý các mối nguy phi tỷ lệ

- Các lựa chọn
 1. Sự phân tầng
 2. Đưa ra đồng biến phụ thuộc thời gian

Xử lý các mối nguy phi tỷ lệ

- Sự phân tầng
 - Đưa ra một hàm nguy cơ cơ sở riêng biệt đối với từng cấp của phân tầng mà xung đột với giả định nguy cơ theo tỷ lệ
 - Dùng để khắc phục vấn đề, nhưng bằng cách này, chúng ta không thể tính được tỷ số nguy cơ cho biến phân tầng vì hiệu ứng của nó được ‘gắn vào’ vào hàm nguy cơ cơ bản.

```
addict.cph04 <- coxph(Surv(stop, status, type = "right") ~  
strata(clinic) + prison + dose, method = "breslow", data = dat);  
summary(addict.cph04)
```

Call:

```
coxph(formula = Surv(stop, status, type = "right") ~ strata(clinic)  
+ prison + dose, data = dat, method = "breslow")
```

	coef	exp(coef)	se(coef)	z	p
prison1	0.376	1.457	0.16889	2.23	2.6e-02
dose	-0.035	0.966	0.00645	-5.42	5.9e-08

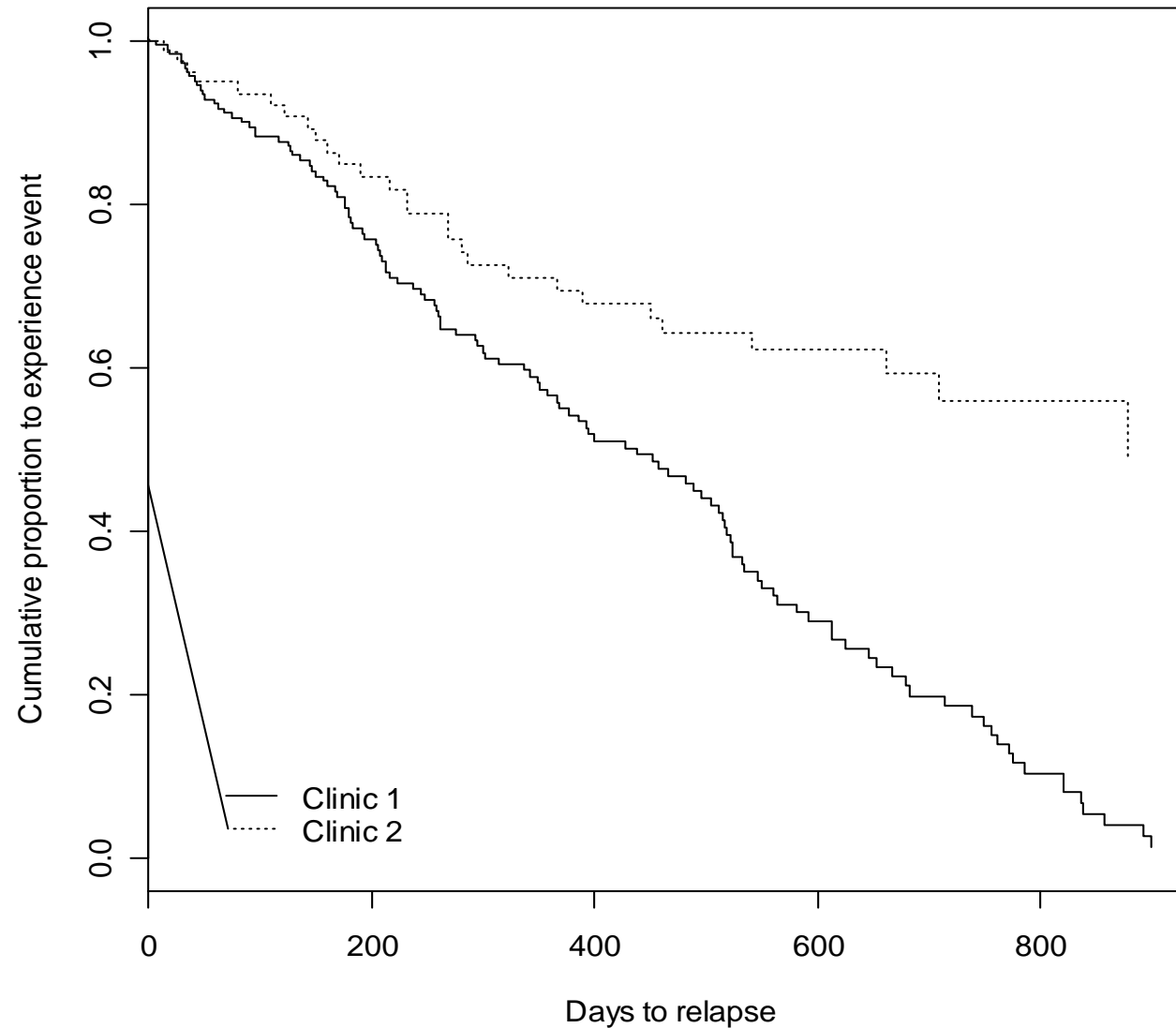
	exp(coef)	exp(-coef)	lower .95	upper .95
prison1	1.457	0.686	1.046	2.029
dose	0.966	1.036	0.953	0.978

Rsquare= 0.131 (max possible= 0.994)

Likelihood ratio test= 33.5 on 2 df, p=5.29e-08

Wald test = 32.3 on 2 df, p=9.73e-08

Score (logrank) test = 33.0 on 2 df, p=6.98e-08



```
plot(survfit(addict.cph04), lty = c(1,2), xlab = "Days to relapse",  
ylab = "Cumulative proportion to experience event")  
legend(x = "topright", legend = c("Clinic 1","Clinic 2"), lty =  
c(1,2), bty = "n");
```

Xử lý các mối nguy phi tỷ lệ

- Các đồng biến phụ thuộc thời gian
 - Nếu một đồng biến được cố định (ví dụ nó không thay đổi theo thời gian nhưng tác động của nó thay đổi theo thời gian), chúng ta có thể đánh giá tác động phụ thuộc thời gian này bằng cách chia thời gian theo dõi thành các khoảng thời điểm khác nhau.
 - Sau đó, chúng ta gắn với các mô hình mối nguy theo tỷ lệ đối với tỷ lệ sinh tồn trong từng khoảng thời gian và so sánh các hệ số cho từng đồng biến trong các khoảng thời gian khác nhau
 - Nếu hệ số thay đổi theo thời gian, chúng ta có bằng chứng về các mối nguy phi tỷ lệ
 - do đó, việc chẩn đoán đối với sự phi tỷ lệ của các mối nguy cũng là một giải pháp

```
library(survival); setwd("D:\\TEMP");  
dat <- read.table("addict.csv", header = TRUE, sep = ",");  
head(dat);
```

id	start	stop	status	clinic	prison	dose
1	0	428	1	1	0	50
2	0	275	1	1	1	55
3	0	262	1	1	0	55
4	0	183	1	1	0	30
5	0	259	1	1	1	65
6	0	714	1	1	0	55

Reformat the data:

id	start	stop	status	clinic	prison	dose
1	0	365	0	1	0	50
1	365	428	1	1	0	50
2	0	275	1	1	1	55
3	0	262	1	1	0	55
4	0	183	1	1	0	30
5	0	259	1	1	1	65

Recode the `clinic` variable to make Clinic 2 (the better performing clinic) the reference category:

```
dat$clinic <- as.vector(ifelse(dat$clinic == 2, 0, 1));
```

First of all we'll consider the period before 365 days. Create a new variable called `t1` such that `t1=1` if the time to event is less than or equal to 365 days and zero otherwise:

```
t1 <- rep(0, length(dat[,1]));  
t1[dat$stop <= 365] <- 1;  
dat <- cbind(dat, t1);
```

Using this coding, the reported hazard for the `clinic * t1` interaction will be for Clinic 1 when time is less than or equal to 365 days.

Next consider the period after 365 days. Create a new variable called `t2` such that `t2 = 1` if the time to event is greater than 365 days and one otherwise:

```
t2 <- rep(0, length(dat[,1]));  
t2[dat$stop > 365] <- 1;  
dat <- cbind(dat, t2);
```

Using this coding, the reported hazard for the `clinic * t2` interaction will be for Clinic 1 when time is greater than 365 days.


```
head(dat) ;
```

id	start	stop	status	clinic	prison	dose	t1	t2
1	0	365	0	1	0	50	1	0
1	365	428	1	1	0	50	0	1
2	0	275	1	1	1	55	1	0
3	0	262	1	1	0	55	1	0
4	0	183	1	1	0	30	1	0
5	0	259	1	1	1	65	1	0

Now fit the model:

```
addict.cph05 <- coxph(Surv(start, stop, event = status, type = '
counting') ~ prison + dose + I(clinic * t1) + I(clinic * t2), method
= 'breslow', data = dat);
summary(addict.cph05);
```

Variable	Subjects	Failed	Coefficient (SE)	P	Hazard ratio (95%)
Prison:				0.03	
Absent	127	81	-		1.00
Present	111	69	0.3650 (0.1684)		1.44 (1.04 - 2.00)
Dose	238	150	-0.0353 (0.0064)	< 0.01	0.96 (0.95 - 0.98)
Clinic × t1	118	87	0.4802 (0.2548)	0.06	1.62 (0.98 - 2.66) ^a
Clinic × t2	120	63	1.8103 (0.3861)	< 0.01	6.11 (2.87 - 13.03)

^a Interpretation: compared with the reference category (patients from Clinic 2) when days on treatment is less than 365, after adjusting for the effect of methadone dose and prison record, patients from Clinic 1 had 1.62 (95% CI 0.98 - 2.66) times the daily hazard withdrawing from the treatment program.

Xử lý các mối nguy phi tỷ lệ

- Kết luận
 - Khi số ngày được điều trị dưới 365 ngày, các bệnh nhân từ Phòng khám 1 có 1.62 lần (95% CI 0.98 đến 2.66) nguy cơ gia tăng tái phát so với bệnh nhân từ Phòng khám 2
 - khi ngày điều trị lớn hơn 365 ngày, bệnh nhân từ Phòng khám 1 có 6.11 lần (95% CI 2.87 đến 13.0) nguy cơ gia tăng tái phát so với bệnh nhân từ Phòng khám 2
 - một khi bệnh nhân đã ở Phòng khám 1 trong hơn 12 tháng, nguy cơ tái phát của bệnh nhân tăng lên rõ rệt

Lộ trình

- Tổng quan
- Xây dựng mô hình
- Kiểm tra giả định các mối nguy theo tỷ lệ
- Residuals
- Goodness of fit
- Trình bày kết quả
- Xử lý các mối nguy phi tỷ lệ



COMMONWEALTH OF AUSTRALIA

Copyright Regulations 1969

WARNING

This material has been reproduced and communicated to you by or on behalf of the University of Melbourne pursuant to Part VB of the *Copyright Act 1968 (the Act)*. The material in this communication may be subject to copyright under the Act. Any further copying or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice.