

Techniques for survival analysis

Survival analysis in descriptive epidemiology

The need for the estimation of survival rates¹ is twofold: the first objective is to describe the outcome, with time, of a given disease in a group of patients. The mortality associated with the disease can then be assessed in terms of setting public health priorities or providing prognostic information for a patient newly diagnosed with the disease. The second objective is to compare mortality between various groups of patients or to study survival according to individual characteristics such as sex, age, diagnosis or type of treatment in order to identify prognostic factors. Cancer registries are primarily concerned with the first, descriptive objective.

From this point of view, it is important to remember that survival rates routinely calculated from incident cases in a population should be distinguished from data of the same kind established from a series of hospital patients or from patients included in clinical trials.

Although the methods used in the two situations are identical, the groups studied in treatment evaluation are submitted to numerous selection criteria (recognized or unrecognized) and only rarely represent the full epidemiological diversity of the patient population. Survival rates estimated from incident cases (from which it is standard practice to remove cases discovered at death) can thus be noticeably different from hospital results, with the extent of difference depending on the selection process which affects the hospital population under study (see page 268). Conversely, survival rates established in an epidemiological context can only be used to estimate the overall efficiency of the health system, which depends not only on the quality of care but also on the accessibility of the system and the consequent likelihood of early diagnosis. For this reason, these rates do not lend themselves to retrospective evaluation of treatment even if the necessary information is available for some or all of the incident cases.

The analysis of survival data, like all epidemiological analysis, requires standardization of the case definitions. In particular, the diagnostic classification, including

¹ The expression "survival rates" will be used in this chapter as meaning survival probability, as it is current practice among clinicians, despite the fact that rate and probability are different concepts in theory.

site, subsite and histological type, may give rise to heterogeneity in the groups being studied. This issue will not be considered here.

We must also acknowledge the fundamental role of the stage of disease at the time of diagnosis. The apparent benefit of an early diagnosis, as measured by an increase in survival time, may be misleading if it only reflects the addition to the survival time of the *lead time* which separates early detection time from the time at which diagnosis is usually made clinically. Thus, screening and early detection may in reality be prolonging the period of morbidity while having no effect whatsoever on the natural history of the disease [1]. In fact, such arguments are not always easy to challenge in the absence of good information on the distribution of the time spent in the preclinical phase of the disease. However, if such were the situation, the survival curve would have about the same shape after the lead time for patients diagnosed early, and their age at death would not be changed [2]. In any case, the standardization of criteria defining date of incidence remains a fundamental objective.

The problems associated with data collection must not be ignored. In the study of survival, as in the study of disease incidence, it is essential to question the reliability of the data. In order to avoid bias, data collection systems should allow not only for the routine registration of death but also for the active verification of the status of cases for which no information on death has been received.

The primary objective of this chapter is to provide the means to calculate as precisely as possible the survival probability as a function of the time elapsed since the occurrence of the event marking the beginning of follow-up. This function of time defines the survival curve of the group under study. The necessary material is covered from page 216 to 222.

Survival data, like incidence data, are subject to sampling variation, that is, they can provide only an estimate of an unknown, underlying reality. For this reason, the degree of confidence to be attached to the results is also considered (confidence interval).

We will then examine (see page 231) methods that have been proposed to take account of the effect of competing causes of death, in order to provide a better understanding of the phenomenon under study: a cancer patient is not protected from other risks of death and adjustment for these is needed for assessing the specific effect of the cancer on the risk of death. Some authors calculate *cause-specific survival*, which only takes into account death due to the disease being studied. Among alternative methods that have been proposed, the most widely used is that of *relative survival*, based on the use of the *life table*, the principal concepts of which have been presented in Chapter 1. The reader will find the methods to overcome the possible lack of published life tables on page 236.

Methods to compare survival in two or more groups are dealt with in the second part of this chapter. These problems are very similar to those which were considered under the comparison of age incidence curves. Just as one might prefer the comparison of incidence curves as a whole to the comparison of cumulative risks alone, comparison involving whole survival curves is preferable to a comparison which only

concerns survival probabilities at a given time point. However, one is often restricted to this latter comparison in the absence of appropriate data, resulting in the need for methods described briefly on page 246; unfortunately, routinely published data do not always provide all the details needed for making this point comparison.

When comparison is carried out over entire survival curves, the assumption is often made that the instantaneous mortality rates are proportional. In this situation, the optimal test, known as the log rank test, is effectively identical to the Mantel-Haenszel test discussed in Chapter 2. We will re-examine it in the context of survival distribution (see page 247).

Populations that are being compared sometimes experience different mortality patterns after the time of diagnosis: some are subject to a very high initial mortality followed by a long remission; others, in contrast, experience a more regular occurrence of deaths. In the former case, the Gehan-Breslow test, also presented in this section, is in principle more appropriate.

In certain situations, if the necessary information is available, survival comparisons can be made taking confounding factors into account. A further section discusses methods of adjustment or stratification which allow us to take account of confounding factors (see page 255).

When individual characteristics which might affect survival ('prognostic factors') are available for each case under follow-up, it is usual to assess the specific role played by each factor in the prediction of survival time. The stratification approach is however rapidly limited by sample size while it is still possible in some situations to use a modelling approach, despite the small number of subjects available. Cox's multiplicative model provides the necessary tool in this context and is discussed from page .

Calculation of long-term survival raises particular problems: because of ageing, the group is subject to an increasing risk of mortality from causes other than the disease under study. In this situation it becomes important to base comparisons on net survival, requiring the application of specific methods (see page 272).

A number of data sets obtained from cancer registries will be used to illustrate methods that have been mentioned: the first set of data refers to survival of incident cases of colon cancer in the French department of Côte-d'Or. This example illustrates the mathematical calculations required to establish survival probabilities by the actuarial method and their confidence limits. Survival of incident cases of skin melanoma in Geneva (Switzerland) will provide an example of the use of the Kaplan-Meier method. Data on colon cancer in Geneva will serve to illustrate the computation of relative survival rates and similar data from the canton of Vaud (Switzerland) will be considered with those from Côte-d'Or and Geneva to show how to carry out a comparative analysis.

Finally, we will use data from the Geneva cancer registry on various cancer sites to compare the results obtained by the relative survival method and those obtained by the cause-specific survival method where deaths not attributable to the disease are taken as censored observations.

Stratification methods will be illustrated by the comparison of survival distribution for men and women with melanoma, taking subsite as the confounding factor.

The more general modelling approach will be illustrated by the analysis of breast cancer survival in Geneva according to the type of health care system providing the treatment.

Estimation of survival distribution

When concepts associated with the study and interpretation of individual follow-up were discussed in Chapter 1, we presented the notion of *censored observations*, that is, observations which are incomplete and which require a specific methodology. In this section, we will illustrate different techniques proposed for the analysis of survival data. Although some of these techniques have been developed for clinical trials, they can be used in the present context with minor adjustments. Nevertheless, the choice of an analytical method and the interpretation of the results require a specific approach, because of the way the data are obtained, the size of the groups under study and the absence of randomization.

When there are no censored observations, survival probability after a given time is estimated simply by the ratio between the number of survivors at a specific date and the number in the group at the beginning of the study. This probability, called by some authors the direct survival rate, obviously cannot be calculated for individuals in the group for whom the period of follow-up is less than the time interval being considered. The group under study is therefore subdivided into subgroups in which the subjects have the same potential follow-up time and the survival probability calculated in each subgroup is assumed to be an estimate of the survival rate for the corresponding length of time. However, as the resulting probabilities have been obtained from different groups of patients, they will not in general provide a consistent survival curve, that is, the survival probability will not necessarily decrease with time.

In fact, observations with incomplete follow-up can still contribute to the estimation of survival for time intervals greater than the duration of their follow-up. This idea is being used in the following two methods which are also described in a more formal context in Chapter 1.

Estimation of crude survival

Actuarial method

The aim of this method is to study the survival of a group of subjects for whom a common event has occurred; for each subject the date of occurrence marks the beginning of follow-up. In the present context the common event which characterizes the group will be the diagnosis of cancer.

Follow-up of individuals comes to an end either at death, or on the date at which the individual is lost to follow-up, or on the date at which the individual is withdrawn from follow-up, for example for the purpose of analysis of the survival data. In fact, the date of withdrawal may be specific to each individual if assessment of status (living, deceased or lost to follow-up) is only carried out at regular intervals following the date of diagnosis; there may also be a single date of withdrawal if this assessment is carried out on the same date for all individuals (see page 227). The time between the date of diagnosis and the end of follow-up is called the *follow-up time*. Figure 4.1 illustrates the two follow-up procedures discussed above. Note that the recruitment period is generally less than the study period, for example, when we only want to consider cases which have a minimum follow-up time. On the graph, e_1 , e_2 and e_3 represent the entry dates of new cases, f_1 , f_2 and f_3 the end of individual observation periods, and the solid circles represent dates of assessment of status.

Theoretically, the method is applicable when new cases are recruited who have been diagnosed before their inclusion in the study group. These cases are only included in the calculation from the time when they are actually under observation and not from the date of their diagnosis. If such a precaution is not taken, mortality

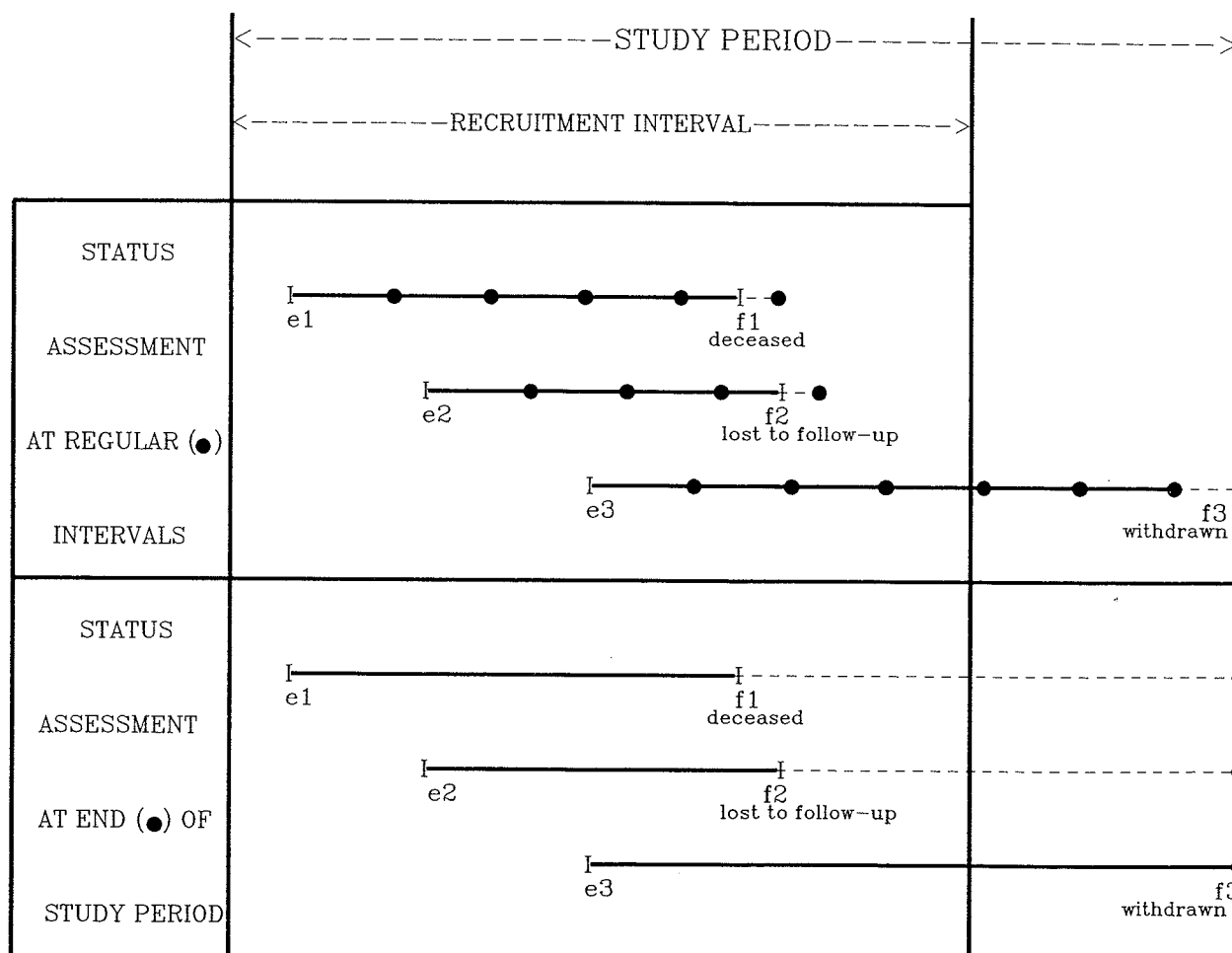


Figure 4.1 Principles of follow-up of cases in a survival study

during the period prior to their inclusion would be underestimated. Apart from the increased complexity in the calculations, it is generally preferable not to take these cases into consideration because of the selection biases to which they may be subject.

The first step of the actuarial method is the subdivision of the maximum observed survival time into intervals; the length of the intervals is set *a priori* taking into account the distribution of deaths over time, so that each interval, on average, has at least some deaths. If necessary, the intervals can be of unequal size.

The second step is the computation of the conditional probabilities in each interval which are defined by the proportion of those living at the start of the interval who were still living at the end of it. However, in order to account for subjects with shorter follow-up time, the number at risk at the beginning of the interval is reduced by half the number of subjects who were withdrawn or lost to follow-up during the interval. The total thus obtained is known as the corrected number of subjects at risk or the *effective number at risk*.

The probability of surviving to the end of a given interval is obtained by multiplying together the conditional probabilities over all the intervals preceding this time point. Survival is obtained by linear interpolation for all other time.

Table 4.1 Calculation of survival probabilities by the actuarial method (incidence of colon cancer in males in Côte-d'Or, France, between 1976 and 1982)

Interval (months)	Number at risk in t_i	Number censored	Death	Effective number at risk	Conditional probability of death	Conditional probability of survival	Survival rate
t_i, t_{i+1}	n_i	r_i	d_i	N_i	q_i	s_i	S_{i+1}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-6	411	1	145	410.5	0.353	0.647	0.647
6-12	265	1	40	264.5	0.151	0.849	0.549
12-18	224	0	32	224.0	0.143	0.857	0.471
18-24	192	2	16	191.0	0.084	0.916	0.431
24-30	174	15	15	166.5	0.090	0.910	0.392
30-36	144	15	4	136.5	0.029	0.971	0.381
36-42	125	15	13	117.5	0.111	0.889	0.339
42-48	97	10	4	92.0	0.044	0.956	0.324
48-54	83	11	0	77.5	0.000	1.000	0.324
54-60	72	13	1	65.5	0.015	0.985	0.319

Columns (1) to (4) give observed data; data in columns (5) to (8) have been calculated.

Let:

- t_i, t_{i+1} be the interval end-points ($0 \leq i \leq I - 1$),
- n_i be the number of subjects surviving and followed up at date t_i (number at risk),
- r_i be the number of subjects withdrawn or lost to follow-up in the interval $[t_i, t_{i+1}]$,
- d_i be the number of deaths in the interval $[t_i, t_{i+1}]$,
- $N_i = n_i - (r_i / 2)$ be the corrected number at risk,

- $q_i = d_i / N_i$ be the estimates of the conditional probability of dying in the interval $[t_i, t_{i+1}]$,
- $s_i = 1 - q_i$ be the conditional probability of being alive at t_{i+1} given that the subject was at risk at t_i , and
- $S_{i+1} = \prod_{j=0}^i s_j$, the survival probability at t_{i+1} .

The method has the advantage of using all available information, as opposed to the direct survival method, which is based only on cases which have a follow-up time at least equal to the time for which survival probability is being calculated. On the other hand, the method relies on the assumption, sometimes debatable, that those for whom observations are censored are subject to the same force of mortality as cases for whom follow-up is complete.

Table 4.1 shows the details of this calculation using data from the Côte-d'Or on colon cancer in males. Figure 4.2 shows the corresponding survival curve. An interval of six months has been used. Cases lost to follow-up and cases withdrawn from follow-up are treated in the same way and considered to be censored observations.

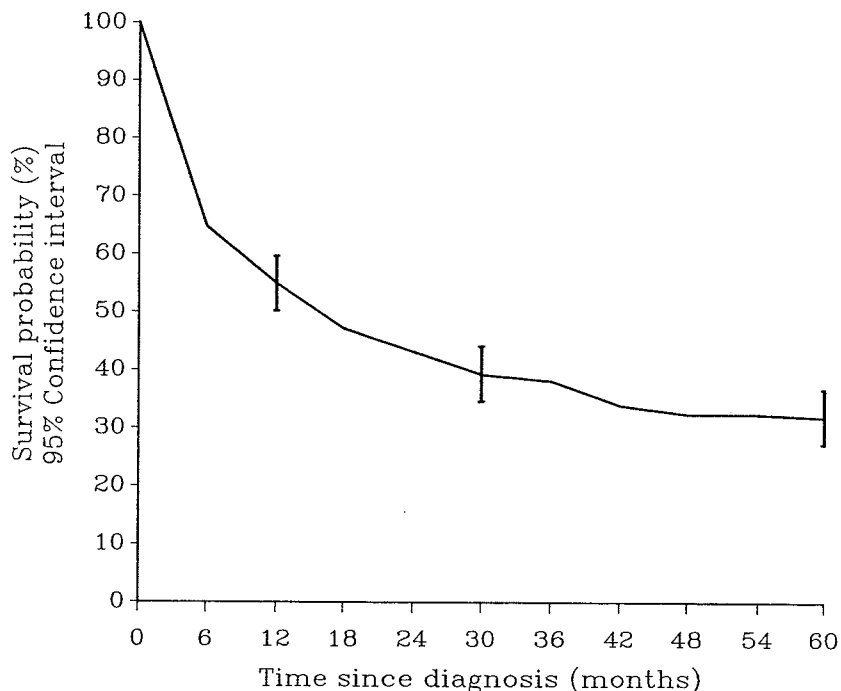


Figure 4.2 Survival of male colon cancer patients in Côte-d'Or, diagnosed between 1976 and 1982

Kaplan-Meier Method

This procedure relies on the same principles as the actuarial method. However, the time intervals are not set *a priori* but are determined by the occurrence of death:

the conditional probabilities of surviving between two dates of death are estimated every time a death occurs. As in the actuarial method, the probability of survival, from the start of follow-up, is obtained by the product of probabilities calculated for each successive interval.

Intervals are usually defined in days. The method requires the calculation of as many survival probabilities as there are deaths, except if several deaths occur on the same day. It is particularly suited to analysing data from small groups. An observation which is censored between two deaths does not affect the cumulative probability of survival, which remains constant in this interval; in principle, it is not necessary to know the exact dates when such observations are censored. However, it is necessary to know the number of censored observations between two deaths which is subtracted from the number at risk at the start of the next interval.

Let:

- t_i be the i th observed time of deaths $1 \leq i \leq I$,
- d_i be the number of observed deaths at t_i ,
- r_i be the number of censored observations in interval $[t_i, t_{i+1}[$,
- $n_i = n_{i-1} - d_{i-1} - r_{i-1}$ be the number at risk just before time of deaths occurring in t_i : to obtain n_i , subtract from n_{i-1} the deaths which took place at t_{i-1} and all other cases for which follow-up ended in the interval $[t_{i-1}, t_i[$
- $s_i = 1 - (d_i/n_i)$ be the conditional probability of being alive in $]t_i, t_{i+1}]$ given the subject was at risk at t_i , and
- $S_i = \prod_{j=1}^i s_j$ be the probability of surviving after t_i . This probability is constant up to date t_{i+1} inclusive.

The details of the computation are shown in Table 4.2 which refers to data from Geneva for skin melanoma in males. The corresponding survival curve is graphed in Figure 4.3.

We can see that the successive conditional probabilities are calculated here from the true number at risk at the moment when death takes place and not from a number obtained by subtracting half those censored in the interval, as is done in the actuarial method. Furthermore, the latter method assumes a constant force of mortality by interval, while the Kaplan-Meier method makes no assumption about the underlying instantaneous rate.

In practice, both methods give very similar results when they are applied to large groups, which is one of the reasons why most registries traditionally use the actuarial method, the application of which in medicine goes back to J. Graunt [4]. The fixed intervals of the actuarial method are perfectly suited to the classic presentation of survival probability for a given number of years after diagnosis, for example, 1, 3 or 5 year survival. However, the existence of modern computer software makes the application of the Kaplan-Meier method much easier than in the past, and it may be better to consider using it more widely, even for large groups,

Table 4.2 Calculation of the probabilities of survival by the Kaplan-Meier method (incidence of skin melanoma for males, Geneva, Switzerland, 1970)

Day of death	Day (number) of censored observations	Number of deaths	Number of censored observations	Number at risk	Conditional survival at t_i	Survival probability between t_i and t_{i+1}
t_i (1)	(2)	d_i (3)	r_{i-1} (4)	n_i (5)	s_i (6)	S_i (7)
373		1		13	0.923	0.923
643		1		12	0.917	0.846
672		1		11	0.909	0.769
975		1		10	0.900	0.692
1173		1		9	0.889	0.615
	1542(1)					
1645		1	1	7	0.857	0.527
	1891(1)					
	1920(1)					
	2055(1)					
	2148(1)					
	2167(1)					
2481		1	5	1	0.000	0.000

Columns 1 to 3 as well as the first line of column 5 (actual size of the group) give observed data; other values are calculated.

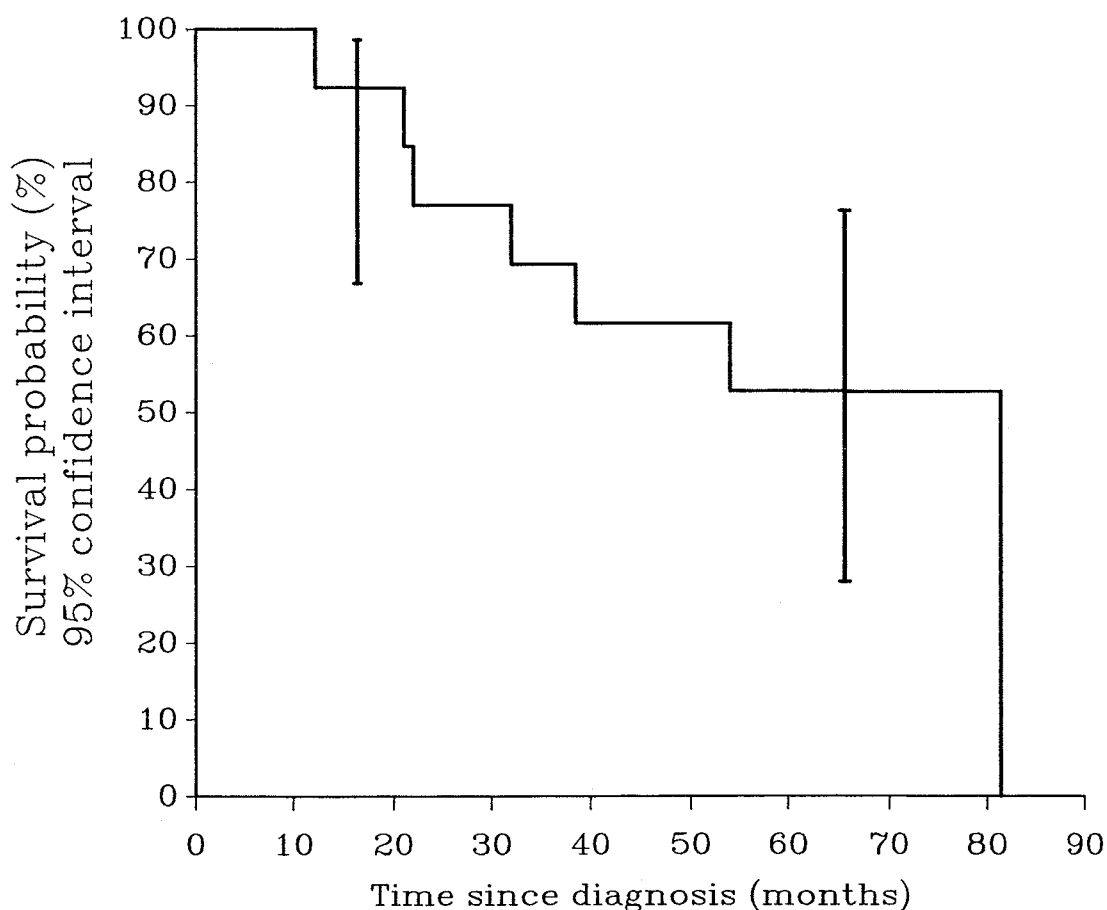


Figure 4.3 Survival of male skin melanoma patients diagnosed in 1970 in Geneva, Switzerland

both for its greater precision and because it is more appealing than the actuarial method in the context of modern tools of survival data analysis.

Confidence interval for a survival rate

The estimation of variability of a survival rate becomes essential when its calculation is carried out on small groups of patients. The computation of confidence limits is usually undertaken for estimating the cumulative survival rates at a given time point. Determining several successive intervals gives an approximate idea of the 'confidence band' within which the real survival curve is taken to be, although, strictly speaking, it does not define the 95% confidence interval of the curve. It is sometimes useful to estimate the confidence intervals of the conditional probabilities of survival, cumulative not from the date of origin but from a given time. For example, one might focus on subjects who have survived for at least two years after diagnosis. The calculations are based on the same methodology.

The estimation of a confidence interval for a cumulative rate S_i is based on the estimation of its variance, which depends in turn on whether the group is *open* or *closed*².

In a closed group, survival at time t_i is the proportion of those surviving: $S_i = n_i/n_0$, where n_0 is the initial number and n_i is the number of survivors at time t_i . In this situation, the variance of S_i is simply:

$$V_i = \frac{S_i(1 - S_i)}{n_0} \quad (4.1)$$

that is, the variance of the estimated probability from n_0 trials among which there are n_i successes (Binomial law).

In the usual situation where the group is open, the number of subjects at risk is decreased over time by censoring; the variance is thus larger than if the entire group had been followed up. Nevertheless, it is smaller than it would be in a closed group which would also have counted n_i survivors and which would have produced the same estimate of survival. In this situation, the initial size of the group would have been

$$n'_0 = \frac{n_i}{S_i} \quad (4.2)$$

Peto [5] suggests using this theoretical number to determine an upper bound for the variance. Thus, the corresponding standard error is

$$E_i = \sqrt{\frac{S_i(1 - S_i)}{n'_0}} = S_i \sqrt{\frac{1 - S_i}{n_i}} \quad (4.3)$$

² See the definition of these terms in Chapter I, page 22

Greenwood [6] based an estimate of the variance of S_i on the estimated variance of the conditional probabilities s_j , which are directly derived as variances of a proportion, since at each time t_j , s_j is the proportion of survivors in the group of subjects at risk of death just before t_j .

Thus:

$$\text{Var}(s_j) = \frac{s_j(1-s_j)}{n_j} \quad (4.4)$$

Furthermore, since:

$$\text{Log}(S_i) = \sum_{j=1}^i \text{Log}(s_j) \quad (4.5)$$

the standard method [7] of calculating the variance of $\text{Log}(s_j)$ based on the variance of s_j (4.4) can be used to show that

$$\text{Var}[\text{Log}(S_i)] = \sum_{j=1}^i \frac{\text{Var}(s_j)}{s_j^2} = \sum_{j=1}^i \frac{1-s_j}{n_j s_j} \quad (4.6)$$

and, consequently, that

$$V_i = \text{Var}(S_i) = S_i^2 \sum_{j=1}^i \frac{1-s_j}{n_j s_j} \quad (4.7)$$

which is the formula that Greenwood established for the actuarial method by replacing n_j , the number at risk at the beginning of the interval, by the effective number at risk N_j . Insofar as this result depends on the estimate of the variance (4.6) being valid for large samples only, this formula can lead to an under-estimation of the variance for long time intervals when the group size is not sufficiently large [8]. This estimate and that proposed by Peto are shown in Table 4.3, in which calculations of variance are made from the data in Table 4.1.

Column 5 of Table 4.3 is obtained by induction:

$$\begin{aligned} V_0 &= 0 \\ V_{i+1} &= V_i s_i^2 + S_{i+1}^2 \frac{(1-s_i)}{n_i s_i} \end{aligned} \quad (4.8)$$

The number calculated in column 6 and the standard error in column 7 are obtained from (4.2) and (4.3).

In general, the confidence interval of size α for a survival rate θ can be calculated by

$$\frac{|S_i - \theta|}{\text{SE}(S_i)} < Z_{\alpha/2} \quad (4.9)$$

Table 4.3 Calculation of standard errors by Greenwood's and Peto's methods

i (1)	n _i (2)	s _i (3)	S _{i+1} (4)	Greenwood √V _{i+1} (5)	Peto n' ₀ (6)	Peto E _{i+1} (7)
0	411	0.647	0.647	0.0236	409.7	0.0236
1	265	0.849	0.549	0.0246	408.0	0.0246
2	224	0.857	0.471	0.0247	408.0	0.0247
3	192	0.916	0.431	0.0245	403.6	0.0246
4	174	0.910	0.392	0.0242	367.1	0.0255
5	144	0.971	0.381	0.0242	328.3	0.0268
6	125	0.899	0.339	0.0242	286.4	0.0280
7	97	0.956	0.324	0.0242	256.2	0.0292
8	83	1.000	0.324	0.0242	222.3	0.0314
9	72	0.985	0.319	0.0244	181.8	0.0346

The numbers in columns 1 to 4 are from Table 4.1 and those in the remaining columns have been derived from formulae 4.8, 4.2 and 4.3.

where θ is the probability of survival to be estimated at t_i and $SE(S_i)$ is the standard error of its estimator S_i . If $SE(S_i)$ is replaced by one or another of the standard errors in Table 4.3 (columns 5 and 7), a symmetric interval is obtained as:

$$[S'_i, S''_i] = S_i \pm Z_{\alpha/2} SE(S_i) \tag{4.10}$$

since when time t is fixed, the estimate of the survival rate at time t is approximately normally distributed.

As Rothman suggests [9], it is also possible to calculate an interval whose

limits always lie between 0 and 1, by substituting the quantity $\sqrt{\frac{\theta(1-\theta)}{n''_0}}$ for $SE(S_i)$ in equation (4.9) with

$$n''_0 = \frac{S_i(1-S_i)}{V_i} \tag{4.11}$$

where V_i is either Greenwood's variance or the maximum variance postulated by Peto. The confidence interval is then obtained by solving equation (4.9) for θ :

$$[S'_i, S''_i] = \frac{n''_0}{n''_0 + Z_{\alpha/2}^2} \left[S_i + \frac{Z_{\alpha/2}^2}{2n''_0} \pm Z_{\alpha/2} \sqrt{\frac{S_i(1-S_i)}{n''_0} + \frac{Z_{\alpha/2}^2}{4n''_0{}^2}} \right] \tag{4.12}$$

The choice between the many different ways of calculating the confidence interval depends on practical considerations and on how conservative an estimate is required. For routine calculations, most scientists in cancer registries use the wider symmetric confidence interval (4.10) with the Greenwood standard error. We prefer

to use (4.12) with the theoretical number n''_0 derived from Greenwood's variance (4.11). In fact it has been shown that this method on average provides the most satisfactory result [10]. Note however that the symmetric confidence interval derived by Peto is an easily-obtained estimate of the magnitude of the variability of the survival probability estimate.

Table 4.4 shows survival probability at 12, 30 and 60 months for cancer of the colon in men in Côte-d'Or between January 1976 and December 1982, and confidence intervals obtained using the three methods discussed. The differences are only noteworthy after 60 months, when Peto's interval is substantially wider.

Table 4.4 Confidence intervals for survival; data from Table 4.1.

Time since diagnosis (months) t_i	Survival probability S_i	Greenwood Rothman	Greenwood symmetric	Peto symmetric
12	0.549	0.501 ; 0.596	0.501 ; 0.597	0.501 ; 0.597
18	0.471	0.423 ; 0.519	0.422 ; 0.519	0.422 ; 0.519
24	0.431	0.384 ; 0.480	0.383 ; 0.479	0.383 ; 0.479
30	0.392	0.346 ; 0.441	0.345 ; 0.440	0.342 ; 0.442
60	0.319	0.273 ; 0.368	0.271 ; 0.367	0.251 ; 0.387

Median survival time

With the preceding methods, the distribution of survival times can be correctly estimated in the presence of censoring without making assumptions about the analytical shape of this distribution. However, by not adopting parametric models, we cannot use the method of maximum likelihood, which is an effective tool for estimating parameters such as life expectancy and its confidence interval. In all the methods previously described, it is accepted practice to estimate quantiles of the distribution of survival. In particular, the median survival time is the time at which the survival rate is equal to 50%. This value can be estimated from the curve calculated using one of the methods previously described. The median survival time is a readily calculated location parameter which provides an easily interpretable summary of the data.

In the actuarial method, a unique value of this median is usually obtained because the survival curve is continuous and, more often than not, strictly decreasing. If S_i and S_{i+1} are respectively the survival rates at the end points of the intervals which include the value 0.5, then the median of the observed distribution is obtained by linear interpolation:

$$m = t_i + \frac{S_i - 0,5}{S_i - S_{i+1}} (t_{i+1} - t_i) \quad (4.13)$$

for data relating to cancer of the colon in Côte-d'Or (Table 4.1), the observed median is

$$m = 12 + \frac{0.549 - 0.500}{0.549 - 0.471} (18 - 12) = 15.77 \text{ months} \tag{4.14}$$

If the survival probability is estimated by the Kaplan-Meier method, there is in general no time for which the observed survival rate is exactly equal to 0.5. The observed median can be taken as the date of the first death corresponding to a survival rate less than or equal to 0.50. For the data on melanoma (Table 4.2), the survival rate is not less than 0.5 before the last observed event, which is a death at 2481 days. Strictly speaking, the median survival is thus 2481 days, or six years and nine months. In fact, in this example, it is clear that the estimate is imprecise because of the small number of observed cases. As a general rule, the variability of this parameter can be substantial. Therefore, it is important that it be accompanied by its confidence interval.

Figure 4.4, based on hypothetical data, shows that a confidence interval for the median can be obtained by simply inverting the functions representing the lower and upper confidence limits of the survival probability. The confidence region R_α for the median is thus constructed from the confidence interval of the survival probability defined as in the previous paragraph. A time point t is included in the confidence region of the median if the confidence interval for S_t includes 0.50:

$$R_\alpha = \left[t \mid \left(S_t - \frac{1}{2} \right)^2 < Z_{\alpha/2}^2 V_t \right] \tag{4.15}$$

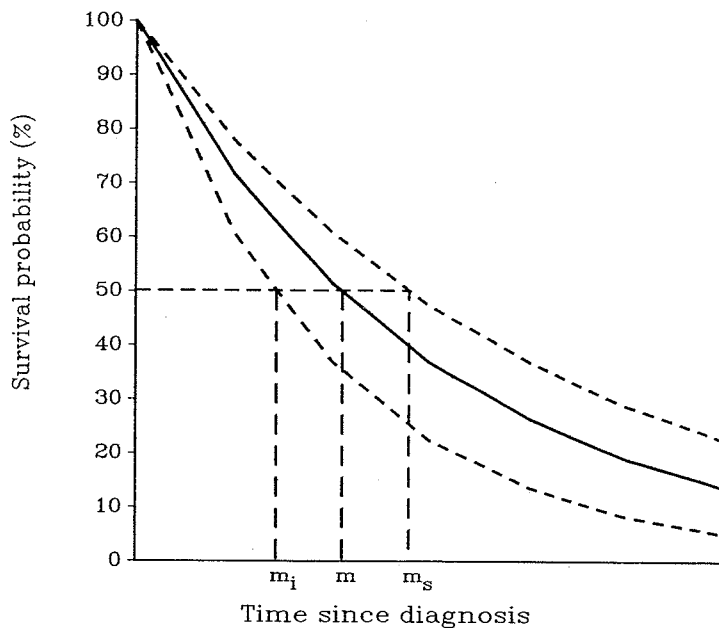


Figure 4.4 Confidence interval of median survival

When the actuarial method is used, the upper and lower confidence limits of the median are determined by interpolation. For example, survival following cancer of the colon in Côte-d'Or is greater than 0.501 at 12 months and greater than 0.423 at 18 months (lower bound of the confidence interval) (Table 4.4); the survival time for which the lower bound of the confidence interval is exactly equal to 0.5 is thus found between 12 and 18 months. This time is obtained by linear interpolation:

$$12 + \frac{0.501 - 0.500}{0.501 - 0.423} \times 6 = 12.08 \text{ months} \quad (4.16)$$

Similarly, survival is less than 0.519 at 18 months and less than 0.480 at 24 months (upper bound of the confidence interval) (Table 4.4). Survival will thus be less than 0.50 at 20.85 months. Therefore the confidence interval of the median is [12.08 ; 20.85].

When the Kaplan-Meier method is used, the upper and lower confidence limits of the median are obtained by determining respectively the first and last date of death for which the confidence interval of survival includes the value 0.50 [11,12]. When this method is applied to the data in Table 4.2, it is seen that the confidence interval of S_t includes 0.50 from $t = 975$ days and that all the subsequent time points are those which correspond to survival probabilities for which the confidence interval includes 0.5. As a result, the confidence interval of the median extends from two years and eight months to infinity, illustrating the imprecision obtained with a small number of cases.

Collecting data for survival analysis

In the usual situation of an open group, we have just seen that the calculation of survival rates requires specific variables for each member of the group:

- date at which follow-up started (date of incidence)
- date when follow-up ended
- status of the subject at this date (dead, living, lost to follow-up).

we have also seen that the date at which follow-up ended for each individual is either the date of death, the last date at which the subject was known to be alive for those who were lost to follow-up, or the date at which follow-up is ended for all subjects as a result of the study being concluded.

This information can be obtained either on an ongoing basis or retrospectively. It comes in general from a variety of sources. These sources are rarely perfect and can give rise to selection biases of different kinds, especially when they involve routine forms of data reporting. For example, failure to notify certain deaths results in the overestimation of survival. The same effect occurs indirectly when the number of individuals lost to follow-up is underestimated, since these cases are then wrongly counted in the numerator and denominator of the survival probabilities. More subtle biases can arise when the amount of information obtained on an individual depends on his or her status. For example, the status of a patient may be better known

simply because of a regular presentation to an oncology department or, conversely, because a death has been recorded. Consequently, no matter how good routine reporting sources may be, it is advisable to carry out a regular update to check the status of all subjects not known to have died, probably using reporting sources independent of those already employed routinely. In other words, for these cases, *passive* follow-up should be complemented by *active* follow-up which may involve searches of official records, direct contact with the patient or the patient's family or other sources.

From an administrative point of view, it is sensible to carry out these updates individually at regular intervals from the date at which follow-up started, for example at every fifth anniversary of the date of incidence. This procedure ensures that the tasks required for active follow-up are ongoing throughout the year rather than all being carried out at once when it is time to analyse the data. Whatever procedure is used, the duration of follow-up should be ended at the date of the last update which ends the follow-up period for the subject.

When the survival probability is calculated for all incident cases, follow-up of subjects who are officially recorded as having left the geographic area covered by registration should be censored at the date of departure, and should be treated as lost to follow-up. Following subjects outside this area can introduce a bias if it is routinely easier to obtain information on death or, conversely, on the surviving patients. A death which occurs outside the registration area obviously should not be counted even if it is known to have occurred. Updates should be organised in such a way that cases lost to follow-up in the literal sense of the term, that is, those cases for which it is not known whether they have left the area or not, are the exception.

Calculating survival probability solely from cases residing in an area for the whole follow-up period has the apparent advantage of characterising the effectiveness of regional medical care. However, when departure from the area is linked to stage of disease (and to the survival probability which follows from it), this way of proceeding might plausibly introduce a bias in the results. For example, suppose that foreign workers with cancer routinely return home, as soon as their condition worsens, to die in their own country. There is unfortunately no ideal solution to resolve this difficulty; only the appraisal of each individual situation ensures that resulting biases are appropriately evaluated.

The mode of follow-up can have a considerable effect on the reliability of results obtained in survival analysis. It is instructive to compare the follow-up procedures used in Côte-d'Or, Geneva and the Swiss canton of Vaud (from where the data used as examples in this chapter have been drawn). Following accepted practice, incident cases in the three regions which are known only from death certificates are excluded from calculation.

In Côte-d'Or, follow-up of patients is based on routine registration of deaths and on an annual update at a fixed time to check the status of all cases. Routine registration identifies approximately 75% of deaths occurring in patients with cancer of the colon. This percentage varies with age, treatment and survival time. The

status of patients for whom there is no report of death is obtained on a fixed date, partially from government registry offices at the place of birth (except for foreign-born subjects) and partially from clinicians (gastroenterologists, surgeons and general medical practitioners). Altogether, in 98% of cases it is thus known whether the patient is living or dead.

In Geneva, follow-up is carried out from two sources of information, one being the updating of the database from routine notification of death, the second resulting from a regular check-up at five-yearly intervals from the date at which follow-up started, for subjects believed to be alive. There are multiple sources of information on deaths: the cancer registry has the right to obtain from the health authority a copy of all death certificates which mention cancer as a cause of death. The registry can be incidentally informed of death when collecting missing data from the hospital records. Finally, information on death is obtained annually by merging records of incident cases with death records at a national level. This link-up of data is based on the date and place of birth. The five-year update is done by manually consulting population records, allowing the registry to note dates of any definite departures and to stop follow-up accordingly. Deaths which might otherwise have escaped detection can also be detected by this manual consultation. It is important to emphasise that cases are only included in the calculation of survival for the period covered by the five-year update. Events which follow this update are not taken into consideration even if they are already registered. In particular, deaths occurring after the time when the update takes place must not be included in the analysis.

In the canton of Vaud, follow-up is based principally on the regular linkage between mortality files held by the Federal Swiss Office of Statistics, and the registry of incident cases. Linkage is achieved by a series of computer processes accompanied by manual verification. The routine update also involves an investigation of cases presumed living, through the municipal population registries in order to determine status and dates of any departures from the canton. Active follow-up is organized and status is determined at a single fixed date for all cases (31 December).

Estimation of net survival

The study of survival in a population subgroup is often motivated by the supposition that the subgroup is subject to a risk of death different from that experienced by other individuals in the population. For example, people with cancer generally experience a much higher mortality than the general population in the years immediately following diagnosis. Insofar as the increase in risk due to the disease being studied tends to diminish, mortality progressively returns to 'normal'. Overall survival probability should therefore be considered as the result of two components, corresponding respectively to the disease being studied and to all other causes taken together.

Thus a *net survival* can be defined as that which might occur if the risks of death other than the cancer were removed. Its complement is none other than the *net probability* of death (see Chapter 1, page 34). Analysing the net survival probability is thus equivalent to the analysis of the excess mortality in the group under study.

To the extent that it is accurately estimated, net survival probability tends to become constant as deaths due to cancer decrease. It then represents the percentage of subjects who can be considered to be cured of the cancer under study. The time period corresponding to the point when net survival probability becomes constant can thus be interpreted as a recovery period. As for all estimates, this value is only valid as an average for the group; it does not exclude the possibility of a fatal recurrence for certain individuals outside this time period.

The justification for calculating the additional risk due to the cancer under study is particularly clear when considering an older age group, because the force of mortality from other causes increases with age. As a group ages over time, net survival probability tends to be decreasingly well represented by observed survival probability, which is increasingly determined by death from other causes. The estimation of net survival probability also responds to the need to make comparisons between subpopulations from the same region or between populations from different regions where the mortality due to other causes may differ such that the comparison of overall survival probabilities might lead to biased conclusions.

The determination of net survival probabilities implicitly assumes that the risk of death from the cancer being studied and the risk of death from other causes are independent, that is, *not interactive* (see Chapter 1, page 34 for the definition of this term). In fact, with cancer, the two risks seem to be positively correlated: the presence of cancer and its treatment cause an increase in the force of mortality from other causes. The inverse situation cannot however be excluded. When survival time increases, the survivors of the group might benefit from a reduction in the risk of death from other causes, because of better medical care or a decrease of risk-related behaviour. It will be seen later that such interactions cannot be totally ignored when deciding which method to use in estimating net survival or when interpreting the results.

There are two classical methods available to estimate the probability of surviving a given disease: the method of *cause-specific survival* and that of *relative survival*.

Cause-specific survival

The principle of this method is simple: the cause of each death is assessed and only those deaths that can be attributed to the disease under study are counted. Other deaths are considered as simply termination of follow-up (in the same way as cases lost to follow-up and observations which are censored at the end of the study). Calculation of cause-specific survival can also be carried out by the actuarial method or by the Kaplan-Meier method. Under both methods, survival rates are

obtained disregarding deaths from other causes. However, if these methods are used, note that it is assumed that the process leading to death from cancer is independent of the process leading to death from other causes, an assumption which is often only approximately true (for example, when a suicide occurs following the diagnosis of advanced cancer).

The method of cause-specific survival is sometimes used in a clinical setting if follow-up is of sufficient quality to distinguish between deaths due to the disease under study and other deaths. The method can only be applied to epidemiological data when the cause of death is routinely recorded, a situation which does not generally hold in population registries. Whatever the situation, the method of cause-specific survival comes up against the difficulties of determining cause of death such as the doubtful reliability of information supplied by the certifying doctor and the arbitrary choice of the primary cause of death when there are multiple or associated pathologies. Moreover, it would seem practically impossible, as much as it would be appropriate, to take into consideration deaths caused by secondary effects of treatment.

Relative survival

The method of relative survival [13] does not require knowledge of the cause of death and thus avoids the difficulties associated with its determination. The method involves calculating, at each time period t , the *relative survival probabilities* defined by the relationship

$$S_c(t) = \frac{S_o(t)}{S_e(t)} \quad (4.17)$$

where $S_o(t)$ and $S_e(t)$ are respectively observed survival and expected survival at time t .

Expected survival corresponds to the mortality of the general population, taking into account the initial distribution in the group of prognostic factors which one wishes to control for. If only age (the effects of which should always be accounted for) is considered, the expected survival $S_e(t)$ is provided by the proportion of survivors that would be predicted at time t in a group having the same initial age structure as the group under study, but subject only to the force of mortality of the general population. The adjustment is thus a standardisation of the indirect type. The calculation of the expected number of survivors is firstly done for each subgroup defined by age at diagnosis in single years or by larger age groups according to the available life tables. Expected deaths are then summed. If n_x is the number of subjects of age x at the beginning of follow-up and $S_{ex}(t)$ is the probability of survival at time t for a subject with initial age x , then the number of survivors at time t for this age group is

$$e_x(t) = n_x S_{ex}(t) \quad (4.18)$$

The total number of survivors at this time is thus:

$$\sum_x e_x(t) = \sum_x n_x S_{ex}(t) \quad (4.19)$$

Consequently, the overall expected survival is

$$S_e(t) = \frac{\sum_x n_x S_{ex}(t)}{\sum_x n_x} \quad (4.20)$$

that is, the average survival probability at time point t , weighted by the initial number.

Remember that survival after t years of a group of individuals aged x at diagnosis whose follow-up can only terminate only with the death of the subjects (ie, a closed group) is equal to the ratio of the number of survivors of age $x+t$ to the number of survivors of age x . Thus

$$S_{ex}(t) = \frac{\ell_{x+0.5+t}}{\ell_{x+0.5}} \quad (4.21)$$

by using the data and expressions of the life table (see page in this chapter and Chapter 1). The half year results from the fact that the table is related to the birthday, while in fact an individual considered to be aged x is, on average, aged $x + 0.5$. The number of survivors at age $x + 0.5$ is obtained by linear interpolation of the values given in the table:

$$\ell_{x+0.5} = \frac{\ell_x + \ell_{x+1}}{2} \quad (4.22)$$

Even if the calculation of expected survival does not have to be very precise (in particular concerning relatively short survival times, for which observed survival is generally substantially less than expected survival), the simplest method is to use a table which gives the number of survivors at each birthday. When the computation is carried out from a table in which the age groups are of a duration Δx greater than one year (known as an abridged life table, for example, by five-yearly age groups), the correction is made in the same way and the value will correspond to the number of survivors at age $x + \Delta x/2$.

The life table which is used should reflect the general mortality of the region at the appropriate time. If the rate of mortality does not change too quickly and if follow-up is over a long time, adopting a life table which relates to the middle of this time period will be adequate. For example, a table based on a population census carried out in 1980 can be suitable for a cohort recruited in 1975 and followed up for ten years. If the time for which survival probability is calculated increases, it might be necessary to adopt mortality rates derived from several successive tables. If the survival of the general population improves, the use of only one table could produce an artificial improvement in the relative survival by underestimating the

expected survival of more recent cohorts. Furthermore, it will be seen in the following section that the construction of a life table does not present any special difficulties when complete data on the population and deaths are available, and that large numbers guarantee precise results.

When prognostic factors other than age and sex are identified, it is preferable to calculate expected survival by taking them into account. Examples of such factors are marital status, ethnicity and socioeconomic status. Their incorporation in the analysis can be achieved if the data necessary to construct the life tables are available as a function of these variables.

As an example, Table 4.5 shows some of the calculations used to establish expected survival for cancer of the colon for males in Geneva between 1970 and 1979.

Table 4.5 Calculation of the expected number of survivors^(a) for cancer of the colon for males in Geneva, Switzerland; incident cases 1970-1979

Age (years)	Survivor function	Number at risk	Expected proportion of survivors ^(b) after:					Expected number of survivors ^(b) after:					
			1 year	2 years	3 years	4 years	5 years	1 year	2 years	3 years	4 years	5 years	
$x, x+1$	ℓ_x	N_x											
1	2	3	4	5	6	7	8	9	10	11	12	13	
15-19	9831	4	0.9989	0.9978	0.9963	0.9945	0.9926	4.00	3.99	3.99	3.98	3.97	
20-24	9777	1	0.9982	0.9963	0.9947	0.9932	0.9918	1.00	1.00	0.99	0.99	0.99	
25-29	9686	5	0.9985	0.9971	0.9958	0.9945	0.9933	4.99	4.99	4.98	4.97	4.97	
30-34	9616	3	0.9987	0.9975	0.9960	0.9943	0.9926	3.00	2.99	2.99	2.98	2.98	
35-39	9556	1	0.9983	0.9966	0.9943	0.9915	0.9886	1.00	1.00	0.99	0.99	0.99	
40-44	9474	19	0.9972	0.9943	0.9907	0.9864	0.9820	18.95	18.89	18.82	18.74	18.66	
45-49	9340	22	0.9956	0.9912	0.9854	0.9783	0.9712	21.90	21.81	21.68	21.52	21.37	
50-54	9136	29	0.9927	0.9854	0.9764	0.9658	0.9552	28.79	28.58	28.32	28.01	27.70	
55-59	8807	34	0.9889	0.9778	0.9634	0.9459	0.9283	33.62	33.24	32.76	32.16	31.56	
60-64	8331	60	0.9811	0.9622	0.9393	0.9124	0.8855	58.87	57.73	56.36	54.74	53.13	
65-69	7579	75	0.9696	0.9392	0.9041	0.8642	0.8243	72.72	70.44	67.81	64.82	61.82	
70-74	6509	72	0.9517	0.9033	0.8515	0.7963	0.7410	68.52	65.04	61.31	57.33	53.36	
75-79	5104	61	0.9256	0.8511	0.7766	0.7022	0.6277	56.46	51.92	47.37	42.83	38.29	
80-84	3501	47	0.8816	0.7630	0.6468	0.5329	0.4191	41.43	35.86	30.40	25.05	19.70	
85-89	1900	16	0.7286	0.4569	0.2905	0.2297	0.1689	11.66	7.31	4.65	3.68	2.70	
90-94	363	5	0.6398	0.2796	0.0895	0.0696	0.0497	3.20	1.40	0.45	0.35	0.25	
Total		454	0.9474	0.8947	0.8455	0.7999	0.7543	430.10	406.18	383.86	363.14	342.43	

^(a) Life table for Geneva 1976-80 (Table 4.9).

^(b) Columns 4 to 8 are derived from column 2; columns 9 to 13 are derived from the preceding columns. For example, the expected proportion of survivors after three years for the age group 50 to 54 years is:

$$\frac{\ell_{50+3+2.5}}{\ell_{50+2.5}} = \frac{\ell_{55.5}}{\ell_{52.5}} = \frac{(4.5 \times 8807 + 0.5 \times 8331) / 5}{(9136 + 8807) / 2} = 0.9764$$

The number of expected survivors is therefore $29 \times 0.9764 = 28.32$. Columns 4 to 8 of the last line are obtained from the last five columns. Expected survival at five years for the whole group is $342.43/454 = 0.7543$.

The above calculations have been obtained from the life table based on five-yearly age intervals shown in Table 4.9. This table has been constructed using the method described below (see page 236). When a life table by single years of age is available, the calculations are based on annual and not five-yearly interpolations and are slightly more accurate. As an example, the expected survival of the same group has also been estimated from the Swiss life table (1978-1983) by single years of age in Annex 1. Initially, the survivors at regular yearly intervals are used and then those at five-yearly anniversaries as if the life table had been abridged. Table 4.6 shows the results using these different methods.

Table 4.6 Expected number of survivors from two life tables for cancer of the colon in men in Geneva, Switzerland (incident cases 1970-1979)

Time since diagnosis (years)	Expected number of survivors		
	Geneva table 1976-1980	Swiss table ^(a) 1978-1983	
	Quinquennial	Annual	Quinquennial
1	430.1	433.1	432.3
2	406.2	412.3	410.6
3	383.9	391.9	389.8
4	363.1	371.8	369.9
5	342.4	352.0	349.9

^(a) See Annex 1.

The results obtained from the Swiss life table show that the use of a five-yearly table hardly changes the estimate of the expected number of survivors obtained from the annual life table. On the other hand, the earlier Geneva life table (1976-1980) gives estimates which are noticeably less than those obtained from the Swiss life table (1978-1983). The observed differences probably reflect real differences in mortality both between Switzerland as a whole and Geneva, and between time periods. However, Table 4.7 shows that the differences between the relative survival probabilities are no more than 2%, which is certainly smaller than the differences between the results obtained with this method and alternative methods described below.

The confidence interval of the relative survival probability is proportional to that of the observed survival if the random variation in expected survival can be assumed to be negligible. The standard error for S_c is thus

$$SE(S_c) = \frac{SE(S_o)}{S_e} \quad (4.23)$$

Table 4.8 gives cause-specific and relative survival for various cancer sites at five years estimated from data from Geneva. For cause-specific survival, deaths have been attributed to the cancer concerned when the first three numbers of the

Table 4.7 Expected and relative survival probabilities from two life tables for cancer of the colon in men in Geneva, Switzerland (incident cases 1970-1979)

Time since diagnosis (years)	Observed survival ^(a)	Expected survival		Relative survival	
		1976-1980 ^(b)	1978-1983 ^(c)	1976-1980 ^(b)	1978-1983 ^(c)
1	0.601	0.947	0.954	0.634	0.630
2	0.479	0.895	0.908	0.535	0.528
3	0.402	0.846	0.863	0.475	0.466
4	0.352	0.800	0.819	0.440	0.430
5	0.324	0.754	0.775	0.429	0.418

^(a) Actuarial survival probability calculated from three-monthly intervals.

^(b) Geneva life table (Table 4.9).

^(c) Annual Swiss life table (Annex 1).

Table 4.8 Cause-specific and relative survival probabilities; Geneva, Switzerland (incident cases 1970-1977)

Site (sex)	Survival probabilities at 5 years	
	Relative	Cause-specific
Stomach (M)	0.16	0.19
Colon (M)	0.35	0.37
Colon (F)	0.37	0.36
Lung (M)	0.09	0.11
Breast (F)	0.69	0.71
Prostate	0.40	0.46
Ovary	0.29	0.30
Bladder (F)	0.38	0.45

ICD 8 code for the primary cause of death (coded by the Registry according to the WHO rules) correspond to the code for the site. Note that under this procedure, carried out in a registry where the causes of death are systematically verified and corrected if necessary, the specific rates for most sites are more optimistic than those obtained by the method of relative survival. This difference can be explained by the fact that, in contrast to cause-specific survival, the method of relative survival attributes to the cancer any deaths resulting from the secondary effects of treatment or from diseases caused by the same risk factors as the cancer.

The method of relative survival is also based on the assumption that the general mortality, as it is described by the life table of the population adequately takes account of all causes of mortality, except for the specific cause under study. This cause is considered to be negligible in comparison to all other causes of death. Only on this condition can relative survival provide an acceptable approximation to net survival. If this assumption does not hold, net survival will be overestimated as a result of the overestimation of the mortality due to other causes. For example, it is known that in women aged 50 to 54 years mortality from breast cancer is responsible for approximately one death in six. If the method of relative survival is

applied to this cause in this age group, net survival will be markedly overestimated. If death rates by cause are available for the general population, it is possible to calculate probabilities of death and then of survival by subtracting the risk of death due to the cancer under study. However, this precaution is rarely taken in practice. Indeed, it has been suggested that this correction is excessive, because the group under follow-up is still at risk of developing a second malignancy at the same site [13]. In addition, the life expectancy of a group of patients may in fact be less than that in the general population, as a result of the factors to which the patients have been exposed. These factors can lead to an increase in the risk of dying from other causes (for example, other diseases related to tobacco in people with lung cancer). In this case, the relative survival could be less favourable because of factors associated with the disease rather than because of the disease itself. The cause-specific survival, on the other hand, would not be affected by the occurrence of these deaths, since they would be attributed to other causes.

As we have indicated above, the determination of the time point t at which relative survival becomes constant leads to an estimate of the percentage $S_c(t)$ of people cured, because t corresponds to the moment at which deaths due to the cause under study are no longer recorded. Graphically, from t onwards the observed and expected survival curves become parallel on a logarithmic scale. In practice, a plateau of this kind in the relative survival is not uncommon even when deaths due to the cancer under study continue to be recorded. Furthermore, the relative survival may even start rising from a certain point onwards. Some authors have attributed this increase to the fact that mortality in surviving patients, who are receiving good medical care, can in the long term be less than that in the general population. This explanation is nonetheless incompatible with the interpretation of relative survival as an estimate of net survival. We will see (page 242) that this effect is more frequently due to a methodological bias [14].

Constructing a life table

Life tables established by national statistical services are usually available to calculate expected survival in the absence of cancer. However, it can happen that there is no life table suitable for the population being studied, either because the last official table is out of date or because the population being followed is too selected for its mortality to be described by the official table available. In this situation, a table can be built from available mortality rates (or calculated specifically for the purpose) provided they are sufficiently reliable and statistically accurate. Frequently, these conditions can only be fulfilled by using a table abridged in five-year age groups, and a long period of follow-up to calculate the mortality rates.

The method described below provides approximate results but sufficient accuracy for the objective. The results differ by definition from those which official statistical services would calculate from similar data; national tables are in fact made by using various methods to smooth the data and, especially in older age groups, are based on mathematical models of mortality. These procedures are not discussed here.

We saw in Chapter 1 that the life table was defined entirely by the probabilities of dying in the different age groups and that these probabilities could be directly estimated from mortality rates (Chapter 1, formula 1.31):

$$q_x = \frac{2\lambda_x\Delta_x}{2 + \lambda_x\Delta_x} \quad (4.24)$$

where λ_x is the annual mortality rate for the age group x and Δ_x the length in years of the age interval. In a five-yearly table, the probability of dying for each age group is obtained by

$$q_x = \frac{10\lambda_x}{2 + 5\lambda_x}$$

It is usual however to break down the first five-year interval by estimating the probability of dying in the first year of life separately. Thus, for the first age group ($x = 0$), and taking into account the concentration of deaths in the first days of life, it may be accepted that

$$q_0 = \lambda_0 \quad (4.25)$$

and, from the formula above, the probability of dying during the four following years is estimated as

$$q_1 = \frac{8\lambda_1}{2 + 4\lambda_1} \quad (4.26)$$

Up until the age group 70-74 years, this formula provides sufficient accuracy. For older age groups, Reed and Merrel's formula is recommended [15]. For $\Delta_x = 5$ this formula gives:

$$q'_x = 1 - e^{(-5\lambda_x - \lambda_x^2)} \quad (4.27)$$

The number of survivors at the beginning of each five-year period can then be calculated by applying the successive probabilities of death starting with the initial number (root of the table), equal for example to 10 000:

$$\ell_{x+5} = \ell_x (1 - q_x) \quad (4.28)$$

As an example, Table 4.9 shows the results obtained from mortality rates observed in Geneva between 1976 and 1980. The calculation of the probability of death q_x is based on formula (4.24) up to q_{75} and on formula (4.27) after that.

Alternative methods

As we have seen, the method described previously for calculating cause-specific and relative survival probabilities raise problems of interpretation which have led many authors to look for alternative solutions to the estimation of net survival.

Table 4.9 Calculations of the probability of death and of survivors by age from mortality rates observed in men in Geneva, Switzerland, between 1976 and 1980

Age x	Mortality rate λ_x	Probability of death q_x	Survivor function ^(a) l_x
0	0.010605	0.010605	10 000
1	0.000623	0.00249	9 894
5	0.003610	0.00180	9 869
10	0.000427	0.00213	9 852
15	0.001100	0.00548	9 831
20	0.001861	0.00926	9 777
25	0.001456	0.00725	9 686
30	0.001250	0.00623	9 616
35	0.001725	0.00859	9 556
40	0.002855	0.01417	9 474
45	0.004409	0.02180	9 340
50	0.007342	0.03605	9 136
55	0.011103	0.05402	8 807
60	0.018901	0.09024	8 331
65	0.030487	0.14116	7 579
70	0.048424	0.21597	6 509
75	0.074263	0.31397	5 104
80	0.119394	0.45731	3 501
85	0.174174	0.59392	1 900
90	0.311673	0.80900	363
95+	0.532710	0.94752	19

^(a) The number of survivors at a given age is obtained by linear interpolation from these numbers. For example, the number of survivors at 33 years is given by:

$$l_{33} = \frac{(2 \times 9616 + 3 \times 9556)}{5}$$

Results obtained by the method of cause-specific survival depend closely upon the quality of information available on causes of death. Systematic national differences in the coding of causes of death make this method unsuited to comparisons between registries. Relative survival, derived from expected survival as in the previous section, is equally subject to methodological biases which will be briefly discussed before presenting other methods.

The group being followed is often heterogeneous with respect to factors influencing both net survival and survival from competing causes, and hence observed survival. Observed survival is consequently an average value which depends not only on the initial structure of the group but also on the changes that the group experiences over time with respect to these different factors. For obvious reasons, the force of mortality acting on the group in the long term is generally closest to the mortality in the group with the longest lifespan [14,16]. The calculation of expected survival as described above does not take into account changes in the group over time, consequently the ratio between observed and expected survival will generally lead to overestimation of net survival over long time periods.

When age is the factor being considered, the consequences of this phenomenon can be easily demonstrated. Individuals in the group under study may end follow-up for one of three reasons, all of which are dependent on age: death from the disease under study, death from other causes and withdrawal from follow-up for reasons other than death. If $\lambda_x(t)$, $\mu_x(t)$ and $\gamma_x(t)$ are the corresponding instantaneous rates for each age x , then the force of mortality to which the group is subject at time t is a weighted average for which the weight $w_x(t)$ changes over time. This can be written

$$v(t) = \sum_{x=1}^g w_x(t) [\lambda_x(t) + \mu_x(t)] \quad (4.29)$$

where $w_x(t)$ is the proportion of individuals aged x years present in the population at time t . Using the notation of formula (4.17), this proportion is equal to

$$w_x(t) = \frac{w_x(0) S_{cx}(t) S_{ex}(t) \Gamma_x(t)}{\sum_{x=1}^g w_x(0) S_{cx}(t) S_{ex}(t) \Gamma_x(t)} \quad (4.30)$$

where $\Gamma_x(t)$ is the probability that an individual aged x years has a potential follow-up time greater than t .

$$\Gamma_x(t) = e^{-\int_0^t \gamma_x(u) du}$$

If the net survival and the potential follow-up time are independent of age, the weighting becomes proportional to the expected number of survivors:

$$w_x(t) = \frac{w_x(0) S_{ex}(t)}{\sum_{x=1}^g w_x(0) S_{ex}(t)}$$

and the net mortality rate comes out from the summation over age:

$$v(t) = \lambda(t) + \sum_{x=1}^g w_x(t) \mu_x(t)$$

Therefore, the observable survival probability can be written

$$\begin{aligned} S_0(t) &= e^{-\int_0^t v(u) du} \\ &= e^{-\int_0^t \lambda(u) du} e^{-\int_0^t \frac{\sum_{x=1}^g w_x(0) S_{ex}(u) \mu_x(u)}{\sum_{x=1}^g w_x(0) S_{ex}(u)} du} \end{aligned}$$

Furthermore, since

$$\mu_x(t) = - \frac{S'_{ex}(t)}{S_{ex}(t)}$$

it follows that

$$S_o(t) = S_c(t) e^{\int_0^t \frac{\sum_{x=1}^g w_x(0) S'_{ex}(t)}{\sum_{x=1}^g w_x(0) S_{ex}(t)} dt}$$

Since the term to be integrated is equal to the derivative of the logarithm of $\sum w_x(0) S_{ex}(u)$,

$$S_o(t) = S_c(t) \sum_{x=1}^g w_x(0) S_{ex}(t) \tag{4.31}$$

This formula implies that the observed survival is the result of two independent forces of mortality and that in this case relative survival correctly estimates net survival.

When the net survival or potential follow-up time depend on age, this relationship no longer holds true. In the long term, relative survival then tends to estimate the net survival of age groups which have the greatest life expectancy. In particular, if the potential follow-up time depends on age when the net survival is constant, the estimate of net survival by relative survival is biased.

In principle, the changes in the group over time can be accounted for in the calculation of expected survival, as in the method proposed by Ederer and Heise [17] and programmed by Rothman and Boice [18]. Instead of calculating expected survival from the initial number in each age group ($N_x(0)$), it is estimated at time $t + \Delta t$ for subjects still living at time t . Conditional expected survival $\bar{s}_e(t + \Delta t)$ is then the average of expected survival probabilities at different ages, weighted by the proportion of subjects still subject to the risk of dying at time t :

$$\bar{s}_e(t + \Delta t) = \sum_{x=1}^g w_x(t) s_{ex}(\Delta t) \tag{4.32}$$

The probability $s_{ex}(\Delta t)$ that an individual aged x years living at time t will still be living at the end of the time period $(t + \Delta t)$ is obtained from the life table. For example, if a five-yearly table is used

$$s_{ex}(\Delta t) = \frac{l_{x+0.5+\Delta t}}{l_{x+0.5}}$$

Expected survival of the group is then given by the cumulative product of the average survival over each five-year interval:

$$S_e(t_i) = \prod_{j=1}^i \bar{s}_e(t_j) \quad (4.33)$$

This method will be termed Ederer II, as opposed to Ederer's method previously described (Ederer I, see page 231).

It can be seen that if the net mortality rate $\lambda_x(t)$ does not depend on age, $S_c(t)$ is factorised as before:

$$\begin{aligned} S_o(t) &= e^{-\int_0^t \sum_{x=1}^g [w_x(u) (\lambda(u) + \mu_x(u))] du} \\ &= S_c(t) e^{-\int_0^t \sum_{x=1}^g [w_x(u) \mu_x(u)] du} \\ &= S_c(t) \cdot S_e(t) \end{aligned} \quad (4.34)$$

where $S_e(t)$ now denotes expected survival after taking into account the group structure at each time point t . This last formula holds whether or not the follow-up time depends on age. This characteristic was noted by Hakulinen [16] who proposed replacing $w_x(t)$ by $w_x^*(t)$ in the calculation of expected survival where

$$w_x^*(t) = \frac{w_x(0) S_{ex}(t) \Gamma_x(t)}{\sum_{i=1}^g w_x(0) S_{ex}(t) \Gamma_x(t)} \quad (4.35)$$

that is, by a weight which corrects for the distortion of age structure caused by censoring but in which net survival no longer plays a role. This expression gives the same result as (4.33) if net survival does not depend on age. It also provides an expected survival which is not affected by the value of net survival when it depends on age. In contrast, the Ederer II estimate derived in (4.32) and (4.33) provides an expected survival which depends on net survival and thus on the cancer being studied.

When $S_{cx}(t)$ does not depend on x but $\Gamma_x(t)$ does, for example when the population ages and increasingly older patients enter the study, Ederer II and the method of Hakulinen are preferable to Ederer I, which can lead to biases of the order of 10% [16].

When $S_{cx}(t)$ depends on x , Ederer II will systematically give results which are less than those of Hakulinen's method if net survival decreases with age. In fact, because they experience a larger number of deaths from cancer, the oldest subjects will be withdrawn more quickly from the calculation of expected survival, which will consequently be overestimated.

The original proponent of relative survival [19] clearly intended to calculate the net survival of a group of patients and not simply to provide an indicator of relative

risk of dying. However, the parameter being estimated when net survival depends on age has never been clearly defined. Implicitly, most authors are in favour of estimating the net survival of the group by the ratio of observed and expected survival. Explicitly taking age into account, this ratio can be written

$$\bar{S}_c(t) = \frac{\sum_{i=1}^g n_x S_{ex}(t) S_{cx}(t)}{\sum_{i=1}^g n_x S_{ex}(t)} \quad (4.36)$$

and thus corresponds to the average of net survival probabilities weighted by the expected number of survivors.

As stated before and shown by formula (4.36), the estimate of $\bar{S}_c(t)$ will always be closest to the net survival of the subgroup that has the greatest life expectancy. Considering the goal of this type of calculation, it would seem preferable to calculate the average of net survival probabilities, weighted by the initial numbers of subjects followed up and not by the expected number of survivors at time t :

$$\tilde{S}_c(t) = \frac{\sum_{i=1}^g n_x S_{cx}(t)}{\sum_{i=1}^g n_x} \quad (4.37)$$

A relative survival estimate based on direct standardisation of relative survival calculated separately for each age group can in theory be obtained in this way. However, it is less likely that this calculation can be made in practice, because of the statistical instability of the estimates for the older age groups. In addition, it can be questioned whether subjects who are diagnosed with cancer at 80 years should play the same role in the calculation of survival in the long term as those who were diagnosed with cancer at 40 years. The expected survival of elderly patients is moreover difficult to estimate from available data.

This discussion shows that the calculation of relative survival should be placed in the context of the classical methods for estimating the parameters of a survival model. This would have the two-fold advantage of providing both a better definition of the concept and a standard method for calculation. It seems natural in this context to estimate the net probability rate using a proportional hazards model (see page 260) where baseline hazard is a step function:

$$\lambda(t, \mathbf{z}) = e^{\beta \mathbf{z}} \sum_{k=1}^m \tau_k I_k(t) \quad (4.38)$$

where \mathbf{z} is a vector of covariates including age which might influence net survival, $I_k(t)$ is the indicator function of the k th interval ($I_k(t) = 1$ if t is in the interval, $I_k(t) = 0$

if it is not), and m the number of intervals being considered in the whole period under study.

Estimating such a model from individual observations (for subject i , t_i is survival time, δ_i the status living or dead in t_i , \mathbf{z}_i the vector of covariates and x_i the initial age) is easily achieved by the maximum likelihood method. The logarithm of the likelihood can be written (see Chapter 1, page 19):

$$L(\boldsymbol{\beta}, \boldsymbol{\tau}) = - \sum_{i=1}^n \Lambda(t_i, \mathbf{z}_i) + \sum_{i=1}^n \delta_i \log [\lambda(t_i, \mathbf{z}_i) + \mu(t_i + x_i, \mathbf{z}_{1i})] \quad (4.39)$$

where $\mu(t_i + x_i, \mathbf{z}_{1i})$ is the mortality rate at age $t_i + x_i$ for a subject in the general population characterized by \mathbf{z}_{1i} , the value of the covariate \mathbf{z}_1 , subvector of the covariate \mathbf{z} for which these data are available (e.g., sex and ethnic group).

After substituting (4.38) for λ in (4.39), the first and second derivatives of $L(\boldsymbol{\beta}, \boldsymbol{\tau})$ are obtained. Then, using the Newton Raphson method, the values $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}$, which maximize $L(\boldsymbol{\beta}, \boldsymbol{\tau})$ are derived. The confidence intervals of these parameters are obtained simultaneously by inversion of the matrix of second derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ at the maximum. It can be shown that in practice this method amounts to describing observed deaths in each interval by the sum of expected deaths based on the risk of overall mortality (μT_k) and of deaths resulting from the disease under study (λT_k), where T_k is the time spent in this interval by patients in the cohort who have respectively rates μ and λ [20].

Applying this method to the survival data for cancer of the colon in Geneva (1970-1979) shows that net survival does not depend on sex but strongly depends on age (appropriate significance tests are given below, see page 274). The mortality rate for age group 65 to 74 years is $e^{0.59} = 1.8$ times bigger than that for age group 0 to 64 years and for age group 75 to 99 years is $e^{0.91} = 2.5$ times bigger (Table 4.10).

The cumulative mortality rate, the corresponding net survival and their confidence intervals are obtained from $\lambda(t, \mathbf{z})$ and from the standard errors of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ by using the formulae already described in Chapter 1 (see for example formula 1.36) as well as the standard procedures for calculating asymptotic variances [7].

If $u_1, u_2 \dots u_m$ are the limits of the intervals used to define the baseline hazard step function and if $t \in [u_\ell, u_{\ell+1}]$

$$\Lambda(t, \mathbf{z}) = e^{\hat{\boldsymbol{\beta}}\mathbf{z}} \sum_{k=1}^{\ell-1} \hat{\tau}_k (u_{k+1} - u_k) + \hat{\tau}_\ell (t - u_\ell) \quad (4.40)$$

$$S_c(t, \mathbf{z}) = e^{-\Lambda(t, \mathbf{z})} \quad (4.41)$$

Table 4.10 gives values calculated for five and ten years and survival probabilities obtained by other methods. It will be seen in the present case that relative

Table 4.10 Survival probability^(a) of patients with cancer of the colon in Geneva, both sexes, 1970-1979

	Age at time 0 (years)			Total
	< 65	65-74	≥ 75	
Numbers				
At risk	322	292	326	940
expected at 10 years ^(b)	287.1	182.5	84.8	554.3
Crude survival^(c)				
5 years	0.51 (0.03)	0.31 (0.03)	0.18 (0.02)	0.33
10 years	0.45 (0.04)	0.19 (0.03)	0.08 (0.02)	0.24
Net survival				
Ederer I ^(c)				
5 years	0.54 (0.03)	0.37 (0.03)	0.32 (0.03)	0.42
10 years	0.50 (0.04)	0.30 (0.05)	0.30 (0.09)	0.40
Likelihood method ^(d)				
$\hat{\beta}$	0	0.59 (0.12)	0.91 (0.12)	–
5 years	0.56 (0.03)	0.35 (0.03)	0.24 (0.03)	0.40
10 years	0.51 (0.04)	0.30 (0.04)	0.19 (0.03)	0.36
Cause-specific survival				
5 years	0.57 (0.03)	0.38 (0.03)	0.25 (0.03)	0.40
10 years	0.55 (0.03)	0.31 (0.04)	0.21 (0.03)	0.36

^(a) The standard error of each estimate is given in brackets.

^(b) Calculations from the Swiss life table (1978-83).

^(c) Crude survival is calculated by the actuarial method with intervals of three months. Relative survival is obtained by the Ederer I method (4.17).

^(d) Estimation of $\hat{\beta}$ is obtained by the maximum likelihood method and survival analysis is deduced from these values and from that of τ (not reproduced here) by using (4.39) and (4.40).

survival is more optimistic than net survival obtained by the maximum likelihood method (Figure 4.5). The difference is particularly noticeable for older age groups. In this situation, the cause-specific survival can be calculated from the same data using causes of death as assessed by nosologists at the Geneva registry. Specific survival is slightly higher for the younger age groups than the estimate obtained by the maximum likelihood method but it is the same for the group as a whole. For the oldest age group, cause-specific survival is closer to the maximum likelihood estimate than the relative survival calculated by the Ederer I method.

Calculations not given here show that in this case the Hakulinen method provides results which are close to those of the Ederer I method (0.42 to 5 years and 0.42 to 10 years for the whole group). The last two methods would suggest that the time at which cancer of the colon can be considered to be cured is five years, because relative survival is then no longer decreasing. The maximum likelihood method is in fact stricter and would reject this assumption. The estimate of the average rate of mortality τ in the interval 5 to 10 years is 0.019 with a standard error equal to 0.008. In fact, the Ederer method overestimates relative survival for

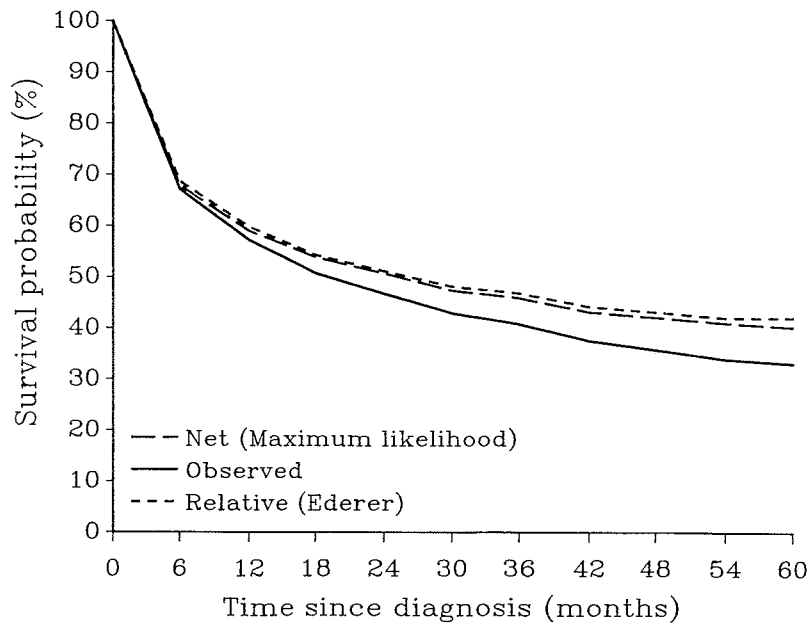


Figure 4.5 Observed and relative survival of colon cancer patients diagnosed between 1970 and 1979 in Geneva, Switzerland; both sexes

the oldest age group on the one hand and overall relative survival on the other hand by weighting each age group with expected survivors

$$(0.5 \times 287.1 + 0.3 \times 182.5 + 0.3 \times 84.8) / 554.3 = 0.40$$

(while the average weighted by the initial numbers would be 0.37).

Other alternatives to the Ederer method have been proposed. Breslow [21], Andersen and coworkers [22], and Hill and coworkers [23] proposed a model in which the presence of the disease multiplies the risk in the general population by a constant. Pocock and coworkers [24] have suggested a model similar to that described in this section, in which the risk added by the disease decreases exponentially with time. We would also suggest that the additive model is more natural in this context. However, the multiplicative model has certain advantages which have already been explained in detail [22].

Methods of comparison

Introduction

In some situations, comparison of survival between two groups can only be made at one time point on the survival curve, for example at five years. The method used in this situation is described below. Generalization from this comparison to the whole survival curve can nonetheless lead to incorrect results : survival probabilities

at five years in the two groups might be the same even though the two survival curves are different as a whole. This is a common situation in clinical practice, especially when two groups are treated with different therapies (e.g., surgery versus radiography) having effects which may be unequally distributed over time. The choice of a specific time point to evaluate survival will thus be arbitrary. Furthermore, when the groups are small in size, a survival rate at a given time point is subject to a relatively large random error and observed differences may appear to be due to chance, while the more powerful overall comparison between survival curves might lead to a significant difference.

Except in particular situations, there is therefore a strong argument to comparing the whole of the two curves, provided that the necessary data are available. The problem is then to summarize all differences between the curves by a suitable indicator and then to use an appropriate test. The methods described below (see page 248) illustrated with data on cancer of the colon in Côte-d'Or, Geneva and the canton of Vaud provide a way to answer these questions.

Furthermore, it is not unusual for the groups being compared to be unequally distributed with regard to age, sex and stage of disease at diagnosis. The comparison should take these prognostic factors into account by a method based on an appropriate stratification, provided that the necessary information is available for each subject. This method will be illustrated by data from Geneva on the survival of subjects with malignant melanoma by site and sex (see page 255).

A substantial amount of research has been carried out over the last two decades in the area of semiparametric models of survival, which simultaneously take into account various prognostic factors defined on an individual basis [25, 26]. Among these, the proportional hazards model has been of particular interest because of its simplicity and effectiveness (see page 260). We will present this model using breast cancer data from the Geneva Cancer Registry to assess the influence of health care systems on survival.

Long-term survival cannot be compared between groups without considering competing causes of death. This consideration leads to the comparison of relative survival between groups, (see page 272). Despite the above-mentioned difficulties, it may be that a point comparison of two rates is appropriate or that, as is often the case, the basic data are not available to make an overall comparison. The method to be used in this situation will be presented briefly below.

Comparison of crude survival probabilities

Test to compare two survival rates

Observed survival probabilities are often only published in the form of an actuarial rate at a given time (e.g., survival rate at five years). Evaluating the statistical significance of the observed differences between two or more populations requires information on the variability of the rates, which unfortunately is rarely

available. Knowledge of the number of cases initially entered in the study is not sufficient to correctly estimate the variability. If the observed number of survivors at a given time is known, the upper limit of the variance of the survival rate can be calculated (see page 222) and a slightly conservative comparison can be made. Implementation of the test is straightforward as soon as an estimate of the variability of the rates is available.

Let $S_1(t)$ and $S_2(t)$ be the survival rates at time t , estimated using one of the two methods described in the first part of this chapter (see pages 216 and 219) and let $V_1(t)$ and $V_2(t)$ be the estimates of their variance. For a given value t_0 of t , the quantity

$$\varepsilon = \frac{S_1(t_0) - S_2(t_0)}{\sqrt{V_1(t_0) + V_2(t_0)}} \quad (4.42)$$

follows an approximately normal distribution and provides an appropriate statistic to test the hypothesis of equality of the rates in t_0 . For example, the five-year survival rates in men following cancer of the colon are obtained using the actuarial method as 0.324 (Geneva) and 0.348 (canton of Vaud). As the standard errors obtained by Greenwood's method are 0.0228 and 0.0216, we have :

$$\varepsilon = \frac{0.324 - 0.348}{\sqrt{0.0228^2 + 0.0216^2}} = -0.76$$

It is therefore concluded from this value that five-year colon cancer survival for men does not differ significantly between the canton of Vaud and Geneva.

When standard errors are estimated from initial numbers (as if there were no censored observations), the standard error for Geneva is 0.0220. The proximity of this value to 0.0228 can be explained by the fact that the majority of subjects were followed up for five years. Conversely, the standard error in the canton of Vaud is 0.0187 instead of 0.0216 because the percentage of censored observations is high among the 650 cases. In this situation, it is particularly important to provide a correct estimate of variance. In fact, the upper limit of the variance calculated solely from the 114 survivors observed at 60 months would be 0.0263 which is a slight over-estimation.

In general, it is highly recommended that survival rates be presented with their standard errors to provide the basis for making more reliable comparisons.

Rank tests to compare survival curves

The objective of these tests is to compare two or more survival curves making optimum use of the available information about the survival of patients in the different groups. They are part of a family of tests which are the extensions to censored data of methods based on ranks [27]. Even if the data were not censored, it is inappropriate to make a simple comparison of average survival times of two populations, because survival times do not follow the normal distribution which is a prerequisite

for the validity of classical tests. Non-parametric procedures (rank tests in particular) allow us to overcome this difficulty.

When observations are complete, the implementation of a rank test requires ordering all survival times within the two groups in ascending order, assigning a score c_i to the observation of rank i , then calculating the observed sum C of c_i for one of the samples and, finally, comparing C to its expected value $E(C)$ under the assumption that ranks are randomly assigned between the two samples. If, in reality the survival times in one of the samples are on average greater than in the other, their ranks will be on average higher and C will be significantly greater than $E(C)$. The scores c_i are chosen in such a way that the resulting tests are efficient against certain alternatives in the framework of the largest possible family of survival distributions. In the Mann-Whitney test, for example, the score c_i of an observation is simply its rank i in the classification. In the Savage test, the score c_i is equal to the expectation of the observation of rank i in a sample from a distribution with a constant mortality rate (that is, an exponential distribution). For n observations, this score is equal to:

$$c_i = \frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-i+1}$$

The extension of these methods to censored data requires the assignment of a score to observations censored between the i th and the $(i + 1)$ th survival time, for which it is only known that survival time exceeds t_i . When the Savage test is extended to censored data, it becomes very similar to the Mantel-Haenszel test described in Chapter II (see page 77). In the framework of survival comparisons it is known as the *logrank test* [28]. The generalization proposed by Gehan and Breslow [29, 30] of the Mann-Whitney test (or the Kruskal-Wallis test for more than two populations) is similar to the logrank test. Both tests are based on a comparison of observed deaths d_i with expected deaths e_i in a group, at each time point t_i where at least one death is observed. The two tests only differ in the weight given to the observed differences $(d_i - e_i)$. Application of the logrank test to the comparison of three populations is given below.

Let:

- $k = 1, 2, 3$ be the three groups being compared,
- t_i be the dates at which at least one death occurred,
- d_{ki} be the number of observed deaths at t_i in group k ,
- e_{ki} be the number of expected deaths at t_i in group k ,
- D_i be the total number of deaths occurring at t_i ,
- n_{ki} be the number of subjects at risk just before t_i in group k , and
- N_i be the total number of subjects at risk just before t_i .

The test is based on the comparison of the total number of observed deaths in each group k , ($O_k = \sum_i d_{ki}$), with the number of deaths ($E_k = \sum_i e_{ki}$) that might have been observed if the force of mortality was the same in the three groups. The

number of expected deaths in each group should be calculated at each time point t_i when at least one death occurred. The comparison takes into account not only the total number of observed deaths in each group before a given date but also the date at which these deaths occurred. It is possible that the overall proportion of deaths is the same in the three groups over the entire time period under consideration, but that there is an unequal distribution of survival times in the groups due to differences between the instantaneous mortality rates. Figure 4.6 shows ten-year survival for men with colon cancer in Geneva, Côte-d'Or and the canton of Vaud. The observed rates are practically identical in the three registries after the fifth year, while survival is better in the canton of Vaud before this date.

Table 4.11 gives the breakdown of counts used in the test at a time point t_i when at least one death occurred.

Under the null hypothesis, the observed number of deaths in each of the three groups would be proportional to the respective numbers of subjects at risk in each group at this date. The numbers e_{ki} of expected deaths under this hypothesis are therefore:

$$e_{1i} = D_i \times \frac{n_{1i}}{N_i} \quad e_{2i} = D_i \times \frac{n_{2i}}{N_i} \quad e_{3i} = D_i \times \frac{n_{3i}}{N_i}$$

Table 4.11 Logrank test. Breakdown of counts at each time point t_i when one or more deaths occurred

	Group 1	Group 2	Group 3	Total
Deceased	d_{1i}	d_{2i}	d_{3i}	D_i
Alive	$n_{1i} - d_{1i}$	$n_{2i} - d_{2i}$	$n_{3i} - d_{3i}$	$N_i - D_i$
Total	n_{1i}	n_{2i}	n_{3i}	N_i

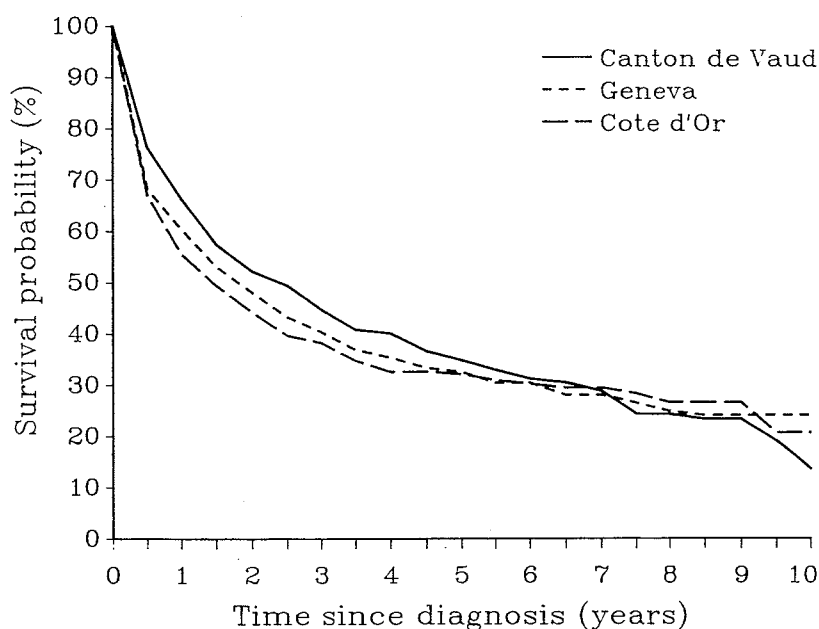


Figure 4.6 Survival of male colon cancer patients in the canton of Vaud, Switzerland (1974-1983), Geneva, Switzerland (1970-1979) and Côte-d'Or, France (1976-1982)

giving the total numbers of expected deaths:

$$E_1 = \sum_i e_{1i} \quad E_2 = \sum_i e_{2i} \quad E_3 = \sum_i e_{3i}$$

The objective of the test is to judge whether the sum of the differences between the observed and expected numbers in each group ($O_k - E_k$) differs significantly from zero. The test is identical in principle to the Mantel-Haenszel test described in Chapter 2 (see page 77). However, the distribution of deaths in this case is multinomial leading to a three-dimensional hypergeometric distribution for d_{ki} when the total number of deaths D_i is fixed, rather than the multinomial distribution which results when the number of deaths follows a Poisson distribution. The variances and covariances of d_{ki} are thus :

$$\text{Var}(d_{ki}) = \frac{D_i n_{ki} (N_i - n_{ki}) (N_i - D_i)}{N_i^2 (N_i - 1)} \quad (4.43)$$

$$\text{Cov}(d_{hi}, d_{ki}) = \frac{-D_i n_{hi} n_{ki} (N_i - D_i)}{N_i^2 (N_i - 1)} \quad (4.44)$$

from this, we derive the symmetric variance-covariance matrix \mathbf{V} of $(O_k - E_k)$ which, when comparing three groups, is of dimension 3×3 ; the diagonal elements are

$$v_{kk} = \sum_i \text{Var}(d_{ki}) \quad 1 \leq k \leq 3 \quad (4.45)$$

and the remaining elements

$$v_{kh} = \sum_i \text{Cov}(d_{ki}, d_{hi}) \quad 1 \leq k \leq 3, \quad 1 \leq h \leq k \quad (4.46)$$

The definition of the expected numbers implies that the sum of O_k is equal to the sum of E_k : it is therefore sufficient to know the deviations $(O_k - E_k)$ in two of the groups in order to know the deviation in the third. To assess the size of the deviation between all pairs of observed and expected values, it is therefore sufficient to retain two of the three groups. The significance is determined from the statistic:

$$T_1 = w_{11}(O_1 - E_1)^2 + w_{22}(O_2 - E_2)^2 + 2 w_{12}(O_1 - E_1)(O_2 - E_2) \quad (4.47)$$

where w_{hk} are the elements of the inverse of the variance-covariance matrix of $(O_1 - E_1)$ and $(O_2 - E_2)$:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix} = \mathbf{V}^{-1} = \frac{1}{v_{11} \times v_{22} - v_{12}^2} \begin{bmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{bmatrix} \quad (4.48)$$

under the null hypothesis of equal survival, the statistic T_1 follows a χ^2 law on two degrees of freedom. Formula (4.47) can be easily generalized to r groups and is expressed in matrix form as:

$$T_1 = (\mathbf{O} - \mathbf{E})' \mathbf{W} (\mathbf{O} - \mathbf{E}) \quad (4.49)$$

where $(\mathbf{O} - \mathbf{E})$ is the vector made up of the $r - 1$ differences between the observed and expected numbers in the first $r - 1$ groups, and \mathbf{W} is the inverse of the variance-covariance matrix of the first $r - 1$ differences of $(O_k - E_k)$. When T_1 exceeds the critical value corresponding to χ^2 on $r - 1$ degrees of freedom, it can be concluded that the r survival curves are different. The test of a more specific alternative hypothesis is carried out in a way analogous to that in Chapter 2, page 90. For example, a trend test is based on the statistic :

$$T_1^* = \frac{\left[\sum_k x_k (O_k - E_k) \right]^2}{\text{Var} \left[\sum_k x_k (O_k - E_k) \right]} = \frac{[\mathbf{X}' (\mathbf{O} - \mathbf{E})]^2}{\mathbf{X}' \mathbf{V} \mathbf{X}} \quad (4.50)$$

where $\mathbf{X}' = (x_1, \dots, x_r)$ is the vector corresponding to the r levels of a quantitative variable such as disease stage or age group which may have an influence on survival. T_1^* follows a χ^2 distribution on one degree of freedom and provides a more powerful test than T_1 for specific alternative hypotheses defined *a priori*. Contrasts, particularly pairwise comparisons, between groups can be tested with the same statistic. It is worth noting however that if a large number of comparisons of this kind are made, the type one error associated with the final result could be seriously underestimated.

The homogeneity of survival rates in 2 groups can be evaluated by a conservative test, approximating formula (4.49) by

$$T_2 = \sum_{k=1}^r \frac{(O_k - E_k)^2}{E_k} \quad (4.51)$$

under the assumption that T_2 is distributed as χ^2 on $r - 1$ degrees of freedom [31,32]. Because of its simplicity, the statistic T_2 is sometimes calculated first when a rapid result is needed; if it gives a non-significant result, more precise tests can still be carried out.

Table 4.12 shows some of the calculations involved in the comparison of five-year survival following cancer of the colon in males in Geneva (incident cases 1970-83), Côte-d'Or (incident cases 1976-80) and the canton of Vaud (incident cases 1974-83).

Table 4.12 Comparison of survival following cancer of the colon among males in the canton of Vaud, Switzerland (incident cases 1974-83), Geneva, Switzerland (incident cases 1970-83), and in Côte-d'Or, France (incident cases 1976-80)

t _i (month)	Vaud		Geneva		Côte-d'Or		Total		Expected		
	n _{vi}	d _{vi}	n _{gi}	d _{gi}	n _{ci}	d _{ci}	N _i	D _i	e _{vi}	e _{gi}	e _{ci}
1	650	80	454	80	411	80	1 515	251	107.7	75.2	68.1
2	568	18	349	18	331	25	1 248	61	27.8	17.1	16.2
3	546	18	329	15	306	15	1 181	48	22.2	13.4	12.4
4	520	15	314	13	291	10	1 125	38	17.6	10.6	9.8
5	501	21	300	5	281	6	1 082	32	14.8	8.9	8.3
6	476	14	295	5	275	9	1 046	28	12.7	7.9	7.4
7	460	9	289	12	265	6	1 014	27	12.2	7.7	7.1
.
.
.
.
57	125	3	128	0	66	0	319	3	1.2	1.2	0.6
58	118	0	128	1	66	0	312	1	0.4	0.4	0.2
59	118	1	127	0	63	0	308	1	0.4	0.4	0.2
60	116	2	127	1	61	0	304	3	1.1	1.3	0.6
Total	O _v = 369		O _g = 292		O _c = 270		931		E _v =406.6	E _g =279.6	E _c =244.8

The logrank test is calculated from the variance-covariance matrix of the observations:

$$V = \begin{bmatrix} 214.55 & -114.33 & -100.21 \\ -114.33 & 183.09 & -68.76 \\ -100.21 & -68.76 & 168.97 \end{bmatrix}$$

If the test is calculated from the first two groups, we obtain (cf formula 4.48):

$$W = \frac{1}{26\ 210.6} \times \begin{bmatrix} 183.09 & 114.33 \\ 114.33 & 214.55 \end{bmatrix}$$

Applying formula (4.47) gives:

$$T_1 = \frac{1}{26\ 210.6} [183.09 \times (-37.6)^2 + 214.55 \times (12.4)^2 + 2 \times 114.33 \times (-37.6) \times (12.4)]$$

$$T_1 = 7.07$$

The simplified test (formula 4.51) gives:

$$T_2 = \frac{(369 - 406.6)^2}{406.6} + \frac{(292 - 279.6)^2}{279.6} + \frac{(270 - 244.8)^2}{244.8} = 6.62$$

which, as stated previously, is slightly less than the result of the exact test.

Overall, survival differs between the three groups ($p = 0.029$). The test T_1^* applied with values of \mathbf{X}' equal to $(1, -1, 0)$ and $(0, 1, -1)$ shows that survival differs between the canton of Vaud and Geneva ($T_1^* = 3.99$, using the first value of \mathbf{X}' , $p < 0.05$), but not between Geneva and Côte-d'Or ($T_1^* = 0.34$, using the second value of \mathbf{X}'). Figure 4.6 showing ten-year survival demonstrates clearly the reason for this apparent contradiction: mortality is initially less in the canton of Vaud, while the five-year survival rate here is almost the same as in Geneva (see the test described above page 247).

The difference between survival curves is often characterized by a constant ratio R between instantaneous death rates. This ratio is a measure of the relative risk of death in one group compared to another. The quantity

$$\tilde{R}_{kh} = \frac{O_k/E_k}{O_h/E_h} \quad (4.52)$$

can be taken as an estimate of this ratio if it is between 0.5 and 2, but it has been shown to be an underestimate when it is too far from unity [21]. In this latter case, estimation of the relative rate by maximum likelihood in the context of the Cox model (see page 263) is preferable. The Mantel-Haenszel estimate can also be calculated from the series of contingency tables obtained for each date of death as in Table 4.11:

$$\hat{R}_{kh} = \frac{\left[\sum_i \frac{d_{ki} (n_{hi} - d_{hi})}{N'_i} \right]}{\left[\sum_i \frac{d_{hi} (n_{ki} - d_{ki})}{N'_i} \right]} \quad (4.53)$$

where $N'_i = n_{ki} + n_{hi}$

The force of mortality in group k is R_{kh} times higher than that to which group h is subjected. The logrank test evaluates the hypothesis $R_{kh} = 1$ against the alternative $R_{kh} \neq 1$. It is optimum in this situation.

When rates are not proportional and the ratio of the forces of mortality decreases over time as in the example above, the difference between curves will be more correctly determined by the Gehan-Breslow test, which only differs from the logrank test in the weighting of the differences between the observed and expected numbers. The principle of this test is based on the calculation of a score for each of the groups being compared. This score summarizes all possible pairwise comparisons of survival times: all follow-up times t_{kj} in group k are compared to all follow-up times t_{lh} in the other groups and:

- $\psi(k,j,l,h) = +1$ if the comparison of t_{kj} and t_{lh} indicates that subject j has a better survival probability, and
- $\psi(k,j,l,h) = -1$ if the comparison of t_{kj} and t_{lh} indicates that subject j has poorer survival;

the value 0 is given to ψ when the comparison is inconclusive, either because of equal survival times or censored observations. The score associated with group k $\sum_j \left[\sum_{l,h} \Psi(k, j, l, h) \right]$ is higher to the extent that survival times in the group are on average greater than those of all the other groups. It is easy enough to verify that, ignoring its sign, this score is equal to

$$C_k = \sum_i N_i d_{ki} - n_{ki} D_i = \sum_i N_i \left(d_{ki} - D_i \frac{n_{ki}}{N_i} \right) \tag{4.54}$$

where the summation extends over all the dates of death. In other words, C_k is again obtained as a sum of the differences between observed and expected numbers over all time points at which a death occurs. However, unlike the logrank test, this sum is weighted by the total number of subjects still at risk at this date. Any given difference will thus have a weight which is higher for short follow-up times, when N_i is still large, while its effect on the score will be negligible for long follow-up times if the number of subjects still under observation is small. Figure 4.6 suggests that this statistic will be particularly sensitive in the first years to large observed differences between the canton of Vaud, Geneva and Côte-d’Or.

The score C_k is calculated for each of the first $r - 1$ groups, and the variance-covariance matrix of the C_k is constructed from:

$$v_{kk} = \sum_i N_i^2 \text{Var} (d_{ki}) \tag{4.55}$$

$$v_{kh} = \sum_i N_i^2 \text{Cov} (d_{ki}, d_{hi}) \tag{4.56}$$

where the variance and covariance are defined by formulas (4.43) and (4.44). When comparing three groups, the statistic T_3 , analogous to T_1 , is written:

$$T_3 = w_{11} C_1^2 + w_{22} C_2^2 + 2 w_{12} C_1 C_2$$

where, as previously described, W_{ij} denotes the terms of \mathbf{W} , the inverse of the variance-covariance matrix of the parameter estimates for the two groups involved in the calculation.

As the value of C_k is generally very high, each score in the numerical application given below has been divided by the number of thousands of conclusive comparisons carried out (with score equal to either +1 or -1). This operation obviously does not change the proposed weighting and reduces the differences to an order comparable to that obtained for the logrank test. The calculations were carried out with data from the canton of Vaud, Geneva and Côte-d’Or; 1 156 446 conclusive comparisons were carried out and the modified score (per 1000) in each of the groups is:

$$C_1 = -50.04 ; C_2 = 18.17 ; C_3 = 31.87$$

which indicates that survival time is on average longer in the first group. To evaluate the significance of this difference, the variance and covariance of these scores are calculated using (4.55) and (4.56) modified by the weighting above. The result is the matrix

$$\begin{bmatrix} 184.74 & -96.85 & -87.89 \\ -96.85 & 155.60 & -58.75 \\ -87.89 & -58.75 & 146.64 \end{bmatrix}$$

\mathbf{W} being the inverse of the matrix \mathbf{V} defined by the first two rows and columns of the matrix above, we have:

$$\mathbf{W} = \frac{1}{19\,365.6} \begin{bmatrix} 155.60 & 96.85 \\ 96.85 & 184.74 \end{bmatrix}$$

and the value of the statistic:

$$\begin{aligned} T_3 &= \frac{1}{19\,365.6} [155.60 \times (-50.04)^2 + 184.74 \times (18.17)^2 + 2 \times 96.85 \times (-50.04) \times (18.17)] \\ &= 14.17 \end{aligned}$$

The three global tests thus lead to the conclusion that survival in the first five years is significantly different between the three registries. Gehan-Breslow's test provides a much larger value ($T_3 = 14.17$) than that obtained for T_1 ($T_1 = 7.07$) or T_2 ($T_2 = 6.62$) as it takes greater account of differences observed at the beginning of the curve and, as Figure 4.6 shows, differences between the Canton of Vaud and the other two registries are present mainly in the initial period. These results suggest that long-term survival is actually the same in the three registries but that artefacts linked to the definition of the date of incidence and to the mechanisms of follow-up may have led to the differences observed in the initial period.

Obviously not every test described in this section needs to be used for each data set under study, nor should a test be selected solely for its convenience. In fact, the choice of a test should, as always, be dictated by a hypothesis made prior to the examination of the survival curves. As Gehan-Breslow's test can behave unpredictably in certain circumstances [26], caution is recommended in its use. In particular, the result of this test should be ignored if it does not produce significance when the logrank test does, especially when there is a large number of censored observations.

Stratified comparison of crude survival

The survival of a group of patients is generally associated with many prognostic factors which are themselves related. When survival of two or more groups are compared, it is crucial to take into account known prognostic factors. Thus, the difference of colon cancer survival between Geneva and the canton of Vaud might

be partially explained by differences in the distribution of clinical stages between the two regions: it would be preferable to compare the two geographic areas 'holding stage constant'. The opposite situation can also occur: a difference in survival between groups may go unrecognized if a major determinant of survival within groups is ignored.

Regardless of the distribution of the prognostic factors in the various groups, a comparison holding these factors constant will tend to increase the power of the test that is used by decreasing the within group variability. Accounting for these factors in the analysis is achieved through an adjustment based on an appropriate stratification as follows.

For each category (or stratum) of the factor under consideration, the approach described in the previous paragraph is used (see page). We calculate, for each group k and stratum j , the number of observed deaths (O_{kj}) and the number of deaths that would occur (E_{kj}) if, in each stratum, the force of mortality had been the same in each of the groups being compared. In other words, the expected number e_{kji} of deaths at date t_i in group k is calculated and summed separately for each stratum j . Evaluation of the difference between observed and expected numbers is based on the sum of results from all strata.

In order to simplify the notation, we first show how to carry out the adjustment for a two-group comparison in the presence of a two-category prognostic factor. Generalization to any number of groups and categories is immediate. The method is illustrated by comparing survival between patients with malignant melanoma of the legs and patients with melanoma at another site (using data from Geneva). As sex is known to be a prognostic factor for melanoma, it should be considered in the comparison.

In the case of two groups, only one needs to be considered. we take the group ($k = 1$) of patients with melanoma of the legs. Thus:

- for males (stratum $j = 1$)

$$O_{11} = \sum_i d_{11i}, \quad E_{11} = \sum_i e_{11i}$$

$$v_{11} = \sum_i \text{Var}(d_{11i})$$

- for females (stratum $j = 2$)

$$O_{12} = \sum_i d_{12i}, \quad E_{12} = \sum_i e_{12i}$$

$$v_{12} = \sum_i \text{Var}(d_{12i})$$

variances in each stratum being calculated using formula (4.43).

Observed and expected number of deaths in the group of patients with skin melanoma of the legs are added over the two strata:

$$O_{1.} = O_{11} + O_{12} ; E_{1.} = E_{11} + E_{12} \quad (4.57)$$

similarly, the variance of $O_{1.}$ (the within-stratum variance) is obtained by addition of the variances of O_{11} and O_{12}

$$\text{Var}(O_{1.} - E_{1.}) = v_{11} + v_{12} \quad (4.58)$$

then the quantity

$$T_1 = \frac{(O_{1.} - E_{1.})^2}{\text{Var}(O_{1.} - E_{1.})} \quad (4.59)$$

is calculated; under the null hypothesis of equal survival, the statistic T_1 follows a χ^2 distribution with one degree of freedom. Comparison of T_1 with the critical value of the corresponding χ^2 distribution tells us if the two survival curves differ overall, after correcting for sex.

An appropriate formula similar to (4.51), based only on observed and expected data in the two groups can also be used:

$$T_2 = \frac{(O_{1.} - E_{1.})^2}{E_{1.}} + \frac{(O_{2.} - E_{2.})^2}{E_{2.}} \quad (4.60)$$

T_2 , under the null hypothesis, approaches (but is greater than) a χ^2 with one degree of freedom. The significance of T_2 is sufficient to conclude that the groups have different survival probabilities, but calculation of T_1 is necessary if T_2 is not significant.

When more than two groups are being studied, the calculation of the within-group covariance using formulae (4.44) and (4.46) and the generalization of (4.58) provides the basis for carrying out the adjusted global test and the associated tests on one degree of freedom given on page 251. In this case, the observed and expected numbers calculated from the obvious generalization of formula (4.57) to more than two groups are used in formulae (4.49) and (4.50).

Table 4.13 gives survival data by site for incident cases of malignant skin melanoma among males and females in Geneva between 1970 and 1982. Melanomas of the legs make up group 1; sites from the rest of the body form group 2. The data are illustrated in Figure 4.7.

The unadjusted comparison of group 1 (legs) and group 2 (other sites) using the logrank test shows significantly better survival in group 1:

$$\chi_{1df}^2 = \frac{(20 - 31.7)^2}{21.3} = 6.4$$

If the comparison is made without taking into account the patients' sex, it could be concluded that the mortality rate is approximately twice as high ($p < 0.01$) in group 2 than in group 1 (4.52).

Table 4.13 Survival of patients with malignant skin melanoma^(a) in Geneva, Switzerland. Incident cases 1970-1982

Year	Group 1 ^(b)				Group 2 ^(b)			
	Number at risk	Observed deaths	Expected deaths ^(c)	Survival (standard error)	Number at risk	Observed deaths	Expected deaths ^(c)	Survival (standard error)
1	80	3	8.0 (5.6)	0.96 (0.021)	180	24	19.0 (5.6)	0.88 (0.023)
3	76	7	19.3 (13.2)	0.92 (0.031)	149	55	42.7 (13.2)	0.73 (0.031)
5	69	14	25.4 (17.2)	0.83 (0.041)	138	66	54.6 (17.2)	0.68 (0.033)
10	22	20	31.7 (21.3)	0.67 (0.068)	37	78	66.3 (21.3)	0.52 (0.048)

^(a) Number of patients living and still under follow-up, cumulative deaths and survival are given for the end of the year under consideration. These values are based on calculations using a month as the time unit.

^(b) Group 1: 83 patients with melanoma of the skin of the leg. Group 2: 204 patients with other skin melanomas.

^(c) Variance in brackets.

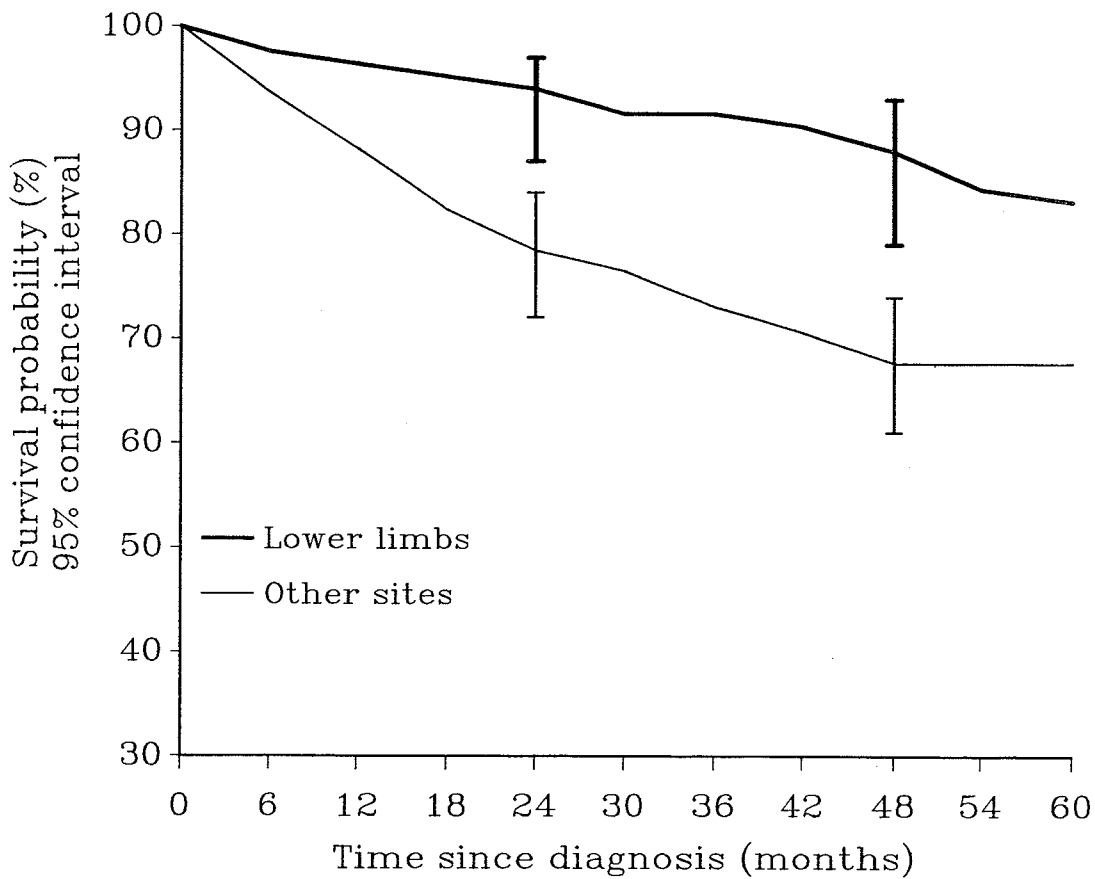


Figure 4.7 Survival of skin melanoma patients by site, Geneva, Switzerland (incident cases 1970-1982); both sexes

In fact, it is known that sex plays a strong prognostic role in the survival of melanoma patients: women have a better survival than men. As melanoma of the legs is much more frequent in women (64/146 = 43.8%) than in men (19/141 = 13.5%), the observed effect can partially be explained by sex.

Table 4.14 gives the same survival data for each sex separately. After adjustment (see formula 4.58), the comparison becomes:

$$T_1 = [(7+13) - (9.9+16.2)]^2 / (8.2+8.7) = 2.2 \quad (p = 0.14)$$

Thus, the hypothesis of equal survival for the two sites can no longer be rejected after taking sex into account. The relative mortality rate is now estimated to be $1.08/0.77 = 1.4$, a value which no longer differs significantly from unity.

Figure 4.8 shows that these results should be examined even more closely since, survival for group 1 (legs) is better than that of group 2 (other sites) regardless of sex or follow-up time. We see that, in men, there are only 19 cases of melanoma of the legs. Thus estimation of survival rates in this stratum is imprecise: for this latter patients the confidence intervals of survival probability at five and ten years

Table 4.14 Comparison of survival of subjects with malignant skin melanoma by site and sex in Geneva, Switzerland. Incident cases between 1970 and 1982^(a)

Year	Group 1				Group 2			
	Number at risk	Observed deaths	Expected deaths ^(c)	Survival (standard error)	Number at risk	Observed deaths	Expected deaths ^(c)	Survival (standard error)
Males								
1	18	1	2.9 (2.5)	0.95 (0.051)	102	20	18.1 (2.5)	0.84 (0.034)
3	16	3	6.4 (5.4)	0.84 (0.084)	82	40	36.6 (5.4)	0.67 (0.043)
5	14	5	7.8 (6.5)	0.74 (0.101)	75	47	44.2 (6.5)	0.62 (0.044)
10	5	7	9.9 (8.2)	0.56 (0.138)	20	56	53.1 (8.2)	0.43 (0.060)
Females								
1	62	2	2.6 (1.5)	0.97 (0.022)	78	4	3.4 (1.5)	0.95 (0.024)
3	60	4	8.6 (4.7)	0.94 (0.030)	67	15	10.4 (4.7)	0.82 (0.043)
5	55	9	12.9 (6.9)	0.86 (0.044)	63	19	15.1 (6.9)	0.77 (0.047)
10	17	13	16.2 (8.7)	0.71 (0.077)	17	22	18.8 (8.7)	0.66 (0.073)

^(a) Number of patients living and still under follow-up, cumulative deaths and survival are given for the end of the year under consideration. These values are based on calculations using a month as the time unit.

^(b) Group 1: melanoma of the skin of the leg (18 male and 64 female patients). Group 2: other skin.

^(c) Variance in brackets.

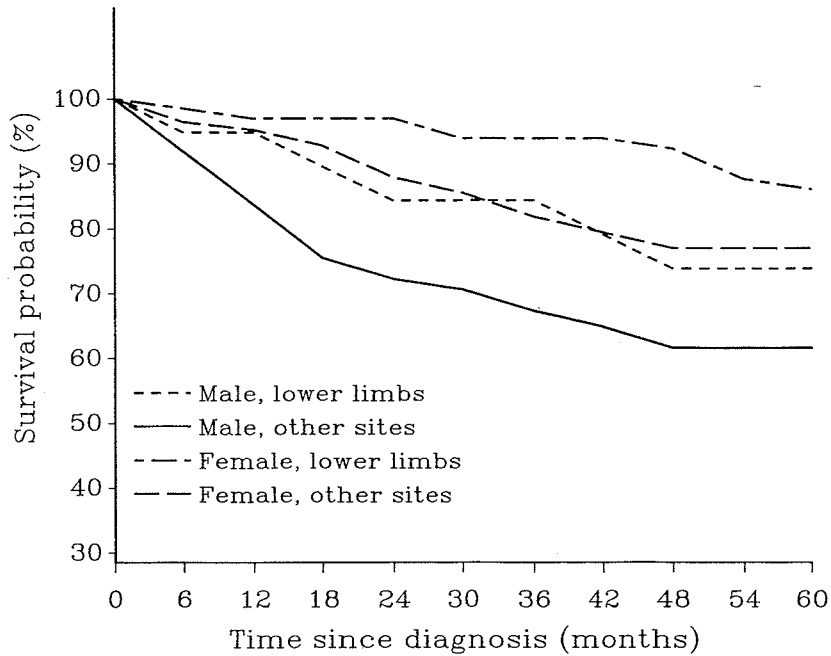


Figure 4.8 Survival of skin melanoma patients by sex and site, Geneva, Switzerland (incident cases, 1970-1982)

are respectively $[0.54 ; 0.94]$ and $[0.28 ; 0.84]$. One might conclude that the role of site can not be definitely excluded, and that the problem requires further research.

Taking account of a prognostic factor in the comparison of survival between groups presumes that there is no interaction between this factor and the group. If, in the example above, survival of subjects with melanoma of the legs was better than that of other sites in men and worse in women, a single test based on summing the differences between the two strata would not have made sense. Only a separate analysis in men and in women would be acceptable. The value of this latter kind of analysis depends however on the number of subjects in each stratum (if the number is small the comparison has less power), the number of categories of the factor to be taken into account (when the number of tests carried out is large, there is a substantial increase in the probability of type one error) and the interpretability of the interaction.

The Cox model

The proportional hazards model

The limits of the stratification method described above are quickly reached as the number of patients under study decreases in relation to the number of strata particularly when the number of covariables or of their levels increase. In this case, there will inevitably be a large number of survival times for which one or more of

the groups will no longer provide subjects at risk of dying, thus making the results from the test increasingly imprecise. In describing all the data by one model, the comparisons to be made are summarized by those between the few parameters which characterize the model itself. The information which was not available in the data is effectively replaced by information provided by the equation defining the distribution of the survival data under study. The benefit is however only real if the model is sufficiently valid.

The classical analytical models (exponential, Weibull; see Chapter 1, page 29) are often too rigid to take into account the diversity of situations which are encountered in practice. Methods with more general applicability, based on principles analogous to those used in linear regression, have therefore been developed. The most frequently used in survival analysis is the proportional hazards model [25], which states that the instantaneous mortality rate depends on covariables through a multiplicative parameter applied to an unspecified background or baseline rate $\lambda_0(t)$:

$$\lambda(t, \mathbf{z}) = \lambda_0(t) e^{\beta \mathbf{z}} \quad (4.61)$$

where \mathbf{z} is a vector of covariables and β a vector of coefficients measuring the intensity of the effect of the components of vector \mathbf{z} . Thus, when \mathbf{z} is the indicator of membership in a subgroup, this model implies that, at each time point t , the mortality rate in the subgroup ($\mathbf{z} = 1$) is proportional to the mortality rate for the rest of the study sample ($\mathbf{z} = 0$):

$$\lambda(t, 1) = \lambda_0(t) e^{\beta} = \lambda(t, 0) e^{\beta} \quad (4.62)$$

where $\theta = e^{\beta}$ is consequently the relative mortality rate of the subgroup with respect to the rest of the sample. θ is also a good approximation of the relative risk of dying at time t when the cumulative mortality rate is small.

As a rule, the ratio of hazard rates for any two values of \mathbf{z} :

$$\frac{\lambda(t, \mathbf{v})}{\lambda(t, \mathbf{u})} = e^{\beta(\mathbf{v} - \mathbf{u})} \quad (4.63)$$

is independent of time, giving rise to the name, *proportional hazards model*. This model is often called semiparametric because the background rate $\lambda_0(t)$, unlike the relative rate, is not specified by a parametric model.

When the factor under study is defined by a qualitative variable with r categories, the variable $\mathbf{z} = (z_2 \dots z_r)$, where the z_i are $r - 1$ indicator variables defining membership of the subgroups corresponding to $r - 1$ categories of the factor, is constructed: for $2 \leq i \leq r$, $z_i = 1$ if the subject falls into the i th category and $z_i = 0$ otherwise. The first category of the factor is characterized by $z_i = 0$, $2 \leq i \leq r$. The choice of the particular role played here by the first category is obviously arbitrary. Thus, if the covariable under study is stage at diagnosis, coded into three classes of increasing severity, two indicator variables are constructed. A subject in stage 1

is coded as $z_2 = 0$ and $z_3 = 0$, a subject in stage 2 as $z_2 = 1$ and $z_3 = 0$ and a subject in stage 3 is defined by $z_2 = 0$ and $z_3 = 1$.

The coefficient β_i corresponding to the indicator variable z_i allows the calculation of θ_i , the relative mortality rate for patients in the category compared to those in the first category. Thus, in the previous example, $\theta_2 = e^{\beta_2}$ is the relative rate of subjects in stage 2 with respect to subjects in stage 1. The relative rates of any two categories can then be derived from the previously described relative rates by simple division, the choice of standard categories being entirely arbitrary. Still using the example of disease stage, θ_3/θ_2 measures the relative rate of the subjects in stage 3 compared to those in stage 2, and similarly for all pairs of categories.

The proportional hazards model has already been discussed in the context of incidence data (Chapter 2). There, it was noted that the ratio of incidence rates in two populations with different cancer risks was frequently independent of age. This variable played the same role as time since diagnosis in the present situation. The constancy of the ratio implied that the age-incidence curves for the two populations were parallel when plotted on a logarithmic scale.

In survival analysis, the hazard rate itself is not usually calculated. Nevertheless, analogous graphical properties can be demonstrated from the estimation of survival probability itself. Since:

$$S(t, \mathbf{z}) = e^{-\Lambda_0(t)} e^{\beta \mathbf{z}} \quad (4.64)$$

we have:

$$\text{Log} [-\text{Log} (S(t, \mathbf{z}))] = \text{Log} [\Lambda_0(t)] + \beta \mathbf{z} \quad (4.65)$$

an equation which shows that the proportional hazards model is also a generalized linear model in which the intercept with the Y-axis is the logarithm of the cumulative background rate at the specified time interval. This relationship also shows that, if the negative of the logarithm of survival probability is plotted on a logarithmic scale for two values \mathbf{u} and \mathbf{v} of \mathbf{z} , the two curves will differ from each other by a translation of value $\beta(\mathbf{v} - \mathbf{u})$. Drawing these two curves in practice helps to assess the validity of the model.

Just as the linear regression method is used in the context of normally-distributed observations to assess the specific effect of a risk factor (see Chapter 3, page 158), fitting the model $\lambda_0(t) e^{\alpha \mathbf{x} + \beta \mathbf{z}}$ to survival data allows the effect of a variable \mathbf{z} to be evaluated after correcting for a confounding variable \mathbf{x} . The regression method is successful where a stratification method would have failed because it assumes that the effect of \mathbf{z} is the same regardless of the value of \mathbf{x} . In contrast, the method of stratification evaluates the effect of \mathbf{z} for all categories of variable \mathbf{x} , which means that a larger number of observations is required for the analysis. When the model is judged to be inadequate, it is obviously possible to construct a model which is slightly modified in each stratum by defining a baseline rate $\lambda_0(t)$ exclusive

to it. Equivalently, interaction terms expressing that the effect of \mathbf{z} is no longer the same for different categories of \mathbf{x} can be introduced into the model. It should however be remembered that overly complicating the model risks sacrificing a large part of its value, as much in its interpretability as in its effectiveness.

Principle of the Cox model estimation

The estimation of the vector β in the model (4.61) has been described in detail in Chapter 1, page 30. In this section, the application of these principles to survival analysis is described for malignant melanoma of the lower limbs, already used to illustrate comparative methods. The covariable under study here is sex, defined by $z = 1$ for female subjects and $z = 0$ for male subjects. The model (4.61) implies that the instantaneous mortality rate for men is $\lambda_0(t)$ and that of women is $\theta\lambda_0(t)$ where $\theta = e^\beta$.

Since under the model, two subjects of the same sex have the same mortality rate, it is clear that there is one chance in two that one will survive longer than the other. Furthermore, since θ is the relative mortality rate for women with respect to men, a man has the probability $1/1+\theta$ of surviving longer than a woman. Similarly, the probability that a woman will survive longer than a man is $\theta/1+\theta$

If at the time point t_i , n_{1i} men and n_{2i} women are still under follow-up and if a subject characterized by the covariable z_i dies, the probability of this death occurring among the $n_{1i} + n_{2i}$ possible deaths is, by extension of the previous concept:

$$p_i(\beta) = \frac{e^{\beta z_i}}{n_{1i} + n_{2i} e^\beta} \quad (4.66)$$

When the mortality rate $\lambda_0(t)$ is not given by a parametric model, the likelihood can only be based on the order of the observations. In this case, it is equal to the probability that the m deaths occur in the observed order and that they had preceded the censored time that followed them. By repeated application of formula (4.66), the likelihood is obtained as:

$$V(\beta) = \prod_{i=1}^m p_i(\beta) \quad (4.67)$$

the estimate $\hat{\beta}$ of β is the value which maximizes the logarithm of $V(\beta)$:

$$L(\beta) = \text{Log}[V(\beta)] = \sum_{i=1}^m [\beta z_i - \log(n_{1i} + n_{2i} e^\beta)] \quad (4.68)$$

To obtain $\hat{\beta}$, the derivative of this function with respect to β is set equal to zero:

$$C(\beta) = \sum_{i=1}^m \left[z_i - \frac{n_{2i} e^{\beta}}{n_{1i} + n_{2i} e^{\beta}} \right] \quad (4.69)$$

known as the *score*. Noting that $\hat{\theta} = e^{\hat{\beta}}$ and $d_{2.} = \sum_{i=1}^m z_i$ is the total number of deaths observed in women, the equation $C(\beta) = 0$ is equivalent to:

$$\sum_{i=1}^m z_i = d_{2.} = \sum_{i=1}^m \frac{n_{2i} \hat{\theta}}{n_{1i} + n_{2i} \hat{\theta}} \quad (4.70)$$

an expression which indicates that the total number of deaths observed in women is consistent with the proportions of female deaths predicted by the model at each observed date of death.

The variance of $\hat{\beta}$ is obtained from the second derivative of $L(\beta)$. Thus, in the context of this example, we have:

$$I(\beta) = -\frac{d^2 L(\beta)}{d\beta^2} = \sum_{i=1}^m \frac{n_{1i} n_{2i} e^{\beta}}{(n_{1i} + n_{2i} e^{\beta})^2} \quad (4.71)$$

Since the likelihood curve is concave around its maximum, its second derivative is negative (see Figure 4.9). The bigger the absolute value of its second derivative, the greater the curvature of $L(\beta)$ and the more precise is the estimate (see Chapter 1,

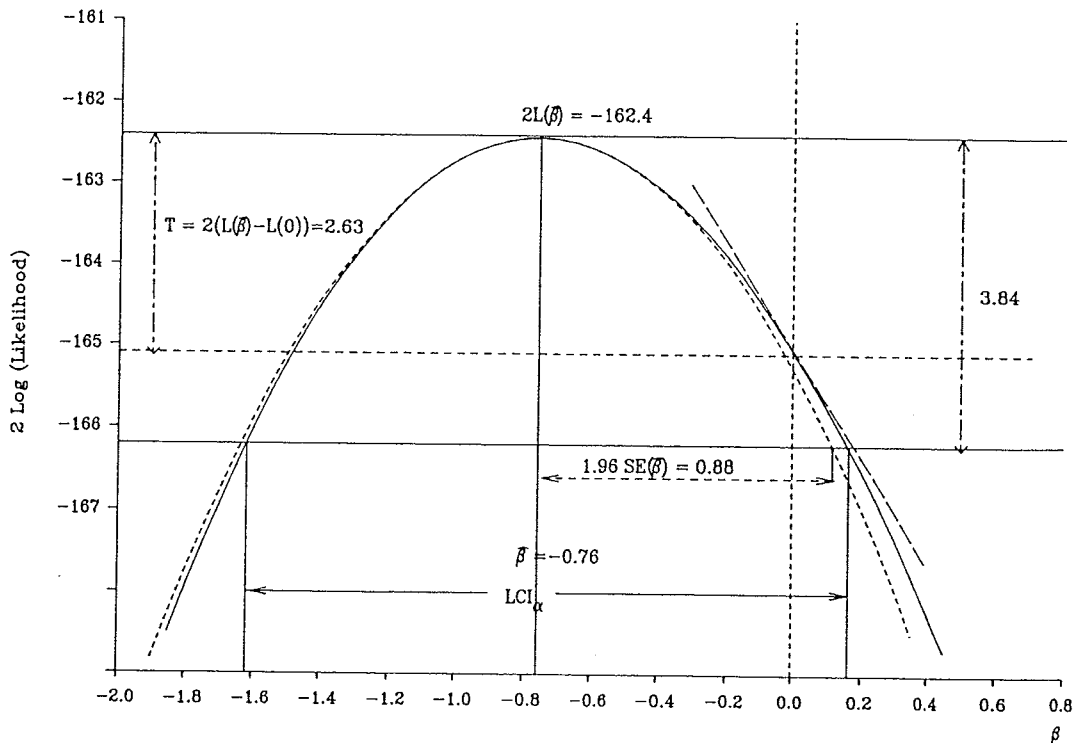


Figure 4.9 Likelihood and associated statistics. Cox model for survival of skin melanoma patients by sex, Geneva, Switzerland (incident cases 1970-1982)

page 17). Similarly, the greater the curvature is, the more $C(\beta)$ is variable in the vicinity of $\hat{\beta}$. It can be shown that the variance of $\hat{\beta}$ is closely approximated by the inverse of $I(\beta)$ and that $I(\beta)$ itself is a good approximation of the variance of $C(\beta)$.

Under the previous reasoning, we suppose that a single death occurs at each time t_i . In practice, this is rarely the case as the measurement of time is never precise enough to allow all deaths to be ordered. This situation is particularly true for data from registries when only the month of death is recorded. When D_i deaths have been observed at time t_i , the available information on survival time could have been generated by the observations defined by all possible orders of the identical survival times. There are therefore $D_i!$ possible configurations. It can be seen that a calculation based on the above principle becomes practically impossible as soon as D_i increases in size past a few units. In practice, the probability $p_i(\beta)$ is approximated by the quantity:

$$p_i^*(\beta) = \frac{\theta^{d_{2i}}}{(n_{1i} + n_{2i}\theta)^{D_i}} \quad (4.72)$$

which is proportional to the probability that d_{1i} male deaths and d_{2i} female deaths are observed out of a total of D_i deaths observed at time point t_i . The equations (4.68), (4.69), (4.70) and (4.71) in our example thus become:

$$L(\beta) = \sum_{i=1}^m [\beta d_{2i} - D_i \log(n_{1i} + n_{2i}e^\beta)] \quad (4.73)$$

$$C(\beta) = \sum_{i=1}^m \left[d_{2i} - D_i \left(\frac{n_{2i} e^\beta}{n_{1i} + n_{2i} e^\beta} \right) \right] \quad (4.74)$$

$$d_{2.} = \sum_{i=1}^m \left[D_i \left(\frac{n_{2i} \hat{\theta}}{n_{1i} + n_{2i} \hat{\theta}} \right) \right] \quad (4.75)$$

$$I(\beta) = \sum_{i=1}^m \left[D_i \frac{n_{1i} n_{2i} e^\beta}{(n_{1i} + n_{2i} e^\beta)^2} \right] \quad (4.76)$$

The likelihood suggested by Cox is defined for each survival time by the probability that the D_i deaths have the observed configuration at this date:

$$p_i^c(\beta) = \frac{\theta^{d_{2i}}}{\sum_{\ell \in C} \theta^{k(\ell)}} \quad (4.77)$$

where C is the totality of choices of d_{2i} deaths among the $n_{1i} + n_{2i}$ subjects present in the study at t_i , and $k(\ell)$ is the number of female subjects in the combination ℓ . The expression (4.72) is an approximation to (4.77), and is used in most computer software for survival analyses.

Table 4.15 Likelihood of the Cox model; prognostic value of sex on survival of subjects with melanoma of the legs in Geneva, Switzerland (Incident cases 1970-1982)

Date	Males		Females		Contribution to the likelihood ^(a)		
	death	at risk	death	at risk	$p_i^*(\beta)$	$\text{Log } p_i^*(0)$	$\text{Log } p_i^*(\hat{\beta})$
2	0	19	1	64	$\theta/(19+64\theta)$	- 4.419	- 4.650
4	1	19	0	63	$1/(19+63\theta)$	- 4.408	- 3.881
12	0	18	1	63	$\theta/(18+63\theta)$	- 4.394	- 4.620
18	1	18	0	62	$1/(18+62\theta)$	- 4.382	- 3.850
23	1	17	0	62	$1/(17+62\theta)$	- 4.369	- 3.829
28	0	16	2	62	$\theta^2/(16+62\theta)^2$	- 8.713	- 9.133
38	1	16	0	60	$1/(16+60\theta)$	- 4.331	- 3.786
43	0	15	1	60	$\theta/(15+60\theta)$	- 4.317	- 4.523
44	1	15	0	59	$1/(15+59\theta)$	- 4.304	- 3.752
52	0	14	1	59	$\theta/(14+59\theta)$	- 4.290	- 4.488
54	0	14	2	58	$\theta^2/(14+58\theta)^2$	- 8.553	- 8.953
55	0	14	1	56	$\theta/(14+56\theta)$	- 4.248	- 4.454
65	1	9	0	28	$1/(9+28\theta)$	- 3.611	- 3.095
66	0	7	1	26	$\theta/(7+26\theta)$	- 3.497	- 3.713
68	0	7	1	22	$\theta/(7+22\theta)$	- 3.367	- 3.610
81	0	6	1	19	$\theta/(6+19\theta)$	- 3.219	- 3.460
102	0	6	1	18	$\theta/(6+18\theta)$	- 3.178	- 3.428
110	1	6	0	17	$1/(6+17\theta)$	- 3.135	- 2.636
155	1	2	0	4	$1/(2+4\theta)$	- 1.792	- 1.353

$$2L(0) = -165.1 \quad 2L(\hat{\beta}) = -162.4$$

^(a) See (4.67), (4.68), (4.72); $\theta = e^\beta$; $\hat{\beta} = 0.76$

Table 4.15 gives dates of observed death, number of subjects under follow-up at each date and the contribution of each death to the likelihood. The approximation (4.72) is used in case of identical survival times. Figure 4.9 shows the function $2L(\beta)$ in the vicinity of $\hat{\beta}$ which, in this case, is -0.76 ($\hat{\theta} = 0.47$). Here the test of the hypothesis $\beta = 0$ has as its objective the comparison of the survival between male and female subjects, and could be carried out by the logrank test. In the context of the maximum likelihood method, there are several classical tests available [33] (see Chapter 1); it is of note that one of them will again lead to the logrank test.

The likelihood ratio test

$$T_1 = 2[L(\hat{\beta}) - L(0)] \tag{4.78}$$

has a χ^2 distribution under the null hypothesis. Its number of degrees of freedom is equal to the dimension of β . In the example given, the number of degrees of freedom is equal to unity and $T_1 = 2.63$ ($p = 0.11$).

Furthermore, $2[L(\hat{\beta}) - L(\beta_0)]$ has this distribution if $\beta = \beta_0$. We define around $\hat{\beta}$ all the values of β which, given the observations, are not rejected by the test T_1 , that is, those for which

$$2[L(\hat{\beta}) - L(\beta)] < Z_{\alpha/2}^2 \quad (4.79)$$

This interval is classically known as the likelihood-based confidence interval, and noted as LCl_α in Figure 4.9.

Wald's test is based on the variance of $\hat{\beta}$, obtained from the matrix of the second derivatives of L . In the example where β is of dimension 1,

$$T_2 = \frac{\hat{\beta}^2}{\text{Var}(\hat{\beta})} \quad (4.80)$$

is a χ^2 variable on one degree of freedom. Here, the standard error associated with the estimate $\hat{\beta}$ is calculated from $1/I(\hat{\beta})$ and is 0.451, therefore $T_2 = (-0.76)^2/(0.451)^2 = 2.84$. indicating that we can not exclude the possibility that β is null.

The variance of $\hat{\beta}$, allows a confidence interval for β to be constructed using the usual expression:

$$Cl_\alpha = \hat{\beta} \pm \left[Z_{\alpha/2} \sqrt{\text{Var}(\hat{\beta})} \right]$$

This interval is derived in the same way as the interval LCI except that $2[L(\hat{\beta}) - L(0)]$ is replaced by its quadratic approximation (see Figure 4.9 and Chapter 1, page 17).

The Score test is based on the vector of first derivatives of the likelihood which, in the example, reduces to the function $C(\beta)$. The statistic

$$T_3 = \frac{C(0)^2}{I(0)} \quad (4.81)$$

assess the magnitude of the slope of the tangent to the likelihood curve $2L(\beta)$ at the value of β being tested (here 0). If it is not possible to reject the hypothesis that this line is horizontal, then it should be concluded that 0 is acceptable as a value of β . In the example, $T_3 = 2.98$. Since, from formula (4.75)

$$C(0) = d_2 - \sum_{i=1}^m D_i \frac{n_{2i}}{n_{1i} + n_{2i}}$$

it can be seen that the score test is based on the same statistic as the logrank. The variance $I(0)$ is exactly that obtained in (4.43) if there are no ties ($D_i = 1$ for each t_i). It is slightly different if there are ties (since each term of the sum defining it is multiplied by $(N_i - 1)/(N_i - D_i)$). Using $p_i^c(\beta)$ (4.77) instead of the approximation $p_i^*(\beta)$ to calculate the likelihood would give exactly the same variance as (4.43) and the score test will thus be identical to the logrank test in all cases.

The previous example illustrated the use of the Cox model to estimate the relative rate for a single dichotomous factor. The principles applicable when more

than one categorical or continuous factor is studied (the case in fact for which the model was developed) have been described in Chapter 1. These principles are exactly the same: only the expression of the model is more complicated. In the construction of these models, the mortality rate $\lambda_0(t)$ has been considered to be a nuisance parameter. However, in practice, it is often desirable to estimate the survival curve corresponding to this background rate. The principle of the Kaplan-Meier method (see formula 1.22) is used here, but the number of deaths at time point t_i is related to a modified denominator to take account of the fact that not all individuals under follow-up at time point t_i are subject to the same risk of death. The estimate of $S_0(t)$ is therefore given by the formula:

$$\hat{S}_0(t) = \prod_{t_i \leq t} \left(1 - \frac{D_i}{\sum_{j \in R_i} \hat{\theta}_j} \right) \quad (4.82)$$

in which each individual in group R_i of subjects at risk of death at time t_i is no longer counted as one but as a value equal to its relative rate.

Example of an adjustment using Cox's model

Survival data based on hospital records are often considered to be a biased representation of the overall situation. In particular, survival rates calculated from these data may be better than those obtained from the ensemble of incident cases. One explanation for this phenomenon is that cases not treated in hospital are often those that are very advanced or that have not been treated at all. Hospital statistics would therefore represent a selection of favourable cases, for which survival is overestimated in comparison to the general population.

The Geneva cancer registry routinely collects the place of treatment of all incident cases (public university hospital or other health care centre). The survival of patients treated in these two health centres can thus be compared. Breast cancer data are given as an example [34]. The 1105 women diagnosed between 1975 and 1982, were allocated to two groups depending on their place of treatment. The two groups were then compared using a Cox model with one covariable z , indicating group membership ($z=1$ corresponds to the group treated in a university hospital):

$$\lambda(t) = \lambda_0(t) e^{\beta z} \quad (4.83)$$

The hypothesis of equal death rates in the two groups ($\beta = 0$) is evaluated by the score test (4.81) which, in this case, corresponds to the logrank test. The value obtained ($T_3 = 29.9$) corresponding to a χ^2 on one degree of freedom is very significant ($p < 10^{-5}$). The likelihood ratio test (4.78) obviously confirms this result ($T_1 = 33.2$).

The estimate of the relative rate obtained by maximizing the Cox likelihood is:

$$\hat{\theta} = e^{\hat{\beta}} = 2.04$$

indicating that the mortality rate is about twice as high for patients of the public university hospital as for those of other health care centres.

Since there is no reason to suspect that health care is inferior in university hospitals, it is important to take into account factors which might lead to the selection of cases with a poorer prognosis into public health care institutions. One possibility may be that the public hospital receives advanced cases which require complicated treatment and have poorer prognoses. There may also be selection by social class associated with the higher financial contribution required of patients in the private sector. Clinical stage, social class and age have therefore been examined in order to assess this assumption. It is known that these three variables have a prognostic role and Table 4.16 shows that they have a different distribution in the two health care systems. Consequently, they should be taken into account in the comparison (see page 255) otherwise all or part of their effect may be attributed to the type of hospital health care system.

Clinical stage, which has a fundamental importance in this context, is used to illustrate the model-fitting for a qualitative variable. This variable has four categories, with an associated vector of indicator variables $\mathbf{x} = (x_2, x_3, x_4)$ defined by:

- Localized only $(x_2 = 0, x_3 = 0, x_4 = 0)$
- Regional involvement $(x_2 = 1, x_3 = 0, x_4 = 0)$
- Metastases $(x_2 = 0, x_3 = 1, x_4 = 0)$
- Unknown $(x_2 = 0, x_3 = 0, x_4 = 1)$

These three variables are then included in the model which is written:

$$\lambda(t, \mathbf{x}, z) = \lambda_0(t) e^{(\alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \beta z)} \quad (4.84)$$

the variable z still being the variable of interest indicating the health care system as defined.

Table 4.16 : Distribution (%) of prognostic variables available by health care system for breast cancer in Geneva, Switzerland (Incident cases 1975-1982)

Age (years)	< 50	50-64	65-74	75+
University hospital (N=808)	25	25	25	25
Other (N=297)	33	31	18	18
Stage	Local	Regional	Metastasis	Unknown
University hospital (N=808)	47	39	10	4
Other (N=297)	60	27	1	12
Social class	Manual workers	Office workers	Executives	Unknown
University hospital (N=808)	29	35	13	23
Other (N=297)	20	39	19	22

The estimate $\hat{\theta} = e^{\hat{\beta}}$ which maximizes the likelihood of observations for this model is the relative mortality rate 'adjusted for stage' for the university hospitals. Its value ($\hat{\theta} = 1.91$) is only slightly less than the previous value, showing a small confounding effect due to the fact that the stage at diagnosis is on average more advanced in public hospitals.

The test of the hypothesis $\beta = 0$ from the likelihood ratio gives a value $T_1 = 22.8$. By the score test, the value obtained is $T_3 = 21.0$. These values are very significant in both cases, indicating that the effect of hospital type remains after taking stage into account.

The logrank test with stratification described earlier (see page 255) could have been used to solve this problem. In fact, it would have provided the same results as a Cox model with stratification specifying:

$$\lambda_{\mathbf{x}}(t) = \lambda_{\mathbf{x}0}(t) e^{(\beta z)} \quad (4.85)$$

This equation implies, as does (4.84), that there is the same relative rate between public hospitals and other health care systems for each stage \mathbf{x} . However, it allows for the possibility that the baseline hazard rates might differ for each stage. The maximum likelihood here is the product of the likelihoods of the proportional hazards model written separately for each stage. To obtain the score test for the hypothesis $\beta = 0$, scores calculated for each stage are first summed and the variance of this sum is calculated by addition of the within-stage information related to β . The test is then the ratio of the two sums and is identical, apart from the way in which ties are dealt with, to the logrank test with adjustment for the stage given page... Its value ($T_3 = 19.4$) is very close to that obtained above, ($T_3 = 21.0$), and the estimate of β ($\hat{\theta} = 1.85$) is little different from that derived from the model (4.84), that is, $\hat{\theta} = 1.91$. This result confirms that the stage at diagnosis is responsible for only a small part of the difference observed between the two types of hospitals.

The adjustment for all other available variables is summarized in Table 4.17. Although social class and age have some prognostic value, they do not qualitatively change the previous conclusion since the estimate of β , after adjusting for these variables, still corresponds to a relative rate of 1.56 which is significantly different from unity.

Figure 4.10 gives survival observed in the two groups. Survival predicted by the model for each of the hospital types is also given. The predicted survival for other health care system is given for a group of patients with an identical distribution of stage, age and social class to that of the patients treated in the public university hospitals. The difference between the curve predicted and observed for patients outside university hospitals is an estimate of the portion of the improvement in survival which can be explained by the most favourable distribution of the recorded prognostic factors in patients treated outside the public sector.

This result suggests that factors other than stage, age and social class as recorded by the registry may lead patients with worse prognoses to be treated in university hospitals. The presence of a large number of cases of unknown stage among patients treated in other health care systems could distort the comparison,

Table 4.17 Cox's model for breast cancer survival data from Geneva, Switzerland (Incident cases 1975-1982)

	$\hat{\beta}^{(a)}$	$SE(\hat{\beta})$	$\hat{\theta}^{(b)}$	Score test ^(c)
Health care system				
University hospital	0.445	0.14	1.56*	29.9
Other	0	—	1.00	
Stage				
Local involvement	0	—	1.00	21.0
Regional involvement	0.801	0.12	2.23*	
Metastases	2.482	0.16	11.97*	
Unknown	1.399	0.19	4.05*	
Age (years)				
< 50	0	—	1.00	11.6
50-64	0.247	0.16	1.28	
65-74	0.440	0.16	1.55*	
75+	1.424	0.16	4.16*	
Social class				
Manual workers	0	—	1.00	10.0
Office workers	- 0.135	0.13	0.87	
Executives	- 0.001	0.18	1.00	
Unknown	- 0.425	0.14	0.65*	

(^a) Estimate obtained when all variables (stage, age and class) are included in the model.
 (^b) The asterisk indicates variables for which the 95% confidence interval excludes 1 ($\hat{\beta}/SE(\hat{\beta}) \geq 1.96$).
 (^c) Test of effect of the health care system after inclusion of the corresponding variable in the model. The first model is the crude test of effect equivalent to the logrank test without adjustment.

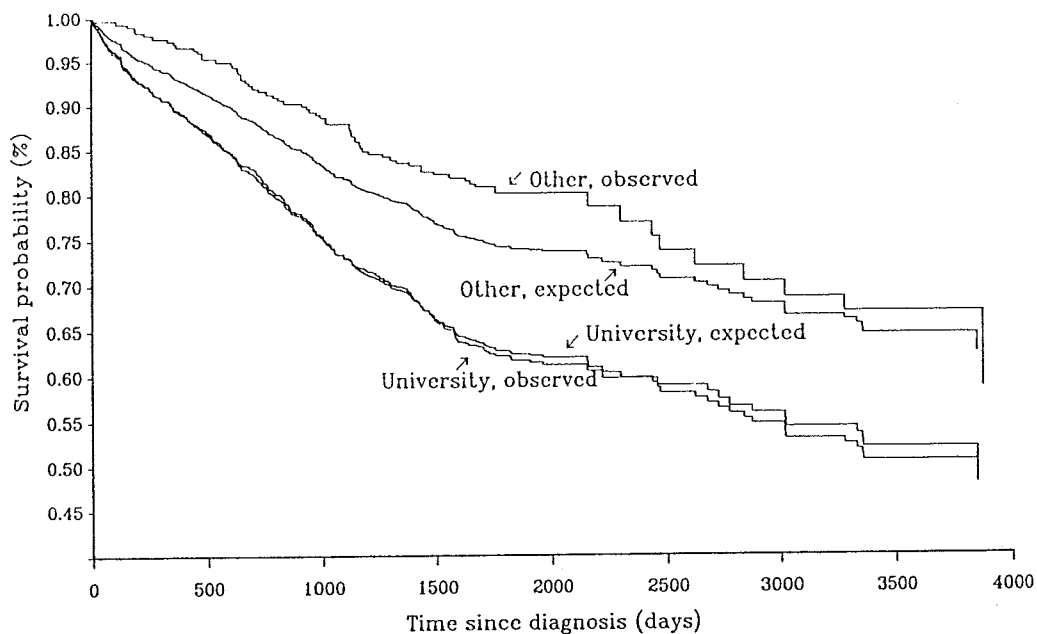


Figure 4.10 Observed and adjusted survival of breast cancer patients by type of hospital – Geneva – Cases diagnosed between 1978 and 1982

as could cases of advanced clinical stages which are almost all treated in university hospitals. It is therefore worth repeating the analysis, limiting the comparison to patients whose tumours were at an early stage, either being localized or only showing regional involvement. The results are not modified substantially although the estimate of $\hat{\theta} = 1.54$ is decreased as is the score test ($T_2 = 7.57$), showing that part of the difference observed for unknown stage was incorrectly attributed to an effect of the health care system.

These results do not correspond to the commonly held notion of better survival in leading hospitals; effects of selection other than those which have been controlled for are probably responsible for the findings. Furthermore, it is known that adjustment by class can leave residual differences with respect to confounding factors because of the phenomenon of within group variation. Classification into three clinical stages only partially accounts for the characteristics of the disease. Fitting stage categories more finely, even though impossible due to the lack of the necessary information, would have improved the results.

Phenomena of this kind have been described by Feinstein [35] to illustrate the paradox attributed to Will Rogers. Convinced of the mediocre intellectual level in California even among higher socio-economic classes, Rogers noted that the average cultural level in California improved even when underprivileged people from Oklahoma emigrated there. This emigration obviously caused a similar phenomenon in Oklahoma. A similar situation occurs with the retrospective reclassification of clinical diagnoses of cancer [35]. Reclassification effectively leads to the worst cases of localized cancer being categorized as regional stage, although these cases have on average a better prognosis than those of the category in which they are now classified. The result is a purely fictional improvement in the survival rate in both categories.

Comparison of net survival

Differences between the survival distributions in several groups are often difficult to interpret because the overall mortality observed is partially due to competing causes of death. These causes, however, are not the object of the comparison and the extent of their effect can vary from one group to another. These difficulties can be very pronounced when a comparison is to be made over a long time period, in particular when the groups being compared do not have the same age distribution or life expectancy. Comparison is then carried out using net survival.

The methods discussed in the previous sections for comparing survival between two or more groups apply without change if the net survival is estimated by the cause-specific survival method (see page 230). In effect, dates of death not caused by the cancer under study are considered as censored observations. However, as noted above, it is not generally advisable to use this method because of inaccuracies in the classification of the causes of death. Other methods such as relative survival should thus be considered to estimate net survival. Unfortunately, the analysis is

then not as simple and should be adapted from principles used above to carry out the comparison of crude survival.

The comparison of two relative survival rates at a given time point (e.g., five years) can be easily carried out by calculating their variances, which are in turn obtained directly from the crude survival analyses. A test analogous to that given in formula (4.42) can then be easily carried out. This approach is however subject to the same criticism as for crude survival.

Note that the use of relative survival does not remove the need to account for age in the comparisons. In fact, as we have seen before (Table 4.10) and as we will confirm below, net survival, like crude survival, is often dependent on age. Relative survival should therefore also be adjusted for the effect of age before any interpretation of observed differences is made. Myers and Hankey [36] and Hakulinen [37] have resolved this problem by calculating relative survival in subgroups which are homogeneous for the main determinants of net survival and by calculating a relative survival rate for the whole group using the method of direct standardisation (see Chapter 2, page 56). Myers and Hankey compared relative survival for different cancer sites between Blacks and Whites after adjusting for age and stage. Hakulinen compared relative survival between Finland and Norway after adjusting for age. However, this method accounts rather imperfectly for the effect of age and depends to some extent on the choice of the standard population. Furthermore, as for crude survival, comparison of survival at specified time points is rarely the only problem of interest. Consequently, methods have been proposed for comparing entire net survival curves and adjusting for confounding variables. We will only describe here the principles underlying these methods and suggest further reading for the reader who wishes to apply them.

One possibility is to adopt a method analogous to the Mantel-Haenszel test, by considering the distribution of deaths in the groups for each interval appearing in the actuarial survival curve (Table 4.11). Brown [38] and Hakulinen [39,40] have constructed tests of this kind, based on the maximum likelihood method. The distribution of the number of deaths d_{ki} for time interval i among the n_{ki} subjects in group k is a binomial distribution:

$$d_{ki} \rightsquigarrow \text{binom}(n_{ki}, 1 - r_{ki} s_{ki}^*)$$

where r_{ki} is the relative survival of the n_{ki} individuals in group k who were living at the beginning of the interval and s_{ki}^* is their expected survival, calculated from the life table, and taking into account the competing forces of mortality. The test of the hypothesis that the r_{ki} are equal is however more complicated than the Mantel-Haenszel test: when n_{ki} and the number of deaths are fixed, d_{ki} has a noncentral hypergeometric distribution with variance and covariance depending on r_{ki} . Although identical in all groups for each interval i under the null hypothesis, the r_{ki} nevertheless remain unknown, and in order to carry out the test, they must be estimated.

Moreover, this method makes implicitly the unrealistic assumption that all individuals at risk of death at the beginning of the interval have the same expected survival s_{ki}^* . Its application would in principle require stratification of the variables

influencing survival so that, for each level j of the stratification variable, the n_{kji} individuals at risk at the beginning of the interval would have the same expected survival s_{kji}^* . Stratification of this kind is always possible for age and sex, for which the existing life tables allow the calculation of expected survival probability s_{kji}^* .

When groups being compared have the same mortality hazard for other causes, this stratification will make expected survival identical for individuals in the different groups being compared within each stratum thus created. The logrank test with stratification then applies without change since, under the null hypothesis, the probabilities of dying are identical. In this special case, stratification by age avoids the need to adjust for competing causes. Adjusting for age using Cox's model is also possible. However, if age classes are included directly in the model equation, the effect of age is implicitly considered as being multiplicative; since a large part of this effect is created by the additive effect of competing causes of death (see formulae 4.29 and 4.39) it might be advisable to fit an age stratified Cox model.

When mortality for other causes cannot be considered equal for the two groups, Brown's and Hakulinen's methods can be used, bearing in mind that the variance of d_{ki} differs from that specified by the binomial distribution. It is relatively easy to take this problem into account in the context of a generalized linear model such as that proposed by Hakulinen [39,40]. However, it seems more efficient to derive the likelihood from individual observations and not from data grouped by interval. This approach has been used by Pocock and coworkers [41], Buckley [42] and in the model presented in the first part of this chapter. Pocock assumes that the rate added to the baseline rate of the general population by the presence of the disease decreases exponentially with time. In contrast, Buckley considers the case where the distribution of net survival is exponential. The extension of his analysis to a proportional hazards model with baseline rate described by a step function is discussed by Esteve and co-workers [20] and summarized by the formulae (4.38), (4.39) and (4.40). This approach leads to a classical test based on the likelihood ratio principle analogous to that described for the Cox model. The strategy discussed for this latter model can be applied without change (see p. 268).

As an example, the method has been applied to colon cancer survival data from Geneva (page 244). Table 4.18 presents the test of the effect of sex on survival in three different contexts. In the first, the comparison is made ignoring mortality from other causes; in the second, the method described above is used and, in the third, cause-specific survival probabilities are compared. Sex has no significant effect regardless of the method used. Nevertheless, it is worth noting that the effect of sex is corrected in the right sense when competing causes are taken into account. The effect of age on net survival can also be evaluated by the likelihood ratio test. The logarithm of the likelihood of model (4.38) with $\mathbf{B}=0$ is calculated to be -1072.25 , giving a χ^2 on two degrees of freedom equal to:

$$2[-1040.02 - (-1072.25)] = 64.46$$

Although the effect of age appears to be weaker for net survival, it still remains very high. The effect of age would have been incorrectly described if it had been

Table 4.18 The influence of age and sex on colon cancer survival in Geneva, Switzerland. (Incident cases 1970-1979)

Context	Estimate of β (Standard error)			Logarithm of the likelihood
	Sex males ^(a)	Age		
		65-74	75 +	
1) Crude survival^(b)	0.101 (0.08)	0.621 (0.11)	0.986 (0.10)	- 4006.28
sex excluded	-	0.616 (0.11)	0.970 (0.10)	- 4007.07
2) Net survival^(c)	0.048 (0.09)	0.593 (0.12)	0.916 (0.12)	- 1039.89
sex excluded	-	0.591 (0.12)	0.909 (0.12)	- 1040.02
3) Cause-specific survival^(d)	0.005 (0.09)	0.613 (0.12)	0.934 (0.11)	- 3364.78
sex excluded	-	0.612 (0.12)	0.934 (0.11)	- 3364.78

(^a) In each context, the second model contains age only in order to allow for the likelihood ratio test to be carried out.

(^b) Proportional hazard models fitted by Cox's method with three months as the time unit.

(^c) Proportional hazard model with a background rate constant for three-month intervals up to three years from diagnosis, six-month intervals up to four years and then an interval of one year and an interval of five years.

(^d) The same models as for (b), but each death not recorded as due to cancer of the colon is censored at the date of death.

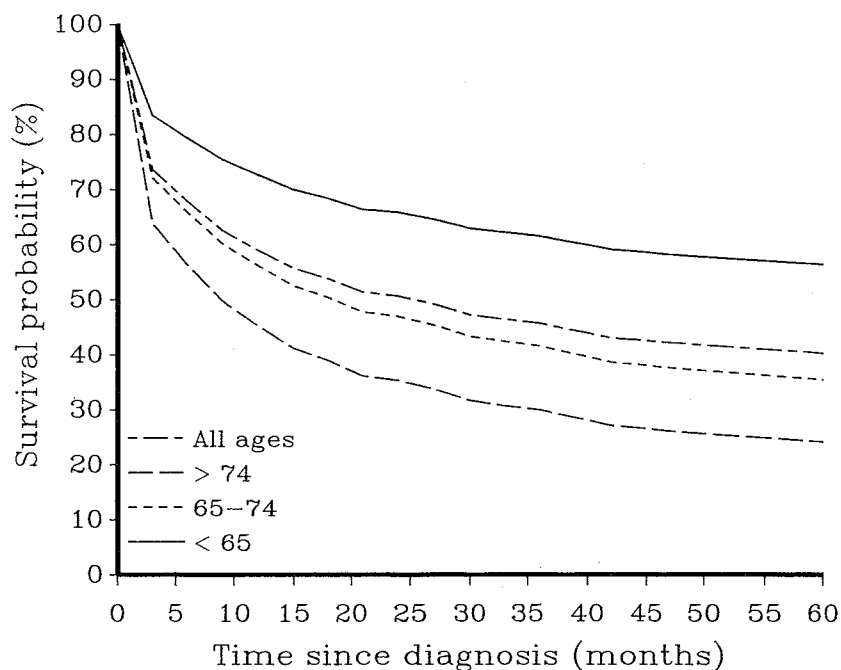


Figure 4.11 Relative survival of colon cancer cases diagnosed between 1970 and 1979 in Geneva, Switzerland, estimated by the method of maximum likelihood

studied for relative survival at five years and it would have been almost unnoticed using relative survival at ten years (Table 4.10). Figure 4.11 shows the net survival of the three age groups and of all groups calculated by the maximum likelihood method.

Bibliographical Notes

Chiang's now classic work [43] dedicated to stochastic processes in the field of human biology, provides a complete, systematic and integrated study of the application of probability theory to the concept of risk and survival which are essential in epidemiology. Although this work is usually reserved for statisticians, it contains many chapters which are relatively accessible mathematically. These sections mainly involve analysis of the life table and the mechanics of its construction, and the problem of competing risks. Chiang has written a manual for the World Health Organization on the life table which is more accessible to epidemiologists [44].

More detailed concepts and methods underlying survival analysis can be found in the more recent, fundamental work of Kalbfleisch and Prentice [26]. Although the concepts and methods are mainly described for mathematically-minded readers, the accompanying examples assist the non-statistician. Of particular note are the sections on parametric models (Weibull, exponential, log-normal, gamma) and the problems presented by the adequacy of the models fitted. Cox's proportional hazards model and its application are the subject of a substantial section in the book, especially with regard to time-dependent covariates. A review of the theory of competing risks and methods for the retrospective analysis of survival can also be found in this text.

The manual on survival analysis by Hill and co-workers is among the more practical texts for non-statisticians. Although it addresses clinical trials rather than epidemiology, this text is nevertheless useful for its detailed review of descriptive and comparative methods of survival study. Problems raised by tests of comparison have an important place in the book; an annex is entirely devoted to the principles underlying rank tests and the choice of the appropriate test. Parametric and semi-parametric models, the evaluation of goodness-of-fit, coding of variables, hierarchical tests and trend tests are treated with equal clarity.

The description of the life table in Pressat's text offers a more demographic view point [45]. Of historical interest, Berkson first proposed the concept of relative survival [19] and the classic article by Elveback presents somewhat didactically the principles behind the actuarial method [46]. The scientific publication on cancer registration published by the International Agency for Research on Cancer (IARC) considers the calculation of survival of incident cases [47].

Cancer registries can calculate survival using the computer program package developed by the Finnish Cancer Registry [48]. These programs enable users to estimate relative and observed survival curves and carry out comparative tests.

Clinicians will benefit from the article by Christensen describing the concepts underlying the Cox model in the context of clinical trials [49]. This article provides practical advice and many examples on how to conduct the analyses and interpret the results.

REFERENCES

- [1] DENOIX P, AMIEL JL, BRULÉ G, et al. *La maladie cancéreuse*. Paris, J.-B. Baillière, 1968, Chap. V
- [2] WALTER SD, STITT LW. Evaluating the survival of cancer cases detected by screening. *Stat Med* 1987 **6** : 885-900
- [3] KAPLAN EL, MEIER P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958 **53** : 457-64
- [4] GLASS DV. Graunt's life table. *Population (Paris)* 1937 **2** : 71-2
- [5] PETO R, PIKE MC, ARMITAGE P et al. Organisation et analyse des essais thérapeutiques comparatifs comportant une longue surveillance des malades. *Rev Épidémiol Santé Publ* 1979 **273** : 171-257
- [6] GREENWOOD M. *A report on the natural duration of cancer*. London, Ministry of Health, His Majesty's Stationery Office, 1926 (Reports on public health and medical subjects, n° 33)
- [7] RAO CR. *Linear statistical inference and its applications* (2nd ed.). New York, John Wiley, 1973 pp. 385-8.
- [8] BRESLOW N. Statistical methods for censored survival data. *Environ Health Perspect* 1979 **32** : 181-192
- [9] ROTHMAN KJ. Estimation of confidence limits for the cumulative probability of survival in life table analysis. *J Chron Dis* 1978 **31** : 557-60
- [10] ANDERSON JR, BERNSTEIN L, PIKE MC. Approximate confidence intervals for probabilities of survival and quantiles in life-table analysis. *Biometrics* 1982 **38** : 407-16
- [11] EMERSON J. Nonparametric confidence intervals for the median in the presence of right censoring. *Biometrics* 1982 **38** : 17-27
- [12] BROOKMEYER R, CROWLEY J. A confidence interval for the median survival time. *Biometrics* 1982 **38** : 29-41
- [13] EDERER F, AXTELL LM, CUTLER SJ. The relative survival rate : a statistical methodology. *Nat Cancer Inst Monogr* 1961 **6** : 101-120
- [14] HAKULINEN T. On long-term relative survival rates. *J Chron Dis* 1977 **30** : 431-43
- [15] REED LJ, MERREL M. A short method for constructing an abridged life table. *Am J Hyg* 1939 **30** : 33-62
- [16] HAKULINEN T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982 **38** : 933-942
- [17] EDERER F, HEISE H. *The effect of eliminating deaths from cancer on general population survival rates*. Methodological note n° 11, End Results Evaluation Section. National Cancer Institute, Aug. 1959

- [18] ROTHMAN KJ, BOICE JD. *Epidemiologic analysis with a programmable calculator*. Washington, US Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health, 1979
- [19] BERKSON J. The calculation of survival rates. In : W Walters, HK Gray, JT Priestley, (eds.): *Carcinoma and other malignant lesions of the stomach*. Philadelphia, Saunders, 1942 pp. 467-84
- [20] ESTÈVE J, BENHAMOU E, CROASDALE M, RAYMOND L. Relative survival and the estimation of net survival : elements for further discussion. *Stat Med* 1990 **9** :529-38
- [21] BRESLOW NE. Analysis of survival data under the proportional hazards model. *Int Stat Rev* 1975 **43** :45-58
- [22] ANDERSEN PK, BORCH-JOHNSEN K, DECKERT T et al. A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics* 1985 **41** : 921-32
- [23] HILL C, LAPLANCHE A, REZVANI A. Comparison of the mortality of a cohort with the mortality of a reference population in a prognostic study. *Stat Med* 1985 **4** : 295-302
- [24] POCOCK SJ, GORE SM, KERR GR. Long term survival analysis : the curability of breast cancer. *Stat Med* 1982 **1** : 93-104
- [25] COX DR. Regression models and life tables. *J R Statist Soc B* 1972 **34** : 187-220
- [26] KALBFLEISCH JD, PRENTICE RL. *The statistical analysis of failure time data*. New York, Wiley, 1980
- [27] In HILL C, COM-NOUGUÉ C, KRAMAR A et al. *Analyse statistique des données de survie*. Paris, Médecine-Sciences Flammarion, 1990 pp. 159-168
- [28] PETO R, PETO J. Asymptotically efficient rank invariant test procedures with discussion. *J R Statist Soc A* 1972 **135** : 185-207
- [29] GEHAN EA. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 1965 **52** : 203-23
- [30] BRESLOW NE. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* 1970 **57** : 579-94
- [31] PETO R, PIKE MC. Conservatism of the approximation $(0-E)^2/E$ in the logrank test for survival data or tumor incidence data. *Biometrics* 1973 **29** : 579-84
- [32] BRESLOW NE, CROWLEY J. A large sample study of the life table and product limit estimates under random censorship. *Ann Stat* 1974 **2** : 437-53
- [33] In C HILL, C COM-NOUGUÉ, A KRAMAR et al. *Analyse statistique des données de survie*. Paris, Médecine-Sciences Flammarion, 1990 pp. 68-73
- [34] RAYMOND L, GASSER A, LEVI F, FIORETTA G. Taux de survie hospitaliers et épidémiologiques : résultats paradoxaux enregistrés à Genève. *Schweiz Krebs-Bull* 1988 **1** : 11-14
- [35] FEINSTEIN AR, SOSIN DM, WELLS CK. The Will Rogers phenomenon : stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* 1985 **312** : 1604-8
- [36] MYERS MH, HAWKEY BF. *Cancer patient survival experience*. Bethesda, US Departement of Health and Human Services, 1980 NIH Publication 80-2148
- [37] HAKULINEN T. A comparison of nationwide cancer survival statistics in Finland and Norway. *World Health Stat Q* 1983 **36** : 35-46
- [38] BROWN CC. The statistical comparison of relative survival rates. *Biometrics* 1983 **39** : 941-8
- [39] HAKULINEN T, TENKANEN L, ABEYWICKRAMA K. Testing equality of relative survival patterns based on aggregated data. *Biometrics* 1987 **43** : 313-25
- [40] HAKULINEN T, TENKANEN L. Regression analysis of relative survival rates. *Appl Stat* 1987 **36** : 309-17

- [41] POCOCK, SJ, GORE SM, KERR GR. Long term survival analysis : the curability of breast cancer. *Stat Med* 1982 **1** : 93-104
- [42] BUCKLEY JD. Additive and multiplicative models for relative survival rates. *Biometrics* 1984 **40** : 51-62
- [43] CHIN LONG CHIANG. *Introduction to stochastic processes in biostatistics*. New York, Wiley, 1968
- [44] CHIN LONG CHIANG. *Life table and mortality analysis*. Geneva, World Health Organization, 1979
- [45] PRESSAT R. *L'analyse démographique, concepts, méthodes, résultats* (2^e ed.). Paris, Presses Universitaires de France, 1969
- [46] ELVEBACK L. Estimation of survivorship in chronic disease : the « actuarial » method. *Amer Stat Ass* 1958 420-440
- [47] JENSEN OM, PARKIN DM, MACLENNAN R, MUIR CS, SKEET RG. *Cancer registration : principles and methods*. (IARC Scientific Publications No 95), Lyon, IARC, 1991
- [48] HAKULINEN T, ABEYWICKRAMA KL, SÖDERMAN B. *A computer program package for cancer survival studies*. Helsinki, Finnish Cancer Registry, 1987
- [49] CHRISTENSEN E. Multivariate survival analysis using Cox's regression model. *Hepatology*, 1987 **7** : 1346-1358