幸存者分析

SURVIVAL ANALYSIS

OBJECTIVE

要点

- What is Survival Analysis?
- What is Kaplan-Meier Approach?
- What is the Log Rank test?
- What is Hazard Rate?
- What is Cox Proportional Hazards Regression Analysis?

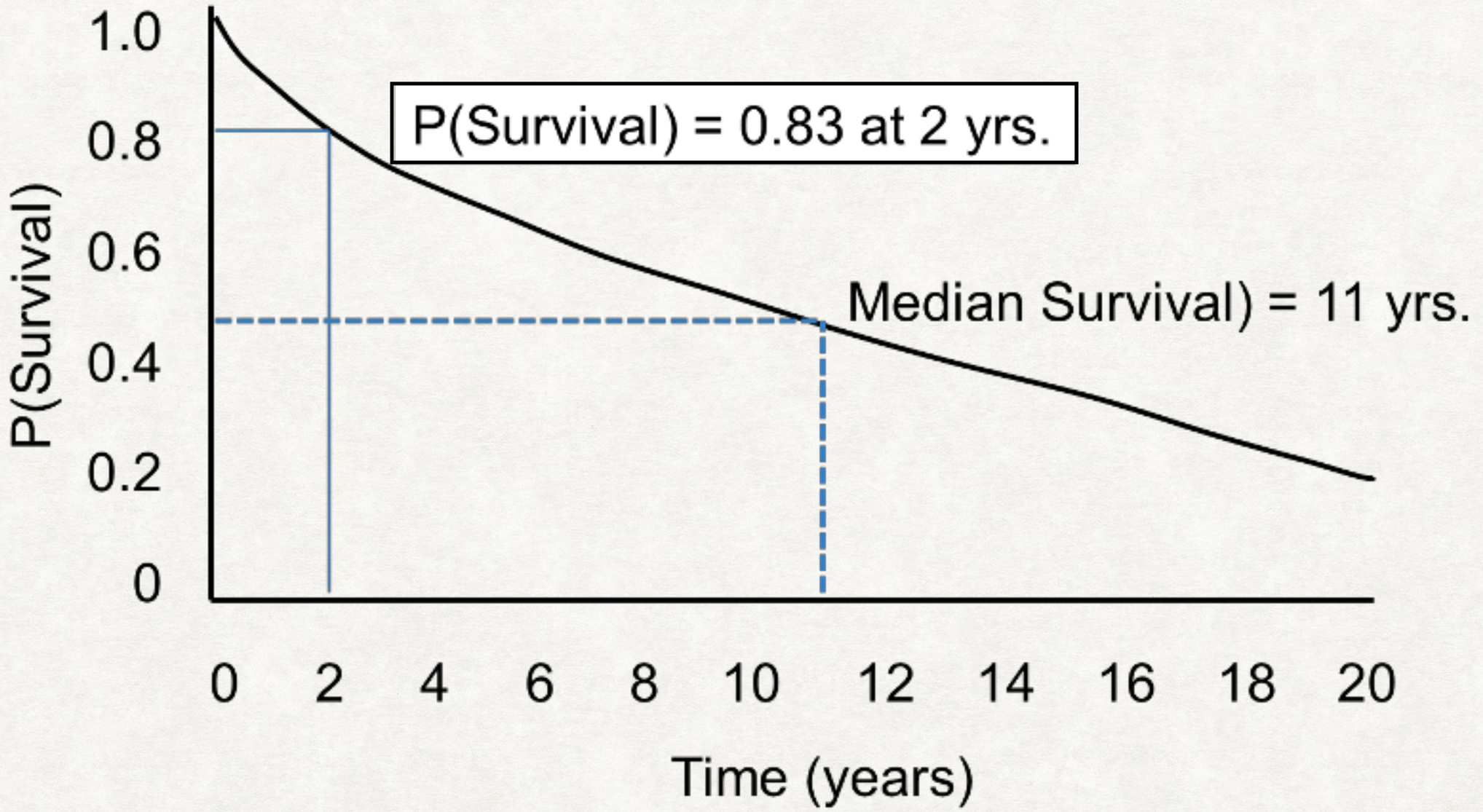
TIME TO EVENT VARIABLE

时间-事件变量

- 生存者分析 (Survival Analysis) 特别的地方在于 *Censoring*（删失，删截，截尾）
- Two important pieces of information:
  - Occurrence of event - 事件是否成立（定义）
  - Follow-up time / last contact - 最近一次接触时间
- Type of Censoring:
  - Left (Before enrolment) / Right (After enrolment)
  - Type I (Time is fixed) / Type II (Proportion is fixed)

SURVIVAL FUNCTION

生存函数



Source: Lisa Sullivan, Boston University School of Public Health

EXAMPLE DATA

样本数据

Participant Identification Number (ID)	Event Occurrence	Last Contact
1		24
2	3	
3		11
4		19
5		24
6		13
7	14	
8		2
9		18
10		17
11		24
12		21
13		12
14	1	
15		10
16	23	
17		6
18	5	
19		9

ESTIMATE SURVIVAL FUNCTION

计算生存概率

- Kaplan-Meier Approach / Product-limit approach
- Censoring is independent of the likelihood of developing the event
- KM is non-parametric

N_t Number at Risk

D_t Occurrence of Event

C_t Number of Censored Data

$$N_{t+1} = N_t - (D_t + C_t)$$

Survival Probability

$$S_{t+1} = S_t * \frac{N_{t+1} - D_{t+1}}{N_{t+1}}$$

LOG RANK TEST

时序检验

- Is there a significance difference between 2 groups?
  2组之间是否有显著差距?

- “The survival curves are identical or not?“
  两组的生存曲线是否一样?

- The log rank statistic is approximately distributed as a chi-square test statistic.
  和卡方检验的统计量分布接近

$$Z = \frac{\sum_{t=1}^{T} (O_{1t} - E_{1t})}{\sqrt{\sum_{t=1}^{T} V_t}} \rightarrow^{d} N(0,1)$$

COX PROPORTIONAL HAZARDS REGRESSION ANALYSIS

COX 比例风险回归模型

- Kaplan-Meier, Log Rank = Univariate Analysis (单变量分析)
- Cox measures **hazard rate (风险概率)**
- **Hazard rate = probability of event per person per unit time**
  e.g. hazard rate 0.2 @ time t and t is in unit month
- Hazard ratio = $\frac{HR_A}{HR_B}$
- Semi-parametric model
- Assumptions: **Relative hazard function is constant over time**

COX PROPORTIONAL HAZARDS REGRESSION ANALYSIS

COX 比例风险回归模型

Default Regression Model

$$h(t) = h_0(t)exp(b_1X_1 + b_2X_2 + \ldots + b_pX_p)$$

Relative Hazard

$$\frac{h(t)}{h_0(t)} = exp(b_1X_1 + b_2X_2 + \ldots + b_pX_p)$$

$$ln\frac{h(t)}{h_0(t)} = (b_1X_1 + b_2X_2 + \ldots + b_pX_p)$$

$h_0(t)$ = *baseline hazard rate*

OBJECTIVE

要点

- What is Survival Analysis?
- What is Kaplan-Meier Approach?
- What is the Log Rank test?
- What is Hazard Rate?
- What is Cox Proportional Hazards Regression Analysis?

REFERENCE

产考

British Journal of Cancer (2003) 89, 232–238
© 2003 Cancer Research UK All rights reserved 0007–0920/03 $25.00
www.bjcancer.com

Tutorial Paper

Survival Analysis Part I: Basic concepts and first analyses

TG Clark*,[1], MJ Bradburn[1], SB Love[1] and DG Altman[1]

[1]*Cancer Research UK/NHS Centre for Statistics in Medicine, Institute of Health Sciences, University of Oxford, Old Road, Oxford OX3 7LF, UK*

British Journal of Cancer (2003) **89,** 232–238. doi:10.1038/sj.bjc.6601118 www.bjcancer.com

Keywords: survival analysis; statistical methods; Kaplan-Meier

INTRODUCTION

In many cancer studies, the main outcome under assessment is the time to an event of interest. The generic name for the time is *survival time*, although it may be applied to the time 'survived' from complete remission to relapse or progression as equally as to the time from diagnosis to death. If the event occurred in all individuals, many methods of analysis would be applicable. However, it is usual that at the end of follow-up some of the individuals have not had the event of interest, and thus their true time to event is unknown. Further, survival data are rarely Normally distributed, but are skewed and comprise typically of many early events and relatively few late ones. It is these features of the data that make the special methods called *survival analysis* necessary.

This paper is the first of a series of four articles that aim to introduce and explain the basic concepts of survival analysis. Most survival analyses in cancer journals use some or all of Kaplan–Meier (KM) plots, logrank tests, and Cox (proportional hazards) regression. We will discuss the background to, and interpretation of, each of these methods but also other approaches to analysis that deserve to be used more often. In this first article, we will present the basic concepts of survival analysis, including how to produce and interpret survival curves, and how to quantify and test survival differences between two or more groups of patients. Future papers in the series cover multivariate analysis and the last paper introduces some more advanced concepts in a brief question and answer format. More detailed accounts of these methods can be found in books written specifically about survival analysis, for example, Collett (1994), Parmar and Machin (1995) and Kleinbaum (1996). In addition, individual references for the methods are presented throughout the series. Several introductory texts also describe the basis of survival analysis, for example, Altman (2003) and Piantadosi (1997).

TYPES OF 'EVENT' IN CANCER STUDIES

In many medical studies, time to death is the event of interest. However, in cancer, another important measure is the time between response to treatment and recurrence or relapse-free survival time (also called disease-free survival time). It is important to state what the event is and when the period of observation starts and finishes. For example, we may be interested

CENSORING MAKES SURVIVAL ANALYSIS DIFFERENT

The specific difficulties relating to survival analysis arise largely from the fact that only some individuals have experienced the event and, subsequently, survival times will be unknown for a subset of the study group. This phenomenon is called censoring and it may arise in the following ways: (a) a patient has not (yet) experienced the relevant outcome, such as relapse or death, by the time of the close of the study; (b) a patient is lost to follow-up during the study period; (c) a patient experiences a different event that makes further follow-up impossible. Such censored survival times underestimate the true (but unknown) time to event. Visualising the survival process of an individual as a time-line, their event (assuming it were to occur) is beyond the end of the follow-up period. This situation is often called *right censoring*. Censoring can also occur if we observe the presence of a state or condition but do not know where it began. For example, consider a study investigating the time to recurrence of a cancer following surgical removal of the primary tumour. If the patients were examined 3 months after surgery to determine recurrence, then those who had a recurrence would have a survival time that was *left censored* because the actual time of recurrence occurred less than 3 months after surgery. Event time data may also be *interval censored*, meaning that individuals come in and out of observation. If we consider the previous example and patients are also examined at 6 months, then those who are disease free at 3 months and lost to follow-up between 3 and 6 months are considered interval censored. Most survival data include right censored observations, but methods for interval and left censored data are available (Hosmer and Lemeshow, 1999). In the remainder of this paper, we will consider right censored data only.

In general, the feature of censoring means that special methods of analysis are needed, and standard graphical methods of data exploration and presentation, notably scatter diagrams, cannot be used.

ILLUSTRATIVE STUDIES

Ovarian cancer data

This data set relates to 825 patients diagnosed with primary epithelial ovarian carcinoma between January 1990 and December

Survival Analysis Part I: Basic concepts and first analyses

Survival Analysis Part II: Multivariate data analysis - an introduction to concepts and methods

Source: T G Clark, M J Bradburn, S B Love & D G Altman

British Journal of Cancer volume 89, pages 232–238 (21 July 2003)

https://www.nature.com/articles/6601118#article-info