
Phân tích theo thời gian:

Giới thiệu chung

Mark Stevenson

Faculty of Veterinary and Agricultural Sciences

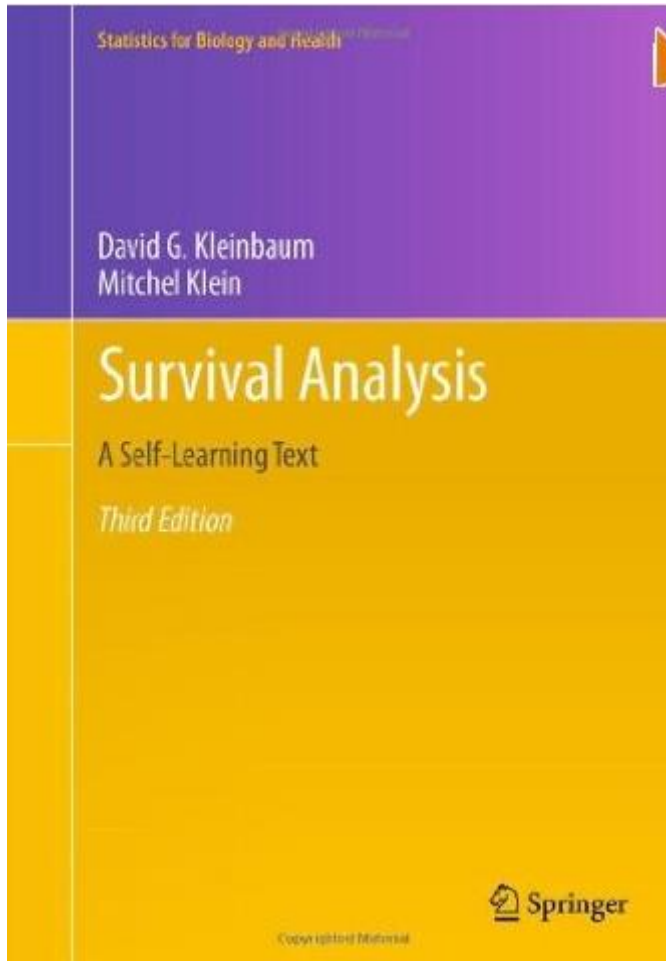
The University of Melbourne, Parkville Victoria 3010 Australia

[mark.stevenson1 @unimelb.edu.au](mailto:mark.stevenson1@unimelb.edu.au)

Lộ trình

- Sách tham khảo
- Tổng quan
- Nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

Sách tham khảo

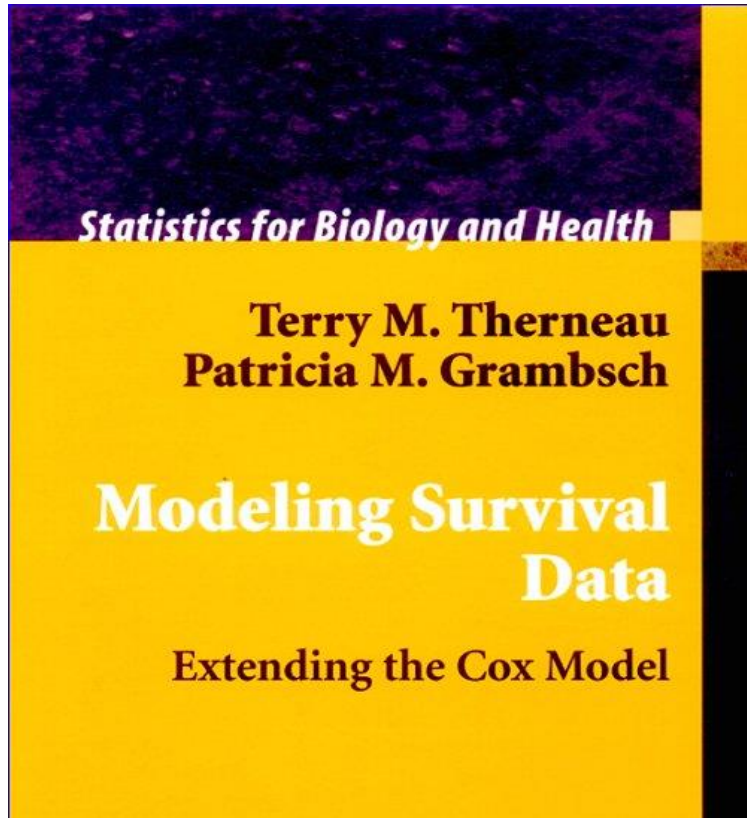


Kleinbaum D, Klein M. Survival Analysis: A Self-Learning Text. New York; Springer-Verlag; 2011.

Thorough coverage of subject. Includes software examples (SAS, SPSS, R).

Rating: +++

Sách tham khảo

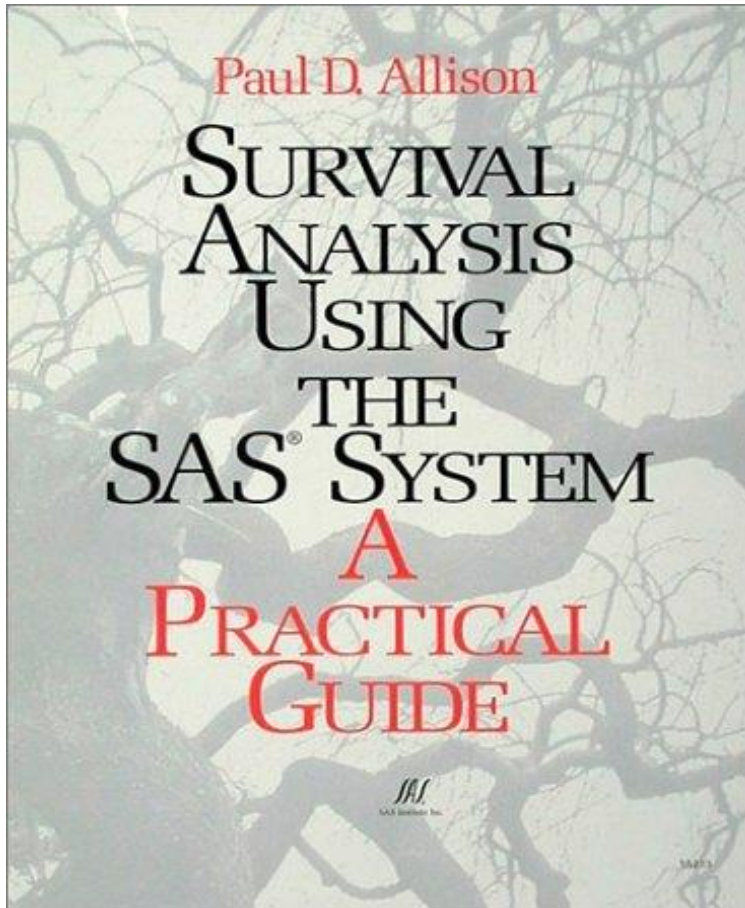


Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. New York; Springer-Verlag; 2001.

Comprehensive coverage of advanced topics.
Therneau is an excellent writer. Gives S-PLUS (\equiv R) and SAS code.

Rating: +++

Sách tham khảo

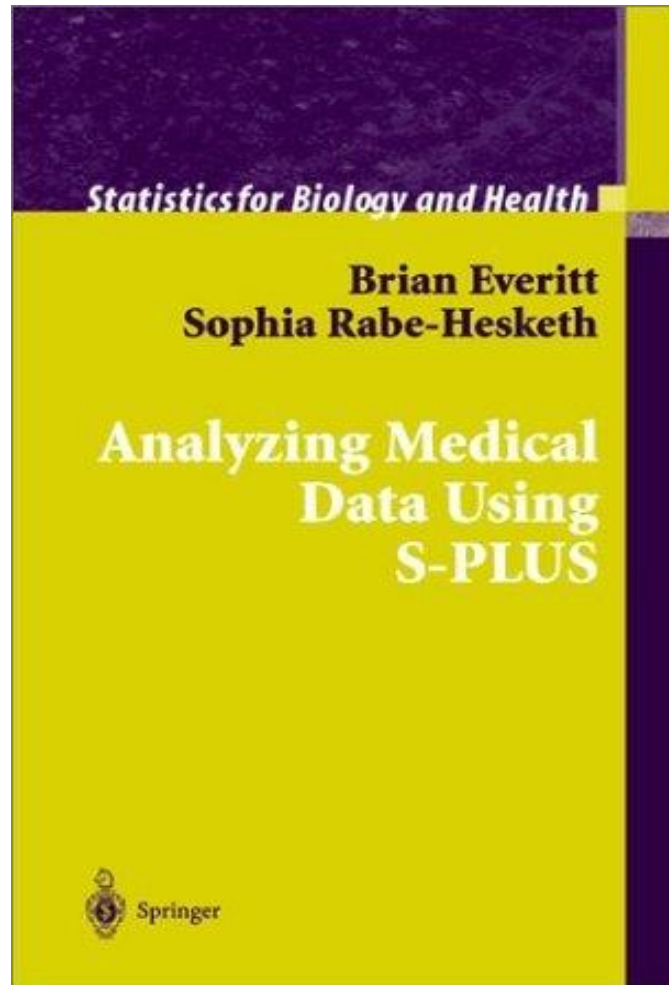


Allison P. Survival Analysis Using the SAS System A Practical Guide. New York: SAS Publishing; 1995.

A little dated now. SAS only.

Rating: ++

Sách tham khảo



Everitt B, Rabe-Hesketh S. Analyzing Medical Data Using S-PLUS. New York; Springer Verlag; 2001.

Practical coverage of a wide range of analytical techniques used in epidemiology. Covers the basics of survival analysis – not advanced topics.

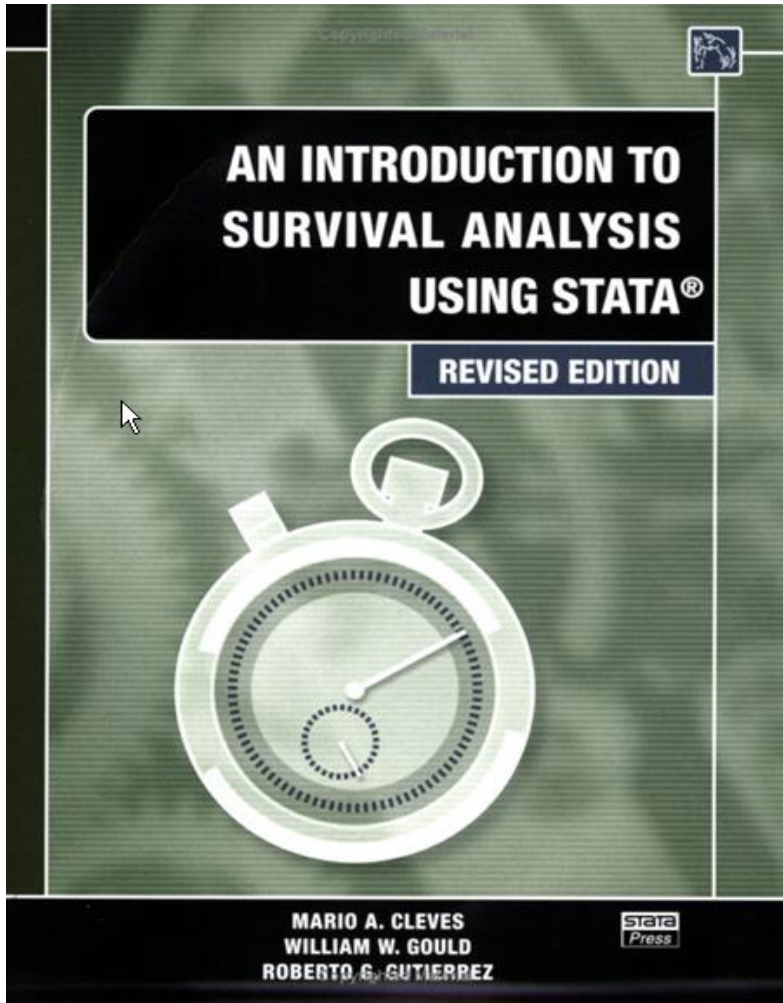
Rating: ++

Sách tham khảo

Cleves MA, Gould WW, Gutierrez RG. An Introduction to Survival Analysis Using STATA. New York; STATA Press; 2003.

Practical coverage of a wide range of analytical techniques used in epidemiology. Good coverage of advanced topics.

Rating: +++



Sách tham khảo

- Các tác giả khác
 - Collett D. Modelling Survival Data in Medical Research. London; Chapman and Hall; 1994.
 - Hosmer D, Lemeshow S. Applied Survival Analysis Regression Modeling of Time to Event Data. London; John Wiley and Sons; 1999.

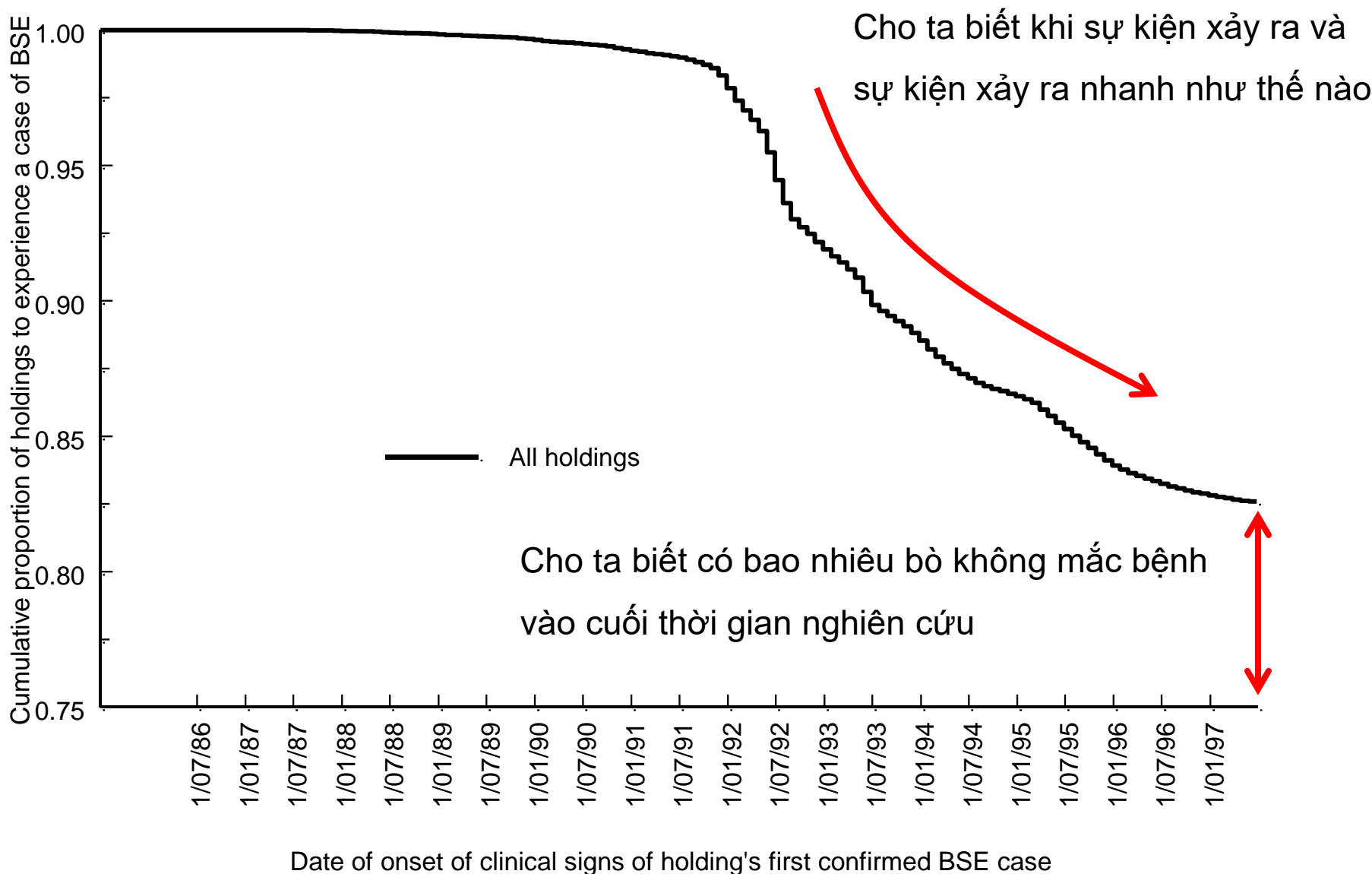
Lộ trình

- Sách tham khảo
- Tổng quan
- nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

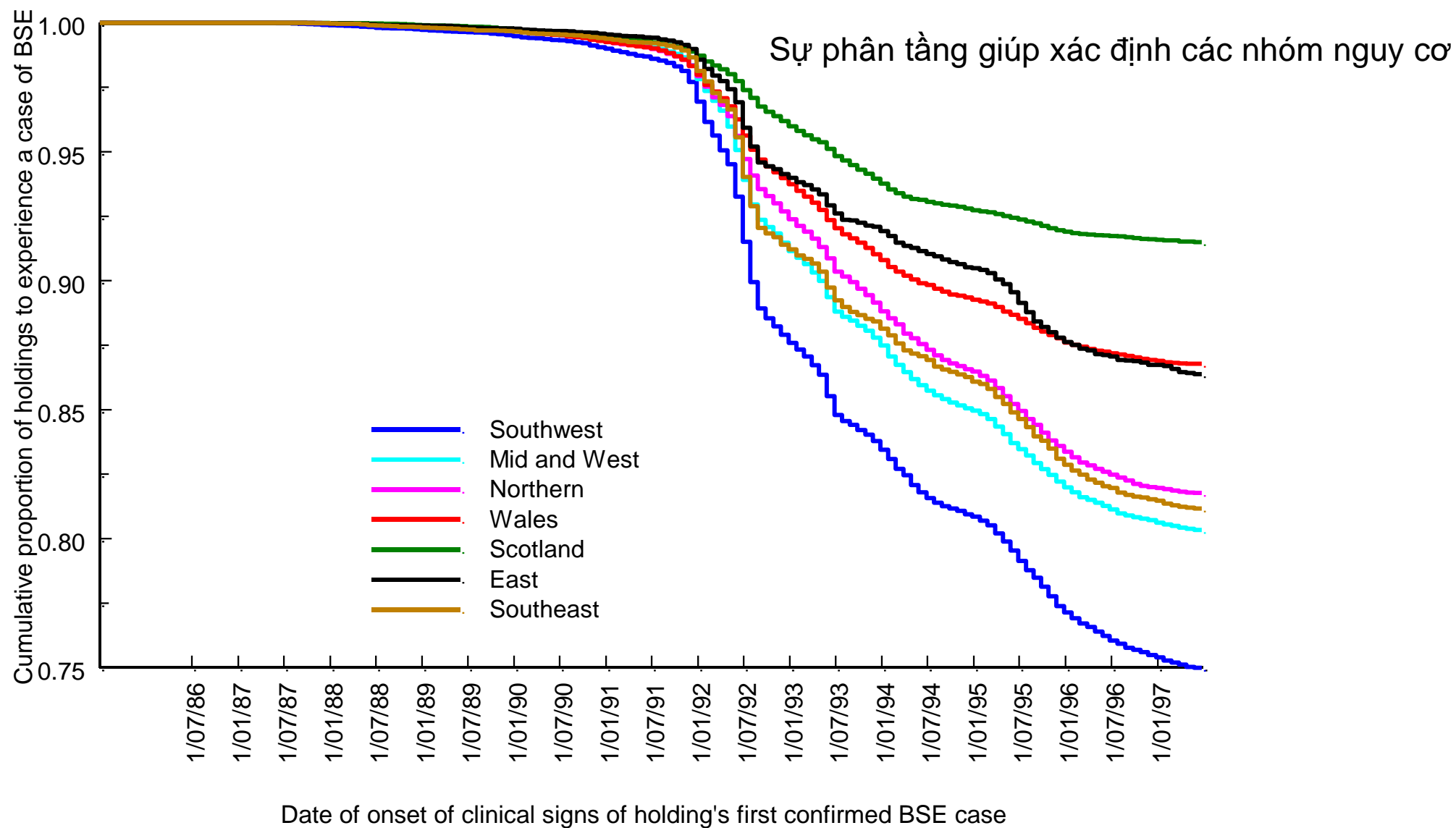
Tổng quan

- Survival analysis \equiv Phân tích theo thời gian đối với dữ liệu sự kiện
- Có nhiều ứng dụng trong dịch tễ học thú y :
 - Tuổi thọ của vật nuôi
 - Ngày xảy ra ca bệnh đầu tiên tại cơ sở
 - Độ tuổi có thể phát hiện kháng thể
- Các thuật ngữ được dùng:
 - Chức năng phân tích theo thời gian (Kaplan-Meier survival curves)
 - Hồi quy các nguy cơ theo tỷ lệ phần trăm của Cox (Cox proportional hazards regression)

Các ngày phát hiện ca chỉ báo bệnh bò điên (BSE) – Tại các trại bò của Anh.



Các ngày phát hiện ca chỉ báo bệnh bò điên (BSE) – Tại các trại bò của Anh.



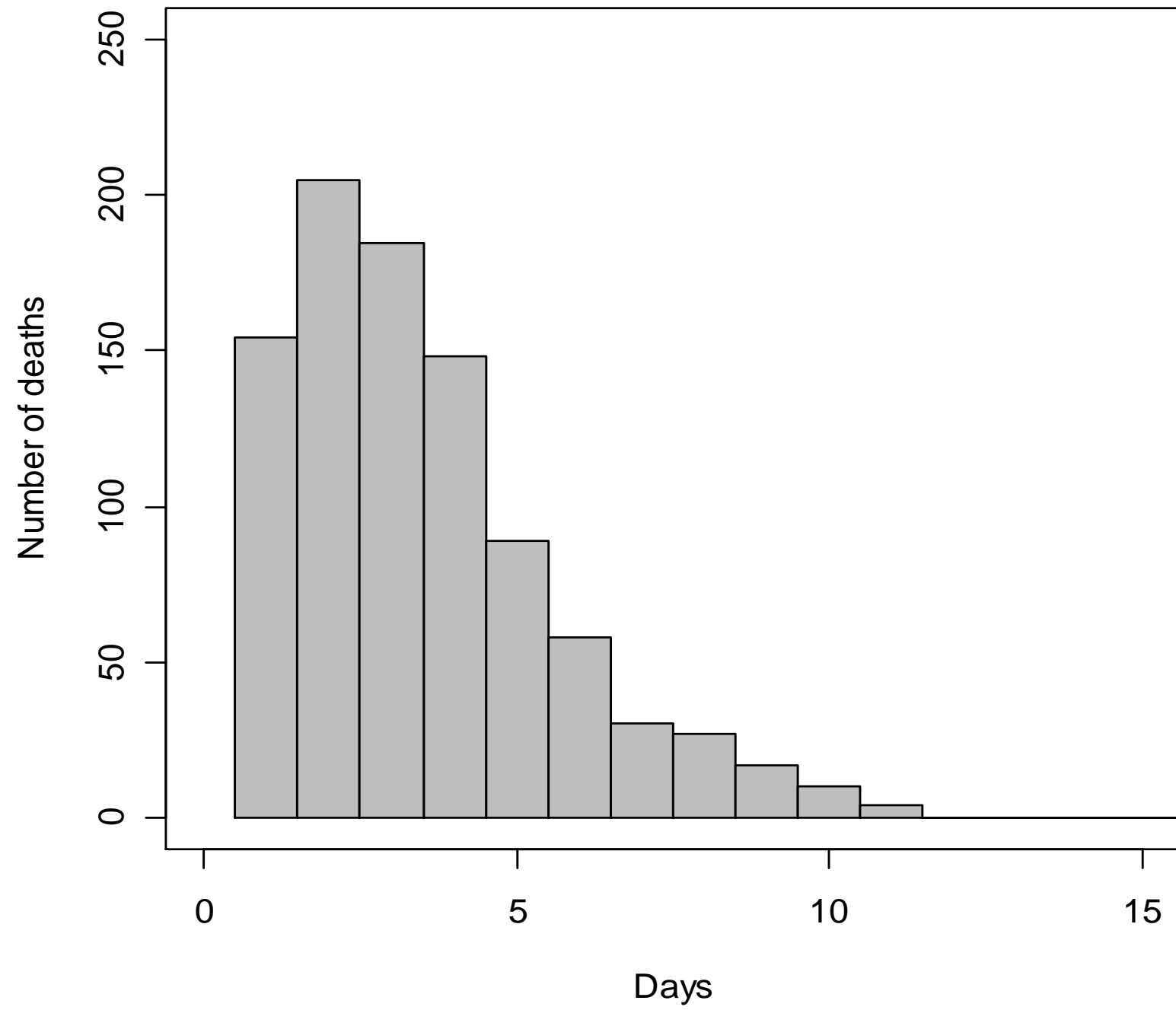
Tổng quan

- Sinh tồn (Survival) \equiv thời gian cần có để sự kiện xảy ra
- Thất bại (Failure) \equiv xảy ra sự kiện được quan tâm
- kiểm duyệt (censoring) \equiv các ca bệnh không hiển thị sự kiện trong thời gian quan sát
 - Sự kiện không xảy ra trong khoảng thời gian quan sát
 - Đối tượng đã bị rút khỏi nghiên cứu vì lý do không liên quan đến sự kiện

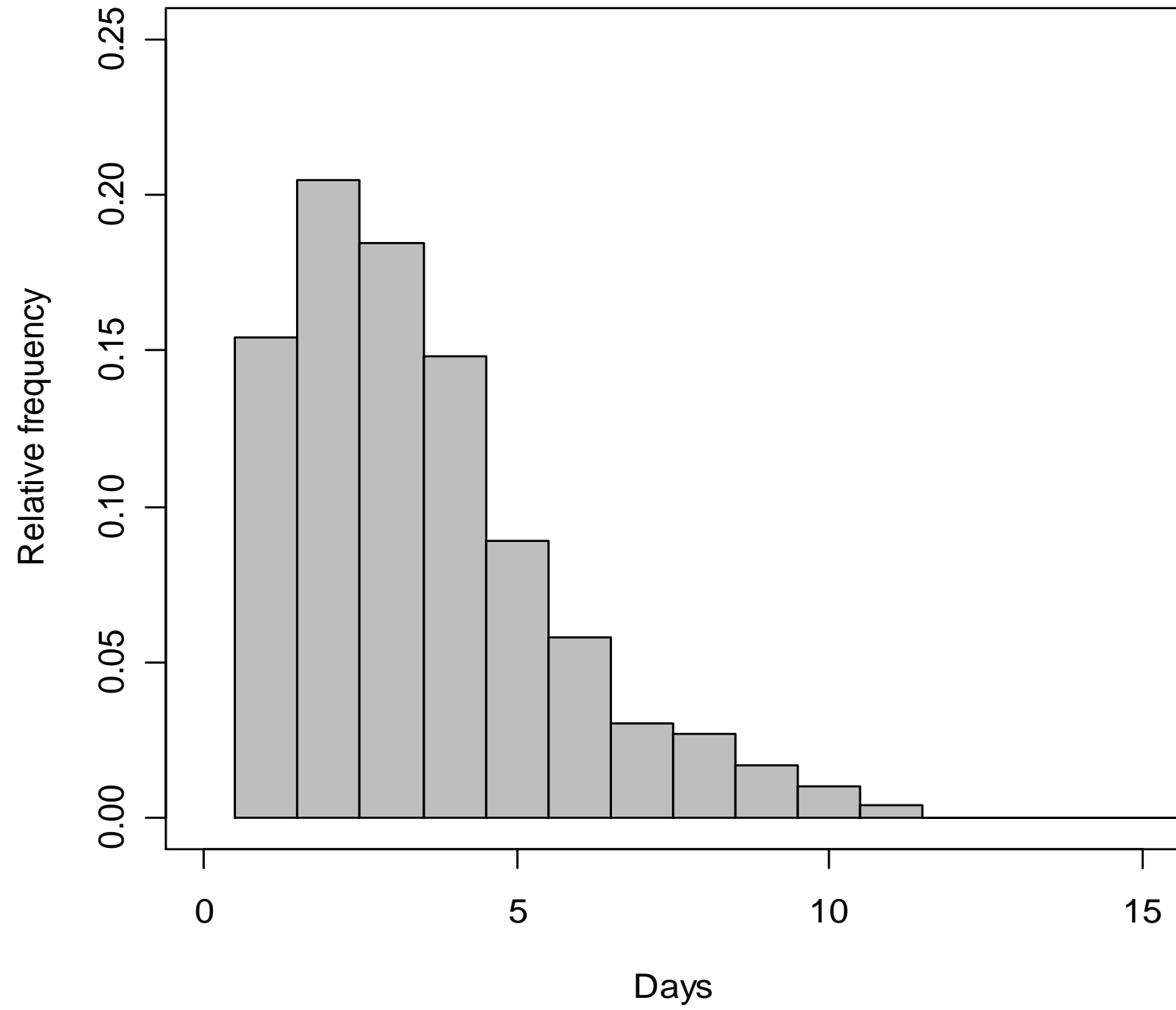
Lộ trình

- Sách tham khảo
- Tổng quan
- Nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

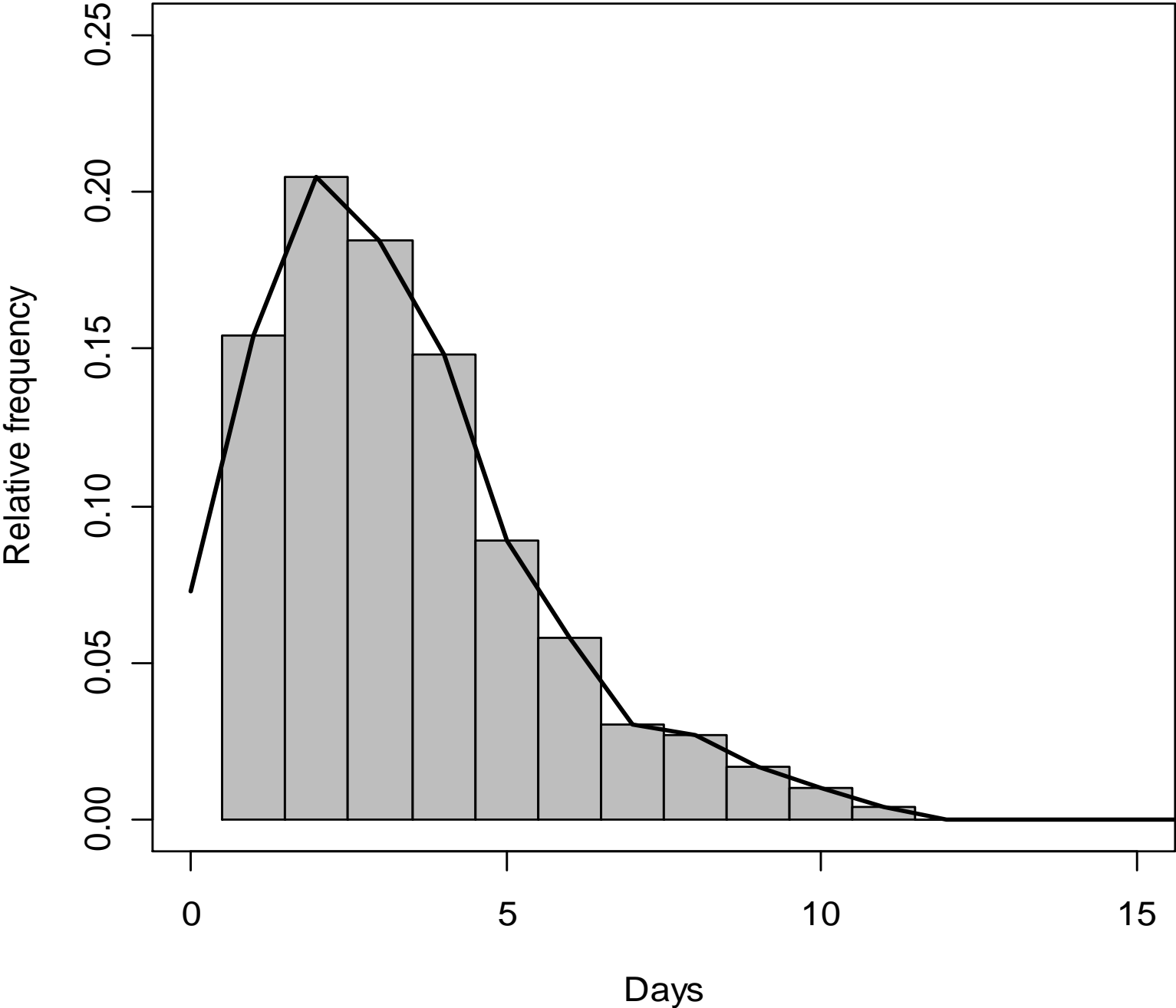
Biểu đồ tần suất cho thấy số ca tử vong là một hàm của thời gian.



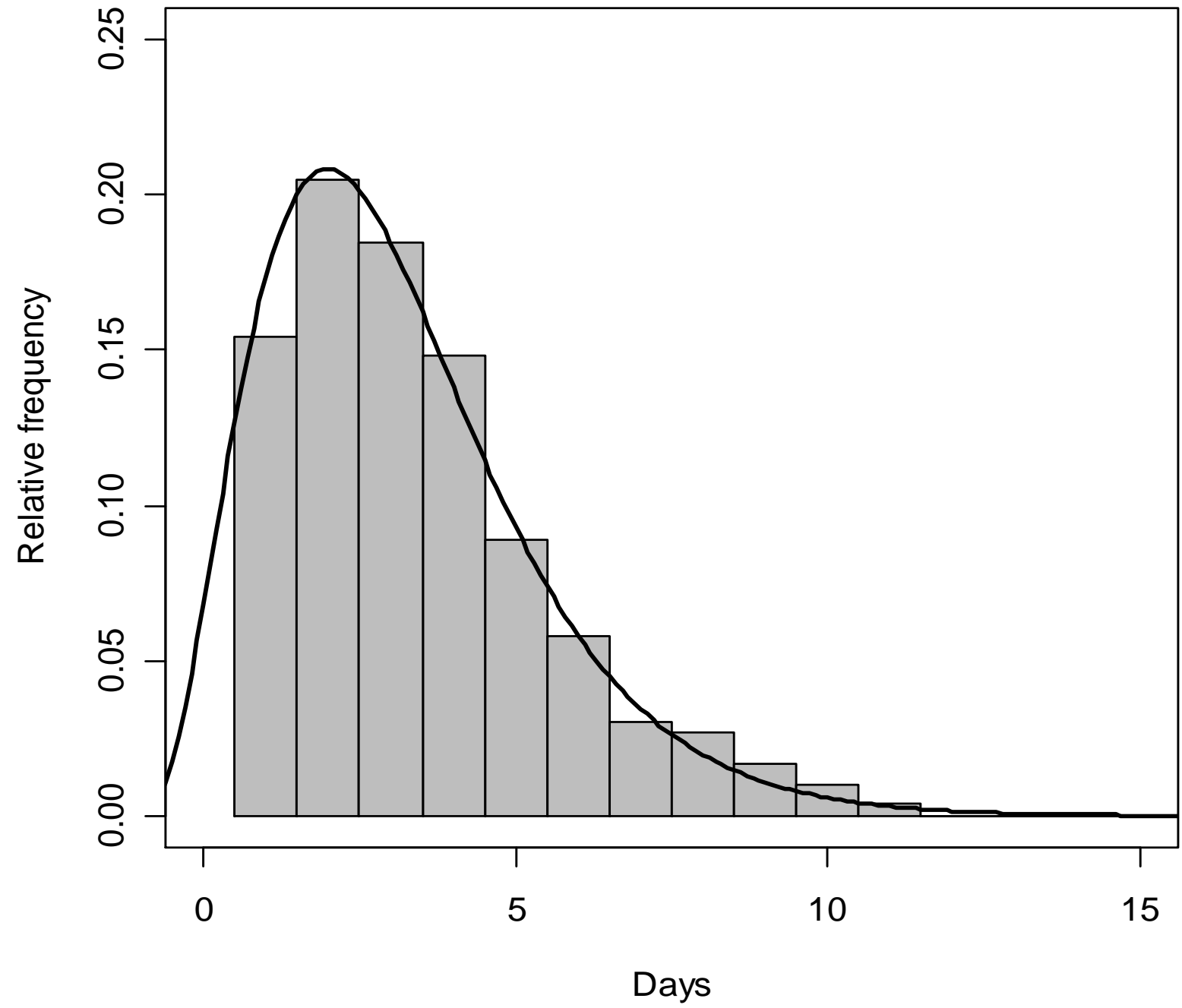
Biểu đồ tần suất cho thấy tần suất tương đối của ca tử vong là một hàm của thời gian.



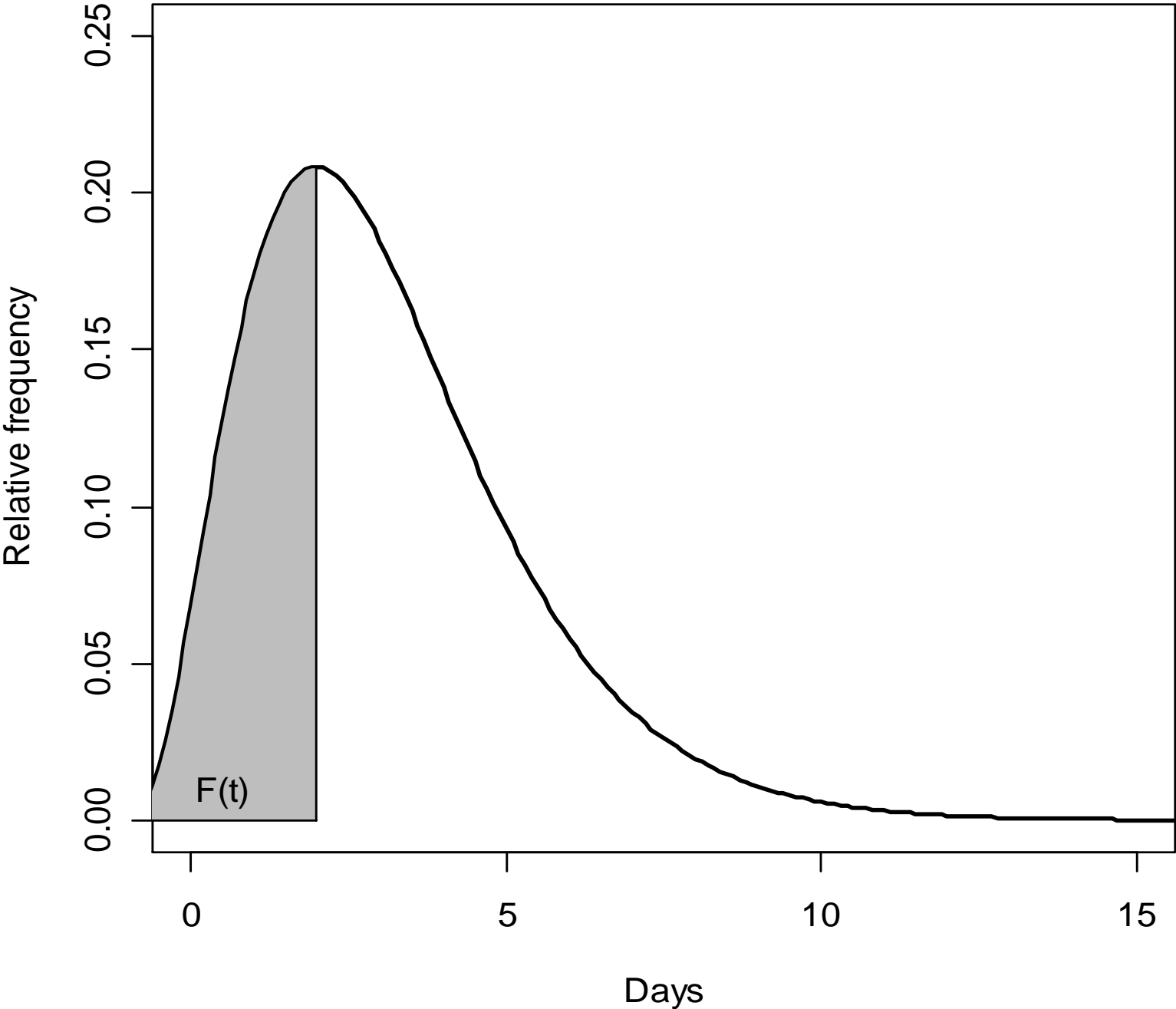
$f(t)$ = hàm mật độ tử vong.



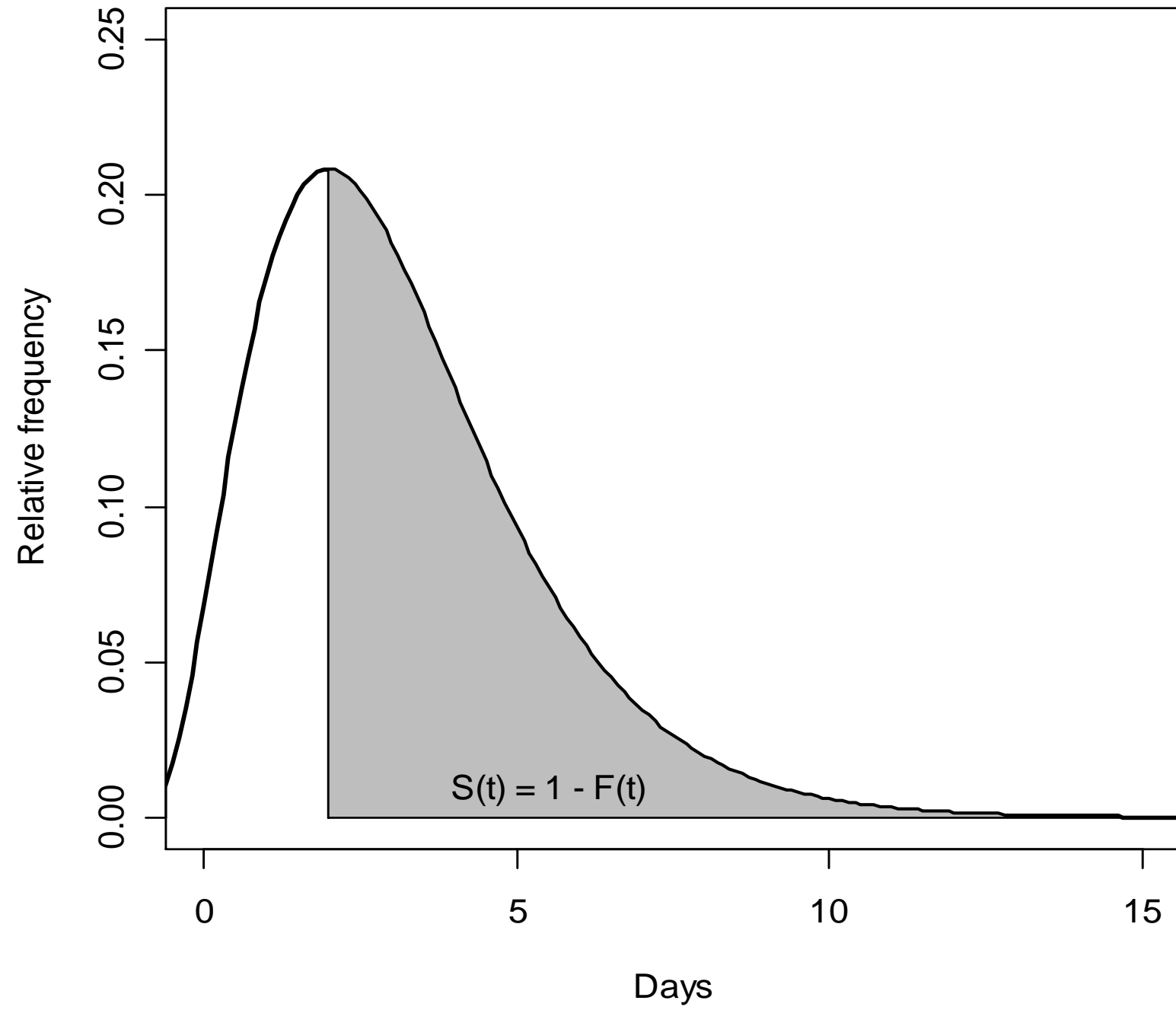
$f(t)$ = hàm mật độ tử vong.



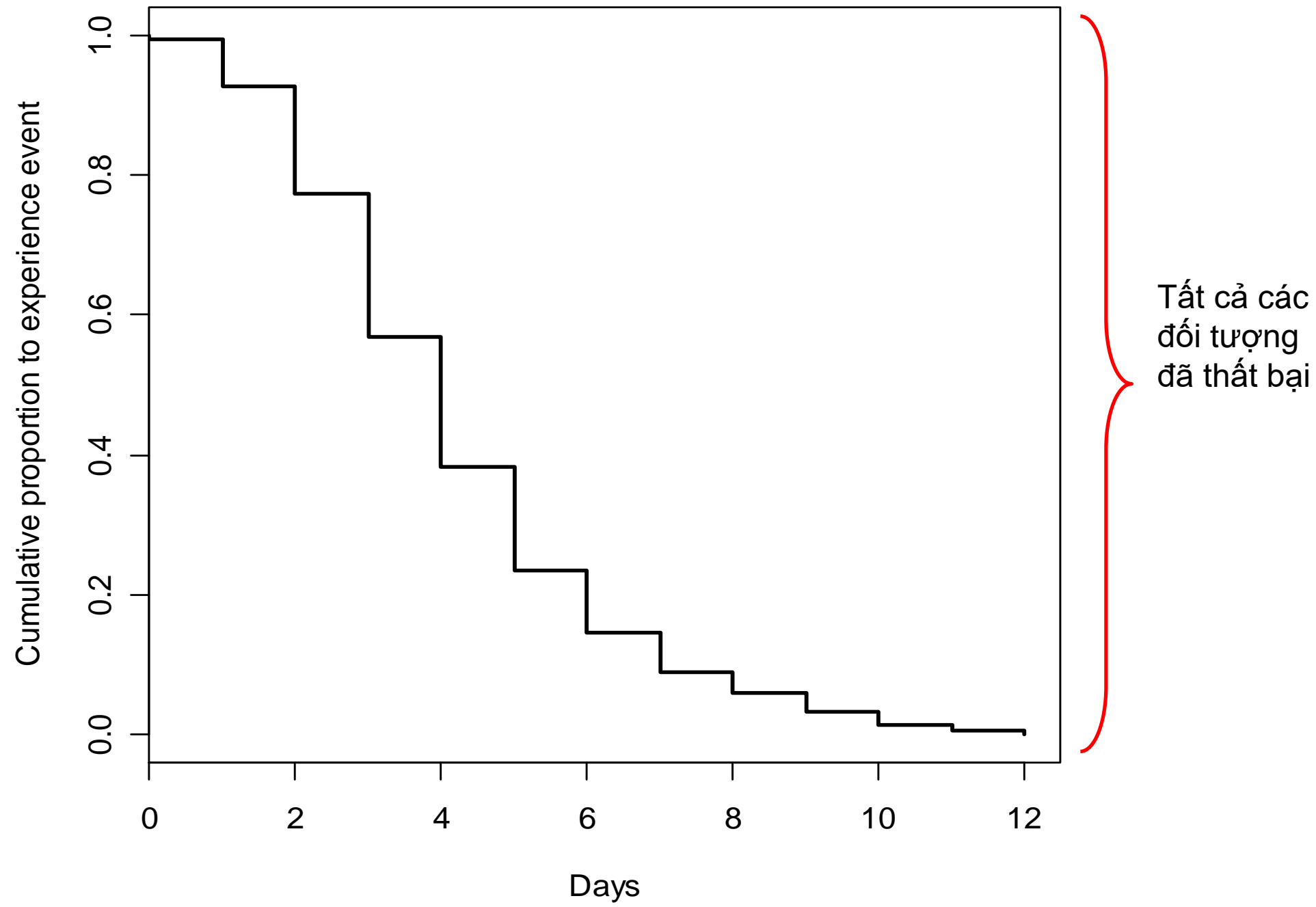
$F(t)$ = hàm thất bại (the failure function) = tỷ lệ không sống sót qua quãng thời gian t .



$S(t)$ = 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 $t = 1 - F(t)$.

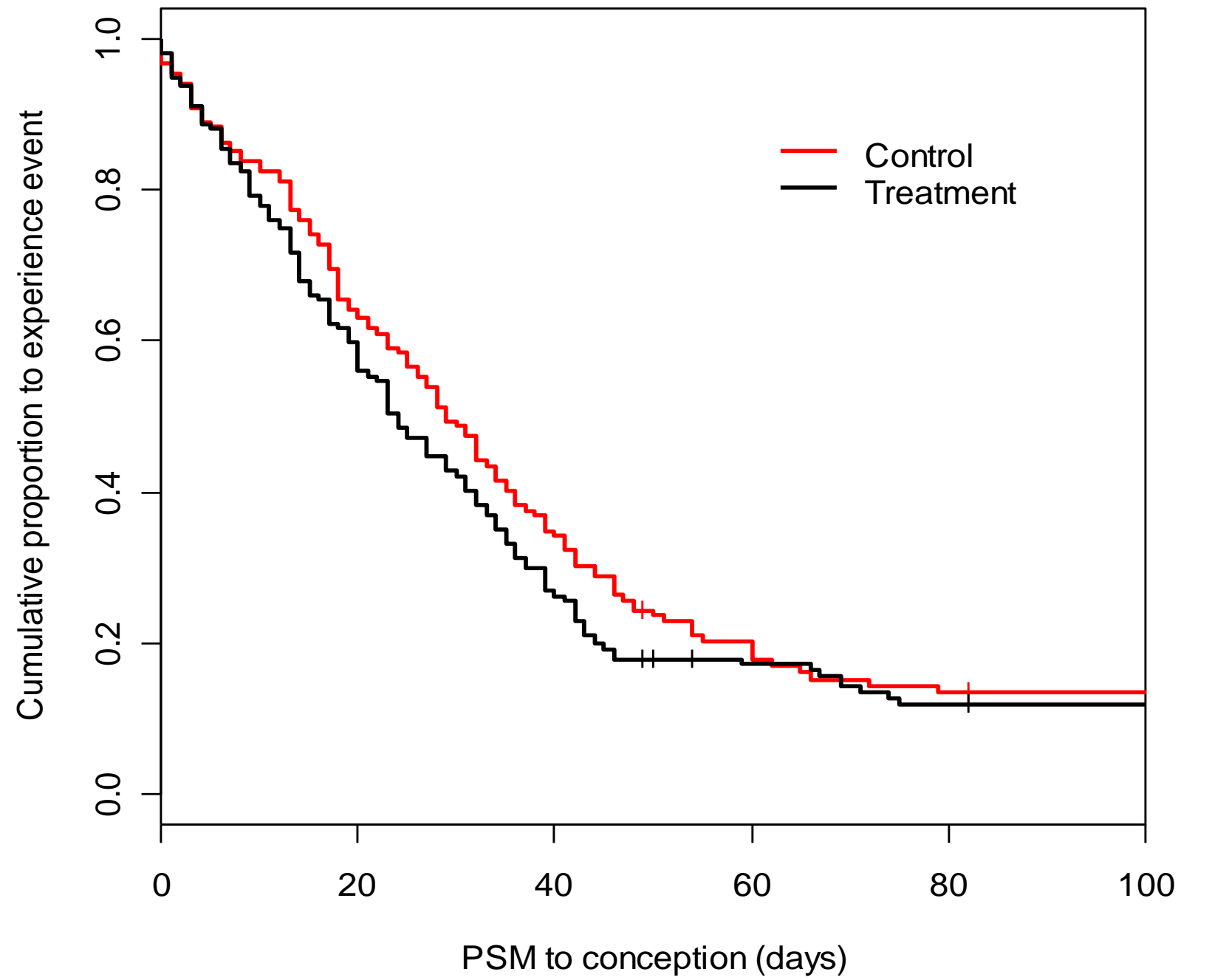


Nguy cơ và sinh tồn

- Hiệu quả của việc điều trị bằng cách cho uống canxi clorua sau sinh đối với khả năng sinh sản ở bò sữa ...



Hiệu quả của việc điều trị bằng cách cho uống canxi clorua sau sinh đối với khả năng sinh sản ở bò sữa .

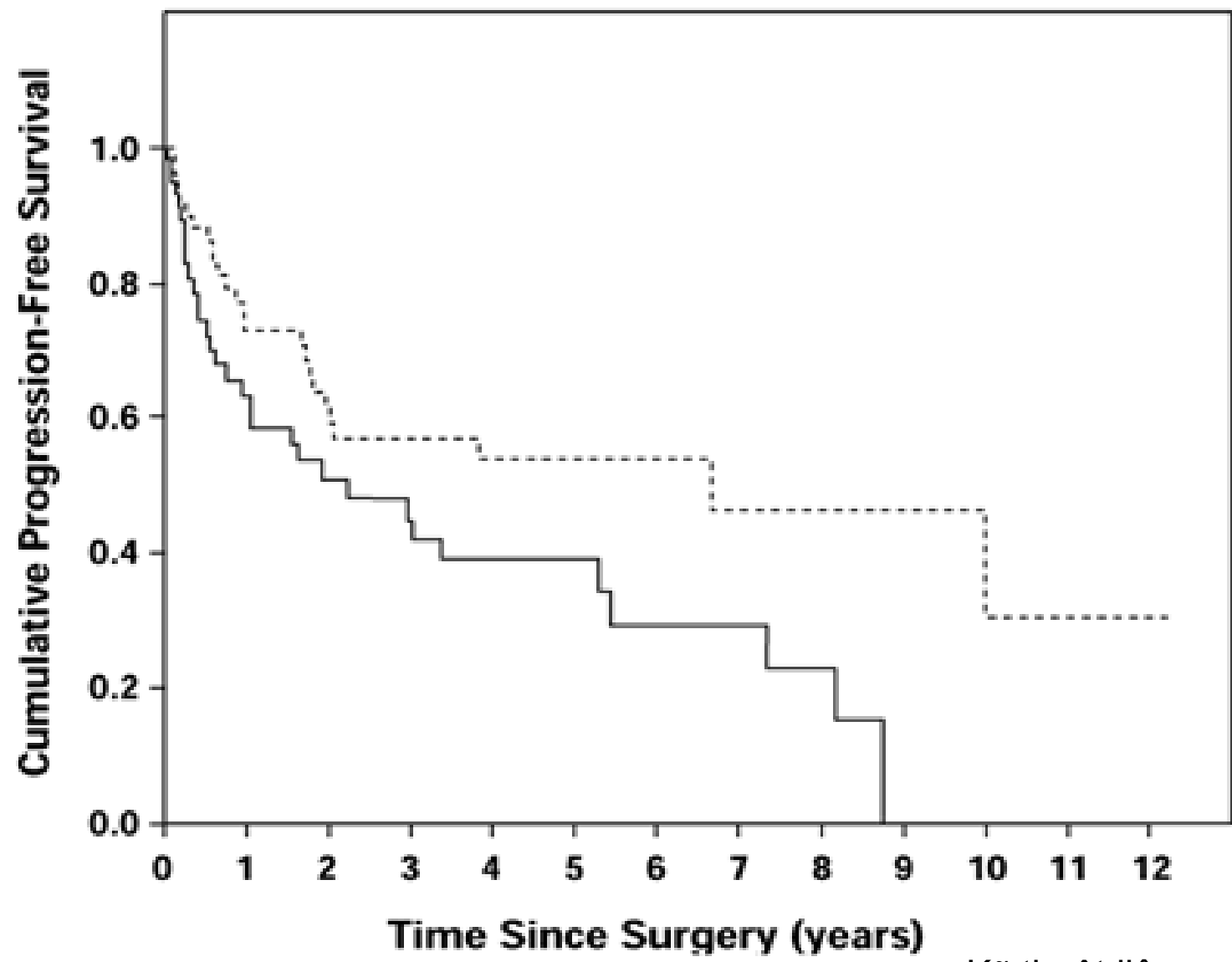


Nguy cơ và biến đổi theo thời gian

- Điều trị mọc mụn tân sinh ở người ...



Tiến trình – thời gian bình phục hoàn toàn sau phẫu thuật.



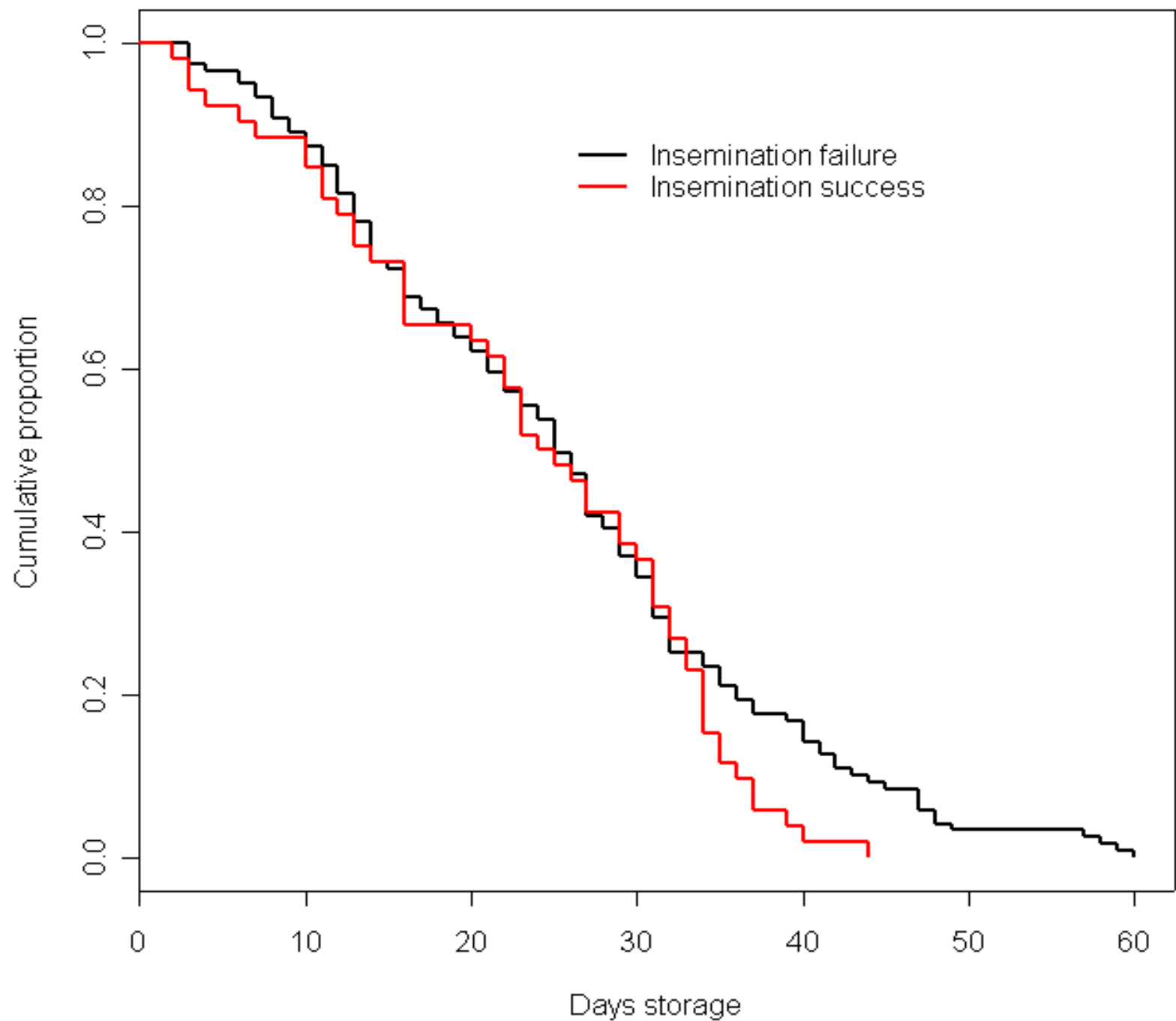
Kỹ thuật liên quan thay đổi cấu trúc dân số: thường không ổn định trong “đuôi” của các hàm survival được (một số lượng nhỏ đối tượng rời khỏi nghiên cứu)

Nguy cơ và sinh tồn

- Kết quả thụ tinh (thành công hay thất bại) là một hàm số về thời gian lưu trữ tinh dịch ...



Tác động của thời gian lưu trữ tinh dịch đến kết quả thụ tinh.



Nguy cơ và sinh tồn

- Vấn đề của việc biến mất các thìa cà phê...

The case of the disappearing teaspoons: longitudinal cohort study of the displacement of teaspoons in an Australian research institute

Megan S C Lim, Margaret E Hellard, Campbell K Aitken

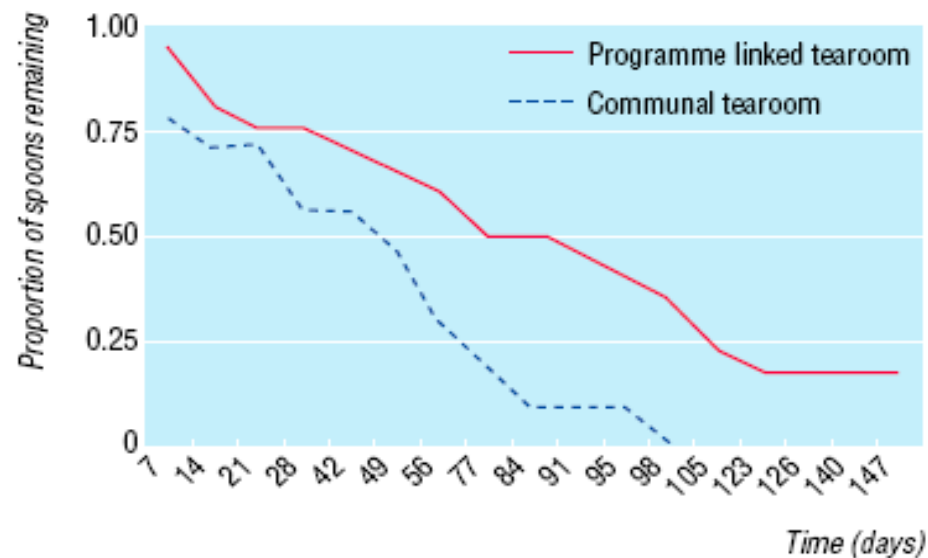


Fig 1 Proportion of teaspoons remaining by final tearoom type

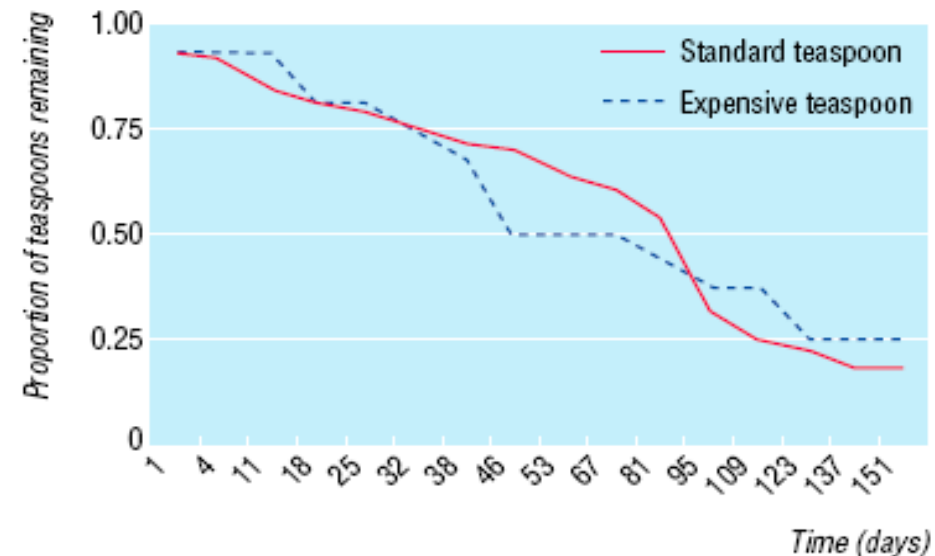


Fig 2 Proportion of teaspoons remaining by teaspoon type

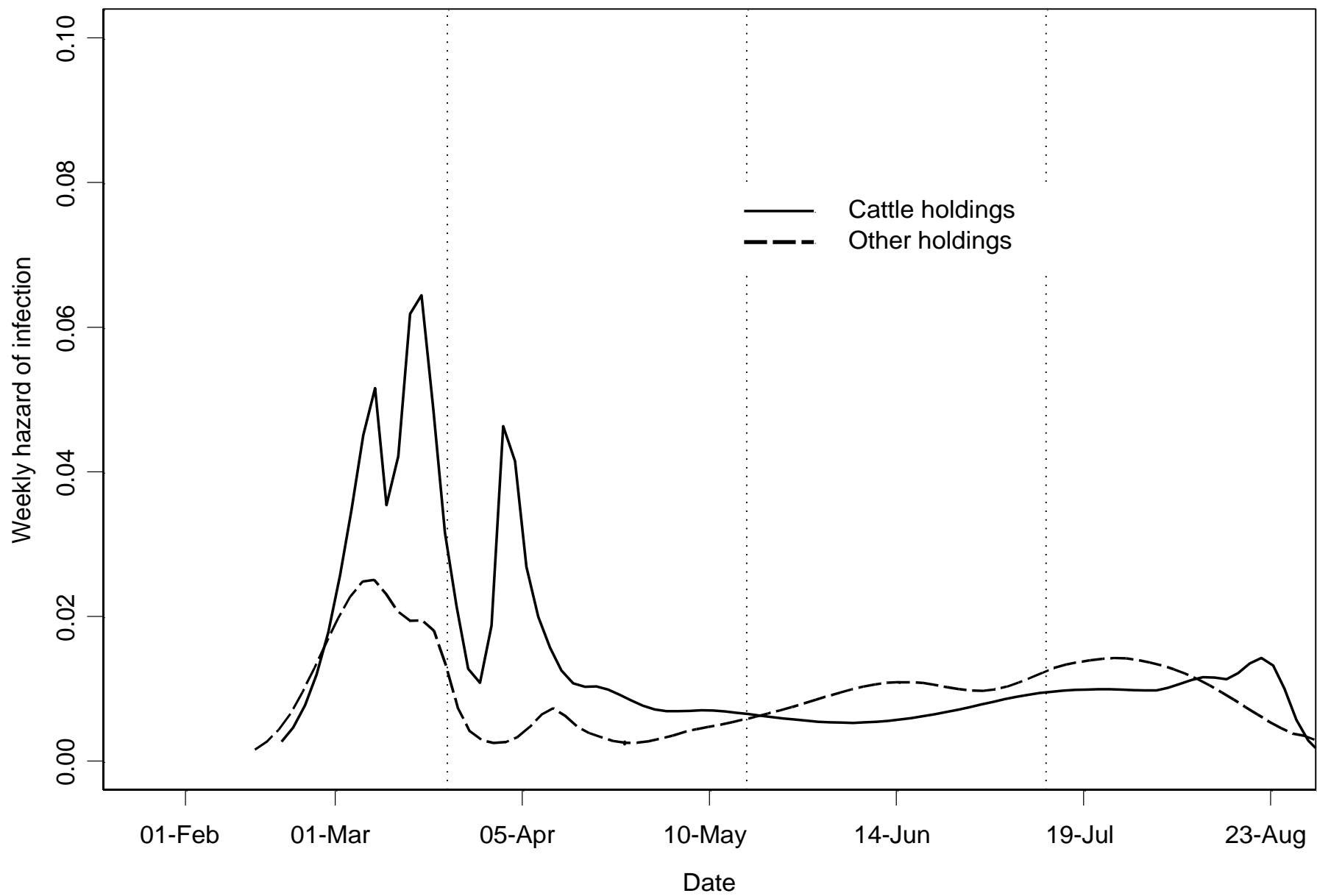


Lim, M., Hellard, M., Aitken, C., 2006. The case of the disappearing teaspoons: longitudinal cohort study of the displacement of teaspoons in an Australian research institute. *British Medical Journal* 331, 1498 - 1500.

Nguy cơ và sinh tồn

- nguy cơ tức thời
 - $h(t)$
 - \equiv ‘nguy cơ’ (Dohoo and Martin)
 - bằng với xác suất của một sự kiện xảy ra tại thời điểm t , với điều kiện là nó chưa xảy ra

Nguy cơ tức thời hàng tuần của bệnh LMLM ở Cumbria, February – September 2001.



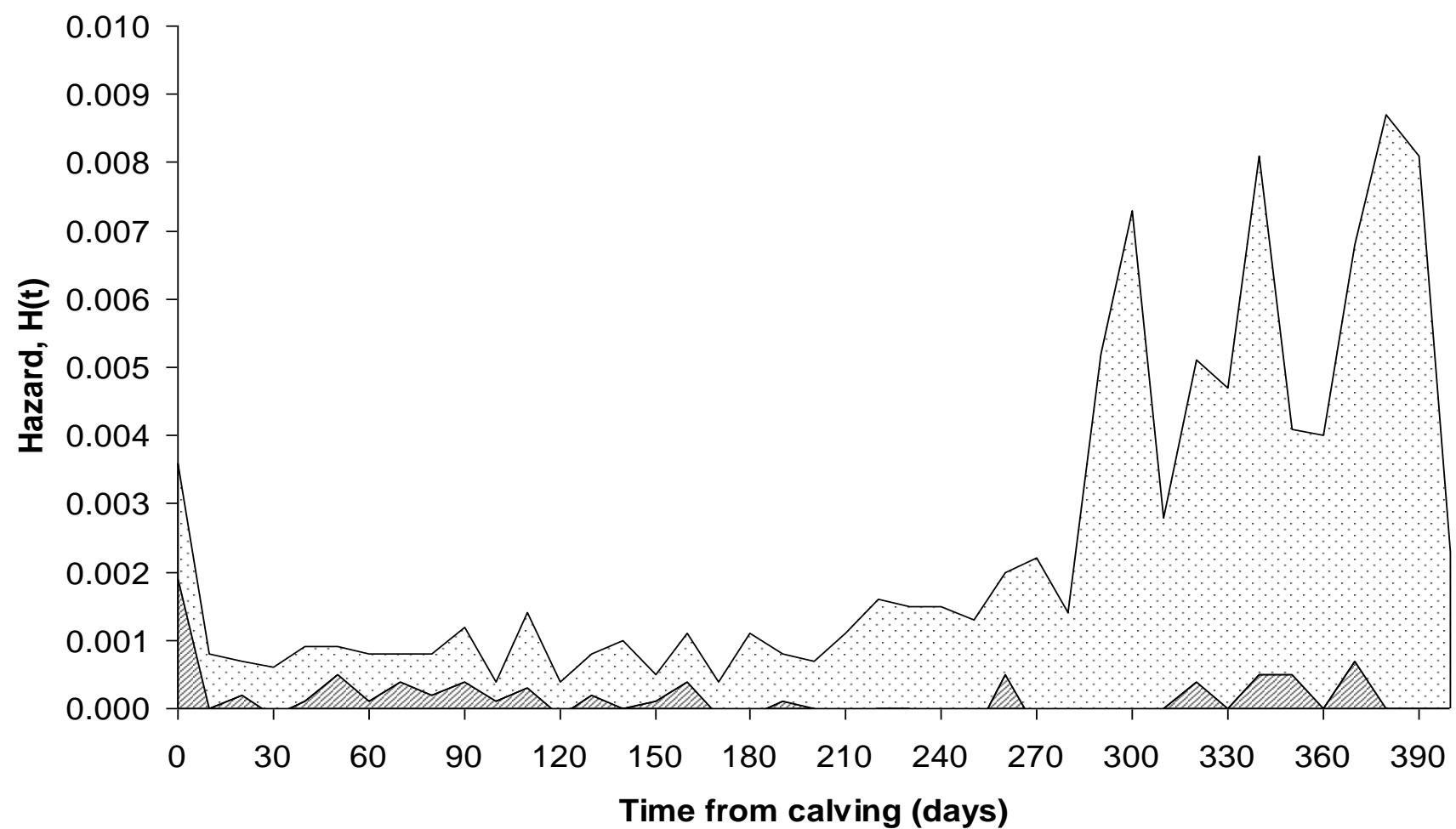
Wilesmith, J., Stevenson, M., King, C., Morris, R., 2003. Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001. Preventive Veterinary Medicine 61, 157 - 170.

Nguy cơ và sinh tồn

- Tuổi thọ của đàn bò sữa ...

Nguy cơ trong việc bán và chết ở bò sữa NSW.

Phần tối đậm là nguy cơ chết, phần tối nhẹ là nguy cơ trong việc bán bò.

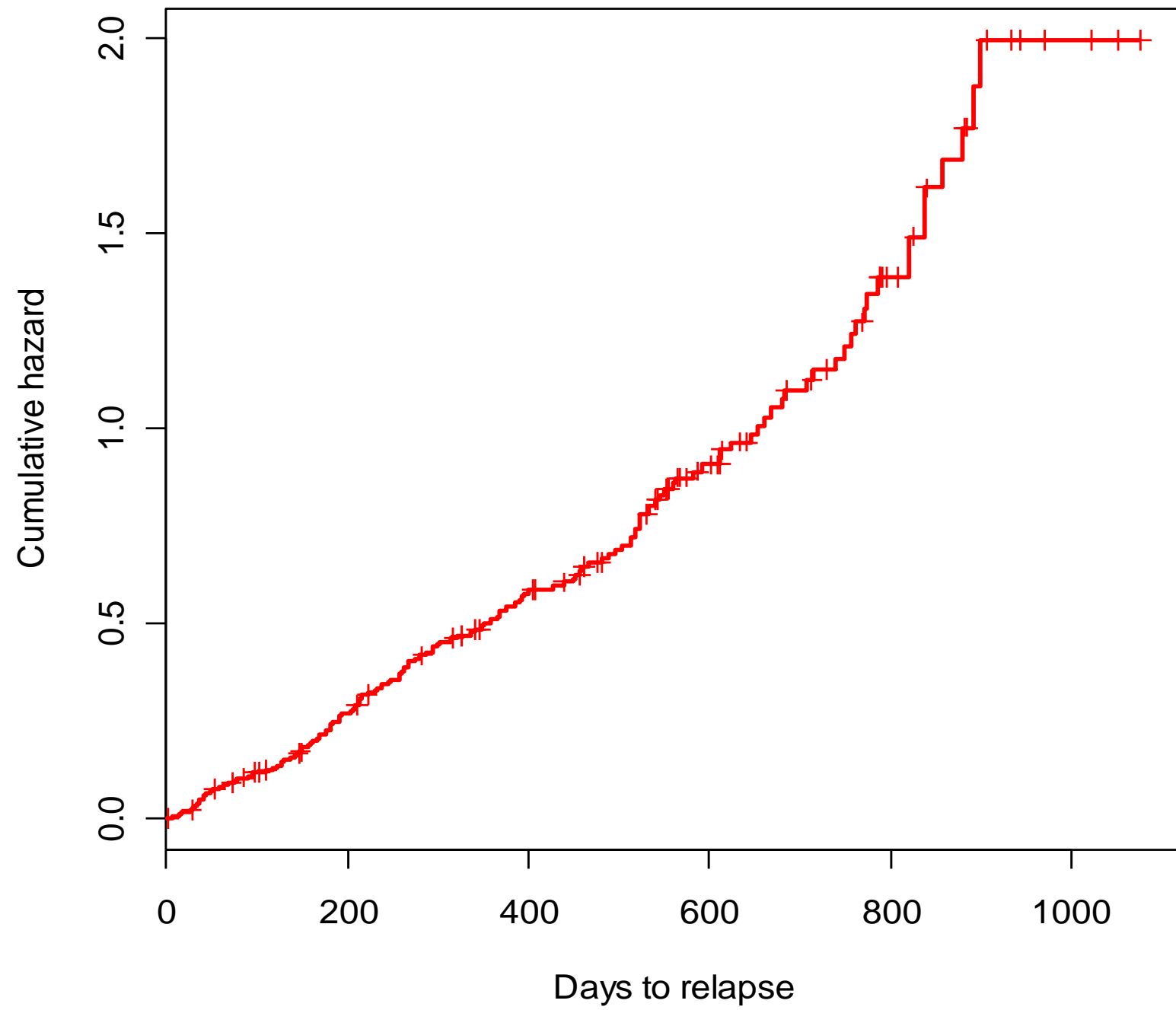


Stevenson, M., Lean, I., 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 482 - 488.

Nguy cơ và sinh tồn

- 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
 - không phải là một xác suất hay tỷ lệ
 - $H(t) = -\log S(t)$

Nguy cơ tích lũy tái nghiện ở người nghiện heroin sau khi xuất viện.



Lộ trình

- Sách tham khảo
- Tổng quan
- nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

Ví dụ

- Các ứng dụng của phân tích theo thời gian trong dịch tễ học thú y
 - Mô tả các hàm sinh tồn (survival) hoặc nguy cơ
 - So sánh các hàm sinh tồn (hoặc nguy cơ) giữa các nhóm
 - Định lượng các yếu tố ảnh hưởng đến sinh tồn



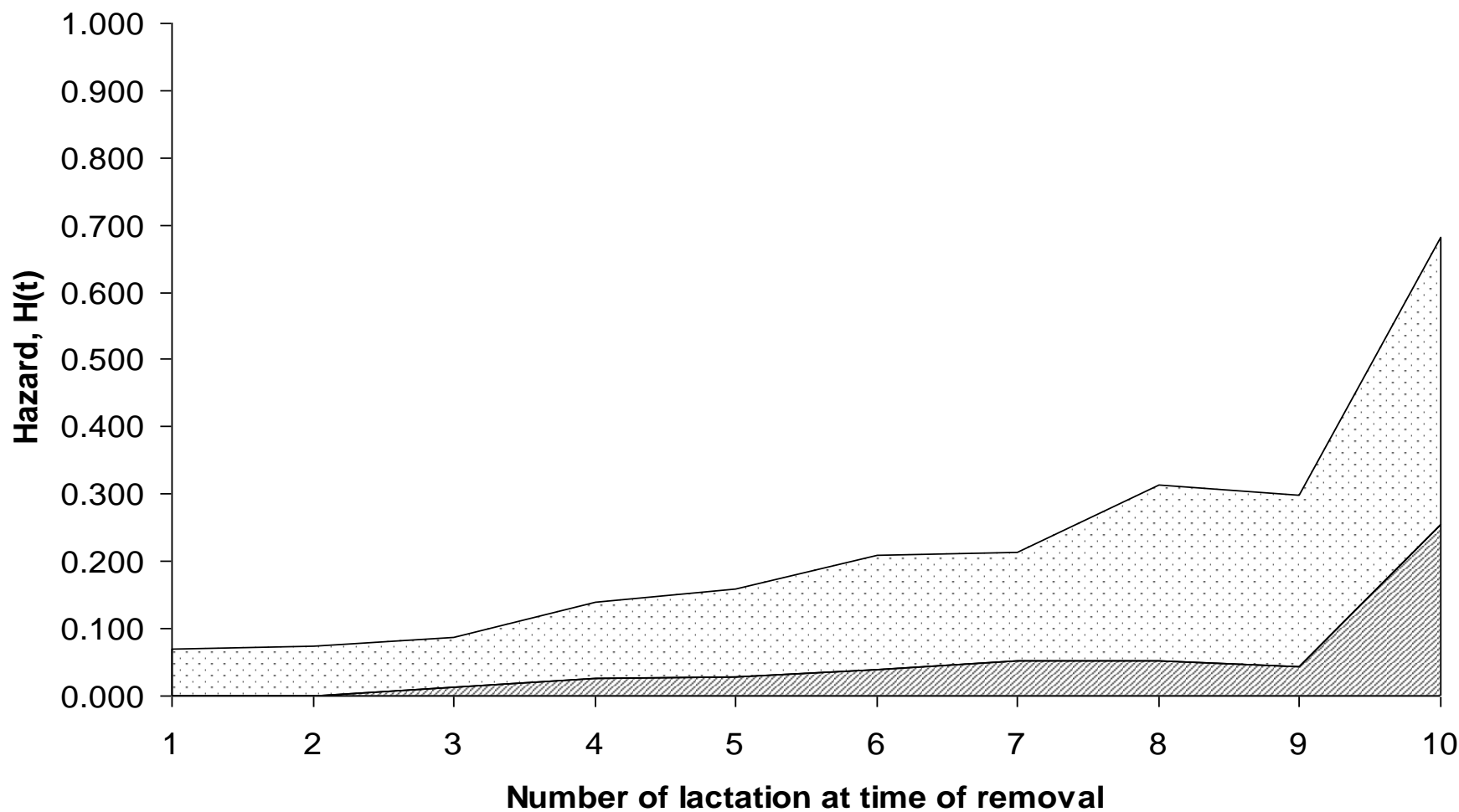






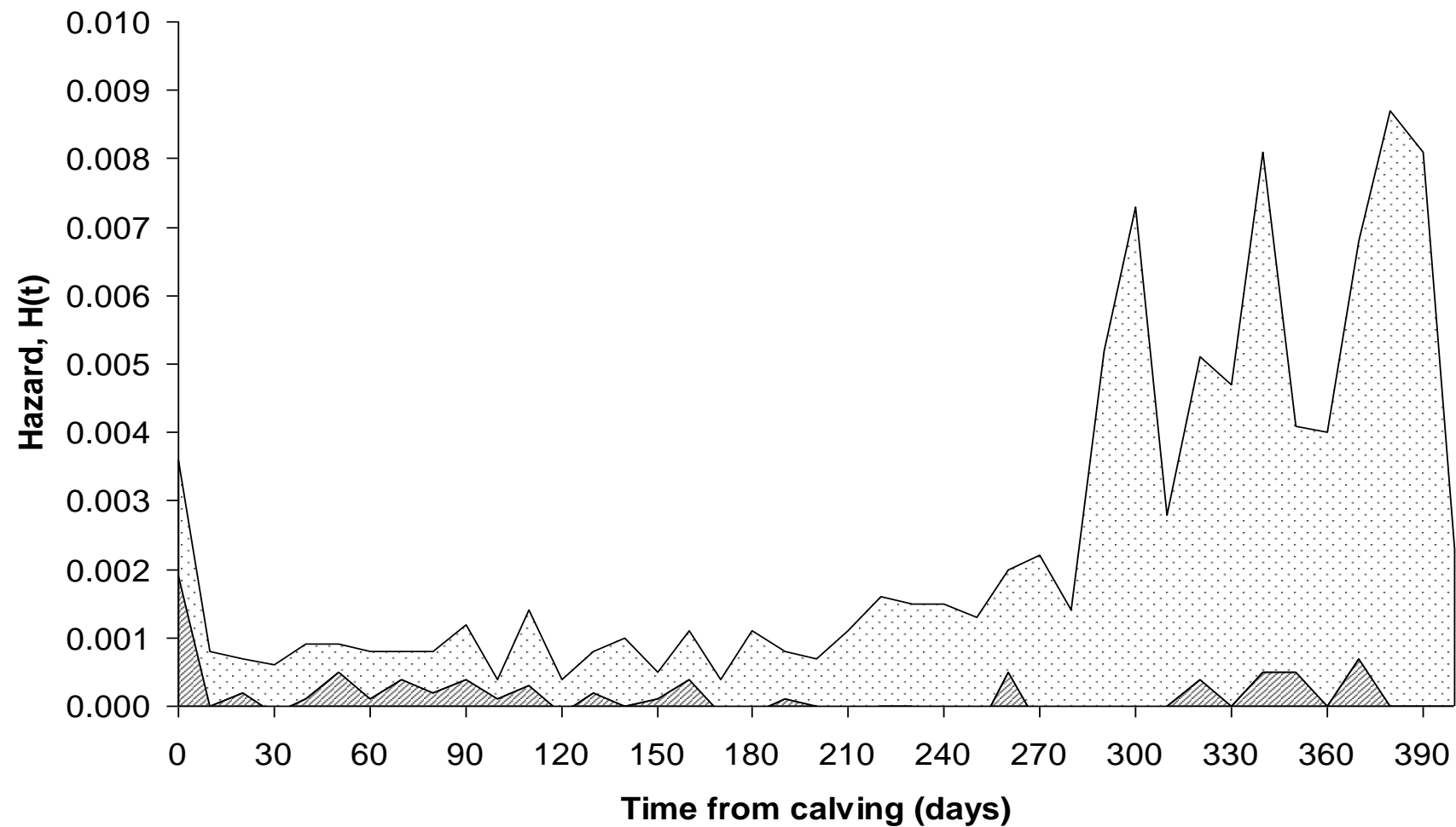


Nguy cơ bán và chết ở bò sữa NSW như là một hàm của số lượng sữa tiết ra.
Phần tối đậm là nguy cơ chết, phần tối nhẹ là nguy cơ trong việc bán bò.



Stevenson, M., Lean, I., 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 482 - 488.

Nguy cơ bán và chết ở bò sữa NSW như là một hàm của số ngày tiết sữa.
Phần tối đậm là nguy cơ chết, phần tối nhẹ là nguy cơ trong việc bán bò.



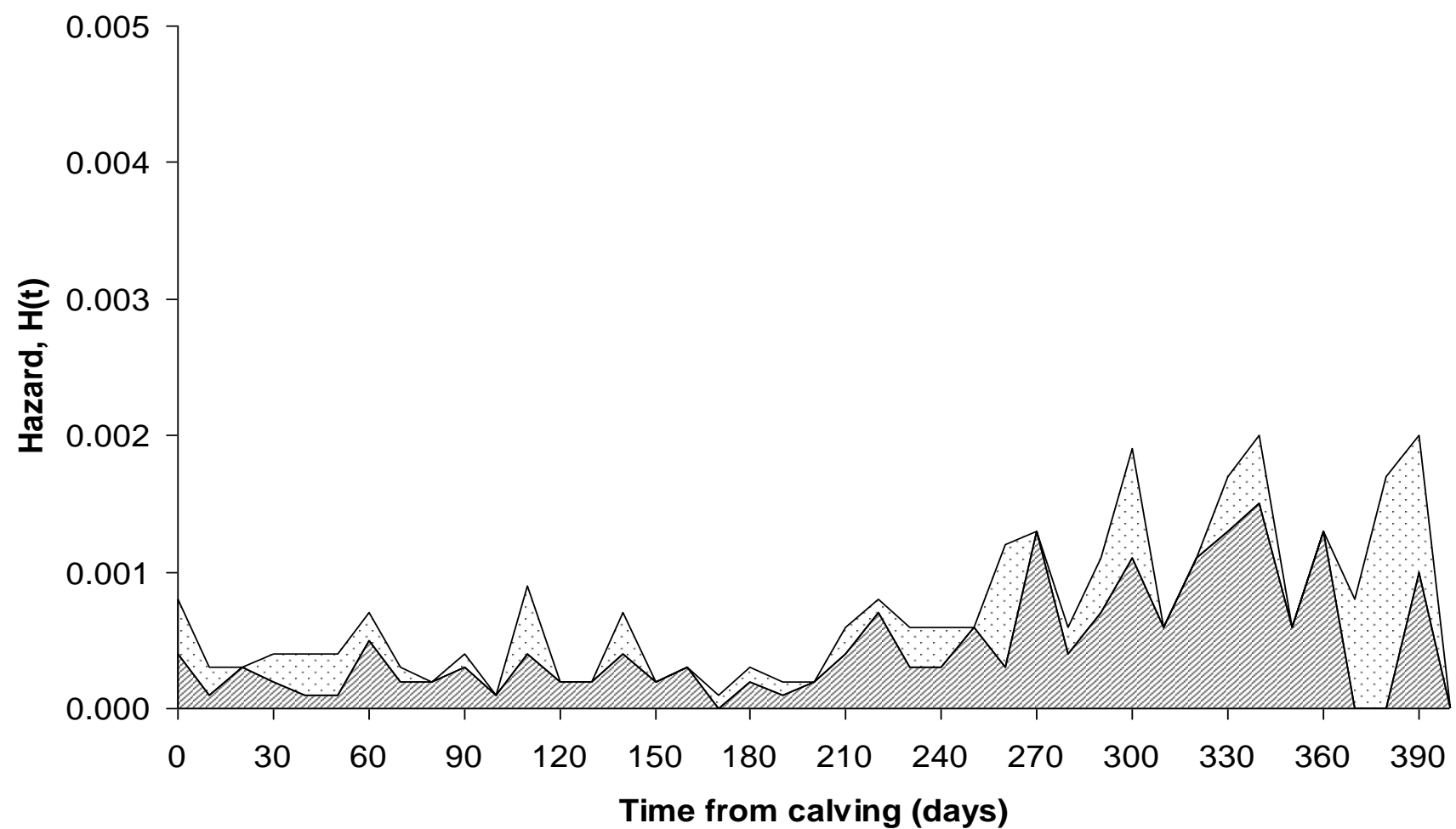
Stevenson, M., Lean, I., 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 482 - 488.

Nguy cơ bị loại do rối loạn bệnh về bầu vú ở bò sữa NSW.
Phần tối đậm là nguy cơ bị loại do viêm vú, phần tối nhạt là nguy cơ bị loại do các bệnh rối loạn khác.



Stevenson, M., Lean, I., 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 482 - 488.

Nguy cơ bị loại do rối loạn bệnh về bầu vú ở bò sữa NSW.
Phần tối đậm là nguy cơ bị loại do viêm vú, phần tối nhạt là nguy cơ bị loại do các bệnh rối loạn khác.

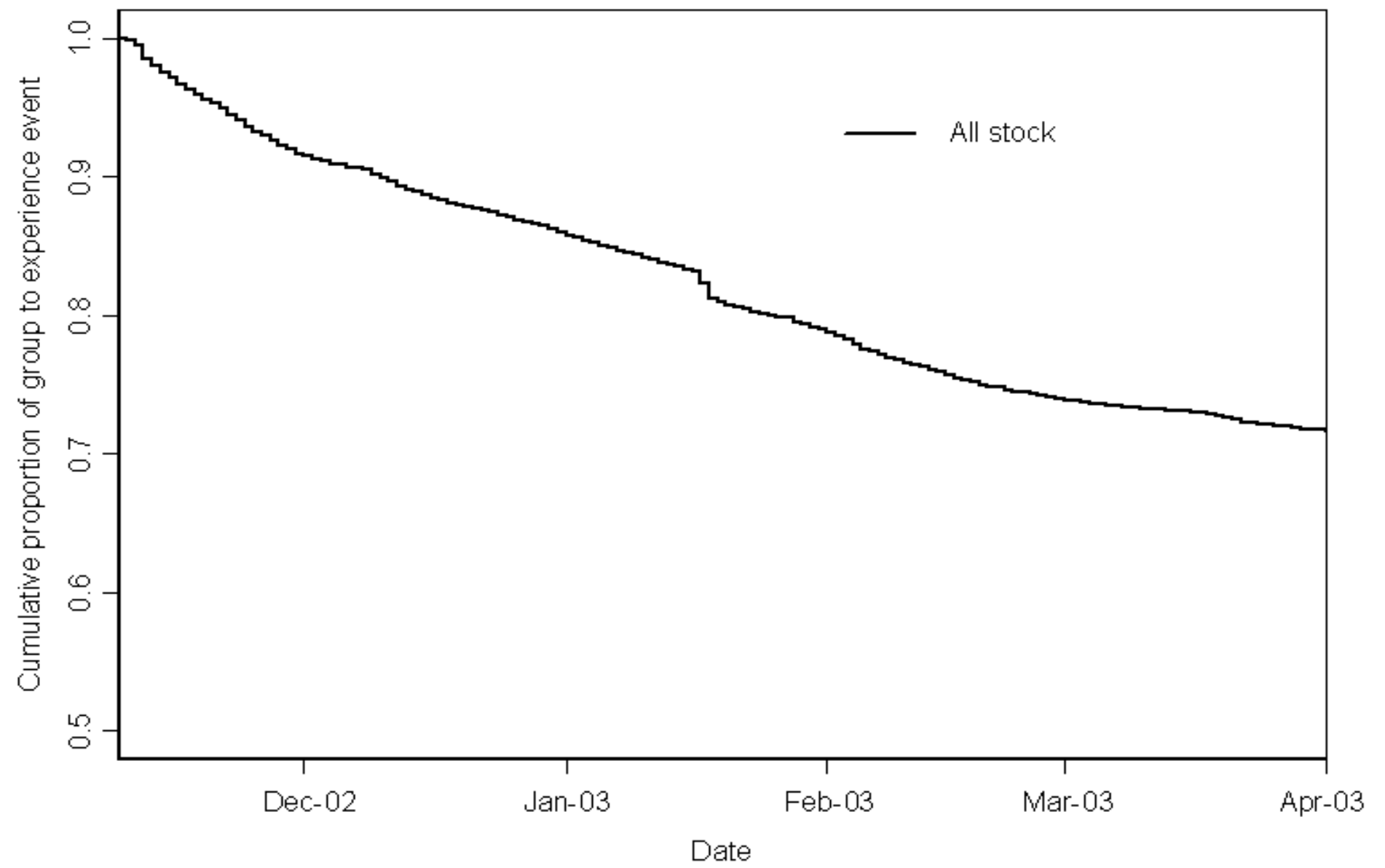


Stevenson, M., Lean, I., 1998. Descriptive epidemiological study on culling and deaths in eight dairy herds. Australian Veterinary Journal 76, 482 - 488.

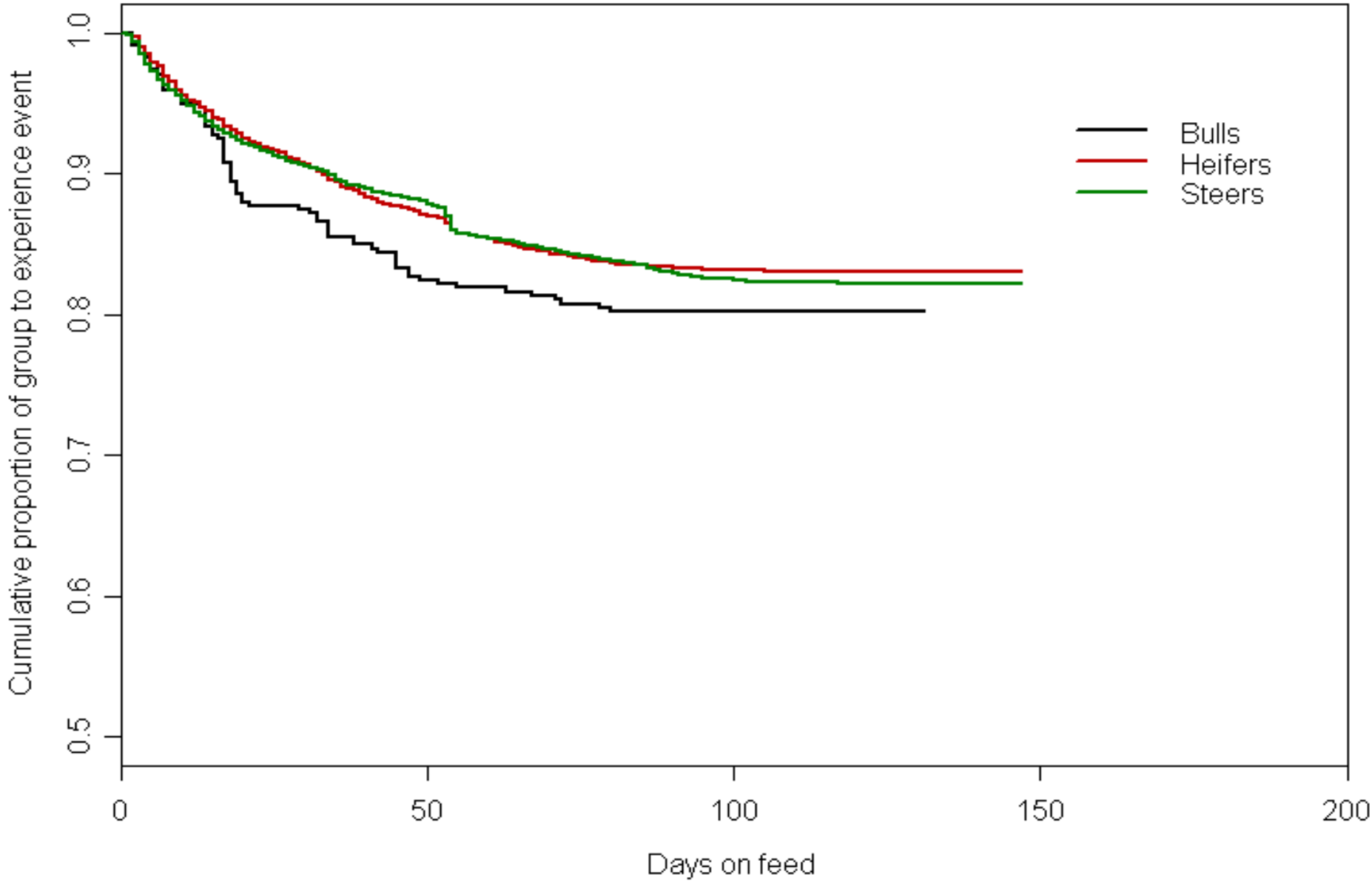
Một trại bò tại Indonesia.



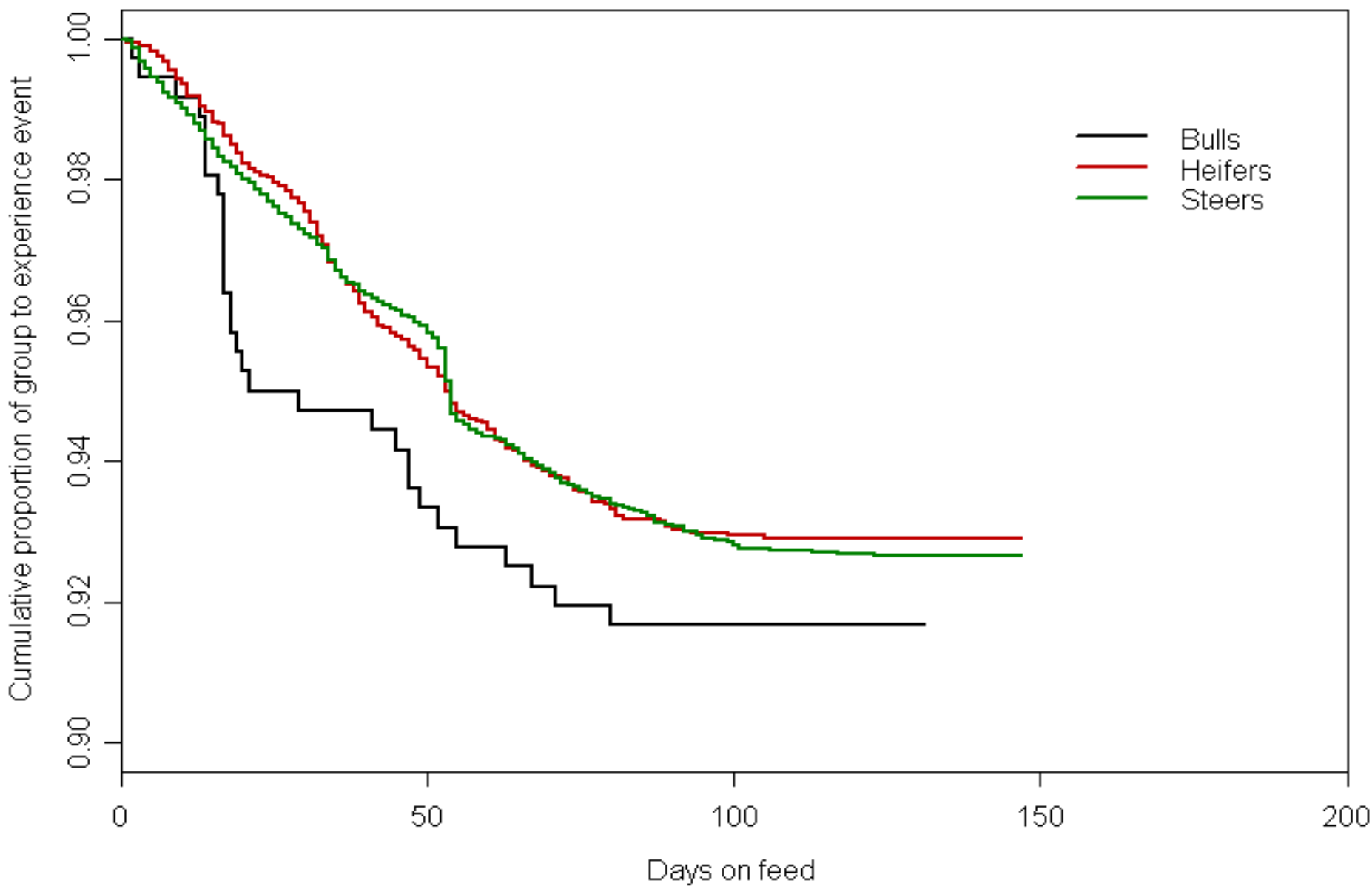
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của động vật bị kéo vì mọi lý do như một hàm số thời gian, áp dụng với tất cả đàn bò.



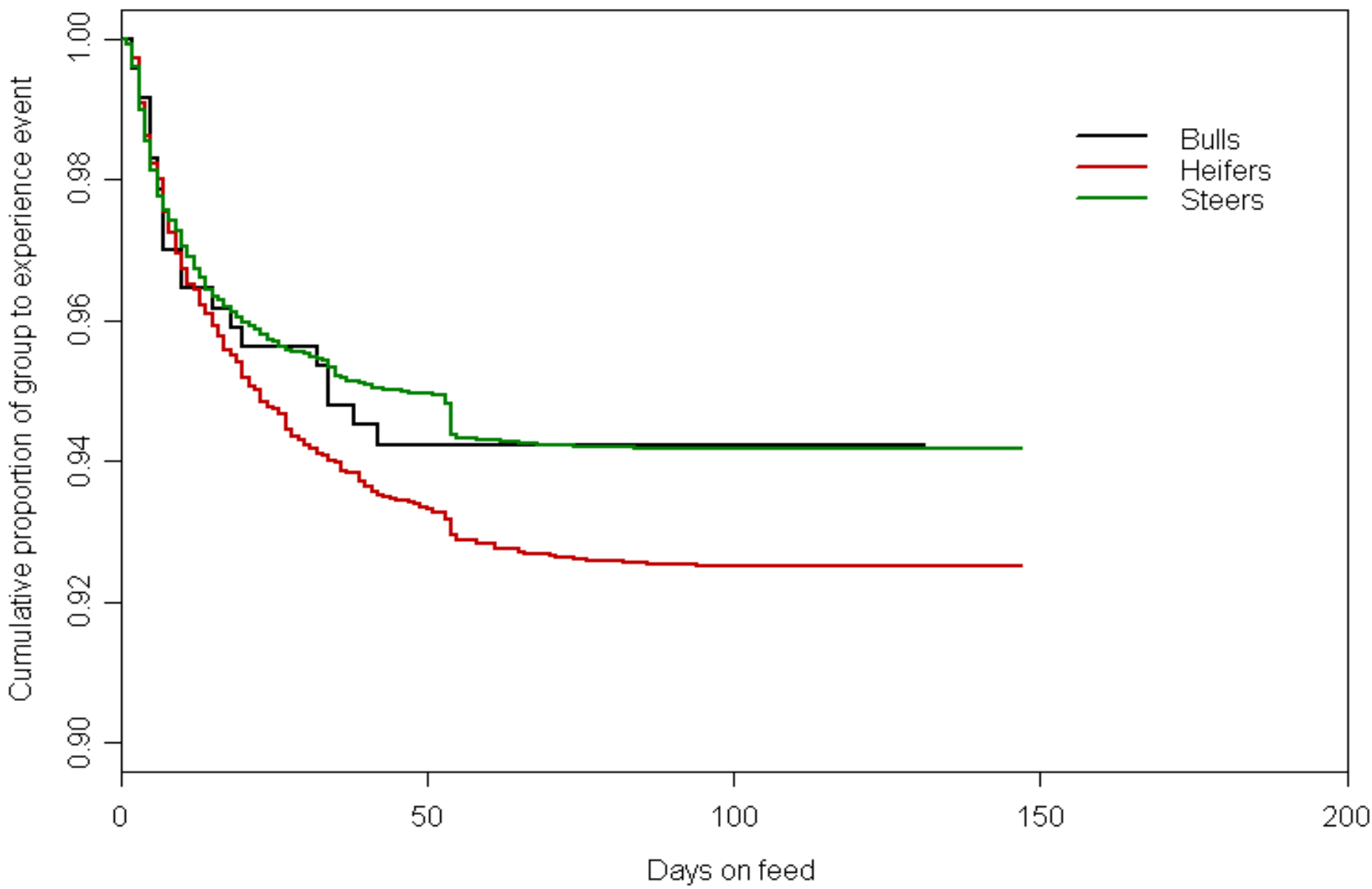
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của động vật bị kéo vì mọi lý do như một hàm số thời gian, theo giống bò



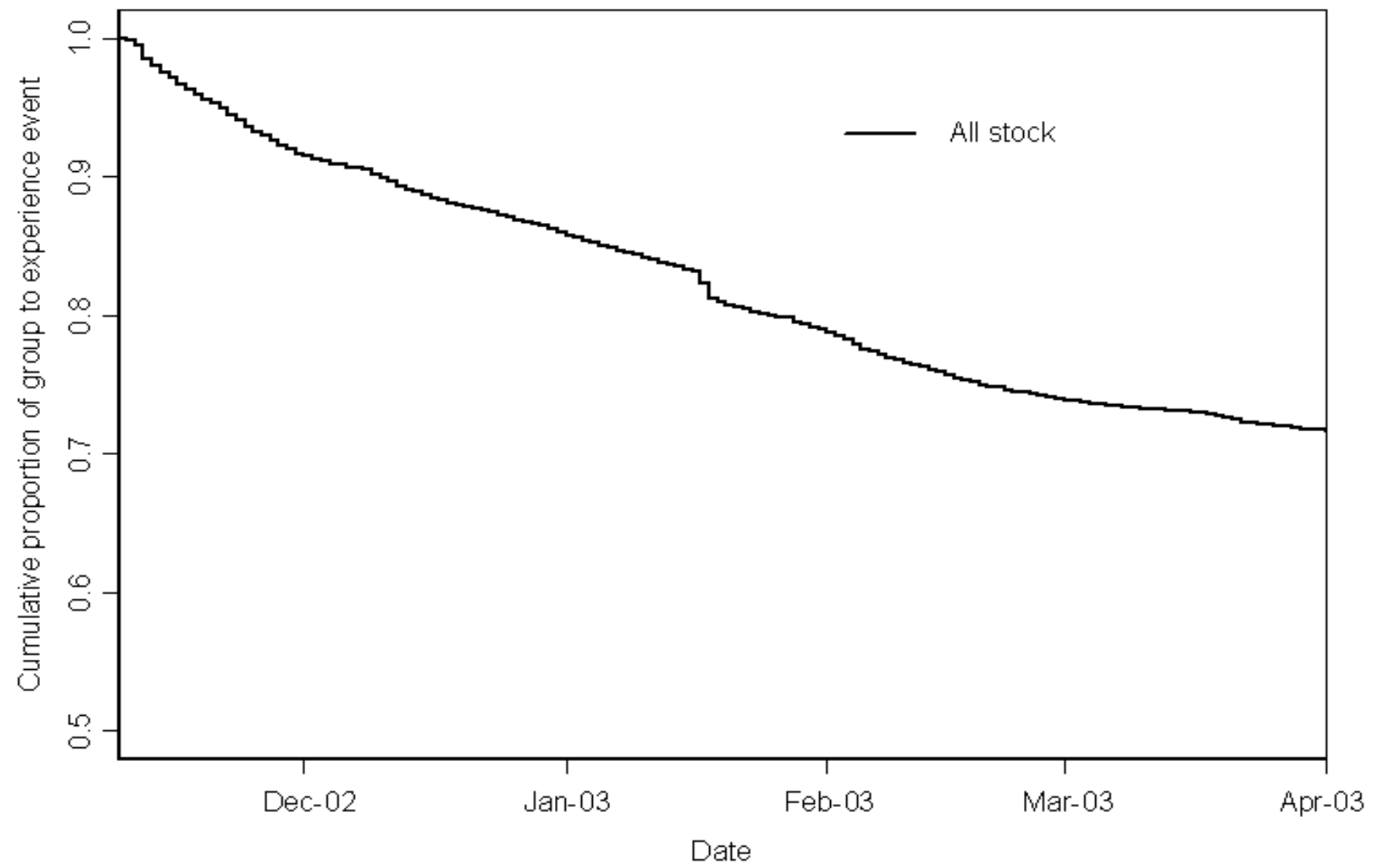
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của động vật bị kéo ra do què quặt như một hàm số thời gian, theo giống bò



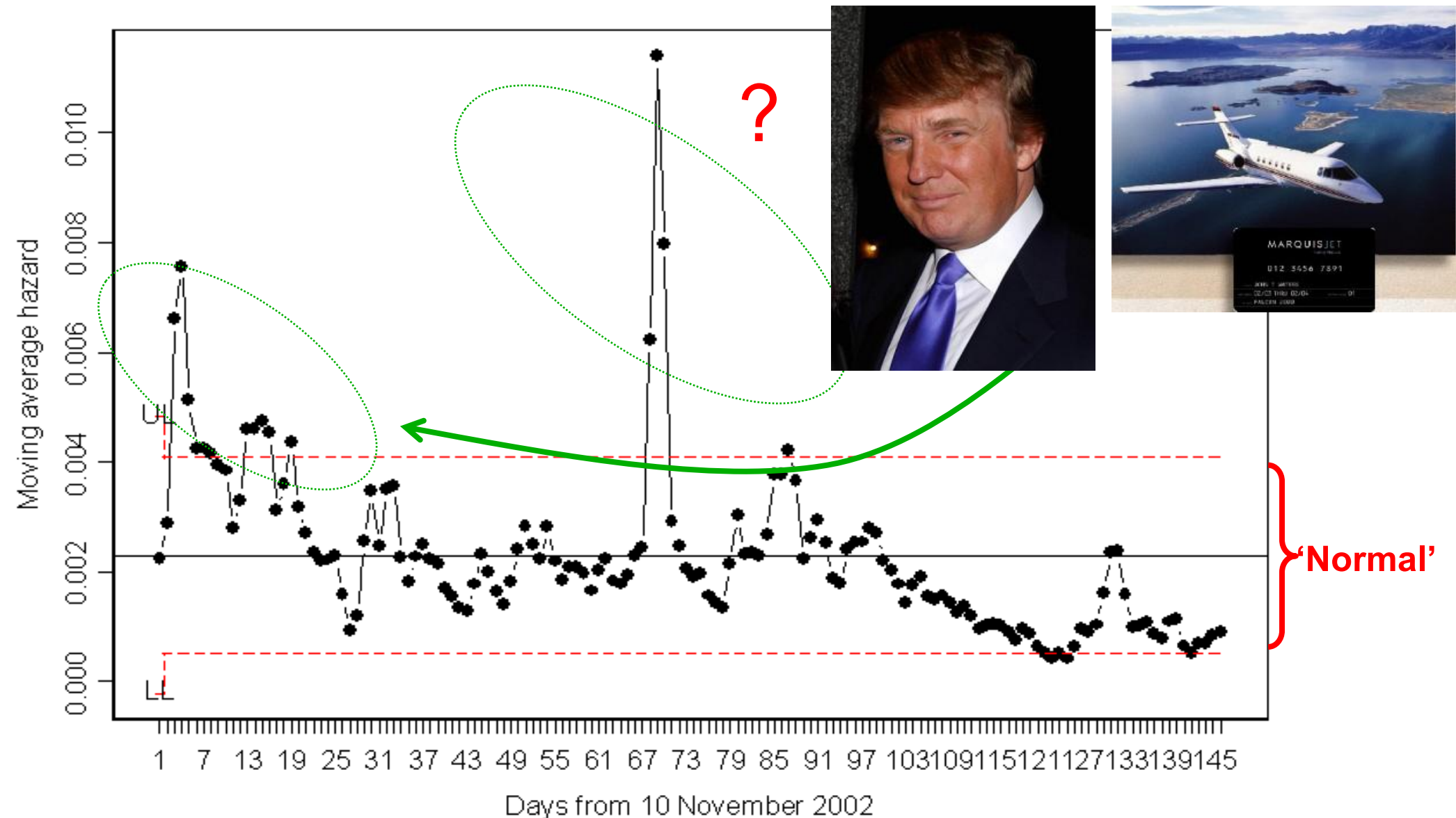
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của động vật bị kéo ra do ốm yếu như một hàm số thời gian, theo giống bò



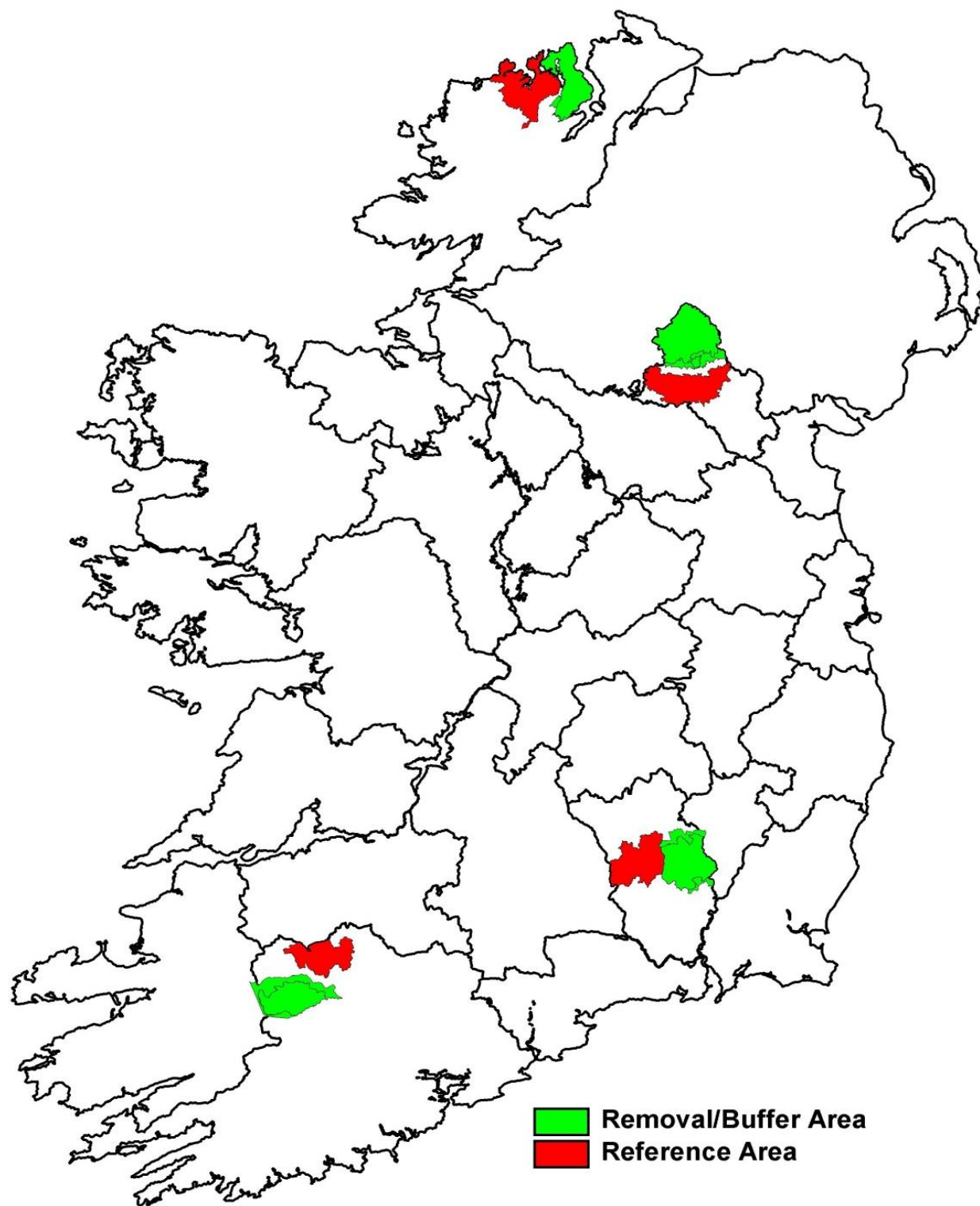
Đường cong sinh tồn của Kaplan-Meier cho thấy tỷ lệ tích lũy của động vật “bị kéo ra” như một hàm số thời gian, tính trên cả đàn bò



Nguy cơ tức thời của việc bị kéo vì tất cả các lý do như là một hàm số thời gian, tính trên cả đàn



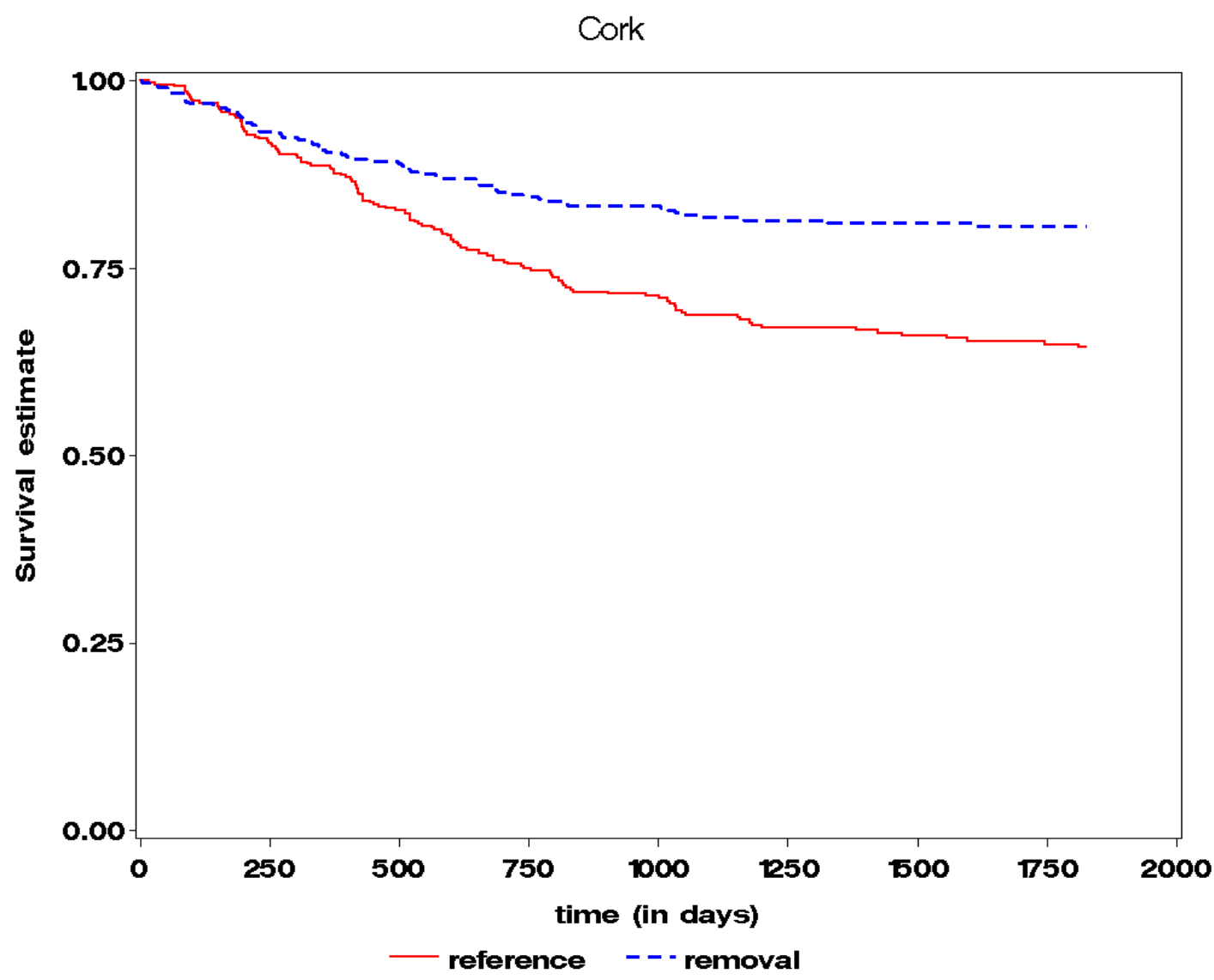
Ireland: four area badger removal project, 1997 – 2002.

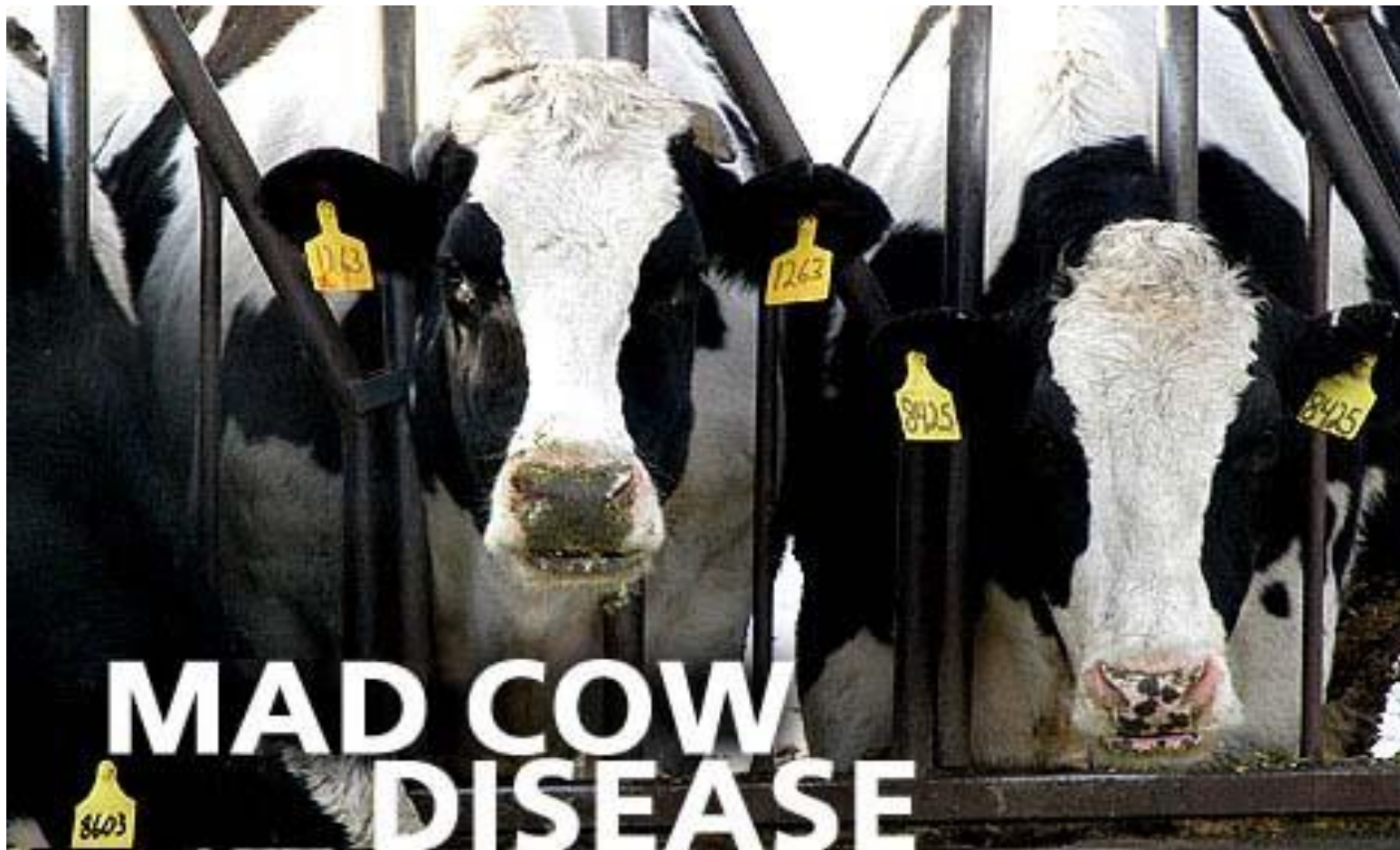


Ireland: four area badger removal project, 1997 – 2002.

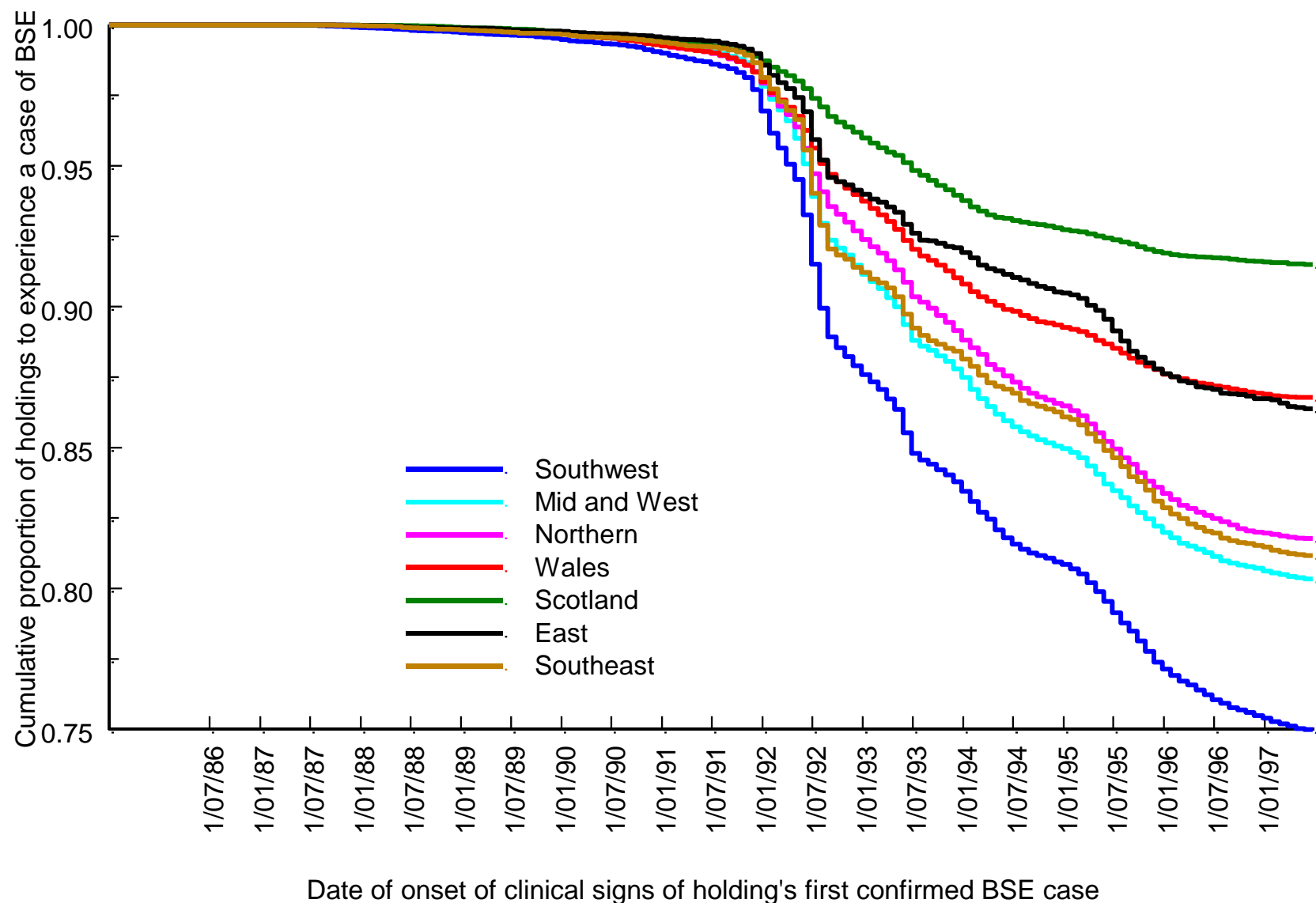


Ireland: four area badger removal project, 1997 – 2002.





Số ngày xuất hiện ca chỉ báo bệnh Bò điên – Các trại bò tại Anh.



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

- Sách tham khảo
- Tổng quan
- nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

Kiểm duyệt

- Kiểm duyệt
 - Trong một nghiên cứu tiếp theo, thời gian sinh tồn chính xác chỉ được biết đến với những cá thể liên quan đến sự kiện trong thời gian nghiên cứu
 - Các cá thể khác (đã được theo dõi đến cuối nghiên cứu) có thể liên quan đến sự kiện sau khi nghiên cứu kết thúc.
 - Những cá thể này được gọi là các quan sát kiểm duyệt
 - Các loại \equiv *phải, trái, khoảng*
 - Các ca kiểm duyệt thường được mã hóa là '0' và ca liên quan đến sự kiện được mã hóa là '1'

Kiểm duyệt

- Kiểm duyệt phải
 - Động vật bị ‘mất đi’ được theo dõi sau khi nghiên cứu được bắt đầu
 - ngụ ý rằng sự kiện đó đã có thể xảy ra sau ngày kiểm duyệt phải
 - Là dạng kiểm duyệt thông thường nhất

Kiểm duyệt

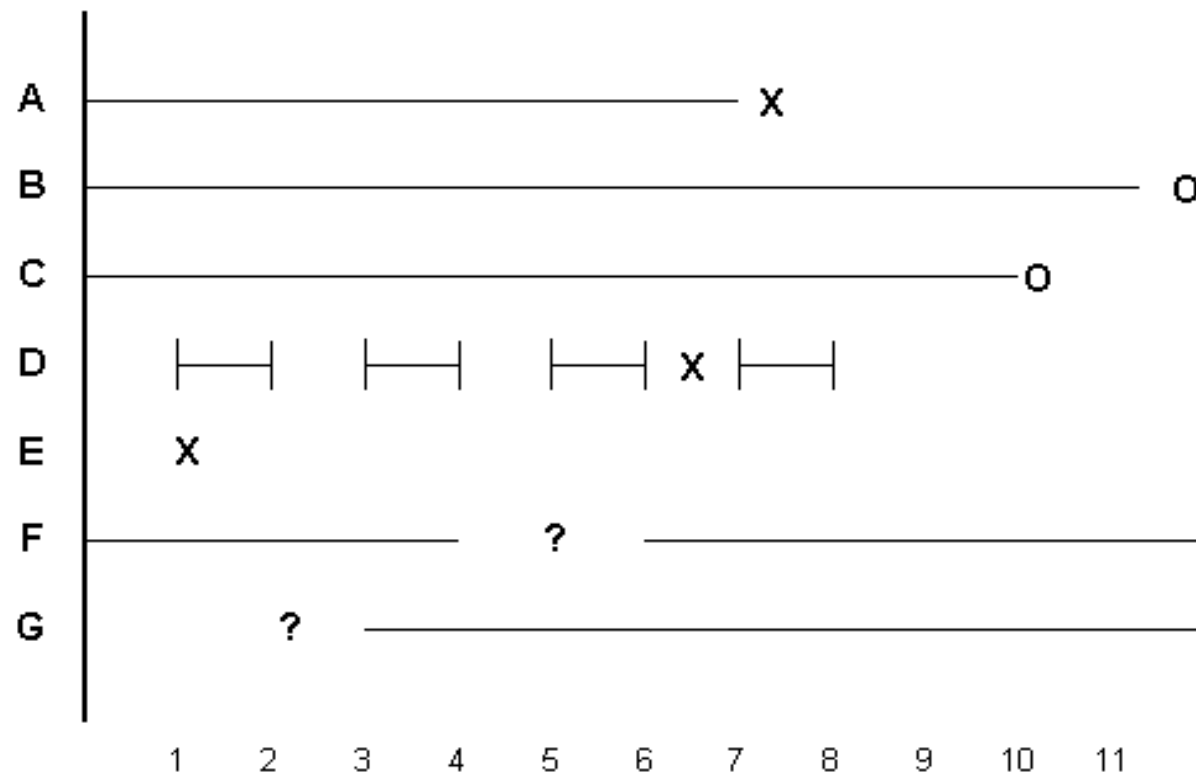
- Kiểm duyệt trái
 - Sự kiện đã xảy ra trước khi bắt đầu nghiên cứu
 - ngụ ý rằng sự kiện đã xảy ra đôi khi trước ngày kiểm duyệt trái
 - Ví dụ:
 - Lập kế hoạch điều tra thời gian bắt đầu động dục lần đầu tiên ở bò sữa
 - thời gian quan sát bắt đầu vào ngày thứ 30 sau tính từ ngày đẻ và kết thúc vào ngày thứ 100
 - kiểm tra tất cả động vật đã tham gia vào nghiên cứu (tại DSC = 30) và thấy rằng một số đã quay vòng - những động vật này sẽ bị kiểm duyệt trái vào ngày thứ 30

Kiểm duyệt

- Kiểm duyệt theo khoảng cách
 - Sự kiện được biết đôi khi xảy ra giữa các khoảng thời gian a và b nhưng chúng ta không biết chính xác khi nào
 - Sẽ hữu ích khi động vật được theo dõi định kỳ
 - kết quả quan tâm = ngày có kháng thể
 - Động vật được lấy mẫu 4 tháng/lần

Kiểm duyệt

- Dữ liệu bị cắt
 - that is, there is no way possible that the event could have occurred
 - một vấn đề riêng biệt từ kiểm duyệt
 - có thể bên trái hoặc bên phải
 - bắt buộc phải tính đến các khoảng thời gian khi đối tượng không bị nguy cơ
 - không có cách nào có thể xảy ra sự kiện
 - Kết quả quan tâm = ngày có ca bệnh chỉ báo trong đàn;
 - những đàn đã rời đi sau khi bắt đầu nghiên cứu bị cắt ngắn về bên trái cho đến ngày đầu tiên tồn tại



Sự kiện ngày thứ 7.

Kiểm duyệt phải vào ngày thứ 12.

Kiểm duyệt phải vào ngày thứ 10.

Kiểm duyệt khoảng cách vào ngày 6.5.

Kiểm duyệt trái vào ngày thứ 1.

Dữ liệu bị cắt ngắn về khoảng cách

Dữ liệu bị cắt ngắn bên trái.

x = sự kiện được quan tâm; o = được kiểm duyệt; ? = không có trong dữ liệu

Lộ trình

- Sách tham khảo
- Tổng quan
- nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

Các kết nối

- Các sự kiện được kết nối
 - Phân tích theo thời gia giả sử thời gian là *liên tục*
 - Kết quả đôi khi không thể đo đạc dựa trên thang liên tục
 - số dịch vụ khi có thụ thai
 - có nghĩa là nhiều sự kiện xảy ra cùng một lúc (tạo ra thời gian sự kiện ‘được kết nối’)
 - phương pháp phân tích đối với phân tích này: xấp xỉ Breslow và xấp xỉ Efron
 - analytical methods to deal with this: Ước lượng Breslow và ước Efron
 - Khi số kết nối lớn, các phương pháp ước lượng sẽ mang lại hệ số sai lệch về 0

Lộ trình

- Sách tham khảo
- Tổng quan
- nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu

Cấu trúc dữ liệu

Ngày kết thúc nghiên cứu = 30 November 2001

ID	Ngày nhập liệu	Ngày xử lý	Ghi chú
001	20 Oct 2001	31 Oct 2001	
002	20 Oct 2001	-	Sold 1 Nov 2001
003	10 Oct 2001	31 Oct 2001	
004	-	10 Nov 2001	Purchased 1 Nov 2001
005	20 Oct 2001	-	
006	20 Oct 2001	22 Oct 2001	

Cấu trúc dữ liệu

ID	Số ngày	Tình trạng
001	11	1
002	12	0
003	21	1
004	10	1
005	41	0
006	2	1

Lộ trình

- Sách tham khảo
- Tổng quan
- nguy cơ và biến đổi theo thời gian
- Các ví dụ
- Kiểm duyệt
- Các kết nối
- Cấu trúc dữ liệu



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.