

Chapter 1

Introduction: Several Classical Data Examples for Survival Analysis

The proportional hazards (PH) model was proposed by Sir David Cox just over 40 years ago (Cox 1972). Today, the Cox model is the most important model in survival analysis, reliability and quality of life research, epidemiology, clinical trials, and biomedical studies. There have also been tremendous applications of the Cox model in demography, econometrics, finance, pharmacology, biology, gerontology, insurance, etc. These have marked the great success of the Cox PH model which further induced extended studies of competitive survival regression models and the corresponding development of semiparametric estimation theory, likelihood principle, counting process modeling and applications.

The developments in reliability and survival analysis have provided the basis and useful methods to obtain general theory. A patient's survival depends on his/her age, sex, fatigue, genetic or physiological damages, the dynamics of body temperature, body weight (or BMI), some physiological or biochemical indices, and also on the presence of chronic disease (like cancer, diabetes mellitus, renal disease, cardiac disease, metabolic syndrome, etc.). In general, these characteristics are coded as the so-called *covariates* or *explanatory variables*; some of them are called degradation processes. We suppose that the *lifespan* of an individual is described by *covariates*. In this case, the survival (or failure) of a patient is characterized by this covariate process and by the random moment of its potential failures. The Cox model is an example which relates the lifetime distributions to a set of covariates by modeling hazard rate.

The popularity and the success of the Cox model is based on the fact that there exist simple semiparametric estimation procedures and that the regression parameter in the PH model is easily interpreted as (log-) hazard ratio. The hazard ratios under different fixed covariates are usually assumed to be constant in time. In practice, the hazard rates may approach, go away from, or even intersect each other. In these circumstances, using the conventional Cox PH model to estimate the hazard ratio leads to biased inference. The phenomenon of *nonproportionality* may be derived from several aspects: First, some authors have considered the heterogeneity effect coming from individuals with unobserved frailty so that extra variations may be present (Hougaard 1984, 1986; Aalen 1988). Second, nonproportionality is part

of the result of the time-varying effect, which could possibly be modeled by the *varying coefficient* Cox model (Martinussen and Scheike 2006). Third, the interaction between time and a qualitative covariate gives nonproportionality (O’Quigley 1991). Finally, some observable covariates contribute both to the mean and to the variance of the lifetime variable or its transformation (Bagdonavičius and Nikulin 1999; Hsieh 2001; Zeng and Lin 2007), and thus produce “nonproportional hazards.” In the last case, stratification by some variables can eliminate part of the nonproportionality. However, stratification is not reasonable if a variable is of continuous type and, in particular, when the sample size is not large. Nevertheless, the Cox model helps to construct dynamic models well adapted to the study of survival functions with cross-effect. The PH model is generalized by assuming that at any moment, the hazard ratio depends on time-varying covariates. Relations with generalized proportional hazards, frailty, linear transformation, Sedyakin and degradation models and cross-effect models have been considered. Using some new flexible regression models, in this monograph, we analyze survival data of the Gastrointestinal Tumor Study Group (Stablein and Koutrouvelis 1985), the Veteran’s Administration lung cancer trials, the data of Piantadosi (1997) on lung cancer patients, the Stanford Heart Transplant data, and a dataset concerning the length of hospital stay of rehabilitating stroke patients.

These data examples illustrate the characteristics of survival data which may be collected from clinical operation (the Stanford Heart Transplant data), hospital registration system (length of hospital stay for stroke patients), and clinical trials (gastric cancer data and lung cancer data). In these data, survival estimates using the Kaplan–Meier method (see Chap. 2) are presented when the characteristics of proportional hazards (see Chap. 3) or nonproportional hazards (see Chaps. 5 and 6) according to different covariate configurations are considered.

1.1 Example 1: The Stanford Heart Transplant (SHT) Data

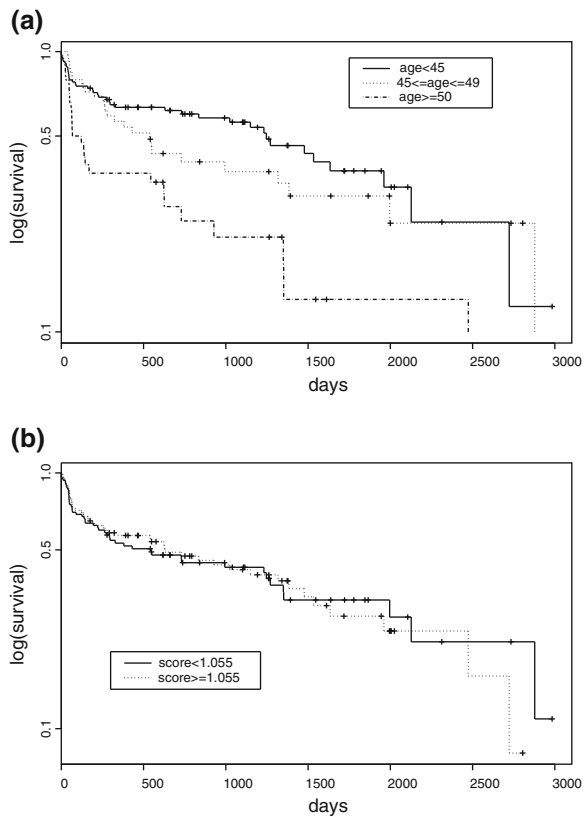
The SHT data reported in Miller and Halpern (1982) contains 184 patients with the following variables: survival time, dead/alive status, age and T5 mismatch scores. Cox and Oakes (1984, Chap. 8) tabulated another version of the SHT data which comprises 249 patients with transplant indicators and waiting times. Here, we consider the data presented in Miller and Halpern (1982). A complete dataset with 154 observations is used. We display the Kaplan–Meier (KM) survival estimates for different age and mismatch score groups. Derivation of the KM estimate and its properties are discussed in Chap. 2.

For the 154 observed times, 102 failed and 52 “right-censored” (explained in Chap. 2) times, the three quartiles of age are 35.0, 44.5, and 49.0. The younger two

groups ($\text{age} \leq 35.0$ and $35 < \text{age} < 45$) have no statistical difference in the lifetimes using the log-rank test (see Chap. 3); these two groups are combined. So we divide the patients into three groups: “ $\text{age} < 45$,” “ $45 \leq \text{age} \leq 49$,” and “ $\text{age} \geq 50$.” The survival estimates are shown in Fig. 1.1(a). The mismatch score measures the tissue incompatibility between recipient and donor; it can be viewed as a continuous random variable. The log-rank test reveals no significant difference in the lifetime distributions among the four groups formed by the quartiles 0.69, 1.05, and 1.49. We simply use the *median* ($T5 = 1.05$) as the cut-off point and plot the KM estimates for the two groups (Fig. 1.1b).

These two figures show that the survivals are significantly different in age, but not in (dichotomized) mismatch score. The group “ $\text{age} \geq 50$ ” has a sudden drop in survival at the early stage (time < 100 days). The other two younger groups have crossings at an early stage and at a time very close to 2000 days. It appears that the “difference between groups” varies with time. With a proportional hazards regression setting (Chap. 3), the effect of age cannot be modeled by a simple *univariate* variable age. As indicated by this example, seeking an alternative model is important.

Fig. 1.1 **a** KM estimates for different age groups. **b** KM estimates for different mismatch score groups. Reprinted from Journal of Statistical Planning and Inference, 139(12), H.-D.I. Wu, F. Hsieh, Heterogeneity and Varying Effect in Hazards Regression, pp. 4213–4222, Copyright 2015, with permission from Elsevier



1.2 Example 2: Length of Hospital Stay of Rehabilitating Stroke Patients in Taiwan

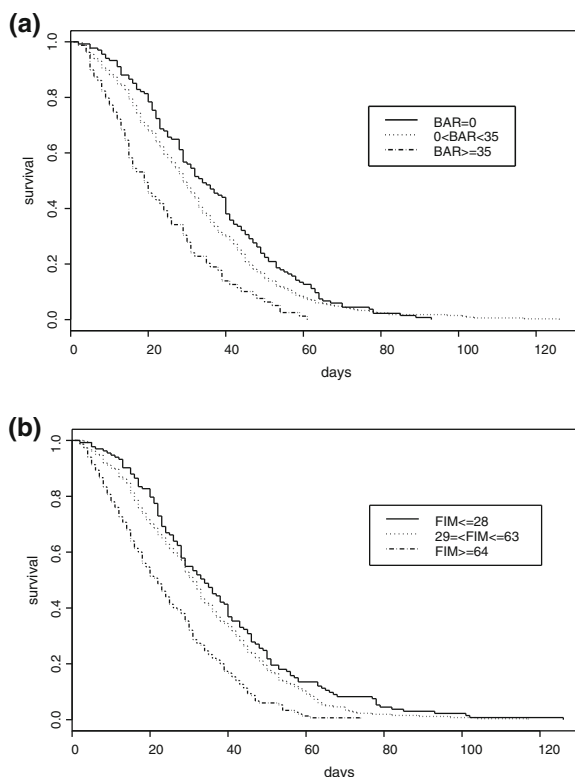
Cerebral vascular disease was among the leading causes of death in Taiwan in recent decades (crude mortalities, 53.5–78.4 cases per 10^5 person-years), and rehabilitating stroke patients often had a long length of hospital stay (LOS). The work of study of the principal factors affecting LOS is essential for the management of health-care costs, of after-discharge home care, and of bed occupancy in hospitals of different levels, etc. Further, LOS is a factor related to short-term prognosis and is also an indicator of long-term survival of patients. These data offer an example for the case of *non-censoring* (see Chap. 2); that is, the time of “discharge” from hospital is treated as an “event time.” The data enrolled 586 patients who experienced their first hemorrhage/infarct strokes and received in-hospital rehabilitations (Lin et al. 2009). The baseline data include age, gender, co-morbidity status, and previous history of stroke and/or severe injury, etc. Modified Barthel index (MBI) and functional independence measure (FIM) questionnaires were administered to patients admitted for rehabilitation. The MBI and FIM are two different scores measuring the severity of disability and functional dependence/independence level of patients. These two scores are highly correlated and both indicative of a patient’s discharge. In this data, 24.6, 60.8, and 14.6 % of the patients had $MBI = 0$, $0 < MBI \leq 30$, and $MBI \geq 35$; and 24.4, 48.0, and 27.6 % had FIM between $[18, 28]$, $[29, 63]$ and $[64, 125]$, respectively.

The KM “survival” estimates for different MBI and FIM groups are displayed in Fig. 1.2. Different Barthel index groups (upper panel, Fig. 1.2a) and different FIM groups (lower panel, Fig. 1.2b) both have the proportional hazards relationship. In Lin et al. (2009), confidence intervals of mean LOS are constructed based on the PH model assumption.

1.3 Example 3: Gastric Carcinoma Data

When analyzing survival data from *clinical trials*, *cross-effects of survival functions* are sometimes observed. A classical example is the well-known data concerning effects of *chemotherapy* (CH) and *chemotherapy plus radiotherapy* (CH+R) on the survival times of *gastric cancer* patients studied by Stablein and Koutrouvelis (1985). The number of patients is 90. Survival times of chemotherapy (group 0 of size 45) and chemotherapy plus radiotherapy (group 1 of size 45) patients are as follows (* denotes right-censored observations). For further details and discussions, see also Kleinbaum and Klein (2005), Klein and Moeschberger (2003), Bagdonavičius et al. (2002), Hsieh (2001), Bagdonavičius et al. (2004), and Zeng and Lin (2007).

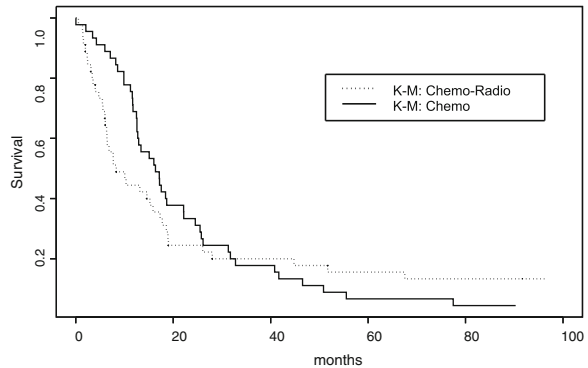
Fig. 1.2 **a** KM estimates for different BAR (or MBI) groups. **b** KM estimates for different FIM groups. Reprinted from Journal of the Formosan Medical Association, 108(8), C.-L. Lin, P.-H. Lin, L.-W. Chou, S.-J. Lan, N.-H. Meng, S.-F. Lo, H.-D.I. Wu, Model-based Prediction of Length of Stay for Rehabilitating Stroke Patients, pp. 653–662, Copyright 2015, with permission from Elsevier



- *Chemotherapy*: 1 63 105 129 182 216 250 262 301 301 342 354 356 358 380 383 383 388 394 408 460 489 499 523 524 535 562 569 675 676 748 778 786 797 955 968 1000 1245 1271 1420 1551 1694 2363 2754* 2950*;
- *Chemotherapy plus Radiotherapy*: 17 42 44 48 60 72 74 95 103 108 122 144 167 170 183 185 193 195 197 208 234 235 254 307 315 401 445 464 484 528 542 567 577 580 795 855 1366 1577 2060 2412* 2486* 2796* 2802* 2934* 2988*.

At the beginning of treatment, the mortality of *CH+R* patients is greater but at a certain moment the survival functions of *CH+R* and *CH* patients intersect, and later the mortality of *CH* patients is greater. That is, if patients survive *CH+R* therapy during a certain period then later this treatment is more beneficial than the *CH* therapy. Doses of *CH* and *R* therapy can be different so regression data can be collected. One will observe (Fig. 1.3) this “cross-effect” phenomena by plotting the *Kaplan–Meier* estimators of the survival function for both treatment groups. The two estimated curves indicate that *radiotherapy* would initially be detrimental to a patient’s survival but become beneficial later on. We shall consider models for analysis of data with cross-effect of survival functions under *constant covariates* in

Fig. 1.3 KM estimates for gastric cancer data



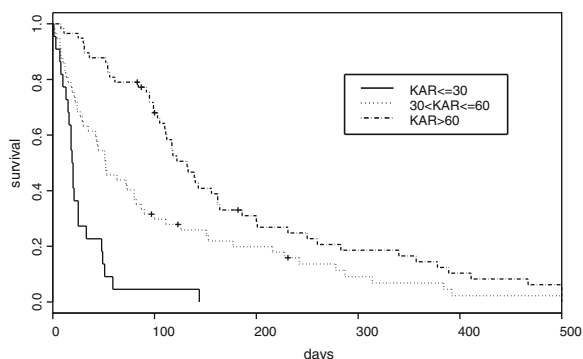
Chaps. 5 and 6. Moreover, we show in Chap. 7 that the conventional log-rank test has low power in this example. The results will be compared between a class of weighted log-rank tests and a score test based on a more flexible model.

1.4 Example 4: The Veteran's Administration Lung Cancer Trials

We studied the survival data of 137 *lung cancer* patients given in Kalbfleisch and Prentice (2002), Bennett (1983), Kleinbaum and Klein (2005), Marubini and Valsecchi (1995), and Therneau and Grambsch (2000), concerning the *Veteran's Administration Lung Cancer Trials*. The dataset includes the following variables: survival time and (right-) censoring status, performance status (Karnofsky rating), cell type of carcinoma (squamous cell, small cell, adeno, and large cell), treatment indicator, months from diagnosis, age, and prior therapy. For ease of illustration, we analyze the *influence of performance status* (Karnofsky rating: 10–30 completely hospitalized, 40–60 partial confinement, 70–90 able to care for self) on the survivals. The *Karnofsky index* is often used to measure the general health status (*degradation*) of a patient (Karnofsky and Burchenal 1949). There are 22 (16.1 %), 57 (41.6 %), and 58 (42.3 %) persons who have the respective Karnofsky ratings (KR): $KR \leq 30$, $30 < KR \leq 60$, and $KR > 60$.

In our example nine observations were censored, i.e., the proportion of censorings is 0.0657. This example illustrates a case when the hazard rates under different values of the covariate do not intersect but the ratios of hazard rates are monotone. That is, the interrelations among these three groups change over time and the proportionalities among the three groups are questionable (Fig. 1.4).

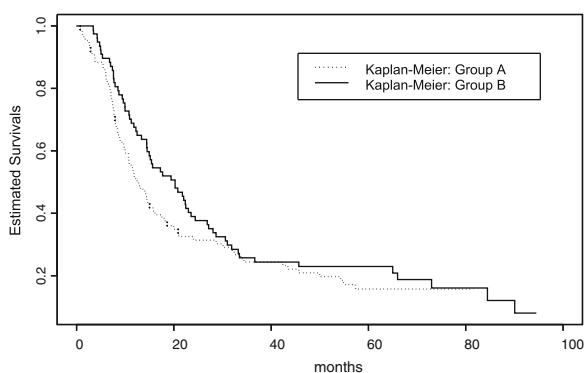
Fig. 1.4 KM estimates for different Karnofsky indices



1.5 Example 5: Other Lung Cancer Data from a Clinical Trial

Piantadosi (1997, Chap. 19, pages 483–488) gives the data concerning the survival times of lung cancer patients. There are 164 patients divided into two groups who received radiotherapy (sample size 86; Group A) or radiotherapy plus “CAP” (sample size 78; Group B). Apart from survival time and censoring status, the variables include: cell type (67 squamous versus 97 non-squamous), performance status (abbreviated as PS: there are 20 “PS = 1” and 144 “PS = 2”), tumor status (abbreviated as TS: there are 19 “TS = 1,” 92 “TS = 2” and 53 “TS = 3”), nodal status (NS: 15 “NS = 0,” 30 “NS = 1” and 119 “NS = 2”), disease-free survival, indicator for recurrence, age, race (24 others and 140 whites), weight loss (WL: 142 “WL = 0” and 16 “WL = 1”; 6 missings), and sex (47 females and 117 males). The variable age has quartiles 52.0 (Q1), 58.0 (Q2) and 64.5 (Q3) with sample mean 57.4. Dichotomizing “age” by its Q2 leads to nonsignificant difference in the lifetime distributions (p -value = 0.536, log-rank test). These data exhibit a nonproportional hazards pattern in treatment, cell type, tumor status, nodal status, weight loss, and

Fig. 1.5 KM estimates for lung cancer data. Reprinted from Wiley Books 2nd edn, S. Piantadosi, *Clinical Trials: A Methodologic Perspective*, p. 125, Copyright 2015, with permission from Wiley



the dichotomized age groups. No significant comparison results can be found in all the above variables using the log-rank test. For illustration, the K–M survival estimates for the two treatments are plotted in Fig. 1.5, in which the K–M estimates cross at 33 months, and the common survival of these two groups is 0.26. We show in Chap. 7 how this data (K–M survivals) can be fitted by two flexible regression models, the Hsieh model and the simple cross-effect model. For the disease-free survivals, proportional hazards assumption seems reasonable and conventional PH analysis applies.

The Cox Model and Its Applications

Nikulin, M.; Wu, H.-D.I.

2016, XIII, 124 p. 25 illus., 18 illus. in color., Softcover

ISBN: 978-3-662-49331-1