
Phân tích theo thời gian:

Sinh tồn tham số và phi tham số

Mark Stevenson

Faculty of Veterinary and Agricultural Sciences

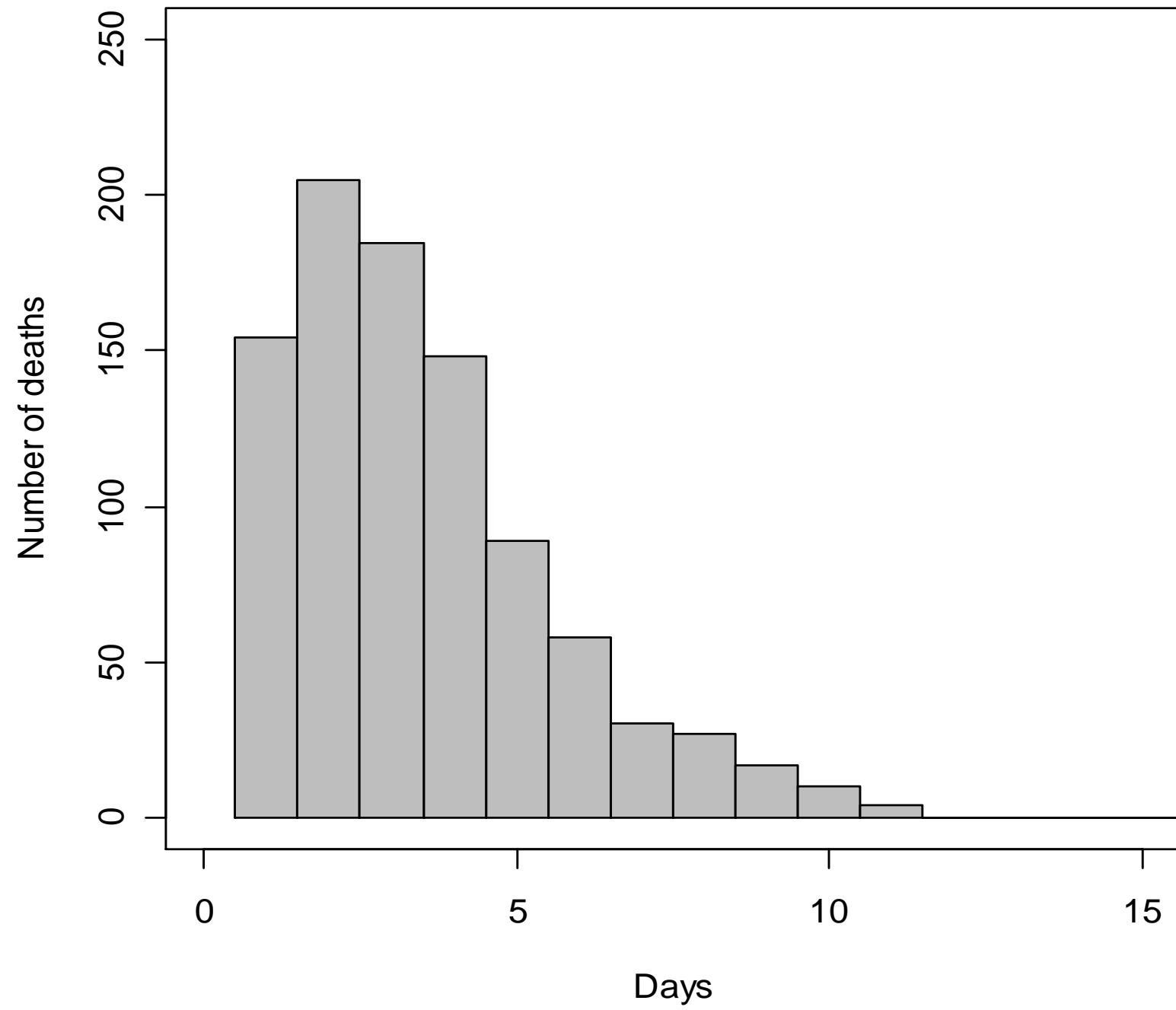
The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au

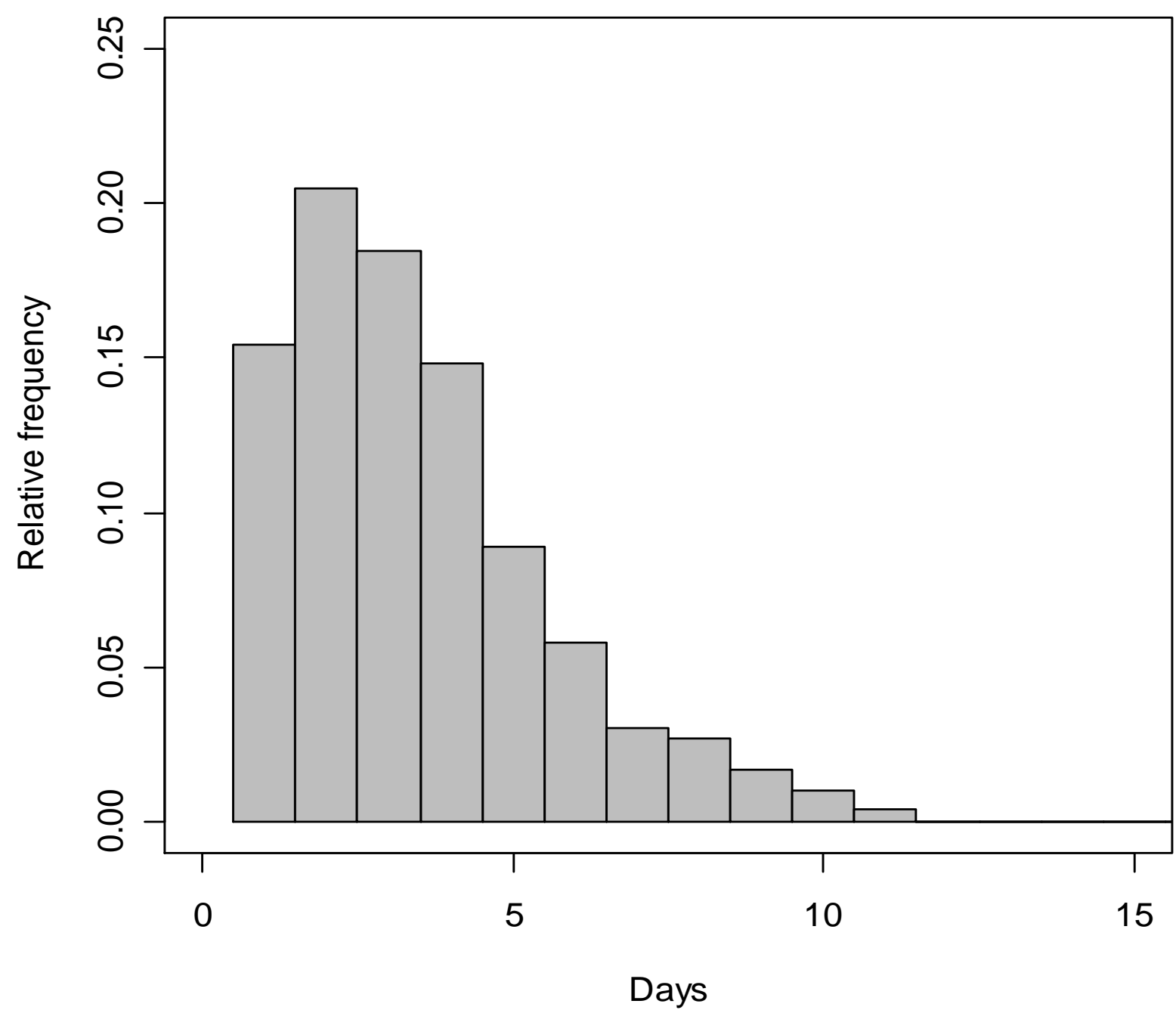
Lộ trình

- Sinh tồn và nguy cơ
- Các hàm sinh tồn có tham số
 - Exponential (số mũ)
 - Weibull
- Các hàm sinh tồn phi tham số
- Trình bày

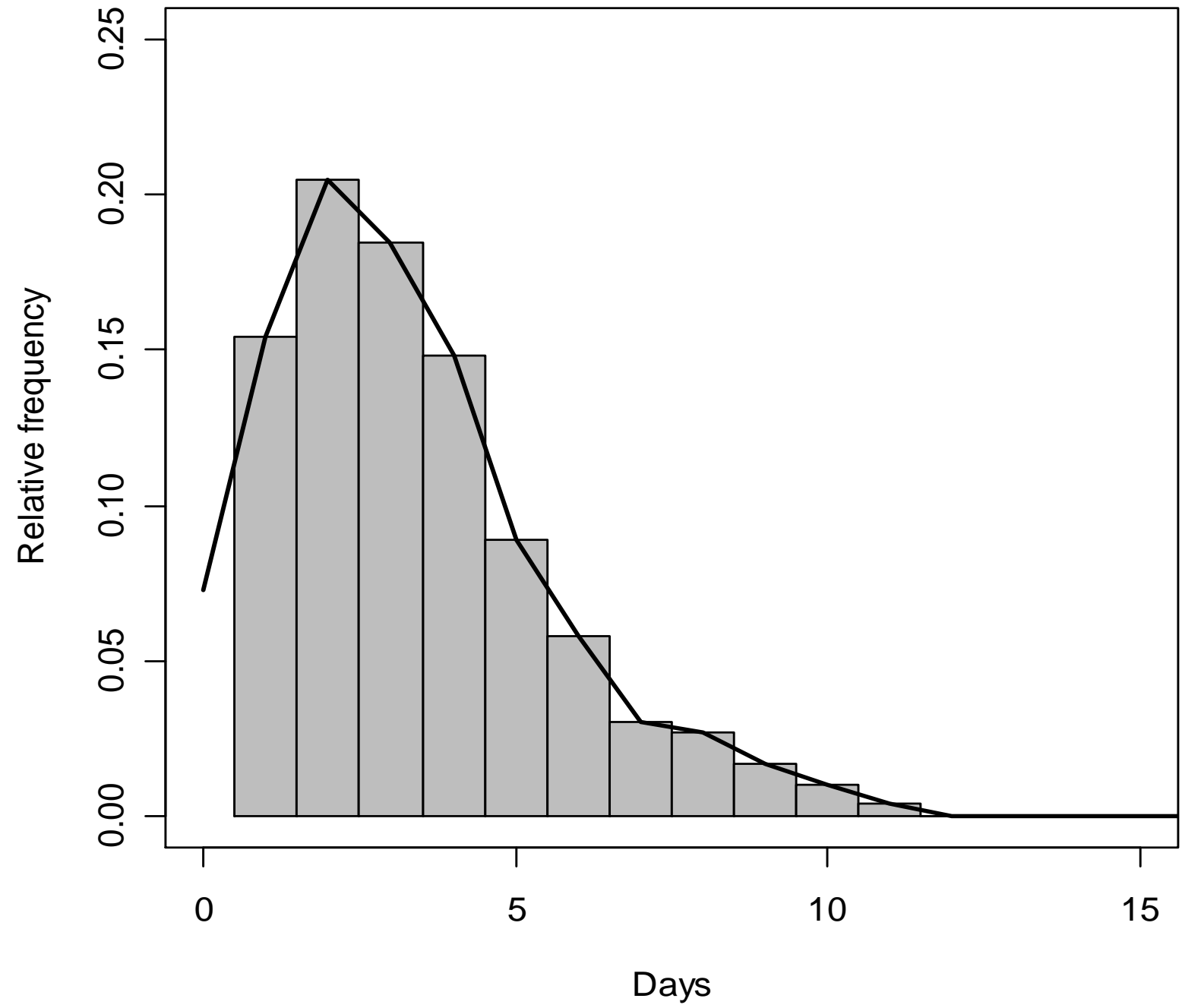
Biểu đồ tần suất cho thấy số ca tử vong là một hàm của thời gian.



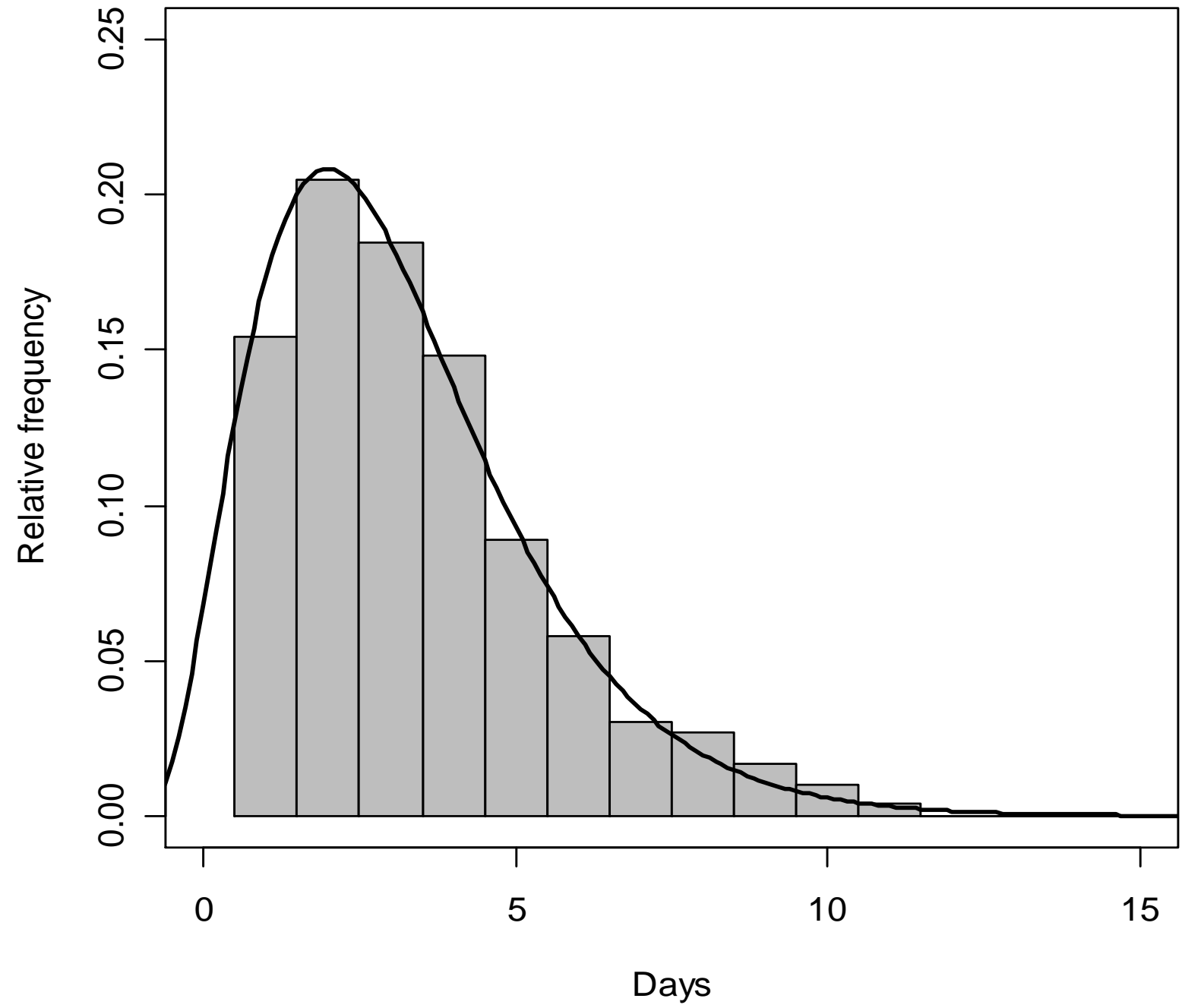
Biểu đồ tần suất cho thấy tần suất tương đối của ca tử vong là một hàm của thời gian.



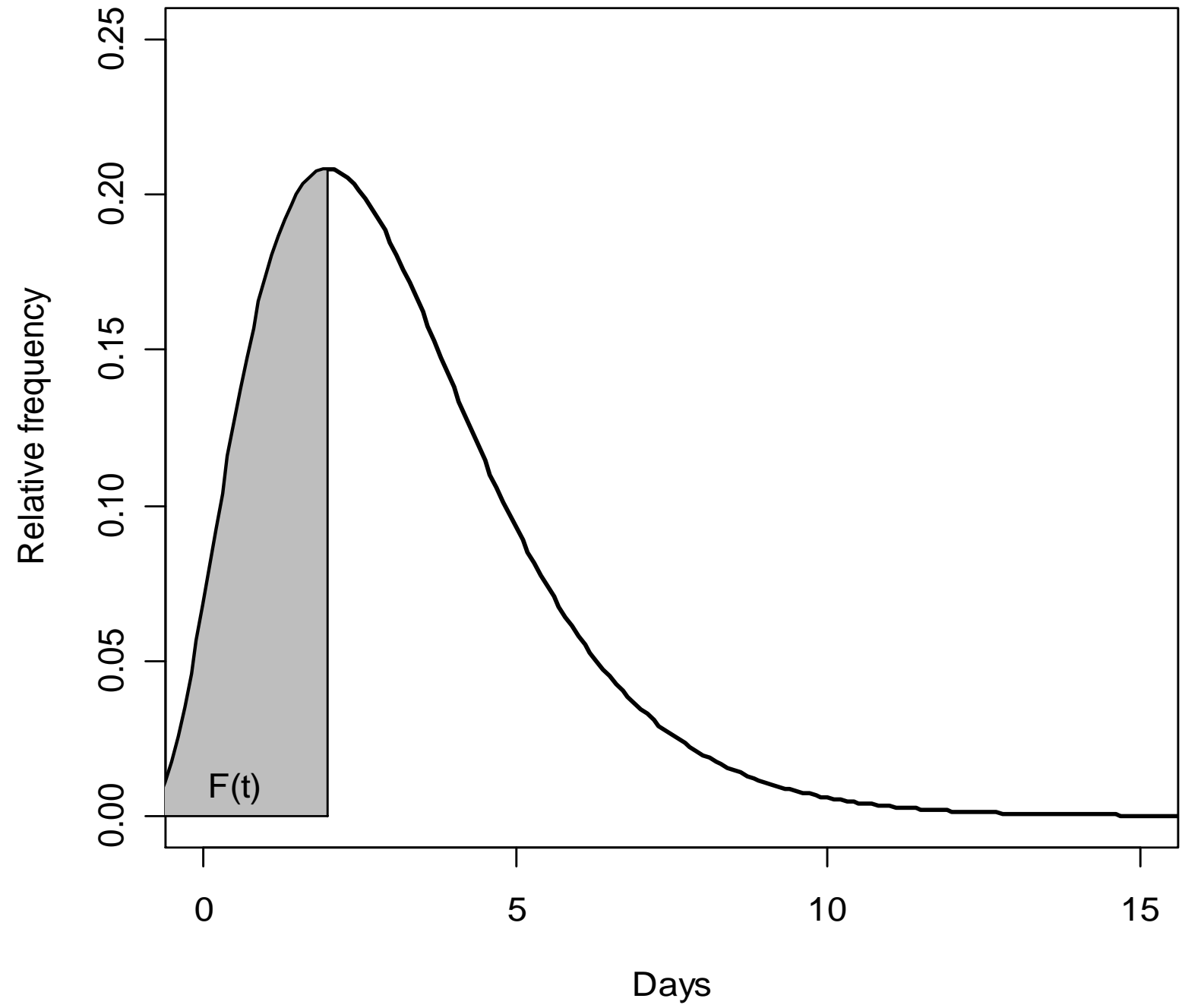
$f(t)$ = hàm mật độ tử vong.



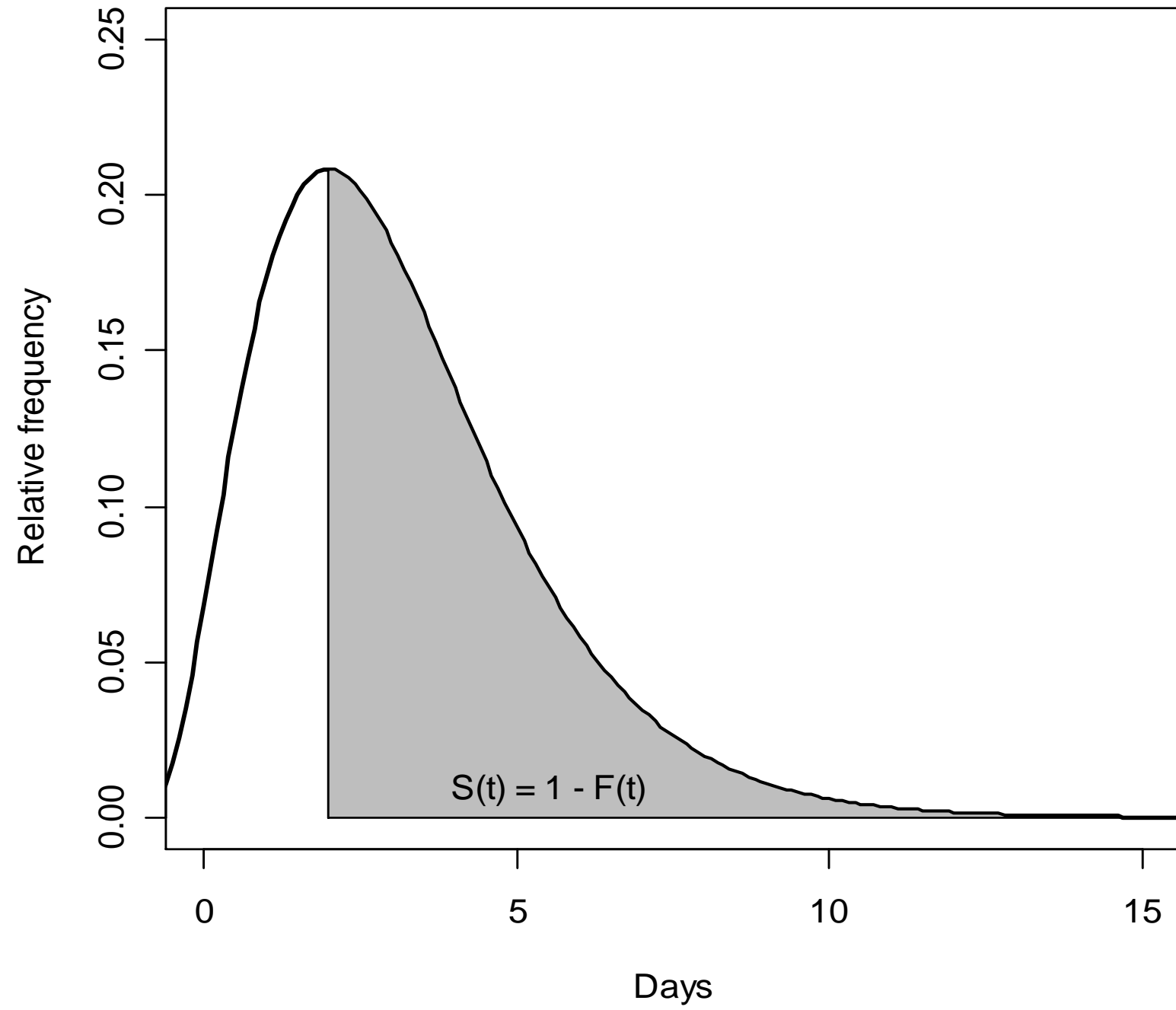
$f(t)$ = hàm mật độ tử vong.



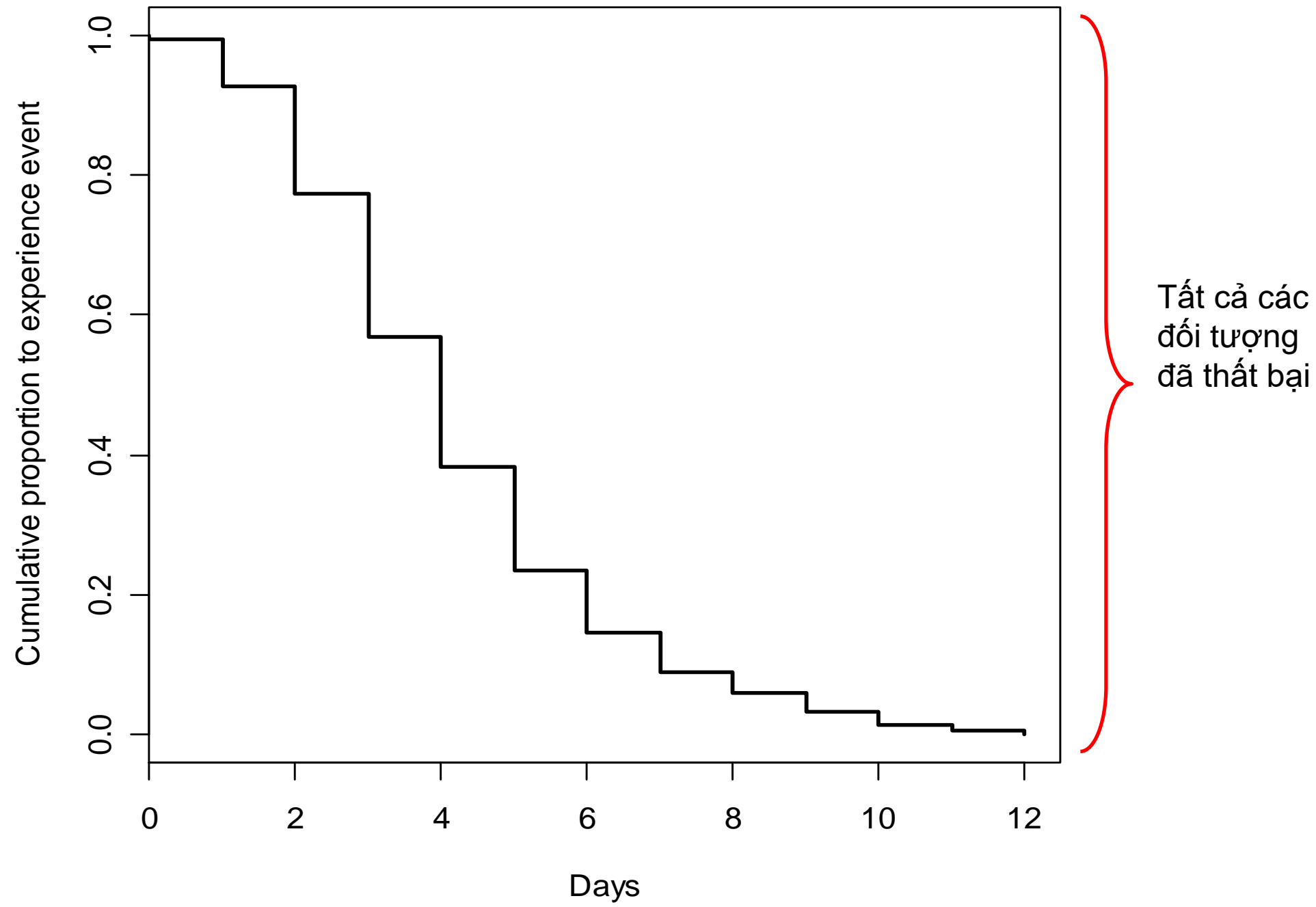
$F(t)$ = hàm thất bại = tỷ lệ không sống sót qua quãng thời gian t .



$S(t)$ = hàm thất bại = tỷ lệ của nhóm sống sót trong quãng thời gian $t = 1 - F(t)$.



$S(t)$ = hàm thất bại = tỷ lệ của nhóm sống sót trong quãng thời gian $t = 1 - F(t)$.



Hàm thất bại: $F(t) = 1 - S(t)$

Khả năng không sống sót trong quãng thời gian t

Tỷ lệ thất bại tức thời: $f(t) = \frac{dF(t)}{dt}$

Độ dốc của hàm $F(t)$

Hàm nguy cơ tức thời: $h(t) = \frac{f(t)}{S(t)}$

Khả năng thất bại ở quãng thời gian t

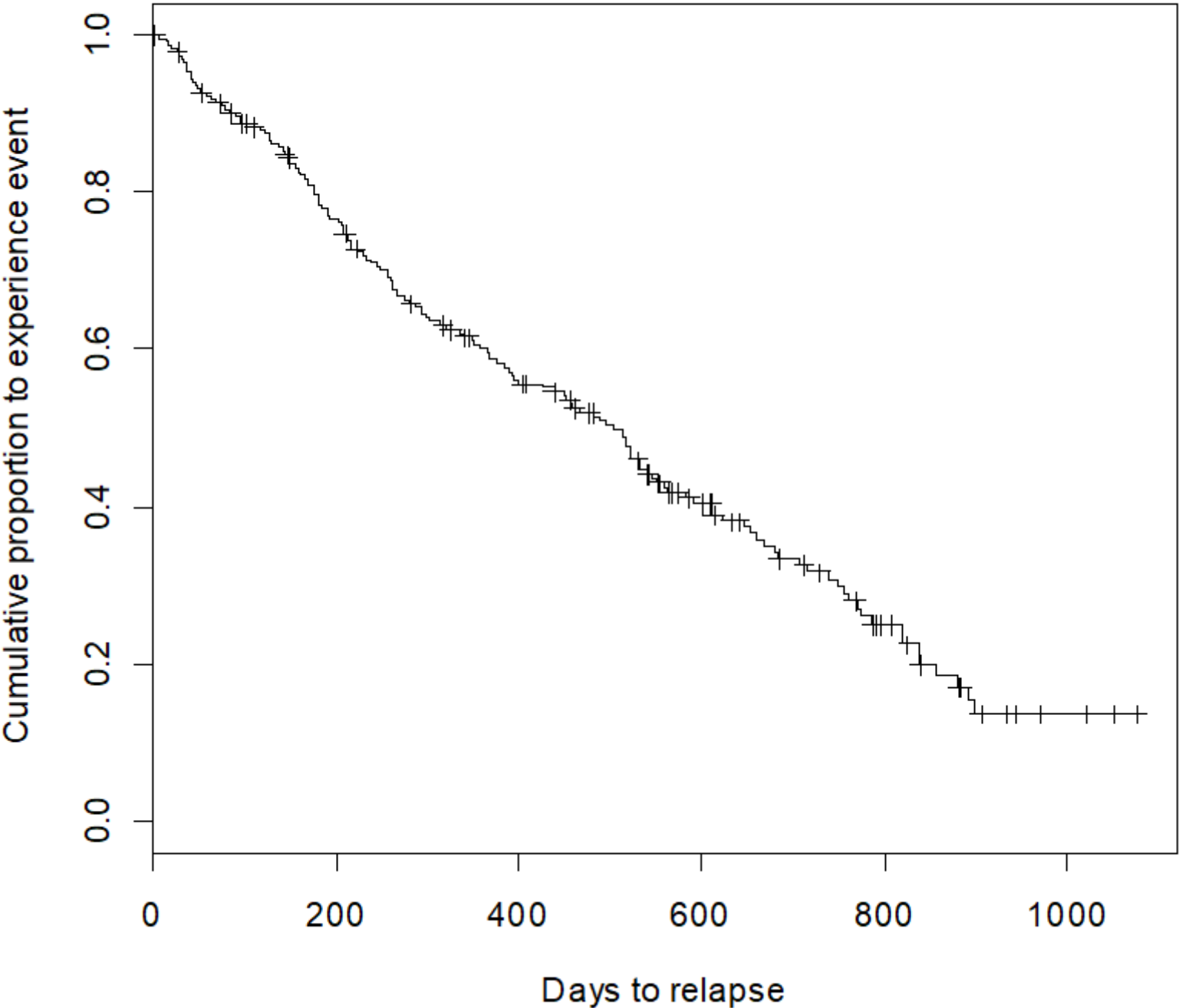
Hàm nguy cơ tích lũy: $H(t) = -\ln S(t)$

Nguy cơ tích lũy theo thời gian

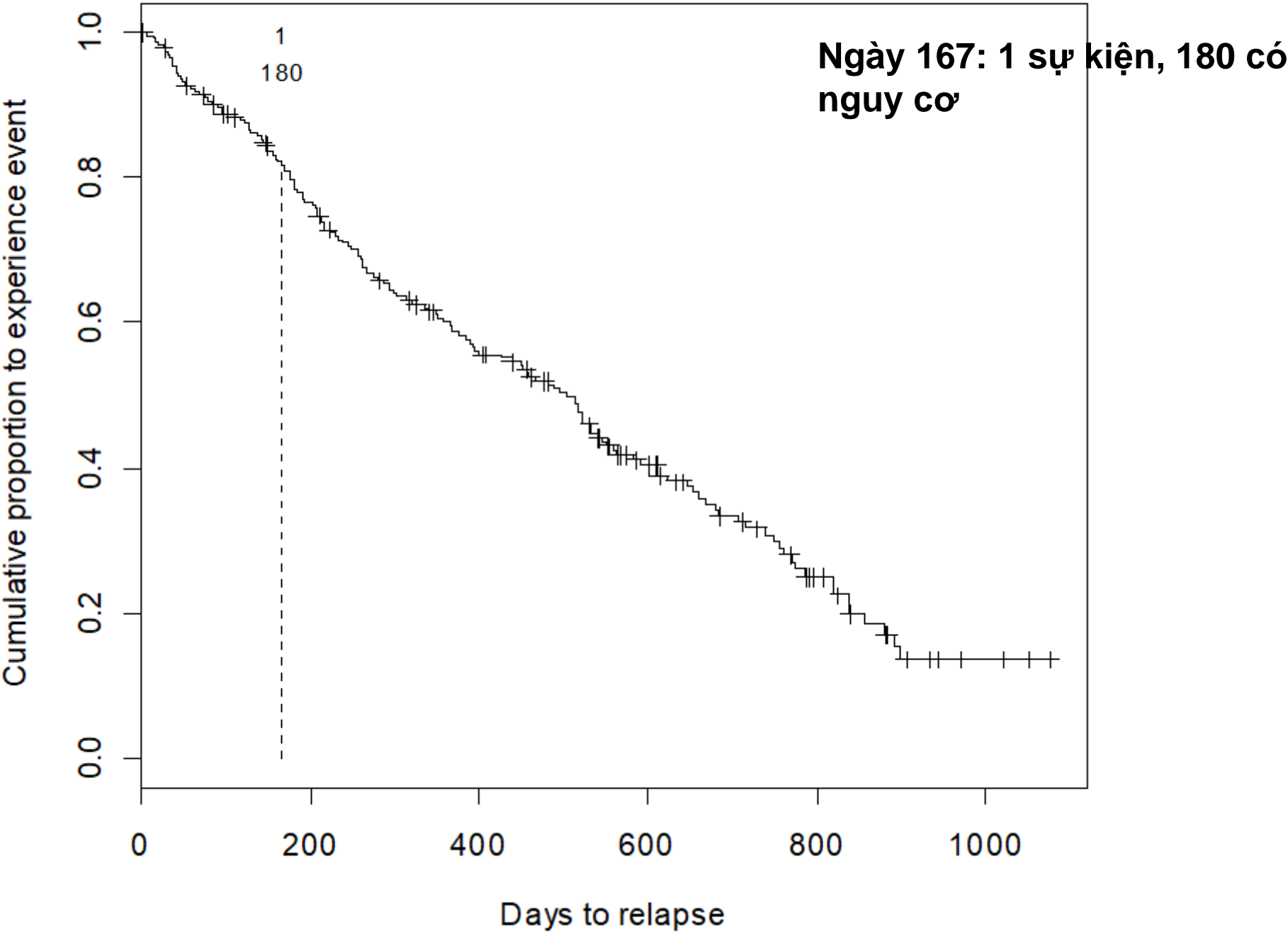
Sinh tồn và nguy cơ

- Nguy cơ tức thời $h(t)$
 - \equiv 'hazard' (Dohoo and Martin)
 - xác suất của một sự kiện xảy ra tại thời điểm t mặc dù thực tế nó chưa xảy ra

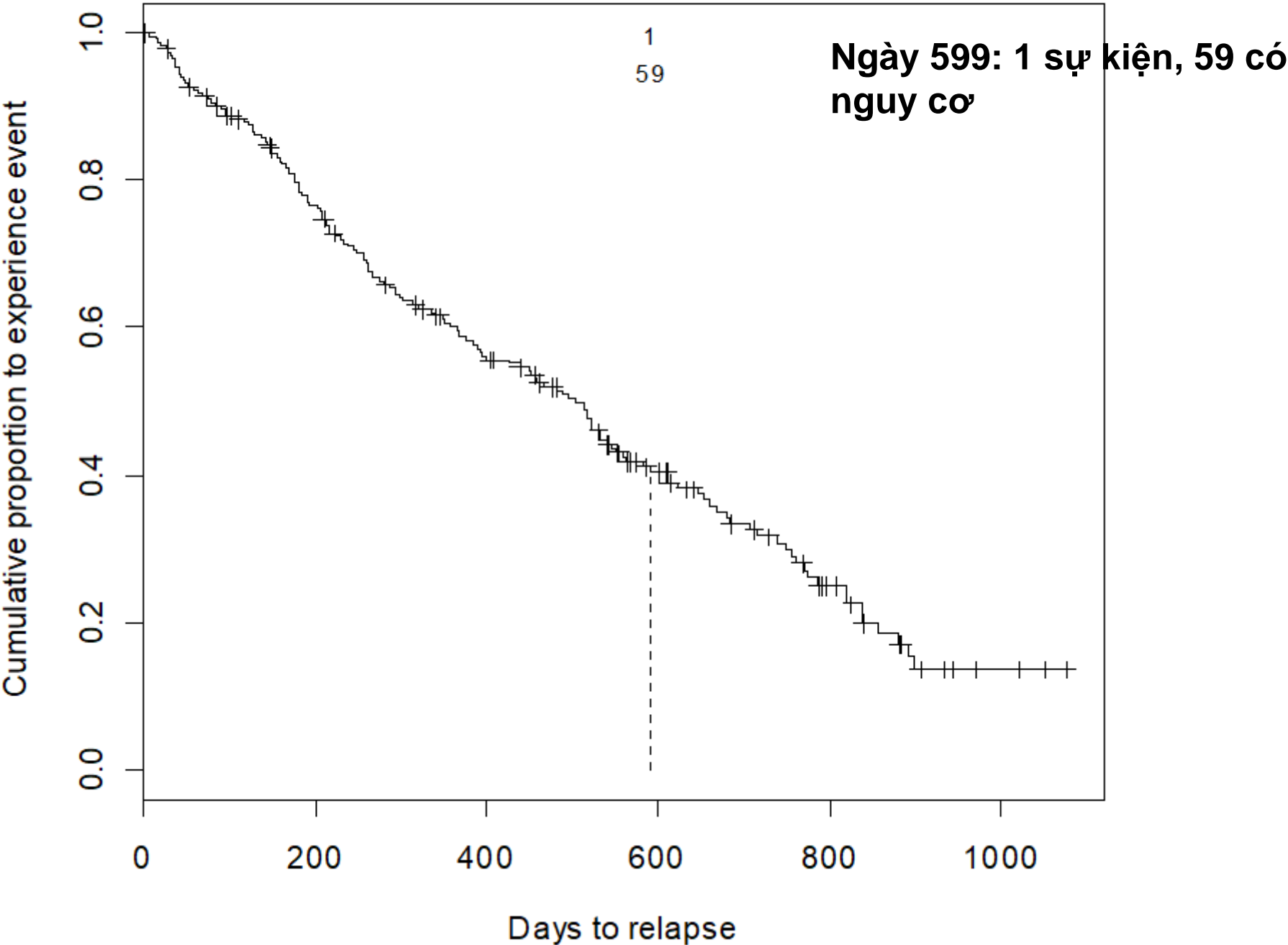
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của người nghiện ma túy đã không tái phát sau khi xuất viện sau khi điều trị như một hàm số thời gian



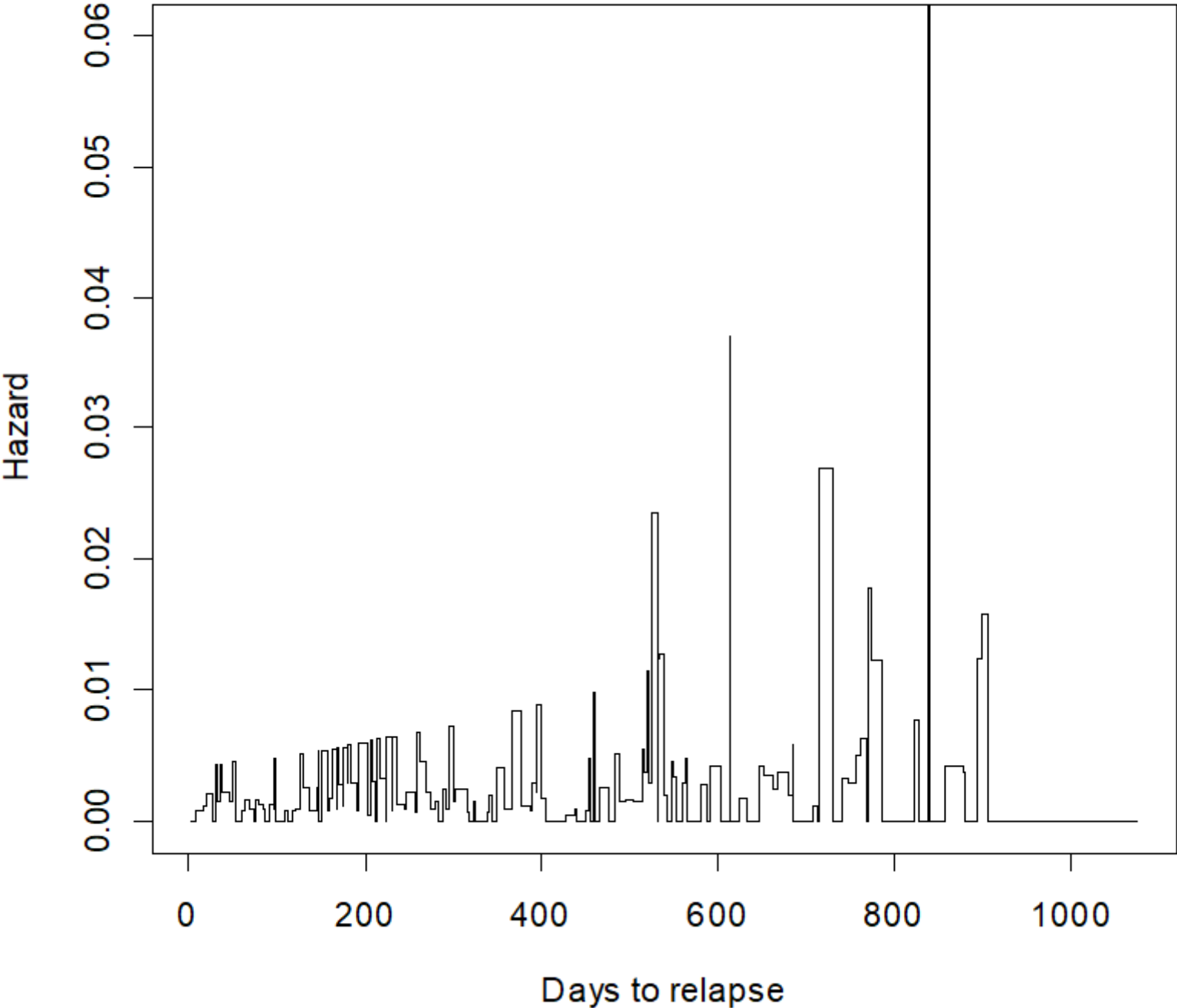
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của người nghiện ma túy đã không tái phát sau khi xuất viện sau khi điều trị như một hàm số thời gian.



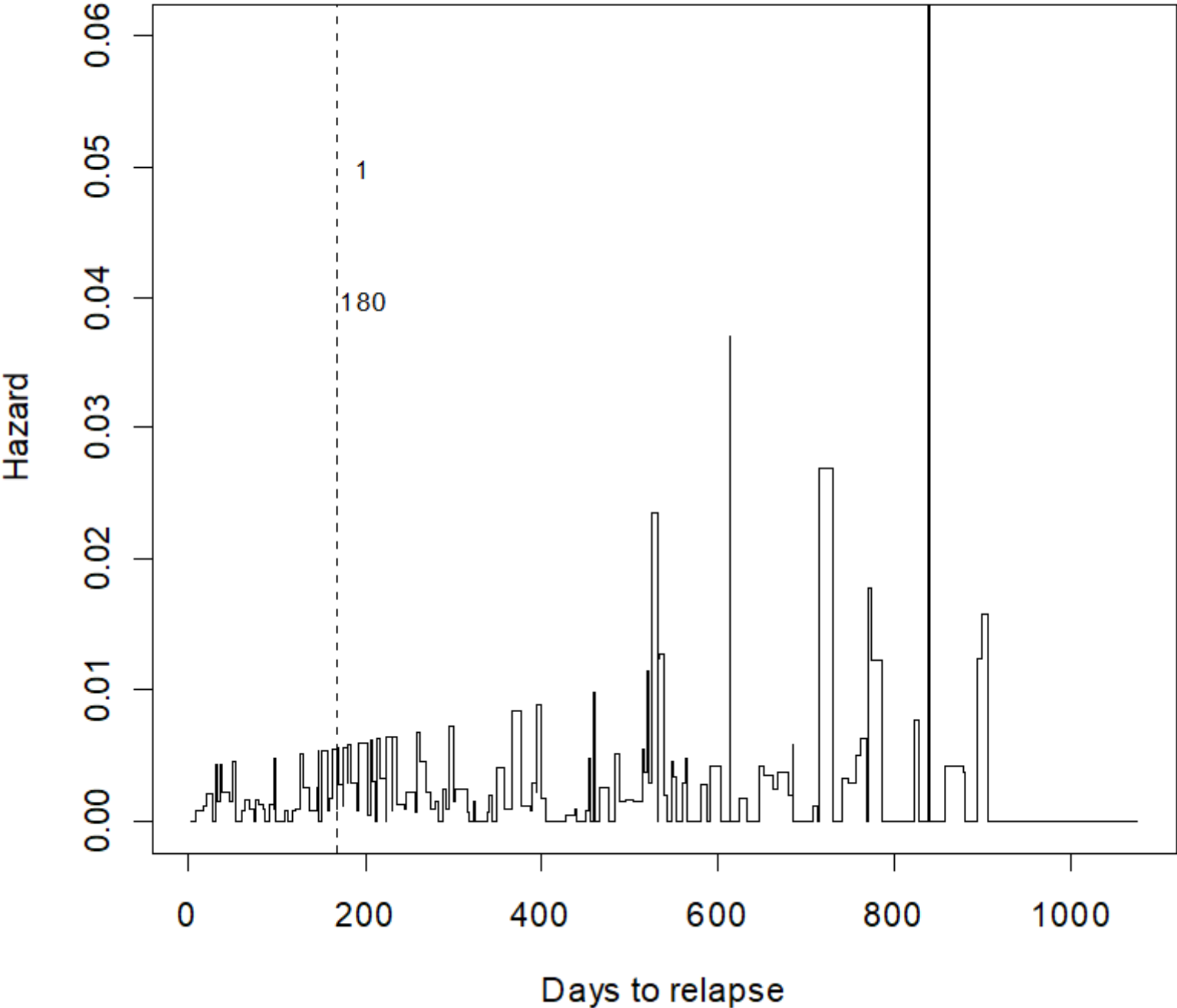
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của người nghiện ma túy đã không tái phát sau khi xuất viện sau khi điều trị như một hàm số thời gian.



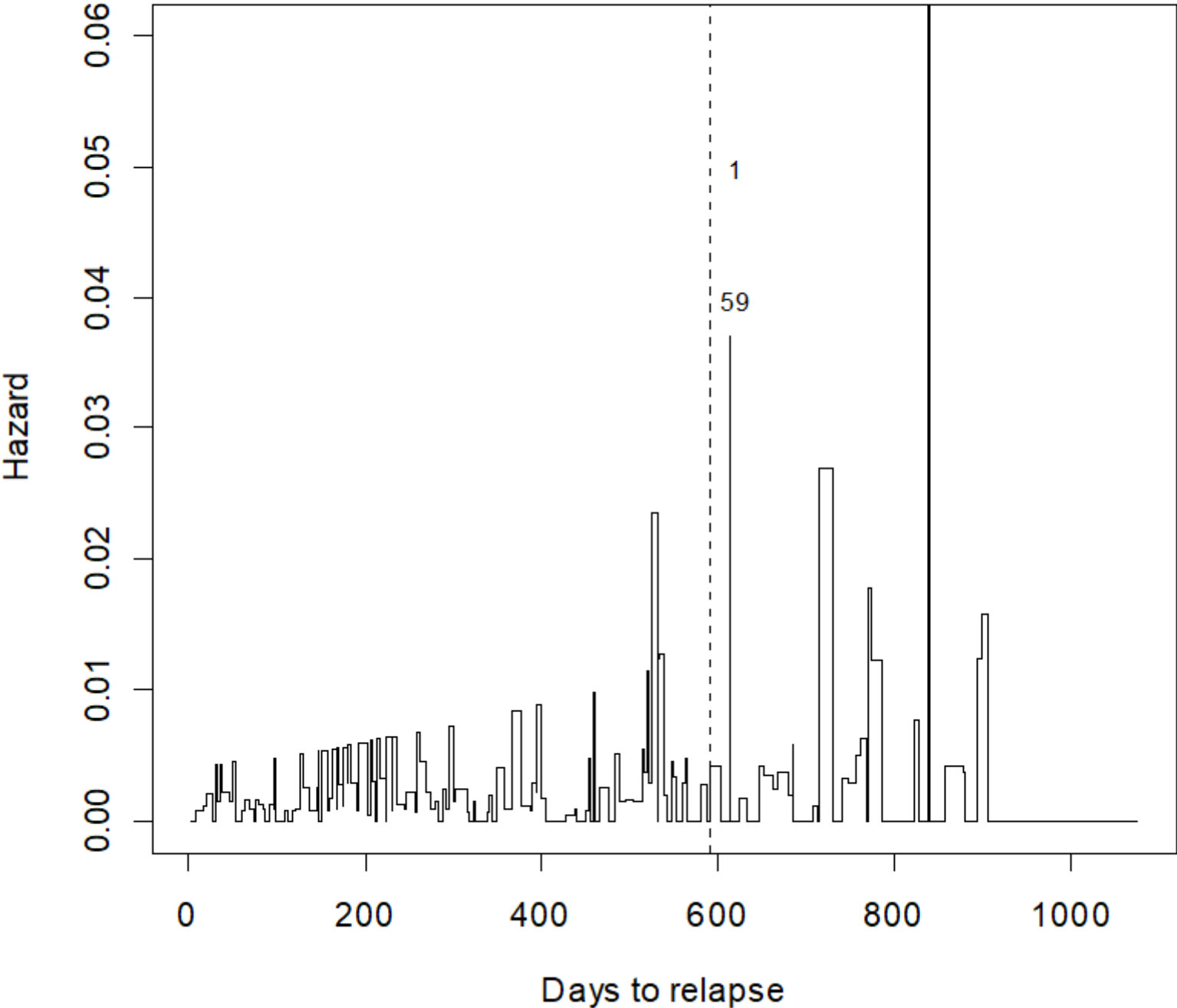
Nguy cơ tái phát tức thời ở người nghiện ma túy sau khi xuất viện.



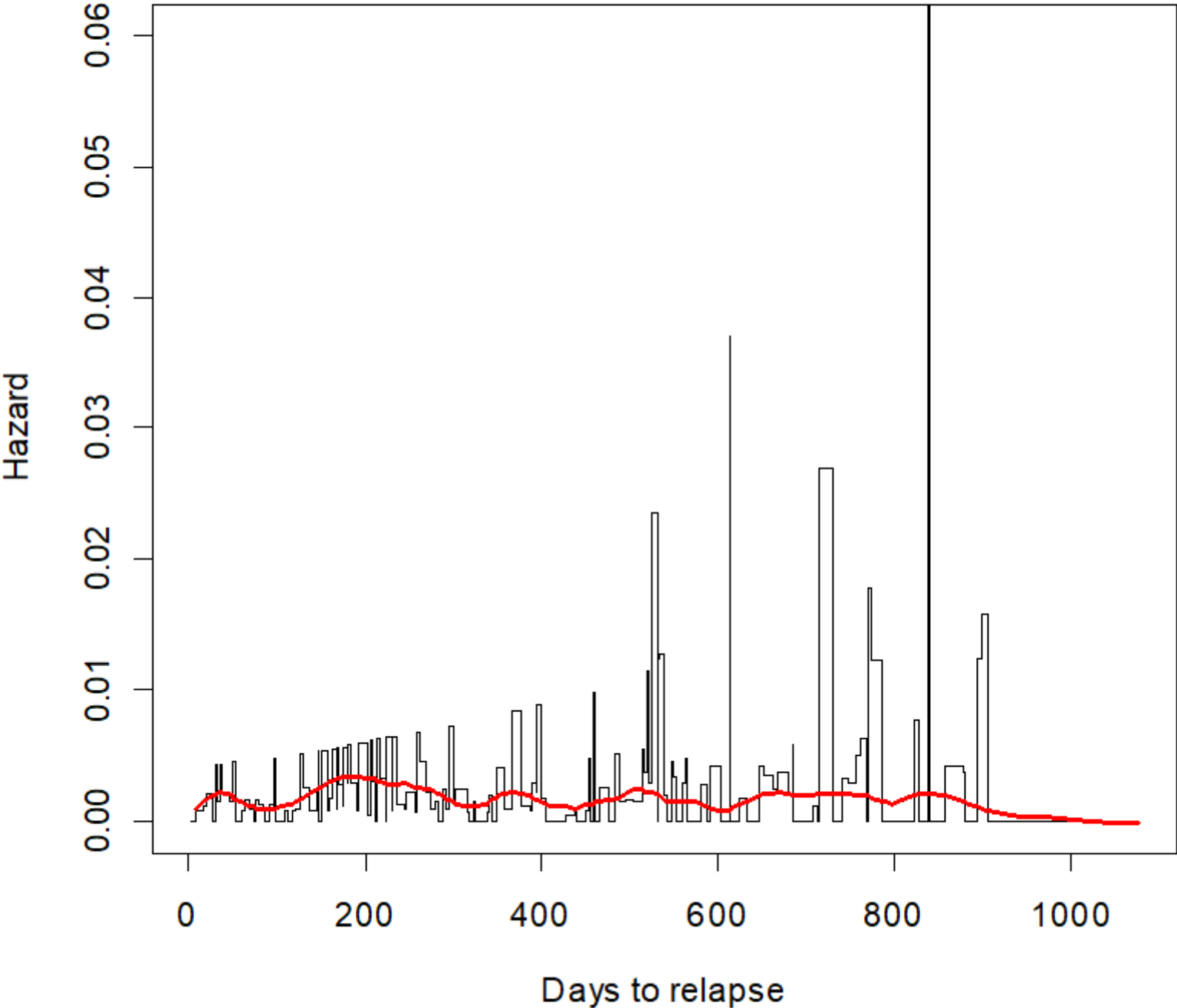
Nguy cơ tái phát tức thời ở người nghiện ma túy sau khi xuất viện.



Nguy cơ tái phát tức thời ở người nghiện ma túy sau khi xuất viện.



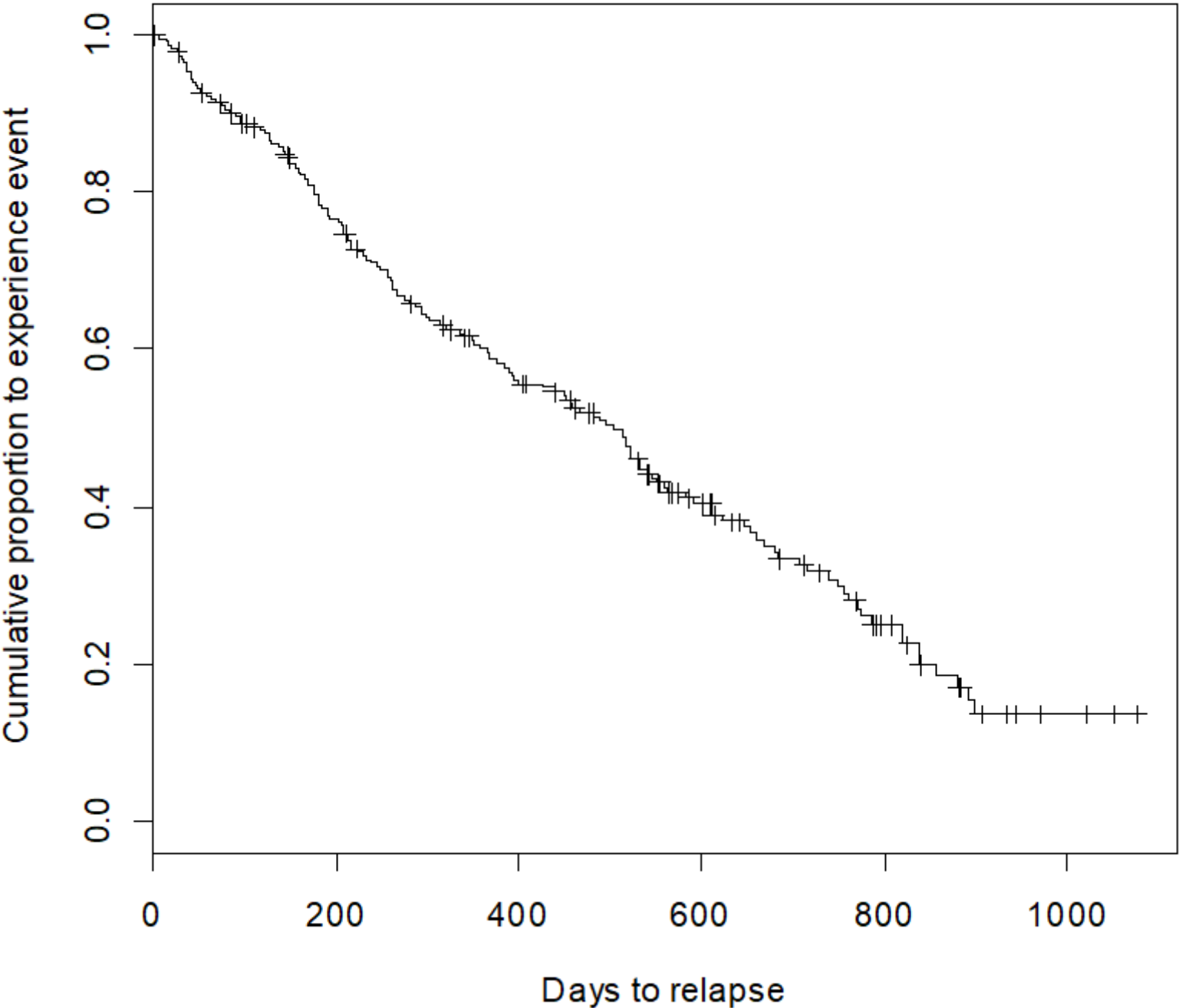
Nguy cơ tái phát tức thời ở người nghiện ma túy sau khi xuất viện.



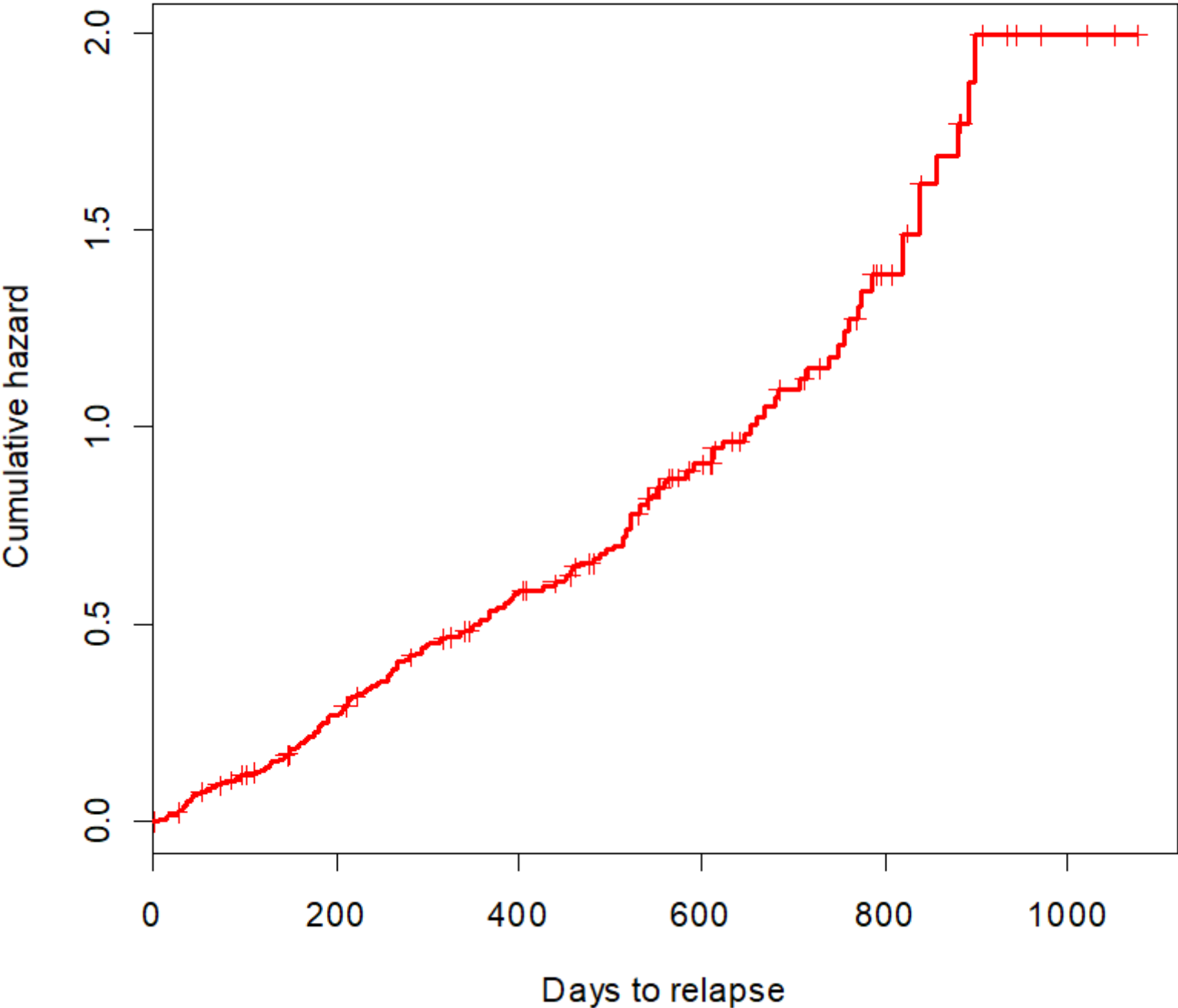
Sinh tồn và nguy cơ

- Nguy cơ tích lũy $H(t)$
 - cũng được gọi là 'nguy cơ tích hợp'
 - bằng tổng số nguy cơ tích lũy mà một cá thể đã gặp phải từ đầu giai đoạn quan sát
 - đại diện cho số lượng sự kiện dự kiến sẽ xảy ra theo quãng thời gian t

Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của người nghiện ma túy đã không tái phát sau khi xuất viện sau khi điều trị như một hàm số thời gian.



Nguy cơ tích lũy của tái nghiện ở người nghiện ma túy sau khi xuất viện.



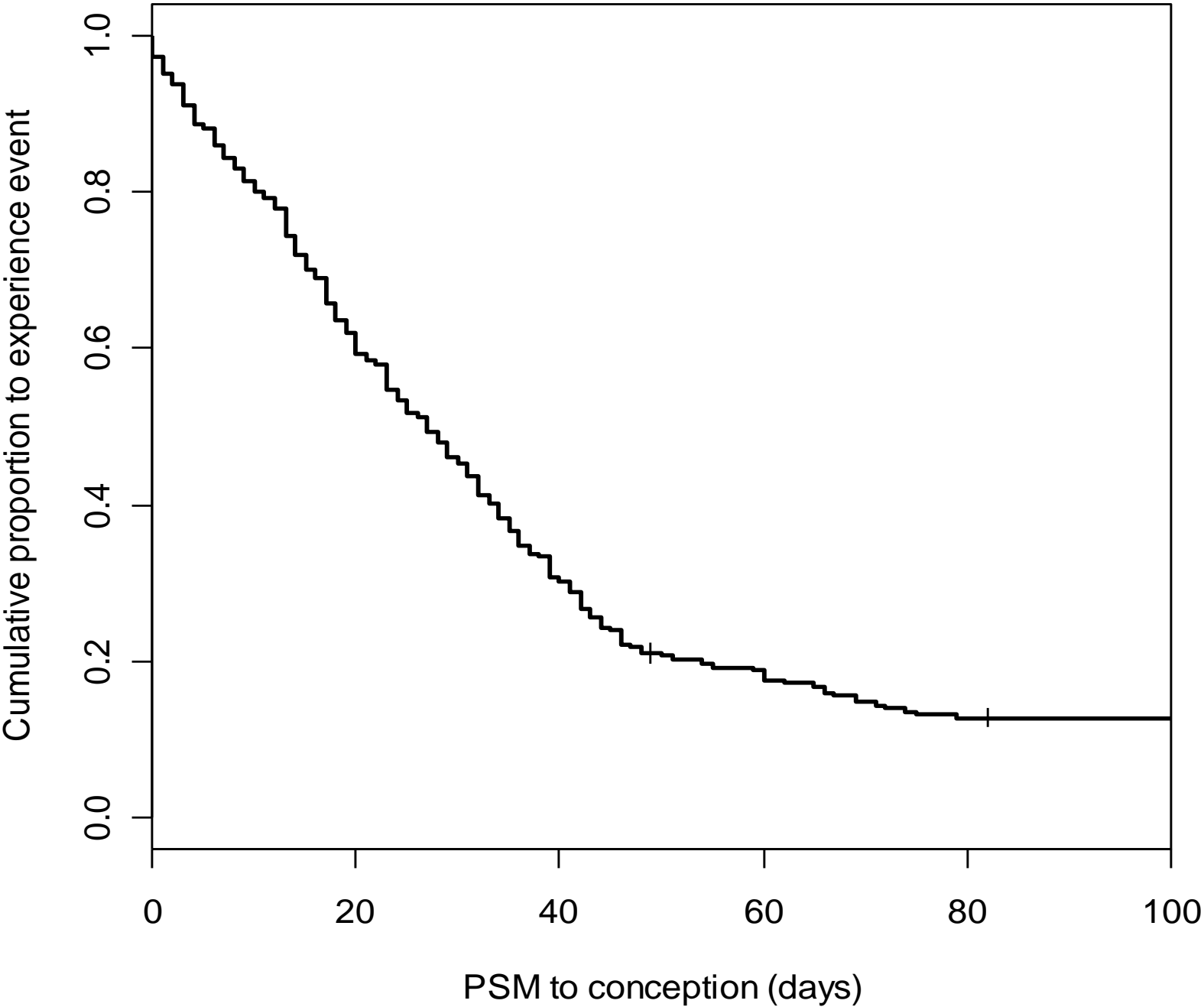
Lộ trình

- Sinh tồn và nguy cơ
- Các hàm sinh tồn có tham số
 - Exponential
 - Weibull
- Các hàm sinh tồn phi tham số
- Trình bày

Các phân bố sinh tồn có tham số

- Trong bài giảng trước, chúng tôi đã nói về những cách để mô tả sự sinh tồn
 - Trục x : thời gian
 - Trục y : tỷ lệ nhóm chưa có sự kiện
- Một số hàm sinh tồn ‘có thể dự báo’ ...

Ngày để thụ thai là một hàm số của các ngày sau ngày bắt đầu kế hoạch giao phối ở bò sữa



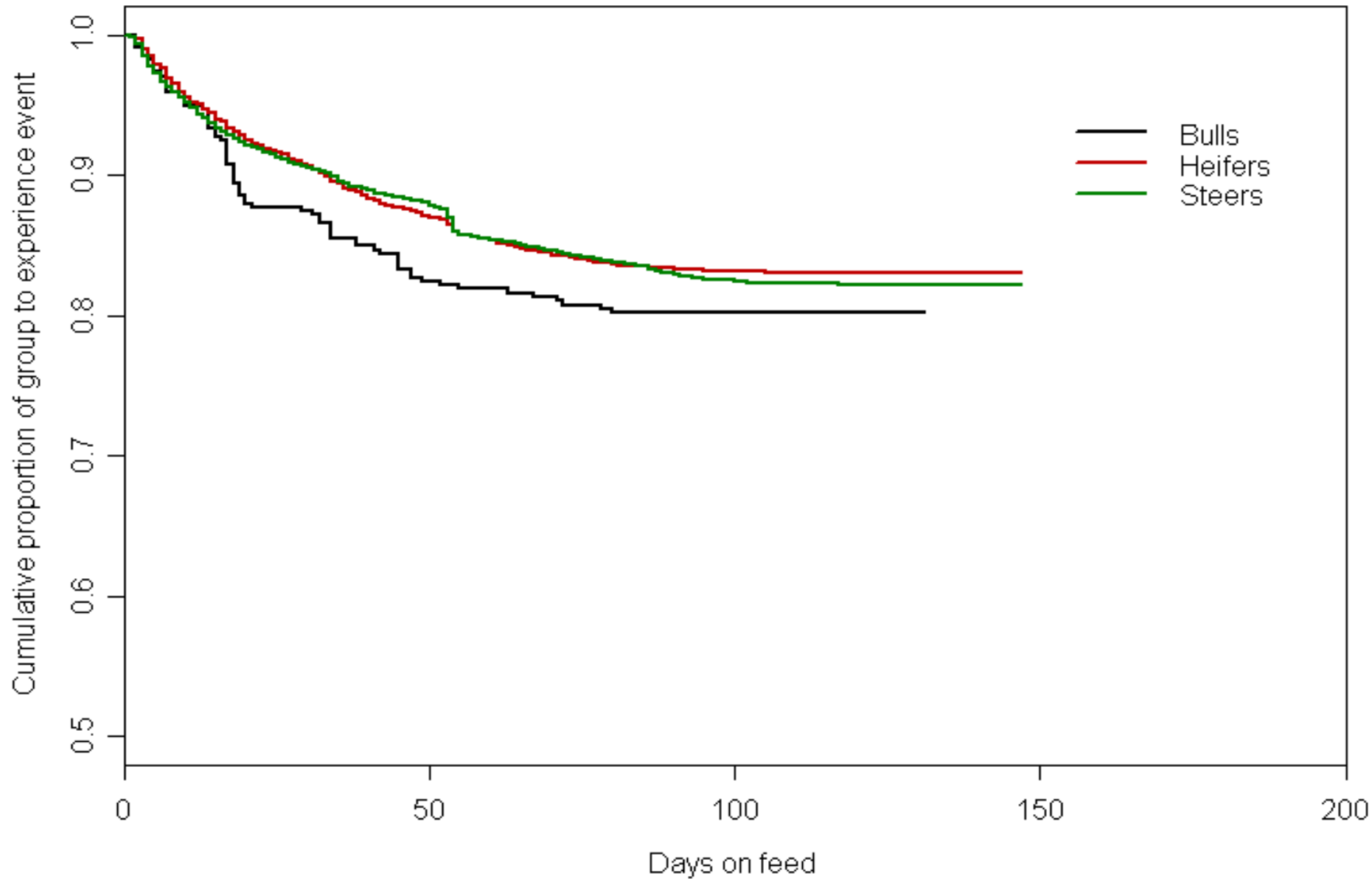
Các phân bố sinh tồn có tham số

- Một số hàm sinh tồn ‘không thể dự báo’ ...

Một trại bò tại Indonesia.



Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của động vật bị kéo vì mọi lý do như một hàm số thời gian, theo giống bò.



Các phân bố sinh tồn có tham số

- Các phân phối đó theo một mẫu đẹp, được xác định rõ có thể được mô tả bằng các phân phối có tham số
- Những phân phối có hình dạng không bình thường chỉ có thể được mô tả bằng các phân phối phi tham số

Các phân bố sinh tồn có tham số

- Ưu điểm của phương pháp tham số là sự sinh tồn có thể dự đoán được (vì khi chúng ta biết rõ đường cong sẽ đưa ra được hành động phù hợp để đến được quan sát cuối cùng)
- Các mô hình phi tham số không phù hợp để dự đoán vì không chắc chắn về bản chất của phân phối cho đến quan sát cuối cùng

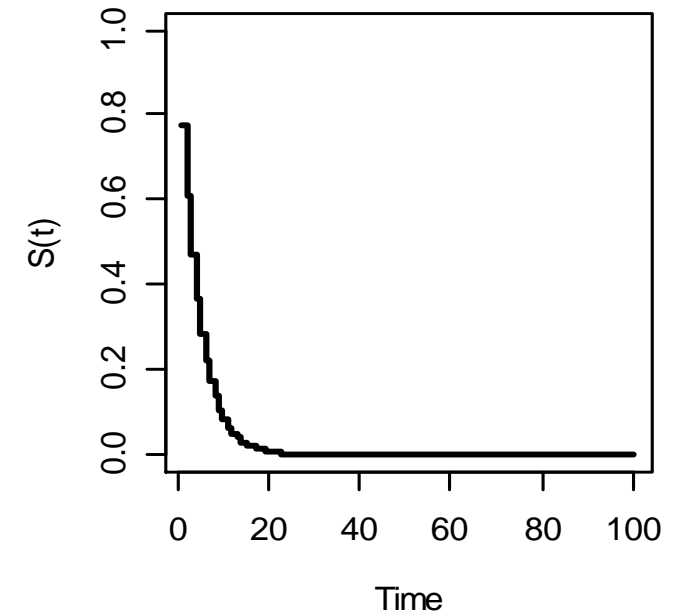
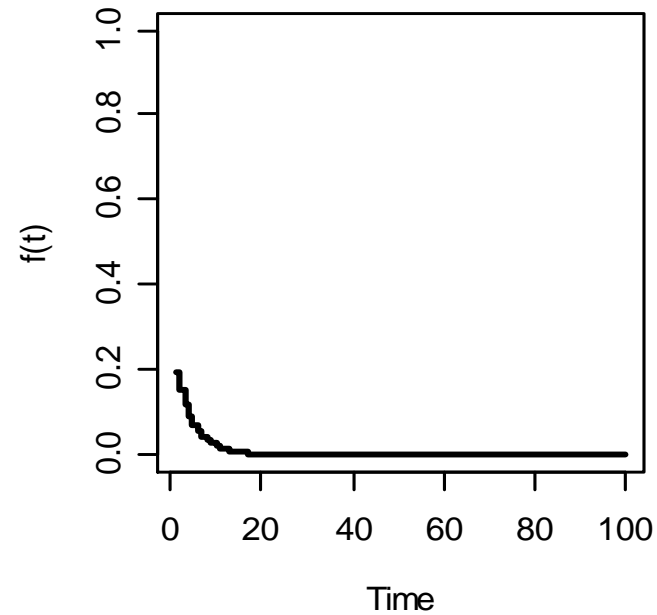
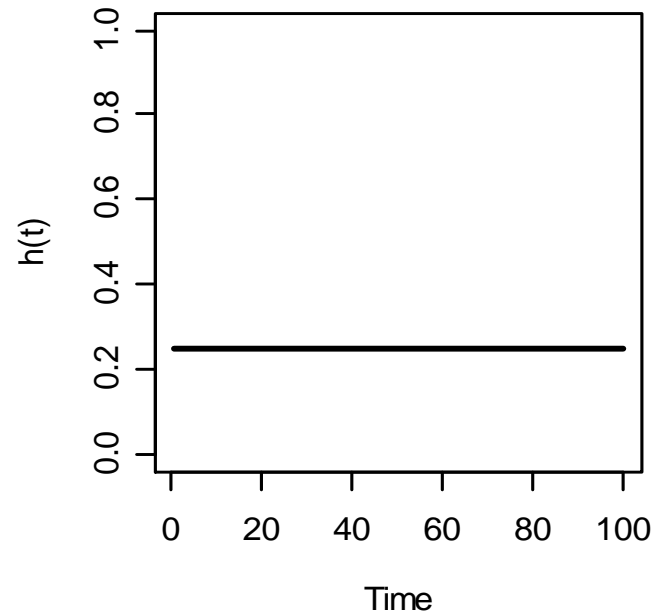
Các hàm số sinh tồn có tham số

- Có nhiều phân phối tham số, mỗi phân phối có các hàm số nguy cơ khác nhau:

Phân phối	$h(t)$	$f(t)$	$S(t)$
Exponential	λ	$\lambda \exp[-\lambda t]$	$\exp[-\lambda t]$
Weibull	$\lambda p t^{p-1}$	$\lambda p t^{p-1} \exp[-\lambda t^p]$	$\exp[-\lambda t^p]$
Log-logistic	$[ab(at)^{b-1}] / [1 + (at)^b]$	$[ab(at)^{b-1}] / [1 + (at)^b]^2$	$[1 + (at)^b]^{-1}$

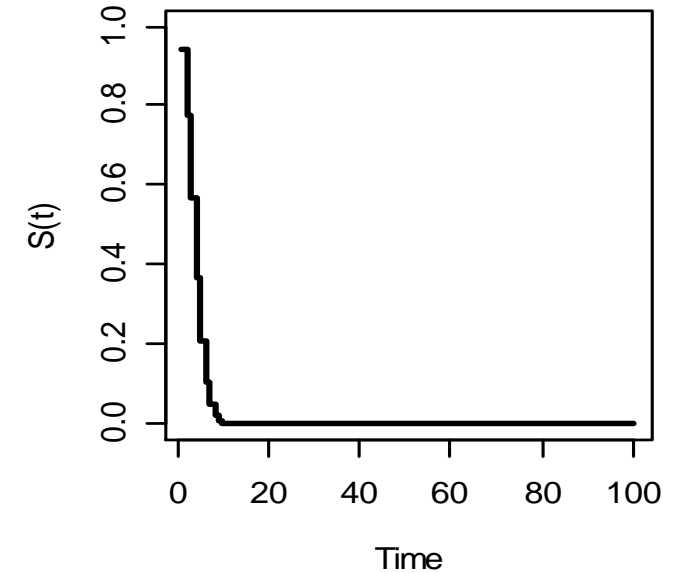
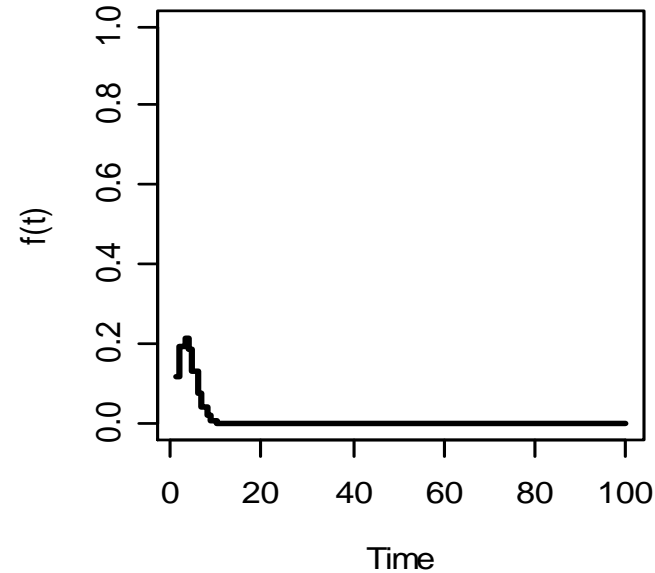
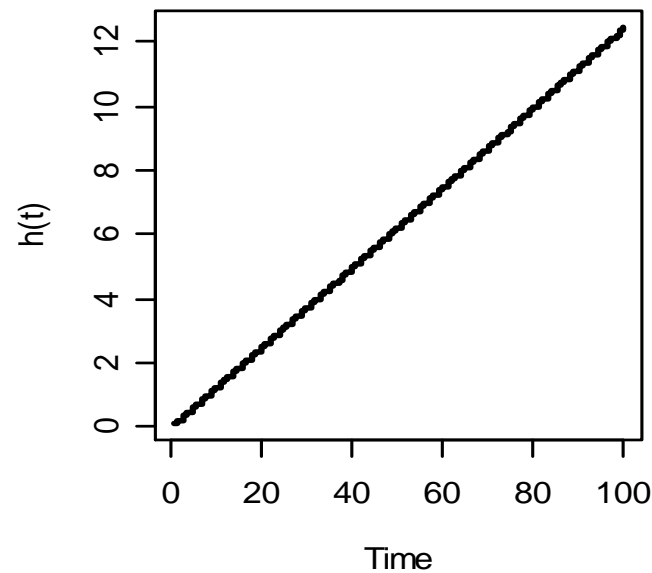
$$f(t) = h(t) \times S(t)$$

Phân bố theo cấp số nhân = hằng số nguy cơ theo thời gian



```
t <- seq(from = 1, to = 100, by = 1)
lambda = 0.25
ht <- lambda
ft <- lambda * exp(-lambda * t)
St <- exp(-lambda * t)
```

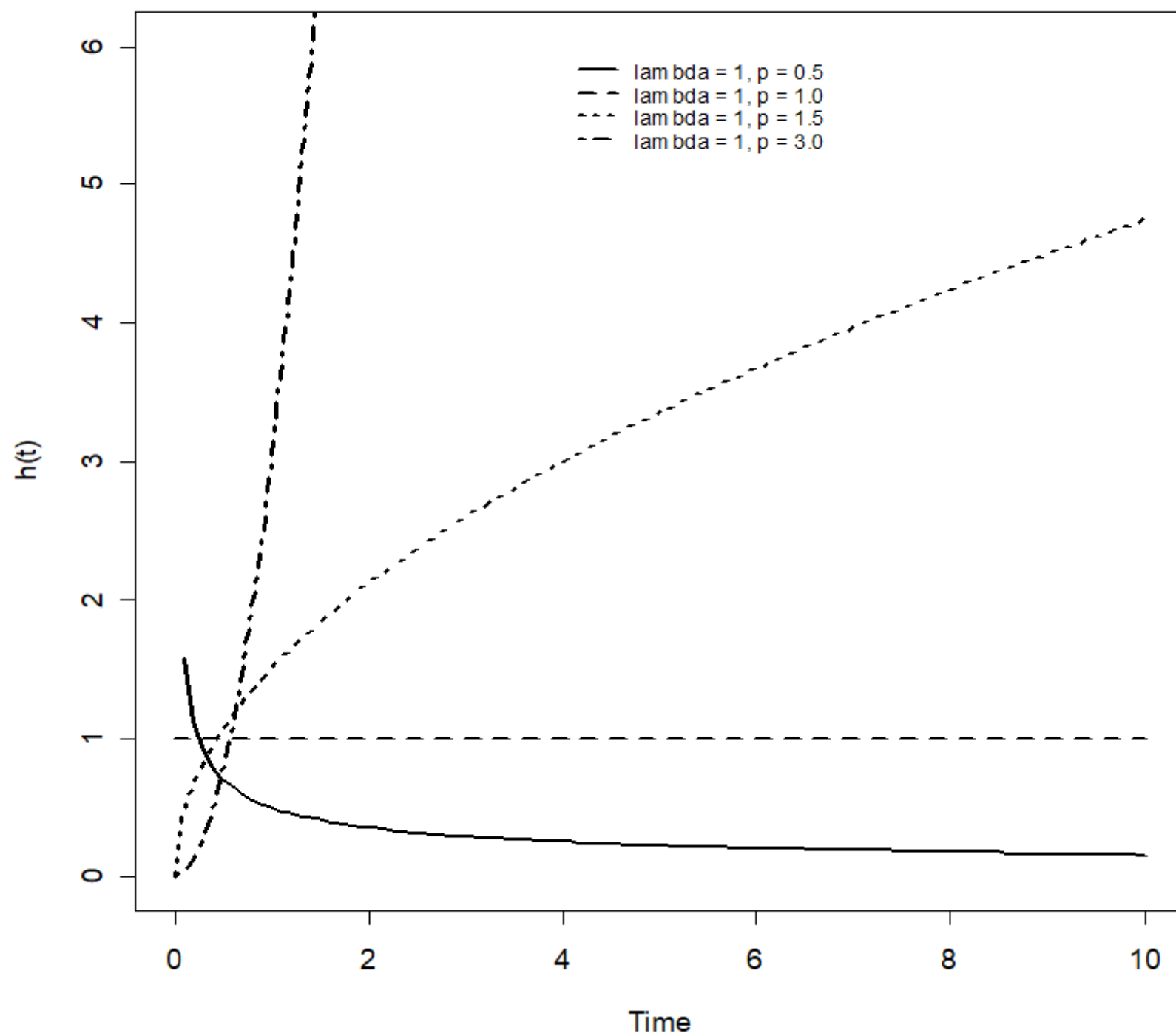

Phân phối Weibull = nguy cơ khác nhau tùy theo lambda và p.



```
t <- seq(from = 1, to = 100, by = 1)
lambda = 0.25; p = 2
ht <- (lambda * p) * (lambda * t)^(p - 1)
ft <- (lambda * p) * (lambda * t)^(p - 1) * exp(-(lambda * t)^p)
St <- exp(-(lambda * t)^p)
```


Các hàm số sinh tồn có tham số

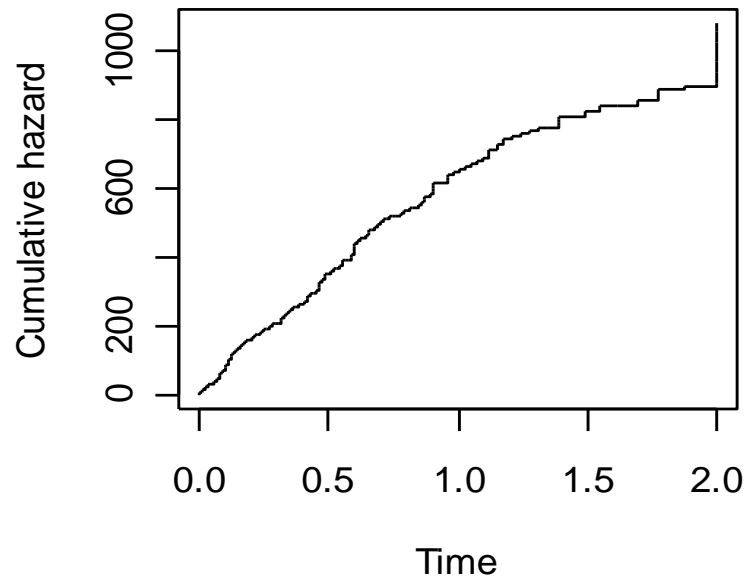
- Thực hành
 - Đối với phân phối Weibull, đánh giá tác động thay đổi λ (tham số tỷ lệ) và p (tham số hình dạng)
 - Điều gì xảy ra khi nguy cơ
 - $p < 1$?
 - $p = 1$?
 - $p > 1$?



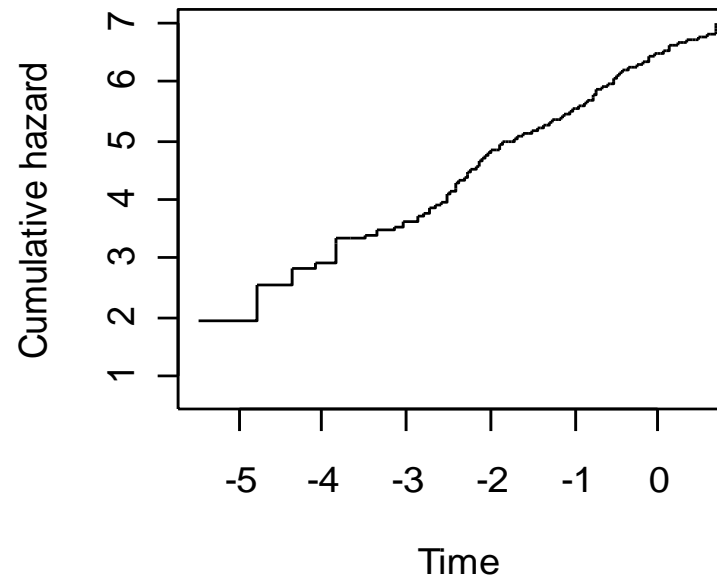
Các hàm số sinh tồn có tham số

- Có là một phân phối tham số thích hợp cho dữ liệu?
 - phân phối theo cấp số nhân
 - vẽ biểu đồ nguy cơ tích lũy như là một hàm của t
 - Phân phối thích hợp sẽ là một đường thẳng
 - Phân phối Weibull
 - vẽ biểu đồ đồ nguy cơ tích lũy log như là một hàm của $\log t$
 - Phân phối thích hợp sẽ là một đường thẳng

Exponential



Weibull



```
library(survival); setwd("D:\\TEMP")
dat <- read.table("addict.csv", header = TRUE, sep = ",")
addict.km <- survfit(Surv(stop, status) ~ 1, conf.type = "none",
type = "kaplan-meier", data = dat)
Ht <- -log(addict.km$surv)
t <- addict.km$time

par(mfrow = c(2,2))
plot(Ht, t, type = "s", xlab = "Time", ylab = "Cumulative hazard",
main = "Exponential")
plot(log(Ht), log(t), type = "s", xlab = "Time", ylab = "Cumulative
hazard", main = "Weibull")
```

Các hàm số sinh tồn có tham số

- Các phân phối khác
 - gamma
 - log-normal
 - log-logistic
- Xem Dohoo and Martin, trang 427

Lộ trình

- Sinh tồn và nguy cơ
- Các hàm sinh tồn tham số
 - Exponential
 - Weibull
- Các hàm sinh tồn phi tham số
- Trình bày

Các hàm số sinh tồn phi tham số

- Các phương pháp phi tham số được sử dụng khi không có phân phối lý thuyết phù hợp với dữ liệu
 - Tần suất (luôn luôn?) được sử dụng trong dịch tễ học
 - Có 2 phương pháp phi tham số đánh giá cách sinh tồn thay đổi theo thời gian :
 - Phương pháp Kaplan-Meier
 - Các phương pháp bảng sống (life table)

Các hàm số sinh tồn phi tham số

- Phương pháp Kaplan-Meier
 - Được biết đến như phương pháp giới hạn sản phẩm (the Product Limit method)
 - dựa trên thời gian sinh tồn của từng cá thể và giả định rằng kiểm duyệt là độc lập với thời gian sinh tồn (lý do là một quan sát bị kiểm duyệt không liên quan đến nguyên nhân thất bại)

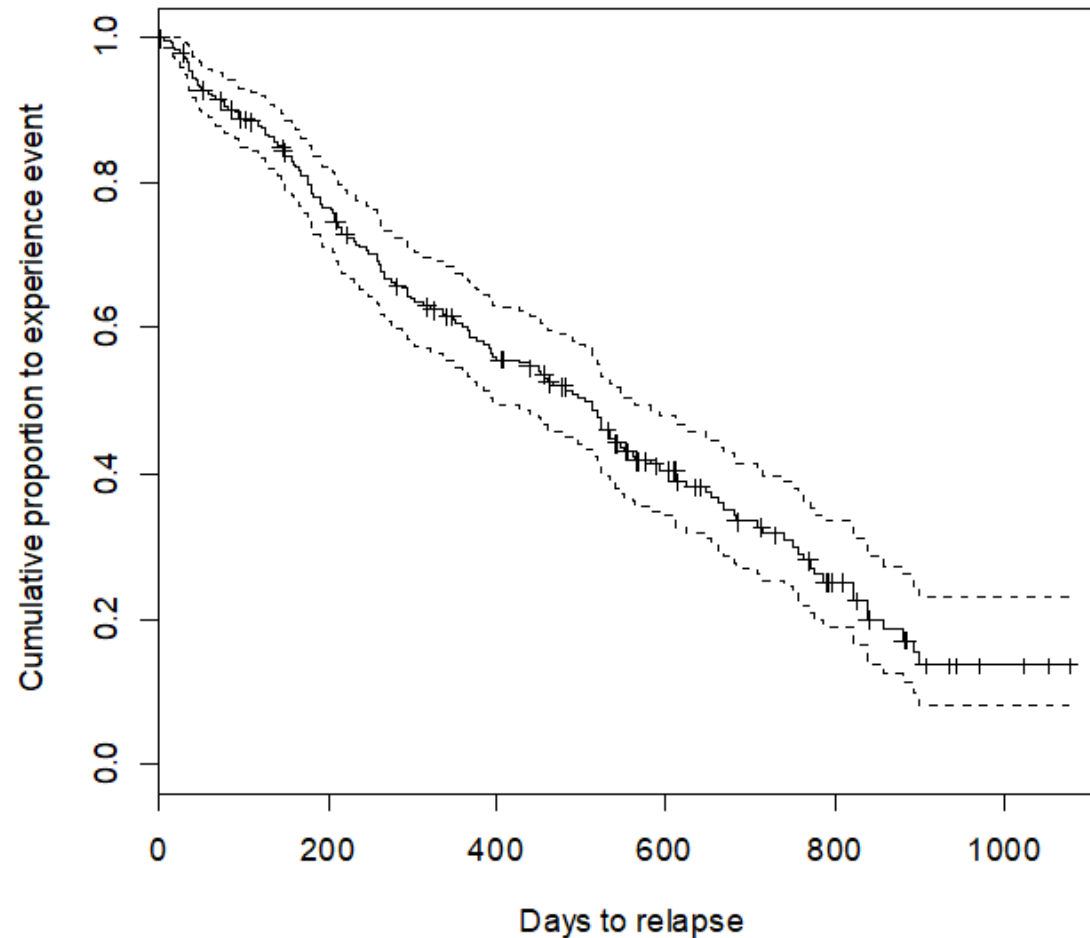
Các hàm số sinh tồn phi tham số

- Phương pháp Kaplan-Meier :

Thời gian	Bắt đầu n_i	Thất bại d_i	Kiểm duyệt w_i	Nguy cơ $r_i = (n_i - w_i)$	Xác suất sinh tồn $p_i = d_i \div (n_i - w_i)$	Sinh tồn lũy kế
0	31	2	3	$31 - 3 = 28$	0.93	$0.93 \times 1.00 = 0.93$
1	26	1	2	$26 - 2 = 24$	0.96	$0.96 \times 0.93 = 0.89$
2	23	1	2	$23 - 2 = 21$	0.95	$0.95 \times 0.89 = 0.85$
3	20	1	2	$20 - 2 = 18$	0.94	$0.94 \times 0.85 = 0.80$
etc						

Các hàm số sinh tồn phi tham số

- Khoảng tin cậy
 - sai số chuẩn của ước tính tỷ lệ sinh tồn có thể được tính toán tại từng thời điểm
 - sai số chuẩn có thể được sử dụng để tính khoảng tin cậy
 - sai số chuẩn có thể được sử dụng để kiểm tra sự khác biệt về tỷ lệ sinh tồn giữa các tầng nghiên cứu



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

addict.km <- survfit(Surv(stop, status) ~ 1, type = "kaplan-meier",
data = dat);
plot(addict.km, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event", conf.int = TRUE);
```

Các hàm số sinh tồn phi tham số

- Phương pháp Life-table (Cutler-Ederer)

Thời gian	Bắt đầu ni	Thất bại di	Kiểm duyệt wi
0 to 1	31	3	4
2 to 3	24	2	4
etc			

Các hàm số sinh tồn phi tham số

- Phương pháp Life-table (Cutler-Ederer)

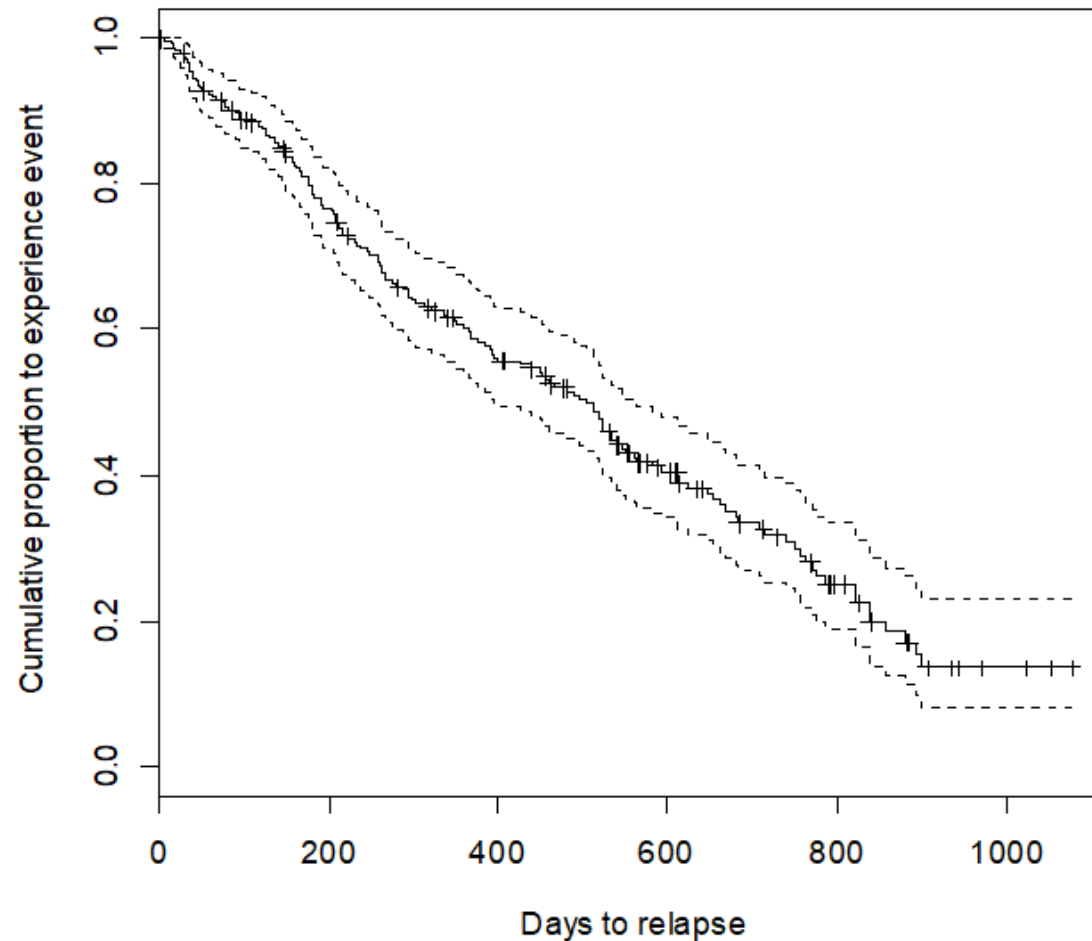
Thời gian	Xác suất thất bại $q_i = d_i / [n_i - (w_i/2)]$	Xác suất sinh tồn $p_i = 1 - q_i$	Sinh tồn tích lũy
0 to 1	0.10	0.90	0.90
2 to 3	0.09	0.91	$0.91 \times 0.90 = 0.92$
etc			

Các hàm số sinh tồn phi tham số

- Phương pháp Life-table (Cutler-Ederer)
 - giả định rằng các đối tượng được rút ngẫu nhiên trong mỗi khoảng thời gian - do đó, trung bình, các đối tượng được rút một nửa trong khoảng thời gian
 - đây không phải là vấn đề quan trọng đối với khoảng thời gian ngắn, nhưng có sai lệch (bias) đối với khoảng thời gian dài
 - thường được sử dụng để tạo ra các bảng sống từ điều tra dân số quy mô lớn (ví dụ như người đăng ký khai tử)
 - Hiện nay không được sử dụng thường xuyên

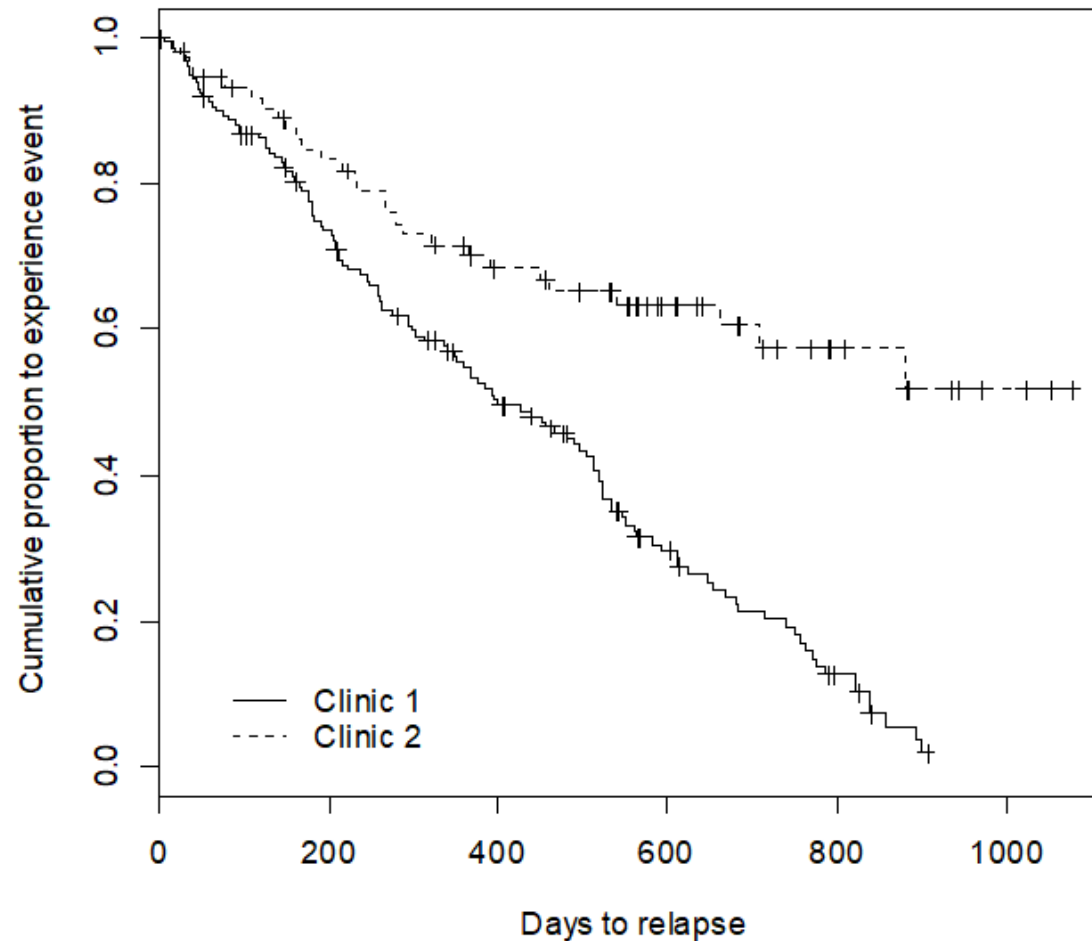
Các hàm số sinh tồn phi tham số

- Phương pháp Caplehorn và các phương pháp khác (1991)
 - Được so sánh bằng việc giữ lại ở hai phòng khám điều trị bằng methadone cho người nghiện heroin
 - Thời gian sinh tồn của bệnh nhân được xác định là thời gian tính bằng ngày cho đến khi bệnh nhân rời khỏi phòng khám hoặc được kiểm duyệt khi kết thúc nghiên cứu
 - Hai phòng khám khác nhau tùy theo chính sách điều trị chung của họ
 - Tham khảo: Caplehorn J et. al. (1991). Methadone dosage and retention of patients in maintenance treatment. Medical Journal of Australia, 154:195 - 199.



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

addict.km <- survfit(Surv(stop, status) ~ 1, conf.type = "none",
type = "kaplan-meier", data = dat);
plot(addict.km, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event");
```

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

addict.km <- survfit(Surv(stop, status) ~ clinic, type = "kaplan-
meier", data = dat)
plot(addict.km, xlab = "Days to relapse", ylab = "Cumulative
proportion to experience event", lty = c(1,2), legend.text =
c("Clinic 1","Clinic 2"), legend.pos = 0, legend.bty = "n")
```

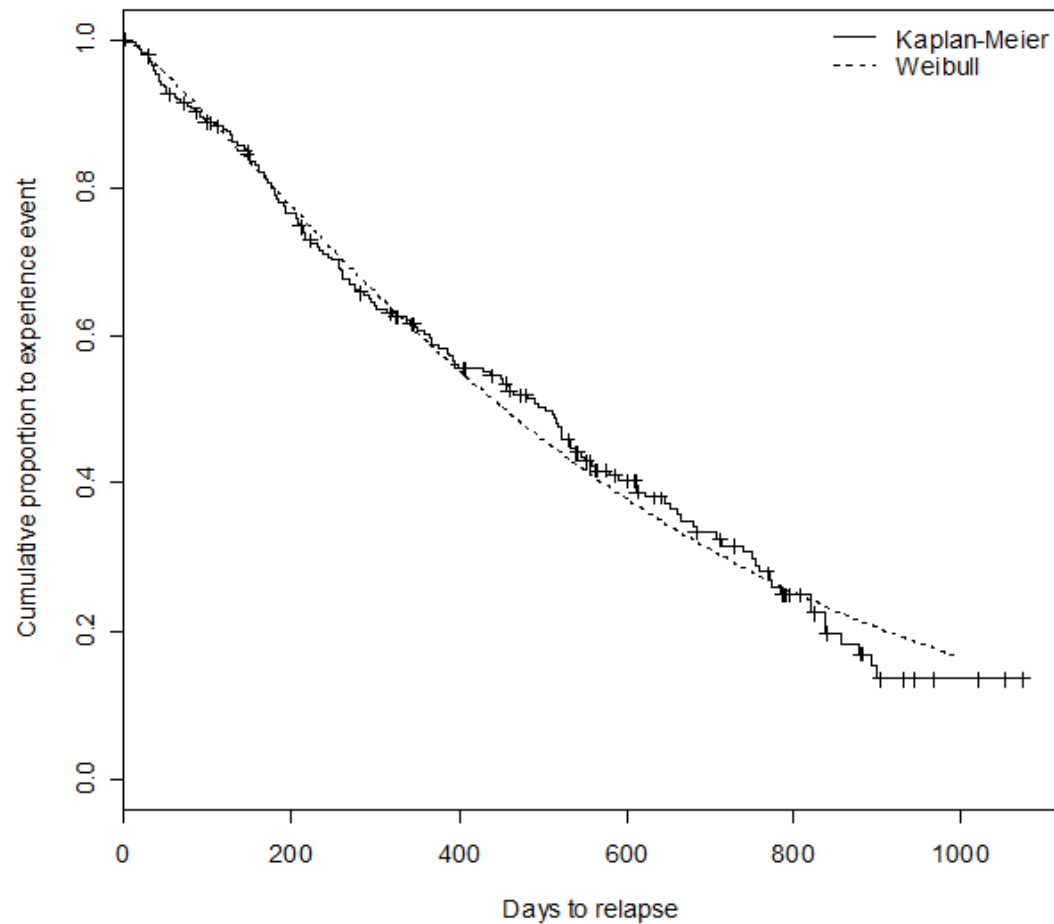
So sánh ước tính tỷ lệ sinh tồn giữa phương pháp Kaplan-Meier và phương pháp Weibull:

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

addict.we <- survreg(Surv(stop, status) ~ 1, dist = "weib", data =
dat)
addict.km <- survfit(Surv(stop, status) ~ 1, conf.type = "none",
type = "kaplan-meier", data = dat)
```

Using the Weibull distribution μ (the intercept) = $-\log(\lambda)$ and σ (scale) = $1 / p$. Thus the scale parameter $\lambda = \exp(-\mu)$ and $p = 1 / \sigma$. See Tableman and Kim p 78.

```
p <- 1 / addict.we$scale
lambda <- exp(-addict.we$coeff[1])
t <- 1:1000
St <- exp(-(lambda * t)^p)
addict.we <- as.data.frame(cbind(t = t, St = St))
```



```
plot(addict.km, xlab = "Days to relapse", ylab = "Cumulative  
proportion to experience event")  
lines(addict.we$t, addict.we$St, lty = 2)  
legend(x = "topright", legend = c("Kaplan-Meier", "Weibull"), lty =  
c(1,2), bty = "n")
```

Các hàm sinh tồn phi tham số

- phân phối Weibull cung cấp sự phù hợp đầy đủ với dữ liệu được quan sát cho đến ngày thứ 500, sau đó nó có xu hướng đánh giá thấp sự sinh tồn

Lộ trình

- Sinh tồn và nguy cơ
- Các hàm sinh tồn tham số
 - Exponential
 - Weibull
- Các hàm sinh tồn phi tham số
- Trình bày

Trình bày

- Trình bày về đường cong sinh tồn:
 - nên là bước thẳng đứng
 - Nên đánh dấu các điểm kiểm duyệt
 - đường cong thường trở nên không ổn định vào cuối thời gian nghiên cứu
 - nên trích dẫn trong tiêu đề của bảng dữ liệu số lượng sự kiện và số lượng quan sát bị kiểm duyệt
 - Trục y: “*Cumulative proportion to experience ...*”
 - Trục x: “*PSM to conception (days)*”

Lộ trình

- Sinh tồn và nguy cơ
- Các hàm sinh tồn có tham số
 - Exponential (số mũ)
 - Weibull
- Các hàm sinh tồn phi tham số
- Trình bày



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.