Population-based cancer survival analysis

Paul W. Dickman and Timo Hakulinen

DRAFT
February 17, 2003

Preface

Patient survival is the most important single measure of cancer patient care (the diagnosis and treatment of cancer). The optimal method for monitoring and evaluating the effectiveness of cancer patient care is through the population-based study of cancer patient survival, which is only possible using data collected by population-based cancer registries. The results of such studies are representative of the entire population, a perspective which is vital for cancer control activities.

Population-based studies of cancer patient survival, the topic of this text, differ in many ways from clinical studies of cancer patient survival. Population-based studies are generally based on a larger number of non-selected patients, who are followed-up less closely with respect to many clinical outcomes over a longer time period. Although conceptually similar, the design, analysis, and interpretation of the two types of studies involve different issues, both practically and statistically. Efficacy of cancer therapy, for example, is best evaluated through the study of patient survival in a (preferably randomised) clinical trial. Patients enrolled in clinical survival studies are often selected in order to answer a specific research question, and are closely followed-up to a well-defined endpoint.

Methods for the statistical analysis of clinical survival studies have been described in several recent texts [1, 2, 3, 4, 5, 6]. None of these texts, however, give more than a cursory coverage to concept of relative survival analysis, which is central to the analysis of patient survival in a population-based setting. The current text focuses on methods for studying cancer patient sur-

vival based on data from population-based registries. Much of the methodology is also appropriate for survival studies which are not population-based, or which study diseases other than cancer. However, in order to provide a stronger focus, we will assume that patient survival is being studied in a population-based cancer registry setting. This enables discussion of the important non-statistical issues which arise in population-based studies of cancer patient survival, such as registration, coding and classification, and follow-up procedures.

The text is aimed at researchers with a non-mathematical background, and the mathematical content is kept to a minimum. Statistical methods for survival analysis are, however, grounded in mathematics, and many of the concepts are best described using mathematical notation. An understanding of basic mathematics is therefore advantageous for the sections describing statistical theory. We assume, for example, that the reader is familiar with natural logarithms and their relationship to exponentials, and can follow formulae involving summations (Σ) and products (Π), although these concepts are described briefly in an appendix. Statistical theory is, by necessity, presented using mathematical formulae, but is also described in a non-mathematical manner so that complete understanding of the methematical detail is not required in order to understand the statistical concepts. Attention is focussed on the application and interpretation of the statistical methodology rather than the mathematical details of the theory. As with most texts, the choice of subjects, and the depth at which they are covered, are a reflection of the authors' own interests and contributions to the field.

Many of the issues which arise in the estimation and comparison of cancer patient survival are non-statistical in nature, such as temporal changes in the definition or coding of diagnostic entities (e.g. stage migration) or whether the survival analysis should include cases registered on the basis of death certificate only or autopsy. Knowledge of cancer registration and cancer biology is therefore advantageous for understanding the content of the text, although, as with the mathematical prerequisites, a brief background of these issues is also given where appropriate.

Paul W. Dickman and Timo Hakulinen

Stockholm and Helsinki

Contents

1

The role of survival analysis in cancer control

1.1 WHAT IS SURVIVAL ANALYSIS?

Survival analysis refers to the collection of statistical procedures used to study the time to occurrence of some event in a population, and is often called time-to-event analysis. The outcome of interest is the elapsed time between a well-defined starting point and a well-defined end point (often referred to as the 'outcome event' or the 'event of interest'). The three basic requirements of time-to-event measurements are

a. an agreed scale for the measurement of time (e.g. time since diagnosis, calendar time, attained age),

b. an unambiguous origin for the measurement of 'time', and

c. a precise definition of 'response,' or occurrence of the event of interest.

Applications of survival analysis can be found in most disciplines, especially in biomedicine, engineering, the social sciences, and marketing, and a wide variety of starting-points and end-points have been studied. Examples of time-to-event measurements in medicine are

- Time from diagnosis of cancer to death
- Time from randomisation to death in a cancer clinical trial
- Time from randomisation to recurrence in a cancer clinical trial

- Time from remission to relapse of leukemia
- Time between two attempts to donate a unit of blood for transfusion purposes
- Time to infection for patients hospitalised with severe burns
- Time to weaning of breast-fed newborns
- Time from HIV infection to AIDS

The study of population-based cancer patient survival, which is the focus of this text, deals almost exclusively with the first of these examples, the study of the time from diagnosis of cancer to death. That is, the time origin is diagnosis of cancer and the event of interest is death. We may choose to define the event of interest as death due to the cancer of interest or death due to any cause. The time scale is time since diagnosis, which is the obvious time scale and also the simplest to work with since every individual enters the study at time zero. Choosing a time scale is generally not an issue in population-based cancer survival analysis; time since diagnosis is the obvious choice. Situations can arise, however, where we might wish to use another time scale, such as attained age. This adds an extra level of complexity to the analysis since individuals no longer enter the study at time zero and the time of entry differs between individuals.

Survival analysis can also be used as a framework for the statistical analysis of epidemiological cohort studies. In such studies the primary interest is not the time to event per se, but rather whether the rate of occurrence of the event (e.g. disease incidence rate) depends on exposures of interest. In such studies it is more common to use a time scale other than time from entry, for example, attained age or calendar time. The primary characteristic of survival studies is that the event of interest is not observed for every individual in the study. It is because of this reason that separate statistical methods are required for the analysis of survival studies. This same characteristic is shared by epidemiological cohort studies, which is why survival analysis methods are used in the analysis of cohort studies. The focus of this text is population-based survival analysis but we will occasionally draw parallels between the concepts and terms used in survival analysis and those used in epidemiology.

It is apparent that not all of the examples of time-to-event measurements listed on page 1 strictly satisfy the 'three basic requirements of time-to-event measurements'. The date of diagnosis of cancer, for example, is often ambiguous, an issue which we will discuss further in this text. Population-based cancer registries (Section 1.2) very rarely collect information on events other than diagnosis or death, and many do not even routinely collect information on date of death. In this text, therefore, we will assume that survival time is measured from the diagnosis of cancer until death. Whether 'death' should

be should be defined as death due to the cancer of interest or death due to any cause is one of the many issues requiring discussion.

The origins of survival analysis can be traced back to work on mortality tables in the seventeenth century. The pressing need for methods to estimate the lifetime of industrial equipment during World War II stimulated research in methods for survival analysis, primarily the development of parametric models. The methodology was further developed after the war and also applied to the study of cancer patient survival. Medical researchers preferred the term 'survival analysis' rather than the terms 'lifetime analysis' and 'failure time analysis' favoured by industrial reliability engineers and also found that not all methods of survival analysis used in other disciplines were appropriate for the study of cancer patient survival.

It is perhaps not surprising that the study of cancer patient survival involves different issues to non-medical applications of survival analysis, such as the study of the lifetime of electrical components or the time between the introduction of a marketing campaign for a new product and consumer awareness of the product. In addition to the non-statistical issues which must be considered, the way in which electrical components age (the relation between the failure rate and the lifetime of the component) is different to the way in which humans age. As such, statistical models used to model the survival time of electrical components are often not appropriate for human populations.

Differences in survival analysis methodology also exist between the various biomedical specialities and even within the field of cancer survival research. Studies of cancer patient survival can be classified into two types; population based-studies and clinical studies (Section 1.4). The methodology used in the two types of study is similar, although sufficient differences exist to warrant a text dedicated solely to population-based studies of cancer patient survival. For example, each of the first two examples on page 1 study the time until death of individuals diagnosed with cancer. In a clinical trial, the time origin is usually the date of randomisation, whereas the date of diagnosis of cancer is used as the time origin in population-based studies. Other major differences exist in the way the patients are identified, recruited, and followed-up, which will be discussed further in Section 1.4.

1.2 THE ROLE OF CANCER REGISTRIES

A cancer registry can be defined as an organisation for the collection, storage, analysis, and interpretation of data on persons with cancer [7]. Hospital-based registries undertake these tasks within the confines of a hospital, or group of hospitals, whereas population-based registries cover all newly-diagnosed cases of cancer in a well-defined population. The purpose of the hospital-based reg-

istry is to serve the needs of the hospital administration, clinicians, the hospital's cancer program, and above all, the individual patient [8]. Hospital-based registries generally collect a broader range of information than population-based registries, information which is often invaluable when conducting randomised clinical trials for treatment evaluation. However, patients registered at hospital-based cancer registries are usually not representative of any population other than the group of patients attending the given hospital(s) so the information collected by hospital-based registries is usually not directly useful for broad-based cancer monitoring or control programs.

Population-based cancer registries, on the other hand, provide the population perspective that is vital to so many areas of cancer control and have a much wider epidemiological role. The use of data from population-based registries includes etiological research, primary and secondary prevention, health care planning, and patient care; thereby benefiting both the individual and society as a whole [9, 10, 11, 12]. The current text is specific to population-based cancer registries, which from this point on will be referred to simply as 'cancer registries'. The accepted role of the cancer registry has been summarised by Jensen and Storm [11]:

> The main objective of the cancer registry is to collect and classify information on all cancer cases in order to produce statistics on the occurrence of cancer in a defined population and to provide a framework for assessing and controlling the impact of cancer in the community.

This role is achieved through the maintenance of an individual-based register in which patients can be identified. Notification of new cancer cases is compulsory in many countries, and most registries receive multiple reports of each new cancer case from hospitals, pathology laboratories, and physicians. Armstrong [13] noted that the cancer registry can provide a 'framework for assessing and controlling the impact of cancer in the community' either indirectly, which is usual, or more directly, by actively participating in epidemiological research and the implementation of cancer control programs. He argues that direct participation is consonant with the basic role and skills of the cancer registry, promotes more rational cancer control, and strengthens the position and standing of the cancer registry.

1.3 THE ROLE OF SURVIVAL ANALYSIS IN CANCER CONTROL

There appears to be no generally accepted definition of cancer control. We favour the definition given by Armstrong (1992) [13] of 'all actions taken to reduce the frequency and impact of cancer', of which he identifies the six components shown in Table 1.1. A vital step in implementing any of these six components is the monitoring and evaluation of the component using an appropriate outcome measure (Table 1.1). The most commonly reported out-

come measures estimated from cancer registry data are incidence, mortality, and survival. None of these measures, however, can be used to evaluate rehabilitation or palliative care, for which a measure of quality of life must be used. A fifth measure, the percentage of cases for which the stage at diagnosis is classified as localised (%localised), has also been included in the table. This measure is sometimes used to evaluate screening or early diagnosis, although we believe that it is best used in conjunction with the preferred outcome measure for these components, mortality and survival respectively. Cancer prevalence, which can be defined as the number of people in the population with a diagnosis of cancer for which excess mortality or some other adverse condition due to the cancer persists, is another important measure of the cancer burden. In order to estimate prevalence defined by the excess mortality condition, we require estimates of both incidence and survival and, in particular, estimates of the time from diagnosis when patients can be considered 'cured'. Using registry data to estimate the proportion of patients considered 'cured' at a given time is a complex task and can only be performed at the group level.

Table 1.1 Components of cancer control as defined by Armstrong (1992) [13] and outcome measures (in decreasing order of utility) we advocate for monitoring and evaluating them.

Component	Outcome measures
primary prevention	incidence, mortality
screening	mortality, %localised, survival
early diagnosis	survival, %localised, mortality
treatment	survival, mortality
rehabilitation	quality of life
palliative care	quality of life

Patient survival is the most important single measure for monitoring and evaluating the early diagnosis and treatment components of cancer control (Table 1.1). Evaluation of these activities at the population level can only be performed using population-based data collected by cancer registries.

Enstrom and Austin (1977) [14] described many of the issues which need to be considered when interpreting cancer survival proportions - issues which are discussed in Chapter 5 of this text. The authors were skeptical of the use of survival proportions for evaluating cancer control and concluded 'If cancer control is related to how many people get and die of cancer, then progress can better be measured by the use of incidence and mortality rates'. It is true that cancer control is related to how many people get and die of cancer, although this does not adequately describe the full range of cancer control activities. We believe this definition of cancer control to be overly restrictive

and believe the definition provided by Armstrong (1992) [13] best describes the components of present day cancer control.

Of the three most commonly reported outcome measures (incidence, mortality, and survival), cancer incidence, the rate at which newly diagnosed cases occur in the population, is the single most important measure of the cancer burden. Incidence rates are published annually by many cancer registries, usually with a two or three year lag between the year of publication and the end of the period covered by the data. The Cancer Incidence in Five Continents series of monographs, published every five years by the International Agency for Research on Cancer (IARC) and the International Association of Cancer Registries (IACR), presents detailed data on the incidence of cancer. The most recent volume [15], published in November 1997, presents data on 170 populations in 46 countries for the years 1988–92.

Cancer mortality, the rate at which deaths due to cancer occur in the population, provides another measure of the cancer burden. Both incidence and mortality are reported as the number of events (incident cases or deaths) per unit of person-time at risk. Cancer incidence and mortality are usually reported as the number of events per 1,000 or 100,000 person-years. Cancer mortality rates are usually reported by the agency responsible for maintaining the central mortality register, rather than by the cancer registry. Cancer mortality data are based on the causes of death reported on death certificates, and the accuracy of these data are not subject to the same level of scrutiny as reports of new cancer cases (see Section 5.4 for further discussion). In many cases, it is difficult for the physician completing the death certificate to determine whether or not cancer is the primary cause of death. Furthermore, cancers which are cured do not appear in mortality statistics, thereby reducing the utility of mortality as a measure of the overall cancer burden. This is not to say that mortality is not a useful and important measure for monitoring and evaluating cancer control activities. For screening programs, such as mammographic screening for breast cancer, the monitoring of mortality rates in the screened population is the principal measure through which the success of the program can be determined. Incidence rates, on the other hand, are related to risk factors and the monitoring of incidence rates is the preferred method for the evaluation of primary prevention programs.

Although improvements in diagnostic and treatment facilities will affect mortality rates, such improvements are best monitored using survival proportions. The survival time for a cancer patient is defined as the elapsed time between diagnosis and death (Figure 1.1) and is the principal measure of the effectiveness of cancer care. Population-based survival proportions are indispensable in that they reflect the average outcome of all cancer patients in the population, without the selection bias present in hospital-based survival proportions or rates estimated for patients selected for clinical trials. Population-based survival proportions are influenced by, and reflect, the effectiveness of the en-

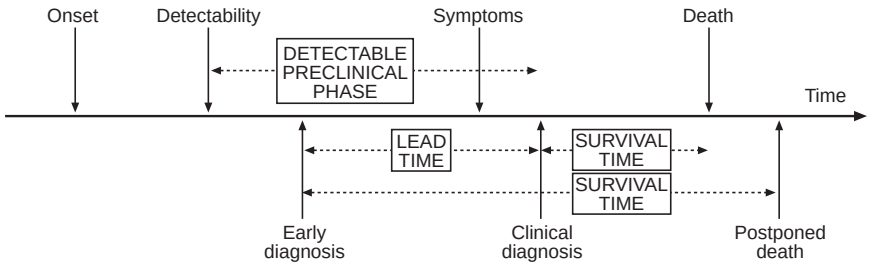

Fig. 1.1 Natural history of chronic illnesses.

tire chain of activities within the health care system, including the success of health education programs in increasing the awareness in the population of cancer symptoms, availability and general acceptance of cancer screening, skills of the physicians (and not only oncologists) in the early diagnosis of cancer, and the efficacy of surgical, oncological, and other treatment, including the aftercare of patients. It is not possible using routine cancer registry statistics, however, to determine the impact of each of these individual components of cancer care on the outcome. The optimal strategy for determining, for example, the efficacy of a treatment or intervention is with a randomised study, where the comparison groups are randomised with respect to other known and unknown explanatory variables.

Patient survival has improved over time for most cancer sites in most populations, primarily due to advances in diagnostic and treatment facilities. It is not possible using population-based data to determine empirically the extent to which the overall improvement is due to improvements in diagnostic methods and the extent to which it is due to improvements in treatment. However, conclusions can sometimes be made based on knowledge of the changes which have occurred in diagnostic and treatment methods along with a detailed study of the data, such as the study of trends in stage-specific incidence and stage-specific survival compared to the overall survival (Section 5.8).

The interpretation and comparison of estimated survival proportions, and time trends in estimated survival proportions, are discussed in detail in Chapter 5. In addition to patient survival, any measure of the effectiveness of cancer patient care should, ideally, also consider patient quality of life. Information on patient quality of life is rarely, if ever, available in population-based studies, although it is often considered in clinical trials [16, 17]. Detailed discussion of this issue is beyond the scope of the current text.

1.4 POPULATION-BASED VS CLINICAL STUDIES OF CANCER PATIENT SURVIVAL

The purpose of population-based survival analysis is to describe patient survival in demographically defined groups in the population in such a way that the results are representative. This is typically achieved by including in the analysis all patients in the population belonging to each group, defined by, for example, sex, age, calendar time, region, and social class. Population-based survival proportions are then estimated for each group [18]. Some clinically relevant prognostic factors such as stage, subsite, and histologic type may also be available. The quality of this information is, however, often not subject to special research efforts but rather reflects the information generally available in routine medical records. Nevertheless, this information is valuable in further stratifying the population-based rates [19].

The purpose of clinical survival studies is usually to evaluate the effect of an intervention (including treatment) on patient survival while controlling for other factors which may affect patient survival. It is not essential that the patients under study be representative of any population or population group. It is, however, essential that the treatment groups being compared are comparable in all aspects other than the intervention, which is best achieved by randomisation. Randomisation assures balance between the treatment groups with respect to known and unknown explanatory variables. Consequently, any major observed differences in the outcome may be attributed to a causal effect of the intervention. Although randomised experimental studies are more common, many clinical studies with survival time as the outcome of interest take the form of observational studies. Observational studies can be conducted with the aim of simply describing the outcome of a group of patients or to evaluate the effect of an intervention where a randomised study is not feasible, for example, where it is unethical to randomise patients according to treatment. Observational studies, including all studies based on population-based cancer registry data, do not provide the same possibilities for causal inference as randomised studies due to the possibility of selection bias. Observational studies are commonly used for evaluating prognostic factors using historical patient data, sometimes supplemented by retrospective biological analyses of archived tissue samples (blood, tumour, etc.). In some instances, clinical survival studies are also conducted for evaluating different diagnostic methods.

In clinical studies, whether they be randomised or observational, patients are followed more closely and accurately with respect to clinical outcomes than in population-based survival studies. As such, information on cause of death, relapses, remissions, side effects, etc. is more readily available than in population-based studies and can be used to define study endpoints other

than death or simple survival, for example, recurrence-free survival and cause-specific death.

All current text books on survival analysis in biomedicine focus on the analysis of clinical survival studies and do not therefore discuss many of the issues or methods relevant to population-based survival analysis. The methods unique to population-based survival analysis, most notably the relative survival ratio, reflect both the scope of the analyses and the quality of the data. These methods are described in individual scientific articles and some appear in texts on cancer epidemiology [20, 21]. The goal of the present text is to present the issues and methods relevant to population-based survival analysis in a single text in a uniform manner. Some of the classical survival analysis methods are also outlined in order to place the population-based methodology in a familiar context.

1.5 AN EXAMPLE OF PATIENT DATA COLLECTED BY THE FINNISH CANCER REGISTRY

Table 1.2 shows an example of data collected by the Finnish Cancer Registry used for estimating patient survival. A list of possible questions we might like to address using such data is given in Section 1.6. These data are a random sample of all cases of colon carcinoma, where stage at diagnosis was known, diagnosed in Finland during the period 1985–1994 and followed-up for deaths until the end of 1995. The complete data set is documented in Appendix A.2 and analysed throughout the text. Many registries, including the Finnish Cancer Registry, collect additional information such as subsite and histopathology, although the data items shown in Table 1.2 are the most important for estimating patient survival. These data items could be considered the minimum requirements for analysing patient survival, the possible exception being information on the clinical stage of the tumour. Clinical stage is a measure of the spread of the tumour and is assessed by the clinician making the diagnosis prior to the commencement of treatment. It is one of the strongest prognostic factors for patient survival and is discussed in detail in Section 5.8. In Table 1.2, stage has been classified as localised (confined to the organ or tissue of origin), regional (regional lymph node metastases), or distant (metastases other organs). Some registries do not collect information on stage, or the collected information is unreliable or incomplete. Lack of information on stage, however, does not preclude the estimation of survival proportions. On the contrary, overall patient survival (i.e. all stages combined) is the most important single measure of patient survival, and should be the primary measure used even for registries where information on stage is available. Stage-specific estimates of patient survival are, however, vital for studying possible explanations for trends in patient survival (Section 5.8). For example, examination of stage-specific trends in patient survival, and comparison to the overall trend, can help determine whether the overall trend is due to improved treatment, earlier diagnosis, or stage migration.

The data shown in Table 1.2 are typical of the data maintained by most population-based cancer registries. The first four columns of the table (sex, age, clinical stage, and date of diagnosis) show information collected at the time the tumour is diagnosed. The patient is then followed-up in order to ascertain the date of death, and survival time calculated as the time interval (usually in completed months or completed years) between the date of diagnosis and the date of death. The methods used for patient follow-up are discussed in Section 5.1. The two basic methods of follow-up are active follow-up, whereby the cancer registry initiates contact with the patients or their relatives in order to ascertain vital status; and passive follow-up, whereby deaths among the patients are identified through death certificates and/or computer-matching with a centralised mortality or population register and patients are assumed to be alive until their death or emigration is registered.

The Finnish Cancer Registry utilises a passive follow-up mechanism involving annual matching with the mortality register, which is actively verified by matching with the files of the Central Population Registry approximately every five years. Follow-up is virtually complete due to the unique personal identification numbers assigned to every Finnish resident. These numbers are used to index all databases.

The data in Table 1.2 have been sampled from all patients in Finland diagnosed during 1985–1994. This is the recruitment period of the material. As everyone was followed-up until the end of 1995, the potential follow-up time has a minimum of 1 year (patients diagnosed in December 1994) and a maximum of 11 years (patients diagnosed in January 1985). The potential follow-up time was longer for patients diagnosed early in the recruitment period, e.g., more than ten years for those recruited in 1985.

The actual follow-up time is equal to the potential follow-up time if the patient did not die or emigrate during the potential follow-up period. The patients in Table 1.2 have been ordered according to their actual survival times. Since none of the patients have emigrated, the patients with "status = alive" at the close of follow-up have survived for their full potential follow-up times, for example, patient 9 survived 13 months (12.95–11.94) and patient 35 survived 108 months (9 years) (12.95–12.86). The actual survival times are shorter than the potential follow-up times for patients who died before the end of 1995, for example, patient 1 survived only two months instead of the potential 82 months (12.95–2.89) and patient 33 survived 103 months instead of the potential 119 months (12.95–1.86). Note that the term 'survival time' is used independently of the outcome. We say that the survival time for patient 1 was 2 months and the survival time for patient 9 was 13 months, even though patient 1 died and patient 9 was censored.

Note that when the study has a common closing date of follow-up, that is, the potential follow-up of all patients end on a fixed day (December 31, 1995 in our examples), all status codes refer to the vital status at this date. Even if additional information indicates that a patient died in early 1996, this information cannot be utilised in the analysis, since not all patients could be followed up into 1996. A patient who, for example, is known to have died in February 1996 is, for the purposes of our study, considered to be withdrawn alive (censored) on December 31, 1995.

Figure 1.2 shows graphically the recruitment pattern of the material and the potential and actual follow-up times. In Figure 1.3 the time axis represents patient follow-up time, which is set to zero at the date of diagnosis of the tumour. In subsequent analyses, the latter time scale, the follow-up time, is used instead of the calendar time, as this best describes the progression of the disease process and the measures to combat it.

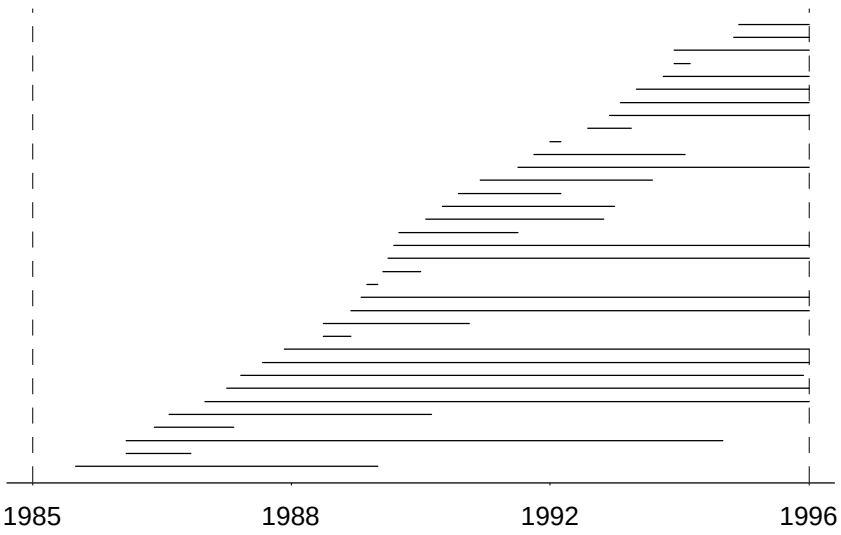

Fig. 1.2 Graphical representation of the survival experience of the 35 patients listed in Table 1.2 where the time scale is calendar time.

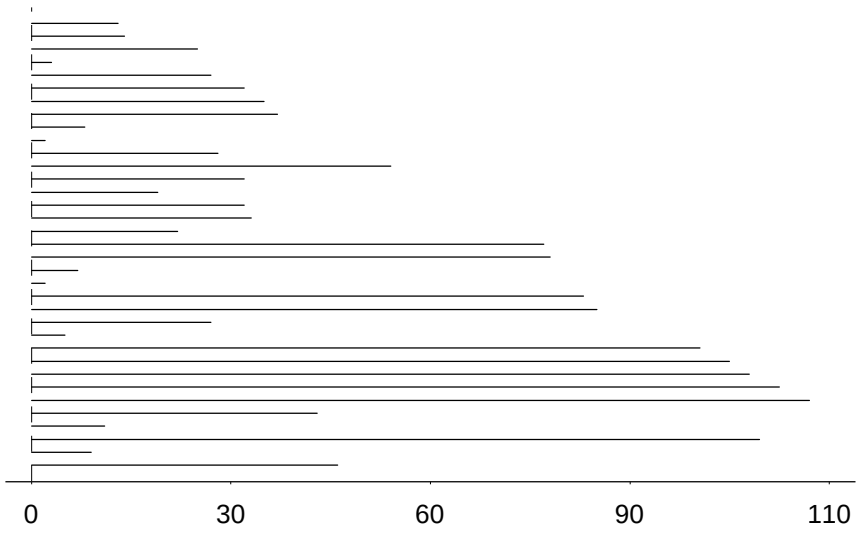

Fig. 1.3 Graphical representation of the survival experience of the 35 patients listed in Table 1.2 where the time scale is time since diagnosis.

Table 1.2 A sample of 35 patients diagnosed with colon carcinoma in Finland during 1985–94, followed-up until the end of 1995, and sorted in order of increasing survival time. Survival time is reported in both completed months (mm) and completed years (yy).

ID	Sex	Age at dx	Clinical stage	dx date mmyy	Surv. time mm	Surv. time yy	Status
1	male	72	Localised	2.89	2	0	Dead - other
2	female	82	Distant	12.91	2	0	Dead - cancer
3	male	73	Distant	11.93	3	0	Dead - cancer
4	male	63	Distant	6.88	5	0	Dead - cancer
5	male	67	Localised	5.89	7	0	Dead - cancer
6	male	74	Regional	7.92	8	0	Dead - cancer
7	female	56	Distant	1.86	9	0	Dead - cancer
8	female	52	Distant	5.86	11	0	Dead - cancer
9	male	64	Localised	11.94	13	1	Alive
10	female	70	Localised	10.94	14	1	Alive
11	female	83	Localised	7.90	19	1	Dead - other
12	male	64	Distant	8.89	22	1	Dead - cancer
13	female	79	Localised	11.93	25	2	Alive
14	female	70	Distant	6.88	27	2	Dead - cancer
15	male	70	Regional	9.93	27	2	Alive
16	female	68	Distant	9.91	28	2	Dead - cancer
17	male	58	Localised	11.90	32	2	Dead - cancer
18	male	54	Distant	4.90	32	2	Dead - cancer
19	female	86	Localised	4.93	32	2	Alive
20	male	31	Localised	1.90	33	2	Dead - cancer
21	female	75	Localised	1.93	35	2	Alive
22	female	85	Localised	11.92	37	3	Alive
23	female	68	Distant	7.86	43	3	Dead - cancer
24	male	54	Regional	6.85	46	3	Dead - cancer
25	male	80	Localised	6.91	54	4	Alive
26	female	52	Localised	7.89	77	6	Alive
27	male	52	Localised	6.89	78	6	Alive
28	male	65	Localised	1.89	83	6	Alive
29	male	60	Localised	11.88	85	7	Alive
30	female	71	Localised	11.87	97	8	Alive
31	male	58	Localised	8.87	100	8	Alive
32	female	80	Localised	5.87	102	8	Dead - cancer
33	male	66	Localised	1.86	103	8	Dead - other
34	male	67	Localised	3.87	105	8	Alive
35	female	56	Distant	12.86	108	9	Alive

1.6 QUESTIONS WE MAY WISH TO ADDRESS

Following is a list of questions related to patient survival we might like to address using data such as those shown in Section 1.5. They are not necessarily the most appropriate questions, and not all of them can be answered using population-based cancer registry data. We will refer back to these questions throughout the book, and comment on the issues involved in studying these questions. In Section 5.14 we briefly discuss the solutions to these questions and provide references to the Sections of the text containing the appropriate methodology.

1. Can we provide a descriptive overview of patient survival and how it varies with follow-up time?
2. What proportion of the patients are alive after 5 years of follow-up?
3. To what extent is the survival proportion affected by deaths due to cancer as opposed to deaths due to other causes?
4. Are any patients cured of the cancer, and if so, after how long?
5. What proportion of patients can be considered 'cured'?
6. Can we accurately predict which individual patients can be cured?
7. Can we estimate how much of the total mortality of the patients is due to cancer?
8. Does the impact of cancer upon patient survival differ according to any of the following factors, and can we determine whether any observed differences are systematic (as opposed to being in the range of what could be expected due to random variation)?
    - sex
    - age
    - calendar period
    - stage
    - morphology
    - social class
    - treatment
    - region of residence (regions covered by the same registry)
    - country
9. If differences exist for the factors above, can we quantify the size of the effect?

10. By how much does a diagnosis of cancer affect expectation of life?

11. How does mortality (both total and cancer-specific) change with follow-up time?

12. What is the prognosis for an individual patient recently diagnosed with cancer?

1.6.1 Life table estimates of patient survival

1.6.2 Adjusting for non-cancer deaths: estimation of net survival

Recall that 8 patients died within one year of diagnosis and the estimated 1-year observed survival proportion was 0.771. However, of the 8 patients who died, death certificates indicated that colon carcinoma was the primary cause of death for 7 patients, while the remaining patient (patient ID number 1) was classified to have died of some other cause. We are generally interested in the effect of the disease of interest (colon carcinoma in this case) on patient survival, and would like to somehow adjust for deaths due to other causes, since these deaths unduly lower the estimated survival proportion as far as the cancer of interest is concerned. A quantity of interest is the proportion of patients who would have survived t years or more following diagnosis in the hypothetical situation where colon carcinoma was the only possible cause of death. This theoretical quantity is known as the net survival rate (Section 2.11). The observed survival proportion underestimates net survival, since both cancer deaths and non-cancer deaths are considered to be outcome events of interest. The size of the bias will be greatest for older patients, who experience higher levels of non-cancer mortality. This bias makes it impossible to directly compare observed survival proportions between patient groups with different age structures.

One method of estimating net survival is to consider only deaths due to the disease of interest as outcome events and consider the survival times of patients who die of other causes to be censored, giving rise to what is called the cause-specific survival proportion (Section 2.12), which we will denote using the symbol s. Although the estimation of net survival is the usual aim when calculating cause-specific survival, cause-specific survival proportions can also be used to summarise that part of the patients' mortality recorded as being directly attributable to their particular cancer (Section 2.11).

Table 1.3 Life table for the 35 patients shown in Table 1.2 with estimates of the interval-specific (s_i) and cumulative ($_1s_i$) cause-specific survival proportions and cumulative observed ($_1p_i$) and relative ($_1r_i$) survival proportions (extracted from Table 2.1 for comparison with $_1s_i$).

i	l_i	d_i	w_i	l'_i	s_i	$_1s_i$	$_1p_i$	$_1r_i$
1	35	7	1	34.5	0.79710	0.79710	0.77143	0.80024
2	27	1	3	25.5	0.96078	0.76584	0.71209	0.76839
3	23	5	4	21.0	0.76190	0.58350	0.54254	0.60914
4	14	2	1	13.5	0.85185	0.49705	0.46217	0.53822
5	11	0	1	10.5	1.00000	0.49705	0.46217	0.55629
6	10	0	0	10.0	1.00000	0.49705	0.46217	0.57409
7	10	0	3	8.5	1.00000	0.49705	0.46217	0.59307
8	7	0	1	6.5	1.00000	0.49705	0.46217	0.61378
9	6	1	4	4.0	0.75000	0.37279	0.25676	0.35141
10	1	0	1	0.5	1.00000	0.37279	0.25676	0.35953

When estimating the 1-year cause-specific survival proportion using the actuarial method, we assume that the survival time for patient number 1 was censored in the middle of the interval, and the estimated 1-year cause-specific survival proportion is $1-7/34.5 = 0.79710$. During the 2nd follow-up interval there was one death due to cancer and 3 censorings (including 1 non-colon cancer death), so the estimated interval-specific cause-specific survival proportion for the 27 patients alive at the start of the interval is $1-1/25.5 = 0.96078$ and the 2-year cumulative cause-specific survival proportion is $0.79710 \times 0.96078 = 0.76584$. Estimated cumulative cause-specific survival proportions ($_1s_i$) are shown in Table 1.3.

Since this text focusses on cancer patient survival, we will use the term 'cancer death' to refer to a death due to the cancer of the patient. A status of 'Dead - cancer' in Table 1.2 indicates that colon carcinoma was classified as the primary cause of death. All other deaths, including deaths due to other types of cancer, are classified as non-cancer deaths, which are sometimes also referred to as deaths due to competing risks. Considering these deaths as withdrawals implies the assumption that the non-cancer deaths occur independently of the cancer deaths within the groups or subgroups under study (e.g. males, aged 45–64 years diagnosed 1990–1994) (Section 2.2).

1.6.3 Relative survival

The estimation of cause-specific survival proportions requires that reliably coded information on cause of death is available. Even when cause of death

information is available to the cancer registry via death certificates, it is often difficult, for several reasons, to determine whether or not the cancer of interest is actually the primary cause of death (Section 5.4). An alternative method of estimating net survival is to use the relative survival ratio, defined as the observed survival proportion in the patient group divided by the expected survival rate of a comparable group from the general population, who are assumed to be practically free of the cancer of interest [22]. A major advantage of this measure is that information on cause of death is not required, thereby circumventing problems with the inaccuracy [23] or nonavailability of death certificates. It is usual to estimate the expected survival proportion from nationwide population life tables stratified by age, sex, calendar time, and, where applicable, race [24]. Although these tables include the effect of deaths due to the cancer being studied, Ederer et al. [22] showed that this does not, in practice, affect the estimated survival proportions. When estimating relative survival, it is assumed that death due to cancer, death due to other causes, and censoring are independent events within each subgroup defined by age, sex, calendar period etc. The relative survival ratio is described in Section 2.13 and the methodology for estimating expected survival in Section 2.18. As with cause-specific survival, the aim of relative survival analysis is to estimate net survival, the survival experience of the patients in the hypothetical situation that the only possible cause of death was the cancer of interest (Section 2.11).

Table 1.4 Life table for the 35 patients shown in Table 1.2 with estimates of the observed (p_i, ${}_1p_i$), expected (${}_1p_i^*$), and relative (${}_1r_i$) survival proportions.

i	l_i	d_i	w_i	l_i'	p_i	${}_1p_i$	${}_1p_i^*$	${}_1r_i$
1	35	8	0	35.0	0.77143	0.77143	0.96399	0.80024
2	27	2	2	26.0	0.92308	0.71209	0.92672	0.76839
3	23	5	4	21.0	0.76190	0.54254	0.89068	0.60914
4	14	2	1	13.5	0.85185	0.46217	0.85869	0.53822
5	11	0	1	10.5	1.00000	0.46217	0.83080	0.55629
6	10	0	0	10.0	1.00000	0.46217	0.80504	0.57409
7	10	0	3	8.5	1.00000	0.46217	0.77928	0.59307
8	7	0	1	6.5	1.00000	0.46217	0.75299	0.61378
9	6	2	3	4.5	0.55556	0.25676	0.73066	0.35141
10	1	0	1	0.5	1.00000	0.25676	0.71415	0.35953

Estimates of the cumulative expected survival proportion (${}_1p_i^*$) and the cumulative relative survival ratio (${}_1r_i$) are shown in Table 1.4. Worked examples of how these estimates were constructed are given in Section 2.13. The cumulative observed (OSR), cause-specific (CSR), and relative (RSR) survival proportions are shown in Figure 1.4. The observed survival proportion is, as expected, lower than the other two rates, since all deaths, irrespective of

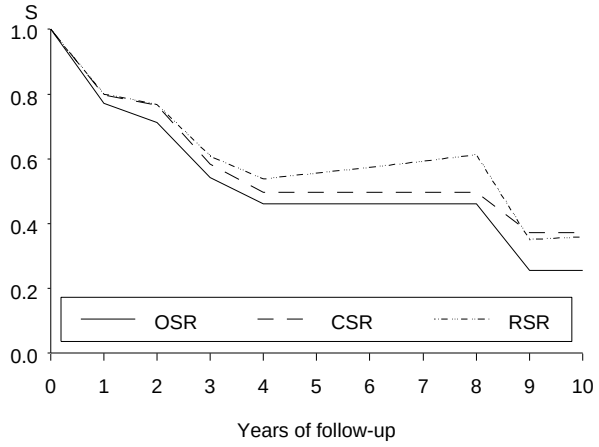

Fig. 1.4 Graph of the cumulative observed (OSR), cause-specific (CSR), and relative (RSR) survival proportions shown in Table 1.3

cause, are considered to be outcome events. The other two quantities are both estimates of net survival, and we would hope that the estimates would be similar. The two estimates are almost identical for the first two intervals, after which there are small differences. Note that the relative survival ratio increases when no deaths are observed in the patient group, whereas the cause-specific survival proportion remains constant. An increasing cumulative relative survival ratio means that the patients were subject to lower mortality during the period than a comparable group from the general population. This may occur due to the 'healthy patient effect' (Section XXX) or artifactually after long follow-up (Section XXX), although in the present example it can be attributed to small sample size. The properties of each of the three survival estimates (observed, cause-specific, and relative) are further discussed in Section 2.17.

The estimates of patient survival presented in this Chapter were based on all 35 patients combined. survival proportions can also be estimated and compared for patient subgroups, defined by, for example, age, sex, and stage. Methods for testing the equality of survival proportions estimated from life tables are presented in Section 2.25. Studying the association between patient survival and prognostic factors is, however, best performed in a model-based framework, where the effect of any one factor can be studied while simultaneously controlling for other factors. Several models for population-based cancer patient survival are described in Chapter 3 with detailed examples using real data presented in Chapter 4.

The validity of any statistical analysis depends on the quality of the data upon which it is based. There are many data quality issues in estimating cancer patient survival, and these are addressed in Chapter 5. One of the most problematic issues, however, is the quality of patient follow-up; patient survival will be biased if the vital status of the patients is inaccurate.

2

Estimating cancer patient survival

2.1 TERMINOLOGY AND NOTATION USED IN SURVIVAL ANALYSIS

This section introduces the central concepts, mathematical terminology, and notation used in survival analysis. The central concepts, such as the survivor function and hazard function, are described in both general terms and by way of mathematical formulae.

2.1.1 Basic terminology

In general, survival analysis refers to the collection of statistical procedures used to study the time to occurrence of some event in a population. In population-based survival analysis we generally study the time from diagnosis of cancer and the event of interest is usually death. Survival time of cancer patients is generally measured in completed months or completed years although some registries record survival time in completed days. A survival time of 2 completed months, for example, means the patient survived 2 months following diagnosis but not 3 months. We say that the survival time of a patient is censored if we are unable to follow-up the patient until the event is observed. A 'survival time' is recorded for all patients irrespective of whether or not the event of interest is observed. The survival time is calculated as the time from diagnosis until either the date of the event or the date of censoring (see, for example, Table 1.2). The outcome of interest in a survival study is therefore

represented by a pair of variables, one variable containing the survival time and a second variable, called a censoring variable or censoring indicator, which indicates whether the event of interest was observed or the survival time was censored.

2.1.2 Random variables

Most statistical methods, survival analysis being no exception, are based on the concepts of random variables and probability distribution functions, which are briefly described here. A random variable is a quantity which, theoretically, may assume a wide range of values, although in any particular realisation we observe only a single value. Measurements are common examples of random variables, such as birthweight or systolic blood pressure. It is standard to use upper case Roman letters to represent random variables and the corresponding lower case letter to refer to the observed value of the random variable. We might therefore talk about the random variable X representing the birthweight of babies born in a given region during a given calendar year (our population of interest). Although the random variable X can theoretically assume a large range of values, some values occur more often than others and it is common to assume a certain probability distribution for the random variable. For birthweight, it may be appropriate to assume a normal distribution with a given mean and variance. Once a probability distribution has been assigned, it is possible to consider such issues as the probability that the birthweight of a particular baby in the population is less than a specified value, say 2500g, which is written as $\Pr(X < 2500)$. Birthweight is a continuous outcome, and X is called a continuous random variable, but the same principles apply for discrete outcomes. We may, for example, be interested in the random variable Y, representing the sex of the newborns in our population, which takes on one of two values. The (discrete) probability distribution function of the (discrete) random variable Y might be $\Pr(Y = \text{male}) = 0.503$ and $\Pr(Y = \text{female}) = 0.497$. [Check these proportions]

2.1.3 The survivor function

The survival time for an individual can theoretically take on any non-negative value and is represented by the non-negative random variable T. The actual survival time for an individual is denoted by t and is assumed to be a realisation of the random variable T. This random variable is generally assumed to have a probability distribution function $f(t)$ and corresponding cumulative distribution function $F(t) = \Pr(T \leq t) = \int_0^t f(x)\,\mathrm{d}x$.

The probability of an individual surviving until at least time t is given by the survivor function, $S(t)$, (sometimes called the survival function),

$$S(t) = \Pr(T > t) = 1 - F(t). \tag{2.1}$$

The survivor function gives the probability that an individual survives longer than some specified time t. That is, $S(t)$ gives the probability that the random variable T exceeds the specified time t. Statistical methods for survival analysis can be classified as either parametric, where assumptions are made about the probability distribution function, or non-parametric, where no such assumptions are made. The majority of methods used in population-based survival analysis are non-parametric – we assume that patient survival time can be described by a probability distribution function but we make no prior assumptions about the form of the distribution.

Estimation of the survivor function is a common method for summarising survival data and methods for estimating $S(t)$ will be presented later in this chapter. Figure 2.1 shows the survivor functions for two groups of patients diagnosed with Non-Hodgkin's lymphoma. The survivor function is a nonincreasing function with value 1 at the time origin (diagnosis, $t = 0$) and value 0 as t approaches infinity.

2.1.4 The hazard function

The rate of decline of the survivor function varies according to the risk of experiencing the event at time t, but it is difficult to determine the essence of the failure pattern, and even more difficult to compare it between groups, simply by studying plots of the survivor function. For example, it can be seen from Figure 2.1 that survival following a diagnosis of non-Hodgkin's lymphoma is superior for those patients whose mothers lived longer than the median. However, conditional on surviving 1000 days, it is the other group which experiences superior survival. It is difficult to compare the failure pattern between the two groups by studying the survivor functions in Figure 2.1.

This pattern is sometimes seen when comparing the survival of patients who received a highly aggressive therapy to those who received a conservative therapy. The patients receiving the highly aggressive therapy may first experience a lower survival than those treated conservatively due to, for example, operative mortality, but the majority of the patients who survive the therapy are cured. Consequently, among the patients who survive, say, 1 year from diagnosis, those who received the highly aggressive therapy experience superior survival.

In Chapter 3 we present methods for modelling cancer patient survival as a function of covariates, where it is usual to model the hazard function rather than the survivor function. The hazard function, denoted by $\lambda(t)$, describes

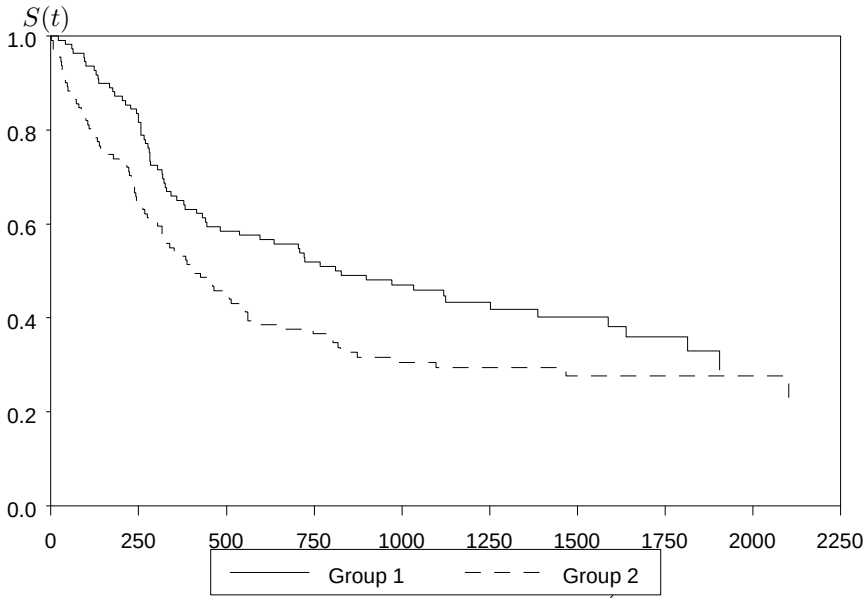

Fig. 2.1 Cumulative observed survivor function, $S(t)$, for patients diagnosed with non-Hodgkin's lymphoma in Sweden 1992–1996, stratified (using a median split) according to the age at death of the patient's mother.

the instantaneous death rate at time t, conditional on survival up to time t. In contrast to the survivor function, which describes the probability of *not* failing before time t, the hazard function focuses on the failure rate at time t among those individuals who are alive at time t. That is, a lower value for $\lambda(t)$ implies a higher value for $S(t)$ and vice-versa. Note that the hazard is a rate, rather than a probability, so $\lambda(t)$ can take on any value between zero and infinity, as opposed to $S(t)$ which is restricted to the interval $[0, 1]$.

The hazard function is known as the force of mortality in demography and the age-specific failure rate in epidemiology (where the time variable is attained age rather than time since diagnosis). Some common generic forms for the hazard function are constant, increasing, decreasing, and bathtub (Figure 2.2). A bathtub-shaped hazard is appropriate for most human populations followed from birth, where the hazard decreases to almost zero after an initial period of infant mortality, and then starts to increase again later in life. A decreasing hazard function is appropriate following the diagnosis of most types of cancer, where mortality due to the cancer is highest immediately following diagnosis, and then decreases with time as patients are cured of the cancer. A constant hazard function is often used for modelling the lifetime of electronic components, but is also appropriate following the diagnosis of some types of cancer, most notably cancers of the breast and prostate, where the level of excess mortality due to the cancer is relatively constant over time and persists even 15-20 years after diagnosis. The cumulative survivor function has the same basic shape (a nonincreasing function from 1 to 0) for all types of data and the hazard function is often a more informative means of studying differences between patient groups.

Another quantity, the cumulative hazard function (also called the integrated hazard function), $\Lambda(t)$, is given by

$$\Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u. \tag{2.2}$$

The distribution function, survivor function, and hazard function are mathematically related such that the knowledge of any one uniquely defines the other two. For example,

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{S(t)} \\ &= -\frac{\mathrm{d}}{\mathrm{d}t}\left[\ln S(t)\right]. \end{aligned} \tag{2.3}$$

Therefore,

$$\begin{aligned} S(t) &= \exp\left[-\int_0^t \lambda(u)\,\mathrm{d}u\right] \\ &= \exp(-\Lambda(t)). \end{aligned} \tag{2.4}$$

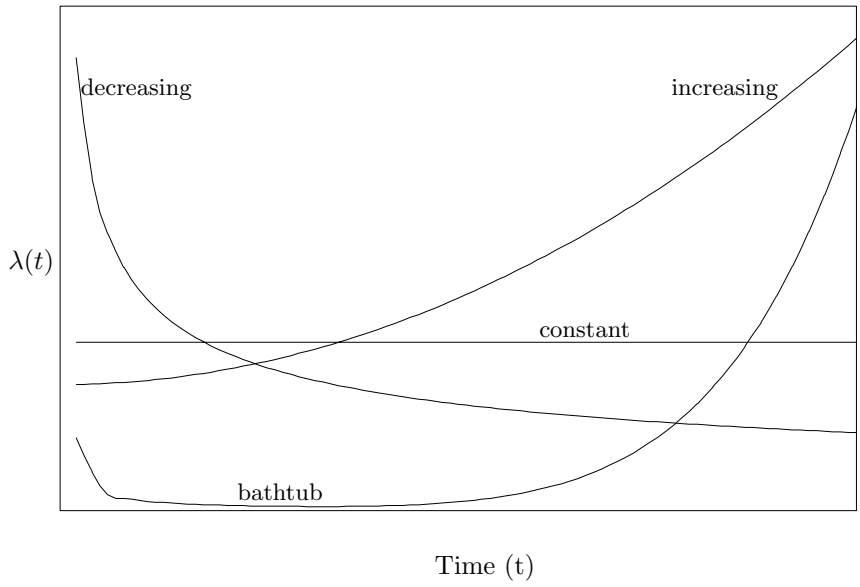

Fig. 2.2 Examples of various generic hazard functions

The cumulative hazard function can therefore be expressed in terms of the survivor function as

$$\Lambda(t) = -\ln S(t). \tag{2.5}$$

So-called parametric survival models generally require assumptions to be made about the form of the hazard function (which implies an assumption about the probability distribution function and the survivor function since they are mathematically related), although such models are less common in population-based cancer survival analysis than in other disciplines. That is, in the methods commonly used for population-based cancer survival analysis, no assumptions are made about the form of the hazard function.

[Discuss the fact that, although time is continuous, we observed it on a discrete scale. Hence we use discrete versions of the above formulae.]

2.2 CENSORING

The presence of censoring is one of the distinguishing features of survival analysis, and refers to the situation where incomplete information is available about the survival time of some individuals. Censoring can be classified into several different types [25], of which right-censoring is the most common in

population-based survival analysis. Observations are said to be right-censored when the survival time of the censored observation is known only to exceed a certain value. We have so far encountered two situations where survival times are censored. There are generally three reasons why censoring may occur in population-based studies of cancer patient survival with death due to the cancer of interest as the outcome event:

1. a patient is alive at the end of the follow-up period (e.g. the patients in Table 1.2 with 'status=Alive');

2. a patient is lost to follow-up during the follow-up period (e.g. due to emigration);

3. a patient is withdrawn from the study during the follow-up period due to death from a cause other than the cancer of interest (e.g. patients 1, 11, and 33). [Note that such patients are considered censored in the analysis of cause-specific survival, but not observed or relative survival.]

Censored observations are one of the main characteristics of survival data and it is important that they be appropriately accounted for in the analysis. Assumptions are often made about censoring patterns (such as the actuarial assumption described in Section 1.6.1). The most crucial of these assumptions is that, conditional on the value of any explanatory variables, the event considered to be the outcome of interest occurs independently of the events classified as censoring. This assumption implies that, at each time, the censored patients can be considered a random sample of all patients at risk. We can then assume that the survival pattern of a censored patient after the censoring time can be estimated by the survival pattern of the non-censored patients, and that the survival estimates can be generalised to all patients. Censoring is discussed further in Section 2.14.

For registries using passive follow-up, such as the Finnish Cancer Registry, losses to follow-up (reason 2 in the list above) occur when a patient is known to have emigrated to another country, thereby precluding the possibility of ascertaining the date of death. Survival time is then calculated from the date of diagnosis to the date of emigration. Among the 15564 colon carcinoma patients diagnosed in Finland during 1985–1994, 4 patients were known to be lost to follow-up (Appendix A.2). For registries using active follow-up, losses to follow-up occur when the patient can no longer be contacted using the contact information available to the registry. In this situation, survival time is calculated from the date of diagnosis to the last date the patient was known to be alive.

2.3 ESTIMATING PATIENT SURVIVAL

The basic goals of survival analysis are to describe the survival experience of the study cohort and possibly also to assess whether survival is associated with explanatory variables. The survival experience of the cohort can be measured using either the survivor function or the hazard function, which are mathematically related (Section 2.1). It is usual to work with the survivor function for descriptive analyses and the hazard function for assessing the association between explanatory variables and survival time, which usually requires statistical modelling.

Consider the task of estimating patient survival for the 35 patients shown in Table 1.2. The simplest measure of patient survival is the survivor function, $S(t)$, which is simply the proportion of patients who survive at least t years following diagnosis. For the moment we will consider all deaths to be events of interest, in which case the survivor function is called the observed survivor function (Section 2.9).

The survivor function $S(t)$ is, theoretically, a continuous function with a real value at every value of time, t. Estimating $S(t)$ involves obtaining estimates at discrete values of t and then interpolating these estimates to obtain estimates over the complete range of follow-up times. We will commence by estimating $S(t)$ at t=0, 1, 2, 3, 4, and 5 years. Estimating $S(t)$ at $t = 0$ is trivial, since everyone must be alive at the start of follow-up. We write $\hat{S}(0) = 1$ where the 'hat' indicates the fact that this is an estimate of $S(0)$.

The survivor function at t=1 year represents proportion of patients who are alive one year subsequent to diagnosis and is often called the 1-year observed survival proportion. From Table 1.2 we see that 8 of the 35 patients died during the first year of follow-up, so the estimated 1-year observed survival proportion is

$$\hat{S}(1) = 1 - 8/35 = 27/35 = 0.771.$$

To estimate $S(2)$ in a similar manner we note that 10 patients died during the first 2 years of follow-up (8 during the first year and 2 during the second year). Estimation of the survivor function is, however, complicated by the fact that the survival times of patients 9 and 10 are censored since these patients could not be followed-up for the full two years. Patient number 9, for example, was diagnosed in November 1994 and was alive at the end of 1995 (the last date of follow-up). We know that this patient survived 13 completed months, so our information is limited to the knowledge that the survival time for this patient must be greater than or equal to 13 months. It is not obvious how to account for the two censored survival times in the estimation of $S(2)$. One (erroneous) approach might be to exclude them from the analysis completely, and base the estimate only on those patients who were followed-up for two years or more. That is, the estimated proportion dying within 2 years of diagnosis would be

10/33 so the estimated 2-year survival proportion would be

$$\hat{S}(2) = 1 - 10/33 = 23/33 = 0.697.$$

This approach underestimates patient survival, since it ignores the fact that the two patients with censored survival times were at risk of death, yet they survived for 13 and 14 months respectively without dying. Another erroneous approach is to say that 10 of the 35 patients died within 2 years and to estimate the 2-year survival proportion as

$$\hat{S}(2) = 1 - 10/35 = 25/35 = 0.714.$$

This approach overestimates the survival proportion, since it presupposes that patients 9 and 10 both survive a full two years. The problem is even greater when we attempt to estimate $S(8)$ since almost half of the survival times are censored. It is apparent that we require a method which allows us to estimate the survival proportion while allowing for the fact that some survival times are censored.

The two most common methods for estimating the survivor function are the actuarial (life table) method and the Kaplan-Meier (product-limit) method. Both of these methods estimate the survivor function as a product of conditional survival probabilities. For example, assume we wish to estimate $S(5)$, the 5-year survival proportion. In order to survive 5 years from diagnosis, one must survive the first year, then the second year, the third year, the fourth year, and, finally, the fifth year. The approach used to estimate $S(5)$ is to first stratify the follow-up into 5 annual subintervals and estimate the conditional survival probability for each of the 5 annual intervals (Figure 2.3). The estimate of $S(5)$ is then given as the product of the 5 conditional probabilities. That is, the probability of surviving 5 years is estimated as the probability of surviving the first year multiplied by the probability of surviving the second year (assuming you have survived the first year) multiplied by the product of surviving the third year (assuming you have survived the first 2 years) and so on.

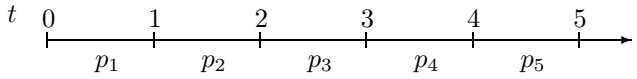

Fig. 2.3 Partitioning follow-up time into subintervals. The intervals are indexed using $i = 1, 2, 3, \ldots$ and the conditional probability of surviving the interval is given by p_i.

The conditional survival probability is the probability of surviving the interval conditional on being alive at the start of the interval. The first conditional probability, denoted by p_1, is simply the probability of surviving the first year following diagnosis and is estimated as 27/35. Since all patients must be alive

at the start of the first interval, this is equivalent the estimated unconditional 1-year survival proportion. The second conditional probability, denoted by p_2, is the probability of surviving the second year following diagnosis, conditional on surviving the first year. It is estimated based on the 27 patients who survived the first year. We note that the survival times of 2 of these 27 patients were censored during the second year so it is not obvious how to estimate p_2. It appears that we have gained no benefit from stratifying the follow-up into subintervals and estimating conditional probabilities. The benefit is that we are able to make assumptions, which are valid within the subintervals, about the survival experience of the patients whose survival times were censored thereby enabling estimation of the survivor function in the presence of censoring. Details are given in the next section where the life table method of estimating patient survival is described.

2.4 ACTUARIAL (LIFE TABLE) ESTIMATES OF PATIENT SURVIVAL

Estimates of patient survival based on large materials are often presented in the form of a life table, such as the one shown in Table 2.1 for the sample of 35 patients diagnosed with colon carcinoma. Note that a group 35 patients is not considered 'large'; we have constructed this life table for 35 patients in order to demonstrate the methodology. Life table estimates of patient survival are also known as actuarial estimates. The approach is to divide the period of observation into a series of time intervals (indexed by $i = 1, 2, 3, \ldots$), giving rise to what are sometimes known as grouped survival data. The intervals need not be of equal length, although they frequently are. Cancer registries often record survival time only in completed years, rather than months or days, so it is common to construct life tables using annual intervals.

Table 2.1 Life table for the 35 patients shown in Table 1.2 with estimates of the interval-specific (p_i) and cumulative ($_1p_i$) observed survival proportions.

i	l_i	d_i	w_i	l_i'	p_i	$_1p_i$
1	35	8	0	35.0	0.77143	0.77143
2	27	2	2	26.0	0.92308	0.71209
3	23	5	4	21.0	0.76190	0.54254
4	14	2	1	13.5	0.85185	0.46217
5	11	0	1	10.5	1.00000	0.46217
6	10	0	0	10.0	1.00000	0.46217
7	10	0	3	8.5	1.00000	0.46217
8	7	0	1	6.5	1.00000	0.46217
9	6	2	3	4.5	0.55556	0.25676
10	1	0	1	0.5	1.00000	0.25676

The notation used for life table quantities is as follows

i interval index (the first interval is assigned index 1)

l_i number of patients alive at the start of interval i

d_i number of deaths during the ith interval

w_i number of individuals whose survival time was censored during interval i

l_i' is the effective number of patients at risk during the interval, given by $l_i' = l_i - \frac{1}{2}w_i$

p_i estimated conditional (interval-specific) survival proportion. That is, the probability of surviving to the end of interval i for those alive at the start of interval i

$_1p_i$ estimated cumulative survival proportion (from diagnosis) at the end of interval i

Note that $_1p_i$ is equivalent to the survivor function, $S(t)$, estimated at the end of the interval.

In the absence of censoring, the conditional probability of surviving interval i is estimated in the usual manner, that is

$$p_i = 1 - d_i/l_i \tag{2.6}$$

For example, the probability of surviving the first interval is estimated as

$$p_1 = 1 - 8/35 = 27/35 = 0.771.$$

In the presence of censoring, it is assumed that censoring occurs uniformly throughout the interval such that each individual with a censored survival time was at risk for, on average, half of the interval. This assumption is known as the actuarial assumption and is discussed further in Section 2.14. In the presence of censoring, the conditional probability of surviving interval i is estimated as

$$p_i = 1 - d_i/l'_i \tag{2.7}$$

where $l'_i = l_i - \frac{1}{2}w_i$. That is, the estimate of the proportion surviving is given by 1 minus the estimate of the proportion dying during the interval. The denominator used in the estimate of the proportion dying is corrected (reduced by $\frac{1}{2}w_i$) to adjust for censoring. Instead of basing the estimate on l_i, the actual number of patients at risk at the start of the interval, the estimate is based on what is called the effective number at risk (l'_i). The correction factor of $\frac{1}{2}w_i$ is reminiscent of the type of correction one would make when estimating person-time at risk and suggests the question of why we don't also reduce the denominator by $\frac{1}{2}d_i$ since it could also be assumed that the patients who die during the interval are each at risk, on average, for only half of the length of the interval. The reason such a correction is not made is because we are estimating a proportion, not a rate (see Section 2.13.2).

When estimated from a life table, the conditional survival proportions are known as interval-specific survival rates. The probability that an individual survives from diagnosis until the end of the ith interval, known as the cumulative survival proportion, is given by the product of each of the preceding interval-specific survival proportions

$${}_1p_i = \prod_{j=1}^{i} p_j. \tag{2.8}$$

The intervals need not be of equal length, although they usually are (annual intervals are common). Life tables produced from cancer registry data are known as cohort life tables and differ slightly from current life tables, which are based on individuals with a common starting time. Life table estimates of patient survival, using annual intervals, for the 35 patients listed in Table 1.2 are shown in Table 2.1.

Life table methods are well-suited to cancer registry data, where data sets are large and exact survival times in days cannot be established with any precision. Unlike date of death, date of diagnosis cannot be precisely recorded to the nearest day since the diagnosis of cancer is generally a process which commences with a suspicion and becomes more definitive through X-rays, endoscopies, and, finally, microscopic examination of the tumour specimen. As such, cancer registries often record only the month and year of diagnosis and survival times of cancer patients are known, at best, to the nearest month, and sometimes only in completed years (Section 5.5).

2.5 AN APPLICATION OF THE ACTUARIAL METHOD

Table 2.2 presents life table estimates of patient survival (observed, expected, and relative survival) for females diagnosed with colon carcinoma in Finland during the period 1985–1994. A description of the notation used in the table is given in Table 2.3.

The life table is constructed for annual intervals, indexed by i. The second column of the life table, labelled l_i, gives the number of patients alive at the start of the ith interval. The next two columns give the number of deaths due to any cause (d_i) during the interval and w_i, the number of patients censored (withdrawn alive) during the interval. Censoring occurs when the patient is known to be alive at the end of the follow-up period or when the patient is 'lost to follow-up' and current vital status is unknown (Sections 2.2 and 2.14). Patients diagnosed near the end of the recruitment period covered by the life table (1985–1994 in this case) are more likely to be censored due to being alive at the last day of follow-up than those diagnosed in the earlier years. Note that these women were followed-up to the end of 1995, an additional year beyond the period of diagnosis, meaning that the minimum potential follow-up time for every woman is at least one year. Consequently, there are no censored observations during the first year of follow-up. Follow-up procedures are extremely efficient in Finland due to the use of unique personal identification numbers and very few (approximately 1 in 5000) are lost to follow-up [26, 27]. As a result, all censoring in Table 2.2 is due to the patients being alive at the end of the follow-up period.

The individuals with censored survival times did not die during the interval, but were nevertheless at risk for part of the interval, so they need to be considered in the denominator when estimating the survival proportion. It is standard to assume that censoring occurs uniformly within the interval (discussed in Section 5.9), such that each censored patient is at risk for, on average, half of the interval. The 'effective' number of patients at risk during the interval is therefore given by $l'_i = l_i - \frac{1}{2}w_i$. The estimated interval-specific observed survival proportion is then given by $p_i = 1 - d_i/l'_i$. The estimated cumulative observed survival proportion from diagnosis to the end of the ith interval, ${}_1p_i$, is given by the product of the estimated interval-specific survival proportions up to the ith interval.

Among the $l_1 = 5285$ women diagnosed with colon carcinoma during 1985–1994, 1784 did not survive one completed year subsequent to the date of diagnosis. Since all patients had a potential follow-up time greater than one year there was no censoring during the first interval and $l'_1 = l_1 = 5285$. The interval-specific and the cumulative estimates of the observed survival proportion are identical for the first interval and are given by $1 - d_1/l'_1 = 1 - 1784/5285 = 0.6624$. This left 3501 women alive at the start of the second interval, of whom 597 died during the interval. That is, 597 women survived

one completed year but not two completed years subsequent to diagnosis. The survival times of an additional 369 individuals were censored during the interval. All of these 369 individuals were diagnosed during the 1994 calendar year and were still alive at the end of December 1995 when patient follow-up ended. The estimated interval-specific observed survival proportion for the second annual interval is given by $p_2 = 1 - 597/3316.5 = 0.8200$. The estimated cumulative two-year observed survival proportion is then given by ${}_1p_2 = 0.6624 * 0.8200 = 0.5432$.

Table 2.2 Life table for 5285 women diagnosed with colon carcinoma in Finland during 1985–1994 and followed-up to December 31, 1995. See Table 2.3 for a description of the notation.

i	l_i	d_i	w_i	l'_i	p_i	p^*_i	r_i	${}_1p_i$	${}_1p^*_i$	${}_1r_i$	2SE(${}_1r_i$)	95% CI for ${}_1r_i$
1	5285	1784	0	5285.0	0.6624	0.9542	0.6943	0.6624	0.9542	0.6943	0.0136	0.681–0.708
2	3501	597	369	3316.5	0.8200	0.9589	0.8551	0.5432	0.9084	0.5980	0.0153	0.583–0.613
3	2535	297	276	2397.0	0.8761	0.9572	0.9153	0.4759	0.8627	0.5516	0.0164	0.535–0.568
4	1962	168	293	1815.5	0.9075	0.9542	0.9510	0.4319	0.8173	0.5284	0.0176	0.511–0.546
5	1501	117	207	1397.5	0.9163	0.9530	0.9614	0.3957	0.7727	0.5121	0.0190	0.493–0.531
6	1177	98	211	1071.5	0.9085	0.9499	0.9565	0.3595	0.7292	0.4930	0.0206	0.472–0.514
7	868	53	198	769.0	0.9311	0.9475	0.9827	0.3347	0.6868	0.4874	0.0225	0.465–0.510
8	617	40	157	538.5	0.9257	0.9450	0.9796	0.3099	0.6455	0.4801	0.0251	0.455–0.505
9	420	27	144	348.0	0.9224	0.9427	0.9785	0.2858	0.6049	0.4725	0.0287	0.444–0.501
10	249	18	129	184.5	0.9024	0.9354	0.9648	0.2579	0.5638	0.4575	0.0356	0.422–0.493
11	102	5	97	53.5	0.9065	0.9281	0.9768	0.2338	0.5220	0.4480	0.0525	0.395–0.500

Effective sample size at termination, ${}_1l^*_i$	953	(Equation 2.32)
Mean (2SE) age at diagnosis	71.35 (0.328)	
Expectation of life for the general population, e^*_1	14.08	(Equation 2.53)
Expectation of life (2SE) for the patients, e_1	7.27 (0.392)	(Equation 2.44)
Proportion of expected life lost, γ_1	48.40%	(Equation 2.57)
Patient life time median	2.64	(Section 2.23)

Table 2.3 Notation for life table quantities.

i	index for the follow-up interval (the example uses annual intervals, so $i = 1$ refers to the first year subsequent to diagnosis)
l_i	number of patients alive at the start of the ith interval. The number of patients entering the life table (the life table radix) is therefore l_1
d_i	number of deaths during the ith interval
w_i	number censored (withdrawn alive) during the ith interval
l'_i	'effective' number of patients at risk for the interval ($l'_i = l_i - w_i/2$)
p_i	estimate of the interval-specific observed survival proportion for the ith interval
p^*_i	estimate of the interval-specific expected survival proportion for the ith interval using the Ederer II method (Section 2.18)
r_i	estimate of the interval-specific relative survival ratio for the ith interval ($r_i = p_i/p^*_i$) (Section 2.13)
${}_1p_i$	estimate of the cumulative observed survival proportion from diagnosis to the end of the ith interval
${}_1p^*_i$	estimate of the cumulative expected survival proportion from diagnosis to the end of the ith interval using the Hakulinen method (Section 2.18)
${}_1r_i$	estimate of the cumulative relative survival ratio from diagnosis to the end of the ith interval (${}_1r_i = {}_1p_i/{}_1p^*_i$) (Section 2.13)
$2\mathrm{SE}({}_1r_i)$	twice the standard error of ${}_1r_i$ (Greenwood's method) (Section 2.20)
CI	confidence interval (estimate $\pm$ 2SE) (Section 2.21)
${}_gp_i$	estimate of the cumulative observed survival proportion from the beginning of the gth interval to the end of the ith interval
${}_gp^*_i$	estimate of the cumulative expected survival proportion from the beginning of the gth interval to the end of the ith interval (Section 2.18)
${}_gr_i$	estimate of the cumulative relative survival ratio from the beginning of the gth interval to the end of the ith interval (${}_gr_i = {}_gp_i/{}_gp^*_i$) (Section 2.13)

2.5.1 An improved estimator for p_i using knowledge of potential follow-up times

The assumption that deaths and censorings contribute person-time equivalent to half the interval length can be improved if we know which patients are due to be censored during the interval. For example, when constructing a life table with annual intervals for patients followed up to the end of 1995, we know that all patients diagnosed in 1994 who survive at least one year are due to be censored during the second annual interval. If these patients survive until the end of follow-up then they will each be at risk, on average, for 6 months during the second interval. If, however, they die during the second interval, then the average time at risk will be less than 6 months and it is possible to improve the estimate of the effective sample size by giving special consideration to those individuals who die during the interval in which they were due to be censored [28]. Chiang's estimator for the interval-specific observed survival proportion, which takes this into account is

$$p_i = \tfrac{1}{4}(l_i - \tfrac{1}{2}f_i)^{-2}\left\{-\tfrac{1}{2}\delta_i + \tfrac{1}{4}\left[\delta_i^2 + 4(l_i - \tfrac{1}{2}f_i)(l_{i+1} + \tfrac{1}{2}w_i)\right]^{\frac{1}{2}}\right\}^2, \qquad (2.9)$$

where δ_i is the number of deaths among those due to withdraw during interval i and $f_i = w_i + \delta_i$. As previously, the cumulative survival proportions are obtained as products of the interval-specific rates. In practice, this refinement has very little impact on the resulting estimates.

2.6 KAPLAN-MEIER ESTIMATES OF PATIENT SURVIVAL

The Kaplan-Meier method for estimating the survivor function is, in essence, the life table method where the interval size is decreased towards zero so that the number of intervals tends to infinity. Each interval is of infinitesimal length, just enough for one event or time increment. The Kaplan-Meier estimate of the survivor function at time t is then given by the product of the conditional probability of surviving each of the preceding intervals. The Kaplan-Meier method is also known as the product-limit method since it is a limit of the life table method where $S(t)$ is estimated as a product of interval-specific survival proportions. The method was first proposed by Böhmer in 1912 [29] although it had little impact until it was 'rediscovered' in the 1958 paper by Kaplan and Meier [30].

In practice, survival time is measured on a discrete scale (e.g. minutes, hours, days, months, or years) so the minimum interval length will be limited by the accuracy to which survival time is measured, rather than the theoretical intervals of infinitesimal length. For cancer registry data, survival time can meaningfully be measured, at best, to the nearest month, so intervals of length

one month are most common. In practice, it is sufficient to take the product over those intervals where a death has occurred, since the interval-specific survival proportion will be one for all other intervals and will therefore not have any influence on the product.

To obtain Kaplan-Meier estimates of the survivor function, the patient survival times are first ranked in increasing order. The times where events (deaths) occur are denoted by t_i, where $t_1 < t_2 < t_3 < \ldots$. The number of deaths occurring at t_i is denoted by d_i. If both censoring(s) and death(s) occur at the same time, then the censoring(s) are assumed to occur immediately after the death time. That is, the individuals with survival times censored at t_i are assumed to be at risk at t_i, an assumption which is discussed in Section 2.7. The Kaplan-Meier estimate of the cumulative survivor function at time t is given by

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t}(1 - \frac{d_i}{l_i}) & \text{if } t \geq t_1 \end{cases} \tag{2.10}$$

A plot of the Kaplan-Meier estimate of the survivor function takes the form of a step function, in which the survival probabilities decrease at each death time and are constant between adjacent deaths times, as in Figures 2.4 and 2.5. If the largest observed survival time (which we will call t_z) is a censored survival time, then $\hat{S}(t)$ is undefined for $t > t_z$, otherwise $\hat{S}(t) = 0$ for $t > t_z$. The standard error of the estimate can be obtained using Greenwood's formula (Section 2.20.1),

$$\text{SE}(S(t)) = S(t) \left[\sum_{t_i \leq t} \frac{d_i}{l_i(l_i - d_i)} \right]^{\frac{1}{2}}. \tag{2.11}$$

2.6.1 Kaplan-Meier estimates for the 35 patients in Table 1.2

As an example, we will use the Kaplan-Meier method to estimate the observed survivor function, $S(t)$, for the 35 patients in Table 1.2. That is, all deaths will be considered as events of interest. The cause-specific survivor function would also be of interest, and could be estimated using an identical approach except that non-cancer deaths would be regarded as censorings when estimating cause-specific survival. We will, however, demonstrate the method by estimating the observed survivor function. The patient survival times in Table 1.2 are already ranked in increasing order and are shown again in Table 2.4 (survival times for censored observations are suffixed with a +) along with Kaplan-Meier estimates of $S(t)$. Although the product in Equation 2.10 can be taken over only those values of t at which an event (death) occurs, it is necessary to keep track of all censorings, as they contribute in Equation 2.10 by decreasing l_i, the number of persons at risk at the time of each event.

Table 2.4 Kaplan-Meier estimates of observed survivor function, $S(t)$, for the sample of 35 patients diagnosed with colon carcinoma in Finland (Table 1.2). The conditional survival probabilities are denoted by p $(= 1 - d/l)$. The standard error (SE) of $S(t)$ is estimated using Greenwood's method.

t	No. at risk (l)	Observed deaths (d)	p	$S(t)$	SE
0	35	0	1.0000	1.0000	–
2					
2	35	2	0.9429	0.9429	0.0392
3	33	1	0.9697	0.9143	0.0473
5	32	1	0.9688	0.8857	0.0538
7	31	1	0.9677	0.8571	0.0591
8	30	1	0.9667	0.8286	0.0637
9	29	1	0.9655	0.8000	0.0676
11	28	1	0.9643	0.7714	0.0710
13+	27	0			
14+	26	0			
19	25	1	0.9600	0.7406	0.0745
22	24	1	0.9583	0.7097	0.0776
25+	23	0			
27	22	1	0.9545	0.6775	0.0805
27+	21	0			
28	20	1	0.9500	0.6436	0.0833
32	19	2	0.8947	0.5758	0.0872
32+	17	0			
33	16	1	0.9375	0.5398	0.0889
35+	15	0			
37+	14	0			
43	13	1	0.9231	0.4983	0.0912
46	12	1	0.9167	0.4568	0.0926
54+	11	0			
77+	10	0			
78+	9	0			
83+	8	0			
85+	7	0			
97+	6	0			
100+	5	0			
102	4	1	0.7500	0.3426	0.1208
103	3	1	0.6667	0.2284	0.1232
105+	2	0			
108+	1	0			

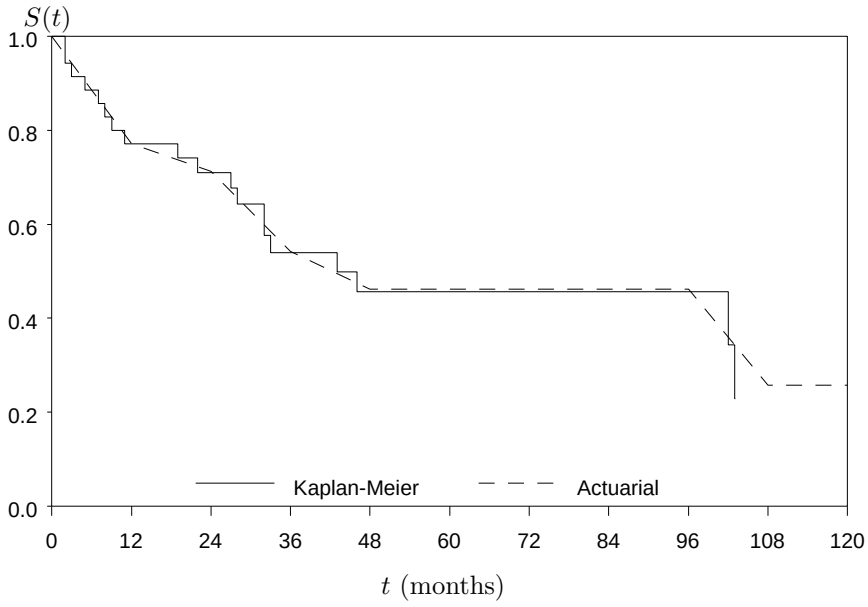

Fig. 2.4 Estimates of the cumulative observed survivor function using the actuarial method (annual intervals) and the Kaplan-Meier method (using survival time in months) for the sample of 35 patients diagnosed with colon carcinoma in Finland (Table 1.2).

The Kaplan-Meier estimates are graphed as a step function, with vertical drops at each death time (Figure 2.4). This curve is sometimes called a 'Kaplan-Meier curve'. The actuarial method provides estimates of the survivor function at the end of each interval, and no estimate of the survivor function is made between these points. It is customary to interpolate the estimates by 'joining the dots', which corresponds to an approximately even distribution of deaths within each interval. Some authors [1] have chosen to graph actuarial estimates as step functions, which misrepresents the data as all of the risk of death is then concentrated at the end of the intervals.

2.7 COMPARISON OF THE KAPLAN-MEIER AND ACTUARIAL METHODS

Estimates of the survivor function can be presented in either tabular or graphical form. For tabular presentations, we rarely require estimates of the survivor function for interval lengths shorter than one year so the actuarial method

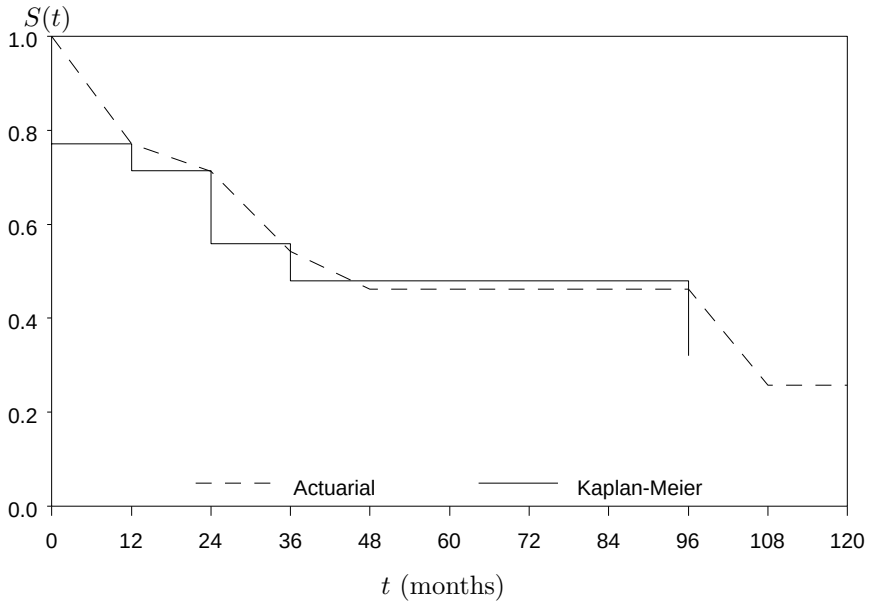

Fig. 2.5 Estimates of the cumulative observed survivor function using the actuarial method (annual intervals) and the Kaplan-Meier method (using survival time in years) for the sample of 35 patients diagnosed with colon carcinoma in Finland.

suffices. We will therefore focus on a comparison of the two methods where the aim is to present graphical estimates of $S(t)$. Since both the Kaplan-Meier and actuarial methods aim to estimate the same quantity, the survivor function, we would expect the estimates to be similar, if not identical. The estimates in Figure 2.4 are similar, although not identical. If the Kaplan-Meier estimates are made using survival time in years (rather than months) as the basis of calculation we see even greater disparity between the two methods (Figure 2.5).

One reason for this difference is the way in which ties (individuals with identical survival times) are handled. If both censoring(s) and death(s) occur at the same time, the Kaplan-Meier method assumes that all of the individuals with censored survival times were at risk at the time of the death(s), whereas the actuarial method assumes that half of these individuals were at risk at the time of the death(s). There is little reason to assume that all deaths should precede all concurrent withdrawals when time is measured on a discrete scale (e.g. in days or longer units). For example, if 27 deaths and 18 censorings occur in a given month of follow-up of a cohort, it is unlikely that all of the deaths occurred before all of the censorings, as is assumed by the Kaplan-Meier method. It would be more reasonable to apply the actuarial

assumption, which assumes that, on average, half of the censored individuals were at risk at the time of each deaths. This would require only a simple modification to Equation 2.10 and would result in the Kaplan-Meier and actuarial methods returning identical estimates when the same time units were used as the basis for each method (such as in Figure 2.5 where survival time in years is the basis for each method). There are no censorings during the first 12 months so the two estimates of $S(12)$ are identical.

The Kaplan-Meier estimator was developed for applications where survival time is measured on a continuous scale, where ties are rare, although the Kaplan-Meier estimator is discrete in nature. Many of the standard methods for survival analysis, such as the Cox proportional hazards model, assume that survival time is measured on a continuous scale and that ties are therefore rare. In population-based survival analysis, however, ties are common. For example, among the 9087 Finns diagnosed with colon carcinoma during 1985–1994, 490 died during the first month of follow-up and 542 during the second month of follow-up (although there were no censorings during these months since every individual had a potential follow-up of at least 12 months). Since survival time cannot realistically be measured with greater accuracy than in months, and sample sizes are often large, ties in the data cannot be avoided. The actuarial method takes account of ties through application of the actuarial assumption, whereas the Kaplan-Meier method overestimates the survivor function (to a very small degree) in the presence of ties by assuming that all deaths precede concurrent withdrawals. This issue of ties is discussed in more detail in Section XX where the various approximations used for estimating the Cox proportional hazards model in the presence of ties are discussed.

Another consequence of the Kaplan-Meier method being developed for continuous time and the actuarial method for grouped data is in the way the estimates are interpreted. The actuarial method provides estimates of the survivor function at the end of each interval, and no estimate of the survivor function is made between the interval endpoints. Estimates of $S(t)$ for values of t between the interval endpoints are obtained by interpolation (as in Figures 2.4 and 2.5), which corresponds to assuming an approximately even distribution of deaths within each interval. This assumption, however, may not always be valid. During the first year of follow-up, for example, mortality is often highest during the first few months, and the first month in particular. The Kaplan-Meier method, on the other hand, provides an estimate of $S(t)$ for all values of t, although the estimate of $S(t)$ is constant between event times. The interpretation of the Kaplan-Meier estimate of $S(t)$ presented in Figure 2.5 is that all patients survive up to 11 months, after which over 20% suddenly die, indicating that the Kaplan-Meier method is clearly inappropriate for heavily grouped data.

For data which is not heavily grouped, such as when survival time is estimated to the nearest month, the Kaplan-Meier method could be considered superior

to the actuarial method with annual intervals. However, the actuarial method based on monthly intervals is an even better alternative, since it accounts for the ties in the data through use of the actuarial assumption. The actuarial method based on annual intervals been popular in population-based survival analysis since it requires fewer arithmetic calculations than would be required for the Kaplan-Meier method with survival time measured in months. With the advent of computers, however, this is no longer a significant advantage and a method based on survival time measured in months is preferable for producing graphical estimates of $S(t)$. Either the Kaplan-Meier method or the actuarial method will suffice, although the actuarial method is preferable if the data contain ties.

It is possible, of course, to use the actuarial method with intervals of varying length, such as shorter intervals during the first year of follow-up, although this is often difficult to implement in standard software. In practice, it is usually simpler to produce a table of estimates of $S(t)$ based on annual intervals and then a separate analysis to obtain graphical estimates based on monthly intervals (using either the actuarial Kaplan-Meier method). We did previously mention a slight bias in the Kaplan-Meier method in the presence of ties, although this can be minimised by ensuring that every individual has a potential follow-up of at least one year. The bias can be avoided completely by using the actuarial method with monthly intervals.

2.8 ESTIMATING PATIENT SURVIVAL FROM POPULATION-BASED CANCER REGISTRY DATA

Survival analysis refers to the techniques used to analyse the time to occurrence of some event for a given population. When studying population-based cancer patient survival, survival time is usually measured from the date of diagnosis to the date of death (Section 5.5). Ultimately, we would like to describe the survival experience of the population in terms of prognostic factors such as age, sex, calendar year of diagnosis, stage, and region of residence. We may or may not wish to adjust for mortality due to causes other than the cancer of interest and such adjustment can be done in different ways. As such, several different measures are available for estimating cancer patient survival which vary in their suitability to different applications. The three basic measures of cancer patient survival (observed, cause-specific, and relative survival) are described in the following sections.

The methodology for analysing patient survival from population-based cancer registry data is distinct from other types of survival analysis (such as clinical trials or engineering applications) due to the large number of cases involved (often exceeding 5,000 cases for a single cancer site) and the long follow-up times involved (usually at least 5 years and sometimes exceeding 20 years).

Consequently, changes in the general mortality experienced by the patients as a result of aging often need to be considered.

2.9 OBSERVED SURVIVAL

The most basic measure of the survival experience of persons diagnosed with cancer is the observed survival proportion from the date of diagnosis to time i, denoted by ${}_1p_i$ and estimated from a life table for the patients [20, 21, 24, 28].

The observed survival proportion is not a particularly useful measure for making comparisons between patient groups since not all deaths will be due to the cancer in question. In particular, elderly patients are more likely than younger patients to die of causes unrelated to the cancer of interest, the so-called competing risks of death. Sections 2.12 and 2.13 describe two commonly used methods for estimating cancer patient survival adjusted for competing risks. The resulting measures, the cause-specific survival rate and the relative survival ratio, both estimate net survival, the survival proportion where the disease in question is assumed to be the only possible cause of death.

2.10 THE HAZARD FUNCTION

The hazard function, which was introduced in Section 2.1, is the central concept in the statistical modelling of patient survival. In practice, the hazard function is modelled when it is desired to study simultaneously the effect of several prognostic factors upon survival (e.g., the Cox proportional hazards model [31]). The cumulative survivor function has the same basic shape (a nonincreasing function from 1 to 0) for all types of data and the hazard function is generally a more informative means of studying differences between patient groups. In the field of epidemiology, the empirical hazard is called the incidence rate and, when death if the outcome, the incidence rate of death, often called the mortality or death rate of the patients. In order to emphasise that we are measuring the mortality of the patients and not, for example, the general population, it is often called the fatality rate or case fatality rate.

The general definition of the empirical hazard is the ratio between the number of deaths and the number of person-time units (e.g. person-years) at risk in the group under observation. As an example, consider the 35 patients in Table 1.2. There were 35 patients at risk at the beginning of the first interval and eight of them died during the interval. By assuming that the deaths occurred, on average, at the midpoint of the interval, the number of person-years at risk during the interval would be $35 - 8 \times 0.5 = 31$. The empirical hazard (hazard estimate) would then be 4/31 person-years = 0.13 per person-year, or 13 per

100 person-years. Note that, unlike the survival proportion, the hazard rate always has a unit of measurement, in this case the person-year. The units are often scaled by a factor of ten and the hazard rate reported as, for example, deaths per 100, 1000, or 100000 person-years.

Since more exact data on the patients are available in Table 1.2, it is possible to improve the calculation on person-years at risk. The total number of person-months lived by the eight patients who died in the first interval was $2+2+3+5+7+8+9+11=47$ making $47/12=3.83$ person-years. A more precise estimate of the number of person-years at risk is therefore $(35-8)+3.83=30.83$, rather close to the first approximation, 31 person-years.

When continuing in the second annual interval, we must also consider the contributions of the censored observations to the person-years. In general, the empirical estimate of the hazard rate is

$$\lambda_i = \frac{d_i}{(l_i' - d_i/2)\Delta_i}, \tag{2.12}$$

where Δ_i is the length of the interval. That is, the hazard rate is estimated as the number of deaths in the interval divided by the number of person-years at risk during the interval, where it is assumed that patients who die or are censored are at risk for half the length of the interval.

In practice, it is more convenient in survival analysis to deal with a quantity which does not have a unit of measurement. This is the interval-specific cumulative hazard, which is the empirical hazard multiplied by the length of the interval. The hazard estimate could be refined in a similar way to the interval-specific survival estimate in Section 2.5.1 to allow for the fact that patients who die during the interval in which they were due to be censored will be at risk for, on average, one quarter, rather than one half of the length of the interval. In practice, this refinement will have little effect on the estimates.

Table 2.5 Interval-specific cumulative hazards (Λ_i), cumulative hazards from diagnosis ($_1\Lambda_i$), and a comparison of count-based (c) and hazard-based (h) estimates of annual (p_i) and cumulative ($_1p_i$) survival proportions for the patients in Table 2.2

i	l_i	d_i	w_i	Λ_i	$_1\Lambda_i$	$p_i(c)$	$p_i(h)$	$_1p_i(c)$	$_1p_i(h)$
1	5285	1784	0	0.4061	0.4061	0.6624	0.6662	0.6624	0.6662
2	3501	597	369	0.1978	0.6039	0.8200	0.8205	0.5432	0.5467
3	2535	297	276	0.1321	0.7360	0.8761	0.8763	0.4759	0.4790
4	1962	168	293	0.0970	0.8330	0.9075	0.9075	0.4319	0.4347
5	1501	117	207	0.0874	0.9204	0.9163	0.9163	0.3957	0.3984
6	1177	98	211	0.0958	1.0162	0.9085	0.9086	0.3595	0.3619
7	868	53	198	0.0714	1.0876	0.9311	0.9311	0.3347	0.3370
8	617	40	157	0.0771	1.1648	0.9257	0.9258	0.3099	0.3120
9	420	27	144	0.0807	1.2455	0.9224	0.9225	0.2858	0.2878
10	249	18	129	0.1026	1.3481	0.9024	0.9025	0.2579	0.2597
11	102	5	97	0.0980	1.4461	0.9065	0.9066	0.2338	0.2355

The life-table estimate of the interval-specific cumulative hazard for the ith interval is

$$\Lambda_i = \frac{d_i}{(l_i' - d_i/2)}. \tag{2.13}$$

Note that Λ_i is the interval-specific cumulative hazard from the start of interval i to the end of interval i. The hazard rate for a point during interval i may be estimated using Equation 2.12. The estimated cumulative hazard from diagnosis to the end of interval i is $_1\Lambda_i = \sum_i \Lambda_i$. The interval-specific cumulative hazards and the cumulative hazards from diagnosis are shown in Table 2.5 for the patients in Table 2.2.

Two methods exist for estimating the cumulative survival proportion ($_1p_i$):

1. Directly based on the counts in the life table; or
2. Through the hazard using Equation 2.4

$$\begin{aligned} {}_1p_i &= \exp(-\sum_i \Lambda_i) \\ &= \exp(-\sum_i \frac{d_i}{(l_i' - d_i/2)}) \\ &= \prod_i \exp(-\frac{d_i}{(l_i' - d_i/2)}) \\ &= \prod_i p_i \end{aligned} \tag{2.14}$$

These two methods give similar, although not equivalent, results. The disparity between the methods becomes smaller as the survival proportions become larger. An empirical comparison of the two methods is shown in Table 2.5 for the patients in Table 2.2.

Equivalently, there are two methods available for estimating the interval-specific cumulative hazard - directly based on counts or on the basis of the interval-specific survival proportions. The latter utilises the relationship $\Lambda_i = -\ln p_i$. In practice, the results of the two methods are very similar.

Standard practice in analysis of cancer patient survival, which we will follow-in this text, is to estimate survival probabilities based on counts and derive the cumulative hazards from the survival probabilities. In the analysis of prognostic factors in Chapter 3, the counts for cumulative hazards form the basis of Poisson regression modelling, whereas the counts for survival probabilities make the basis of the binomial models. In practice, the differences in results between the two methods are generally small.

2.11 NET SURVIVAL

The definition of net survival is not consistent in the statistical and epidemiological literature. In this text, the term net survival proportion at time t is defined as the proportion of patients who would have survived t years or more following diagnosis in the hypothetical situation where the disease of interest were the only possible cause of death. It is a hypothetical quantity which can be estimated using, for example, the cause-specific survival rate (Sections 2.12) or the relative survival ratio (Sections 2.13).

This definition of net survival is consistent with the use of the term 'net probability' in the theory of competing risks [32, 33]. The study of competing risks refers to the study of mortality patterns in a population of individuals, all subject to the same $k \geq 2$ competing risks or causes of death. The study of population-based cancer patient survival can be viewed in this framework. Each individual diagnosed with cancer can potentially die of any one of thousands of different causes. We are often interested in estimating the risk of death due to the specific cancer with which the patient has been diagnosed, but are generally not interested in studying the effect of any other causes of death. In this text, we will use the term 'competing risks' to refer to all causes of death other than death due to the specific cancer of interest. When studying, for example, the survival of patients diagnosed with cancer of the colon, lung cancer is classed as a 'competing risk' together with, for example, cardiovascular disease and motor vehicle accidents. The net survival probability can therefore be defined as the survival probability which would be observed

in the cohort in the hypothetical situation where the cancer of interest were the only possible cause of death (i.e. if competing risks were eliminated).

Although cause-specific survival and relative survival were originally developed as a means of estimating net survival, these two measures have other important uses and interpretations. They also have theoretical counterparts not necessarily relating to a hypothetical elimination of competing risks. The cause-specific survival proportion, for example, is a summary measure of cause-specific mortality in the presence of competing risks. The relative survival ratio is a summary measure of the excess mortality of the patients in the presence of competing risks. As these summary measures do not refer to any hypothetical situation it might be more useful to think of the estimates in these terms rather than more hypothetical quantities.

As mentioned previously, net survival is a hypothetical concept. Net survival times of individuals are not observable so estimates of net survival must be based on the observed survival times. The methods for estimating net survival presented in this text, cause-specific survival and relative survival, assume independence between mortality due to the cancer of interest and mortality due to competing risks. This assumption of independence is conditional on the covariates (age, sex, and calendar period) upon which competing risk mortality is known to depend. That is, 70 year old men diagnosed with localised colon cancer are assumed to be subject to the same competing risks mortality as 70 year old men diagnosed with metastatic colon cancer during the same calendar period. This assumption is discussed in more detail in Section 2.16.

2.12 CAUSE-SPECIFIC SURVIVAL

Cause-specific survival proportions can be used to estimate the probability of surviving a given cancer, independent of mortality due to competing risks. The term corrected survival proportion is sometimes used in place of cause-specific survival proportion, although we discourage this terminology, since survival proportions can be corrected for many different factors in many different ways. For example, age-standardised rates could be considered corrected. When estimating cause-specific survival proportions, only those deaths which can be attributed to the cancer in question are counted as deaths, while all other deaths are treated as censorings. For the sample data sets described in Appendix A, patients with status codes 0,2, and 4 are all considered censored when estimating cause-specific survival, with only patients with status=1 considered as outcome events. When estimating observed survival proportions (Section 2.9), patients with status codes 1 and 2 are counted as outcome events, while the survival times of patients with status codes 0 and 4 are considered censored.

The estimation of cause-specific survival proportions requires reliably coded information on cause of death. Even when cause of death information is available to the cancer registry via death certificates, it is often vague and difficult to determine whether or not cancer is the primary cause of death (Section 5.4). It is unclear how to classify, for example, deaths due to suicide, or to the secondary effects of treatment for the cancer in question. Another problem is that the cause of death is sometimes recorded as cancer in the distant site, rather than the primary site. For example, if colon cancer has metastasised to the liver, cancer of the liver may sometimes be recorded as the cause of death, instead of the correct classification of cancer of the colon.

Another problem with cause-specific survival is that it assumes that cancer mortality is independent of competing risk mortality, an assumption which is often only approximately true, although stratification by age will usually prevent serious problems of this type (Section 2.14). Cause-specific survival is not as well suited to data from population-based cancer registries as it is to clinical studies where more effort is made to distinguish between deaths due to the disease under study and deaths due to competing risks. Cause-specific survival is, however, useful for the estimation of social class specific cancer patient survival from cancer registry data, since social class-specific general population life tables are often not available for the calculation of relative survival ratios (Section 2.13) and it is thought that cause-specific survival provides a suitable method of adjusting for social class differences in general mortality.

2.13 RELATIVE SURVIVAL

Relative survival has become the preferred measure for the analysis of patient survival based on data from population-based cancer registries. A major advantage of this measure is that information on cause of death is not required, thereby circumventing problems with the inaccuracy [23] or nonavailability of death certificates. The relative survival ratio is defined as the observed survival in the patient group divided by the expected survival of a comparable group from the general population, matched to the patients with respect to the main factors affecting patient survival and assumed to be practically free of the cancer of interest [22]. It is usual to estimate the expected survival proportion from nationwide population life tables stratified by age, sex, calendar time, and, where applicable, race [24]. Although these tables include the effect of deaths due to the cancer being studied, Ederer et al. [22] showed that this does not, in practice, affect the estimated survival proportions.

An estimate of the expected survival proportion is required before the relative survival ratio can be estimated, although we have chosen to first discuss the relative survival ratio before presenting the methodology for estimating

expected survival, which is introduced in Section 2.18. The expected survival proportion is, in principle, a simple concept and the methodology used for its calculation is not necessary to the understanding of the relative survival ratio. In practice, the estimation of the expected survival proportion is non-trivial and several different methods have been proposed. In keeping with our aim of describing cancer survival analysis methodology in an applied setting, we have chosen to first discuss the relative survival ratio, although some readers may wish to first study the methodology for estimating expected survival in Section 2.18.

The cumulative relative survival ratio (RSR) from diagnosis to the end of the ith interval, denoted by ${}_1r_i$, is calculated as the ratio of the observed survival proportion ${}_1p_i$ to the expected survival proportion ${}_1p_i^*$. Examples of these calculations are shown in Table 2.2. The cumulative relative survival ratio can be interpreted as the proportion of patients alive after i years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death. The validity of this interpretation is dependent on the accurate and appropriate estimation of the expected survival proportion and the assumption that non-cancer mortality is independent of cancer mortality (Section 2.14). Even without the independence assumption, the RSR can be interpreted as the ratio between the observed and expected proportions of survivors, the latter being based on the survival of the general population [34]. The strength of the relative survival ratio is that it provides a measure of the excess mortality experienced by patients diagnosed with cancer, irrespective of whether the excess mortality is directly or indirectly attributable to the cancer. Deaths due to treatment complications or suicide are examples of deaths which may be considered indirectly attributable to cancer. When estimating cause-specific survival, it is necessary to ascribe each death as being 'due to the cancer of interest' or 'not due to the cancer of interest'. It is not possible to allow for the fact that cancer may be a partial contributing factor, as can be done in relative survival analysis.

2.13.1 Interpreting relative survival estimates

As mentioned in the previous section, the cumulative relative survival ratio, ${}_1r_i$, can be interpreted as the proportion of patients alive after i years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death. The validity of this interpretation is dependent on the accurate and appropriate estimation of the expected survival proportion and the assumption that non-cancer mortality is independent of cancer mortality. When estimating relative survival, all mortality experienced by the patients in excess of the expected mortality, which is usually based on general population mortality, is attributed to the cancer. If not all of the excess mortality is due to the cancer, then the relative survival ratio will underestimate the true net

survival proportion. For example, patients diagnosed with smoking-related cancers will experience excess mortality, compared to the general population, due to both the cancer and other smoking related conditions.

Should the patients be a selected group from the general population, for example, with respect to social class, a general population group might not be an appropriate group for comparison. In this case, the relative survival ratio would be too high or low, depending on the composition of the group. If all mortality in excess to the baseline mortality in the general population is attributable to cancer then no problems arise due to dependence between causes of death. Otherwise, one solution is to correct the general population mortality by accounting for social class or region [35, 36].

The life table is a useful tool for describing the survival experience of the patients over a long follow-up period. In particular, an interval-specific relative survival ratio equal to one indicates that, during the specified interval, mortality in the patient group was equivalent to that of the general population. The attainment and maintenance of an interval-specific RSR of one indicates that there is no excess mortality due to cancer and the patients are assumed to be 'cured'. Graphs of the (usually annual) interval-specific RSRs can be used to ascertain the 'point of statistical cure' [37]. In Table 2.2 the annual relative survival ratios, r_i, increase with follow-up time but do not reach a value of 1.00 during the eleven years of follow-up. The annual relative survival ratio during the 11th year of follow-up of 0.977 indicates a 2.3% annual excess mortality rate. Since all estimates are subject to random error, with the effect being even larger during latter intervals, one should study the general pattern in the interval-specific RSRs rather than a single estimate. The concept of statistical cure, along with methods for estimating the proportion cured and the cure point, is discussed in Section 5.12.

An interval-specific relative survival ratio of unity is generally not achieved for smoking-related cancers, such as cancers of the lung and kidney, since, compared to the general population, the patients are subject to excess mortality due to other conditions caused by smoking, such as cardiovascular disease, in addition to excess mortality due to the cancer. The problem is that the general population is not strictly comparable to the patients with respect to mortality due to causes other than cancer. The choice of an appropriate comparison group for estimating expected survival proportions also becomes an issue when relative survival ratios are estimated for socioeconomic groups and is discussed in Chapter 6.

It is also possible that the mortality rates of the general population overestimate the non-cancer mortality of the patients. This can occur when the cancer in question is more common in the higher socioeconomic classes, or due to the 'healthy patient effect'. The 'healthy patient effect' is sometimes observed among patients diagnosed with low-fatality cancers, such as localised melanoma of the skin. Cumulative relative survival ratios greater than one

can be observed for such patient groups since very few patients die of the cancer and they experience lower mortality due to other causes as a result of having greater than average contact with the health system. The exact causal mechanism governing this is, however, impossible to determine; it is possible, for example, that patients who have regular health checks, and are accordingly in better health, are more likely to have melanoma diagnosed at an earlier stage, or it may be that the contact with the health system as a result of being diagnosed with melanoma has a positive effect on a patient's general health. Isolated cases can be ascribed to random variation, but it may also occur that those patients who survive the cancer may be systematically healthier than a comparable group from the general population of a similar age and sex.

The most likely reason, however, for RSR estimates greater than one is poor quality follow-up. Patient survival will be overestimated if deaths among the cancer patients are not registered, which can result in RSR estimates greater than one. As such, studying interval specific relative survival ratios is a useful method for examining the quality of follow-up (Section 5.3).

Figure 2.6 shows some further examples of interval-specific relative survival ratios extracted from a recent study of cancer patient survival in Finland [19]. The figure on the left shows that excess mortality due to cancer of the stomach is very high during the first year of follow-up and decreases with each additional year survived. The interval-specific RSR is equal to 1 for the eighth year of follow-up, indicating no excess mortality due to a diagnosis of stomach cancer was experienced by the patients during this period. It can therefore be assumed that patients who survive 7 years following a diagnosis of stomach cancer in Finland can be considered cured. The pattern of excess mortality observed for stomach cancer is typical of the shape observed for most cancer types, where excess mortality is highest during the first year and decreases with each year survived. For most cancer types, however, the excess mortality during the initial years is not as high as that observed for stomach cancer. Cancer of the female breast is, along with cancer of the prostate, one of two notable exceptions where a relatively constant level of excess mortality is observed throughout the follow-up [38]. A similar level of excess mortality is observed for breast cancer for up to 20 years of follow-up and longer, although estimates of long-term survival lack precision due to small numbers of cases.

Figure 2.7 shows corresponding plots of the cumulative relative survival ratios for the data shown in Figure 2.6. Note how the plots is constructed by 'joining the dots', which is the appropriate method for plotting life table estimates of patient survival, as opposed to Kaplan-Meier estimates of patient survival, which should be plotted as a step function. Plots of cumulative relative survival ratios are useful for showing, for example, the cumulative probability of surviving up to a given point in time, but it is more difficult through studying these plots to ascertain the pattern of excess mortality due to the cancer.

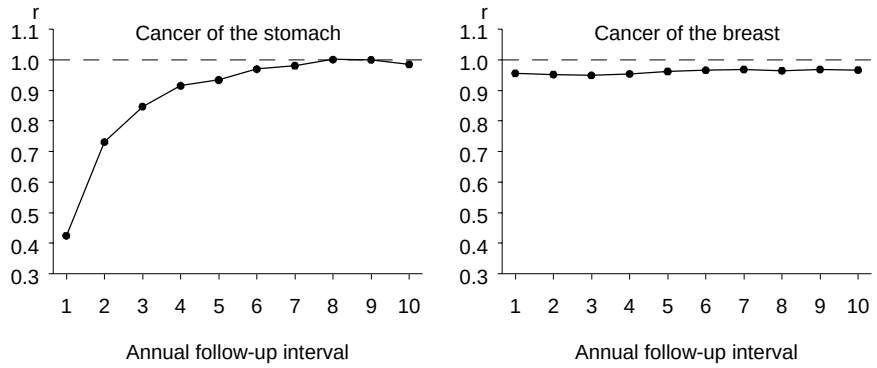

Fig. 2.6 Plots of the annual (interval-specific) relative survival ratios (r) for cancer of the stomach (both sexes) and cancer of the female breast diagnosed in Finland 1985–1994 and followed up to the end of 1995.

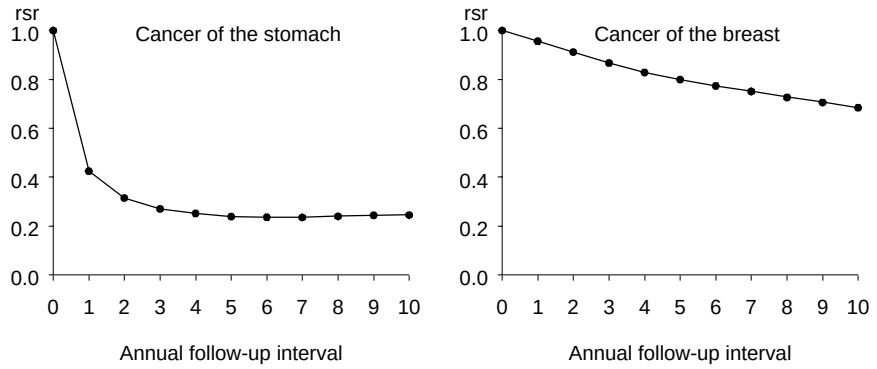

Fig. 2.7 Plots of the cumulative relative survival ratios (rsr) for cancer of the stomach (both sexes) and cancer of the female breast diagnosed in Finland 1985–1994 and followed up to the end of 1995.

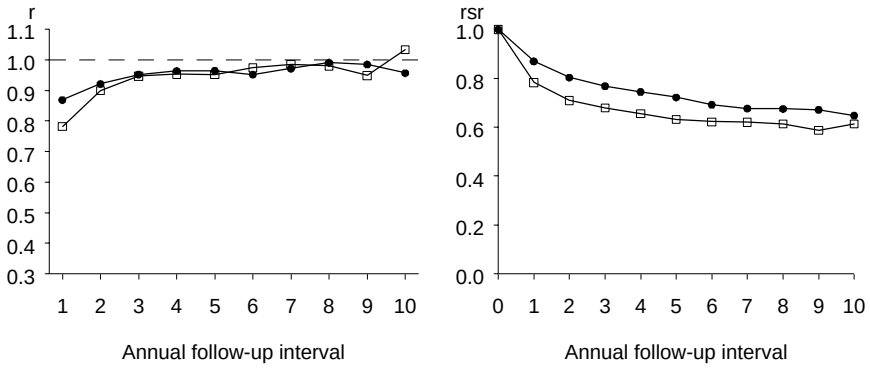

Fig. 2.8 Plots of the interval-specific (r) and cumulative (rsr) relative survival ratios for cancer of the urinary bladder for males (●) and females (□) diagnosed in Finland during 1985–1994 and followed up to the end of 1995.

The plot of cumulative relative survival ratios will 'flatten' when the interval-specific rates are equal to one and this criteria is sometimes used to identify the cure point. It is especially difficult to study differences in the pattern of excess mortality between patient groups by studying plots of cumulative survival proportions. From the plot of interval-specific relative survival ratios for cancer of the urinary bladder shown in Figure 2.8, it is clear that females experience lower survival (higher excess mortality) than males during the first year of follow-up, but there is little difference in survival between the two sexes during subsequent years. In contrast, this pattern is not so easily discernible from the plot of the cumulative relative survival ratios (Figure 2.8).

The cumulative RSR after five years of follow-up is often used as a single measure of the survival experience of a population. This is called the five-year relative survival ratio and is interpreted as the proportion of patients alive after five years of follow-up in the hypothetical population where the cancer in question is the only possible cause of death. For the example shown in Table 2.2, the five-year relative survival ratio is 0.5121. The five-year relative survival ratio is, however, sometimes erroneously interpreted as the proportion of patients cured. Such an interpretation would require the cure point to be reached before 5 years of follow-up, which is often not the case (Figures 2.6 and 2.8, for example).

It is possible to compare survival in different patient groups by tabulating or graphing the five-year RSR. Although it is possible to see general trends (e.g. RSR decreases with increasing age), these methods do not provide conjoint information on the impact of factors such as age, sex, period of diagnosis, stage, and histology on survival. The use of regression models enables estima-

tion of the effect of each factor on survival while simultaneously controlling for all other factors (Chapter 3).

2.13.2 What's in a name? Relative survival ratio or ratio?

In the strictest sense, a *ratio* is the result of dividing one quantity by another. In the sciences, however, it is mostly used in a more specific sense, that is, when the numerator and the denominator are two separate and distinct quantities [39]. A *proportion* is a type of ratio in which the numerator is included in the denominator, while a *rate* is a measure of change in one quantity (y) per unit of another quantity (x) on which y depends. The 'survival proportion' of a group of patients over a specified time period is therefore not strictly a rate, but a proportion. Similarly, the 'relative survival proportion' ($r = p/p^*$) is not actually a rate, but a ratio of two proportions. Since the terms survival proportion and relative survival proportion belong to standard terminology, they are employed here instead of the correct terms.

2.14 CENSORING

We will now return to the concept of censoring, which was introduced in Section 2.2. The survival time of an individual is said to be censored when we do not know the exact time from the start of follow-up to the occurrence of the event of interest. In population-based survival analysis we generally know the date of start of follow-up (date of diagnosis) and censoring occurs when we are not able to follow-up the patient for a sufficient length of time to observe the event of interest. That is, we are able to follow up the patient to time t without observing the event of interest, so know only that survival time is greater than t. This type of censoring is known as right censoring and the three main reasons for its occurrence in population-based survival analysis were listed on page 27. All of the methods presented in this text were developed for estimating patient survival in the presence of censoring. These methods require the important assumption that, conditional on the prognostic factors, censoring is independent of the outcome event (Section 2.16).

Another form of censoring is interval-censoring, which occurs when we know that the event of interest has occurred between two time points but do not know the exact date, for example, HIV infection between two test dates. We will only consider right-censoring in this text as other forms of censoring are rarely encountered in the analysis of population-based cancer survival. Further information on other forms of censoring can be found in most survival analysis texts (the text by Klein and Moeschberger [6] covers this extensively) or the paper by Leung et al. [25].

2.15 TRUNCATION

Survival data may also be subject to truncation, a characteristic which is sometimes confused with censoring. Censoring refers to the situation where we are unable to ascertain the exact survival time (from the time origin to the event time) for individuals in the study. Truncation, on the other hand, refers to the situation where we may be unaware of the existence of certain individuals who should have been part of the study (and consequently are also unaware of their survival times). Truncation is generally not an issue in the analysis of population-based cancer registry data, although we will briefly discuss the issues here in order to highlight the strengths of population-based data and some of the potential problems with survival data which are not truly population-based.

Left truncation occurs when individual are observed only if they are event free after a certain follow-up time. Population-based cancer registry data are never subject to left truncation since all newly diagnosed cases of cancer are reported to the registry. Left truncation can occur in clinical survival studies due to, for example, late entry to the study. In order to enter the study late, an individual must survive from the time origin up to the entry date and the methods used for analysing the study must take account of this.

Right truncated data occur when only individuals who experience the event of interest are included in the study. For example, using the mortality register to identify all individuals who died of prostate cancer during a given calendar period and then ascertaining the date of diagnosis (and hence survival time) from medical records would be an example of a study with right-truncated data. Special methods of analysis are required, such as use of a conditional likelihood or a method which uses a selective risk set (see Klein & Moeschberger (1997) [6]). Nevertheless, there are examples in the literature where right truncated cancer survival data have been analysed using standard methods (see Altman and Bland (1998) [40] for some examples). Such an approach, for example, led to the dubious finding that left handed people die, on average, seven years younger than right handed people [41]. Those individuals in the cohort who died, for example, while aged in their 70s, belong to a birth cohort where left handedness was less prevalent since many natural left handers were forced to become right handed. Those who died, for example, during their 20s, would be much more likely to be left handed (than those who died during their 70s). A naive method of analysis therefore gave the erroneous impression that left handed people die younger.

[HIV example]

2.16 INDEPENDENCE OF CENSORING AND THE OUTCOME EVENT

The estimation of observed, cause-specific, and relative survival all require the assumption that censoring is independent of the outcome event. That is, the probability of being withdrawn from the study is independent of the risk of death due to the outcome event of interest at the time of withdrawal. This is known as noninformative censoring or random censoring. When censoring is associated with the outcome event it is called informative censoring and results in biased estimates of survival proportions. In order to illustrate the necessity of the noninformative censoring assumption, consider the life table shown in Table 2.2. The estimate of the 10-year cumulative observed survival proportion is ${}_1p_{10} = 0.2579$, which is constructed as the product of the interval-specific rates, $p_1 \dots p_{10}$. The estimated 10-year survival proportion is interpreted as the estimate for the entire cohort of 5285 women, yet the estimates of the interval-specific rates are based on only those women who were alive and under follow-up at the start of the interval. For this interpretation to be valid, we require that each interval-specific estimate be representative of the survival proportion we would have observed is we were potentially able to follow-up all women in the cohort for 10 years. That is, for the women censored at time t, we assume that their survival experience after time t can be approximated by the survival experience of the women who could be followed up after time t. If the prognosis of censored individuals is systematically different from those who remain under follow-up (i.e. censoring is informative) then the survival experience of those who remain will not be representative of the entire cohort and survival estimates will be biased.

If censoring is noninformative then there is no reason to differentiate in the analysis between the different reasons for censoring listed on page 27. We will now explore each of the reasons for censoring and whether or not it is reasonable to assume it is noninformative.

2.16.1 Censoring due to close of follow-up

This is the most common form of censoring in population-based survival analysis and is usually noninformative. Censoring due to close of follow-up can, however, be informative if the age range of patients in the cohort is wide, although this problem can be easily overcome by stratifying the patients by age and estimating the survival proportion within each stratum. As an example, consider the task of estimating observed survival in a cohort of cancer patients containing both young and old patients at diagnosis. Patients are recruited over a number of years and follow-up is terminated at a common closing date. Old patients become (relatively) more common over time due to aging of the population. This leads to a situation in which the old patients

have, on average, shorter potential follow-up times than the young patients. When the survival times of the elderly patients are censored, their subsequent survival experience will be predicted by that of patients with longer potential follow-up time, i.e., a group where young patients are overrepresented. Consequently, the estimated overall observed survival proportion overestimates the rate which would have been observed if we were able to follow-up all individuals until death.

This problem can be easily overcome by stratifying the patients by age and estimating the survival proportion within each stratum. An overall estimate of the survival proportion is then obtained by taking a weighted average of the stratum-specific estimates with weights proportional to the number of patients at the beginning of follow-up in each stratum. In practice, however, age is a continuous variable and no age strata are completely homogeneous. Experience has shown that four age strata are sufficient in most practical applications.

In order to further explore the implications of this independence assumption, consider a group of 70 year old men diagnosed with localised colon cancer. The general health at diagnosis of these men will differ. During the subsequent year, some of the men will die of colon cancer and some will die due to competing risks. Is it reasonable to assume that the risk of death due to colon cancer is independent of the risk of death due to competing risks? It could be argued that the men whose general health is poorer at diagnosis will be more likely to die of competing risks and also more likely to die due to colon cancer since they are less receptive to treatment. This violates the independence assumption. The assumption of independence between mortality due to the cancer of interest and mortality due to competing risks is probably never strictly true. However, provided the departure from independence is only slight, we can estimate net survival without concern for large bias.

2.16.2 Censoring due to emigration

In order for survival estimates to be unbiased it is necessary that the follow-up mechanism is identical for all patients and, in particular, that the probability a death (due to any cause) is registered, given that it has occurred, is the same for every individual. If a cancer patient moves outside the catchment area for death registration (e.g. due to emigration) then the survival time of the patient must be censored at the date of emigration as it will no longer be possible to register the death of the patient. Such patients are referred to as being 'lost to follow-up'. For registries using active follow-up, losses to follow-up occur when the patient can no longer be contacted using the contact information available to the registry. In this situation, survival time is calculated from the date of diagnosis to the last date the patient was known to be alive.

Losses to follow-up do not cause any problems provided they are known to the registry and the reason for the loss is noninformative. Unfortunately, losses to follow-up are not always known to the registry, which leads to overestimation of survival proportions since the patient is assumed to be alive when they may, in fact, be dead. The problem can be controlled by actively contacting all patients or effective record-linkage with the population register, provided that the latter exists and is of good quality. This issue is discussed further in Section 5.3, which details how data quality can be assessed and how estimates can be adjusted to account for data quality problems.

If the cancer registry obtains information on deaths from neighbouring areas, it is important that information is obtained on all deaths for all cases. If it becomes known to the registry that one of the patients recorded as lost to follow-up due to emigration has subsequently died in another country, it may be tempting to update the registry and calculate survival time up to the date of death. This is not appropriate, however, unless the correct vital status of all individuals recorded as being lost to follow-up was also obtained.

Censoring due to emigration is usually uninformative, although this is not always the case. For example, in Geneva, a large proportion of the patients are migrants and have a tendency of returning to their (generally less affluent) home countries after a diagnosis of cancer, particularly if the state of illness is very serious [42]. For example, patients resident in Geneva less than 10 years at the time of diagnosis are approximately 10 times more likely to emigrate following a diagnosis of cancer than patients born in Geneva. It is likely that the prognosis for the migrants is not the same as that of the non-migrant Swiss patients. Nevertheless, the known high Swiss survival is extrapolated to the cases lost to follow-up, resulting in overestimation of the survival proportions. This may, however, be somewhat compensated by the fact that extremely ill patients are probably unable to travel. On the other hand, migrants in Geneva are likely to be in a very good state of health to start with, and thus have a potential for higher survival than the unselected population on average.

2.16.3 Censoring due to death from a cause other than the cancer of interest

Censoring due to death from a cause other than the cancer of interest may or may not be informative, depending on the cause of death. Deaths due to motor vehicle accidents could safely be assumed to be noninformative, although deaths due to suicide or operational mortality would usually be informative.

Patients with a smoking history have a higher mortality from cardiovascular diseases than non-smoking patients. In population-based cause-specific survival analyses, however, there is usually no way to stratify the material

with respect to smoking. A stratification by age, sex, and calendar period, nevertheless, decreases the within-stratum heterogeneity in competing risk mortality and the problem of non-validity of the independence assumption, but does not remove it. The basic problem remains: the survival times of smoking patients are more likely to be censored due to competing risks of death (primarily cardiovascular disease). The potential survival of these patients with respect to cancer will be predicted by what will be observed for patients whose risk of dying from competing risk is lower, i.e., a group where non-smoking patients are overrepresented. Thus, the overall cause-specific survival proportion overestimates the true value.

Table 2.6 Number of deaths due to colon carcinoma (d_i(cancer)), number of deaths due to other causes (d_i(other)) and estimated 5-year observed ($_1p_i$), cause-specific ($_1s_i$) and relative ($_1r_i$) survival proportions for 5285 women diagnosed with colon carcinoma in Finland during 1985–1994 and followed-up to December 31, 1995.

i	l_i	d_i(cancer)	d_i(other)	d_i(total)	w_i	$_1p_i$	2SE($_1p_i$)	$_1r_i$	2SE($_1r_i$)	$_1s_i$	2SE($_1s_i$)
1	5285	1626	158	1784	0	.6624	.0130	.6943	.0136	.6877	.0129
2	3501	503	94	597	369	.5432	.0139	.5980	.0153	.5819	.0139
3	2535	222	75	297	276	.4759	.0142	.5516	.0164	.5271	.0144
4	1962	115	53	168	293	.4319	.0144	.5284	.0176	.4932	.0148
5	1501	67	50	117	207	.3957	.0147	.5121	.0190	.4692	.0152
6	1177	44	54	98	211	.3595	.0150	.4930	.0206	.4494	.0157
7	868	19	34	53	198	.3347	.0155	.4874	.0225	.4381	.0161
8	617	8	32	40	157	.3099	.0162	.4801	.0251	.4313	.0166
9	420	7	20	27	144	.2858	.0174	.4725	.0287	.4224	.0176
10	249	3	15	18	129	.2579	.0201	.4575	.0356	.4153	.0191
11	102	4	1	5	97	.2338	.0274	.4480	.0525	.3839	.0349

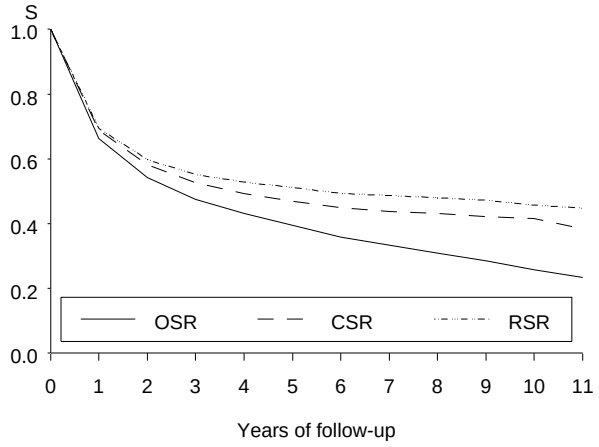

Fig. 2.9 Graph of the cumulative survival proportions shown in Table 2.6

2.16.4 Checking the assumption of noninformative censoring

Reverse the values of the censoring indicator, and treat censoring as the outcome event of interest, while considering deaths as being censored. Look at the risk of being censored in relation to background factors.

2.17 COMPARISON OF OBSERVED, CAUSE-SPECIFIC, AND RELATIVE SURVIVAL

The observed survival proportion is, as expected, lower than the CSR and RSR since deaths due to other causes are considered events of interest when estimating the OSR (Figure 2.9 and Table 2.6). The RSR is always greater than the OSR, since it is obtained by dividing the OSR by a quantity smaller than one (the expected survival proportion). The disparity between the OSR and the other two rates depends on the background mortality due to other causes of death. In young patients, the OSR is similar to the CSR and RSR, since mortality due to other causes is negligible, whereas the OSR will be markedly lower than the other two estimates for elderly patients.

The RSR and CSR are both estimates of the same quantity, net survival, and should therefore, in general, be similar. In the example shown in Figure 2.9 and Table 2.6, the RSR is consistently higher than the CSR. Differences between the RSR and CSR are generally caused by two factors:

1. The general population mortality rates used to estimate the expected survival proportion are not representative of the baseline mortality of the group of cancer patients, leading to biased estimates of the RSR.

2. The cancer being studied is systematically over or under recorded as a cause of death on death certificates, leading to biased estimates of the CSR.

As an example of the first reason, colon cancer is more common among the upper social classes, so the use of nationwide general population mortality rates will underestimate the expected survival proportion for a group of colon cancer patients, leading to an overestimate of the relative survival ratio. The estimated RSR in Figure 2.9 is higher than the CSR, although we cannot be certain that this difference is due to the RSR being overestimated due to disparity in the distribution of social class. Both the RSR and the CSR are estimates of net survival and the estimates are influenced by a number of factors. Although the data shown in Figure 2.9 are consistent with the explanation that the CSR estimates the net survival proportion without bias and the RSR slightly overestimates net survival, we cannot be certain that this is the case. In fact, the size of the bias in the RSR caused by the patient group having a different social class distribution to the general population is rarely as large as the difference between the CSR and RSR in Figure 2.9, indicating that other forces are also operating.

For colon cancer, the size of this overestimation in Finland has been shown to be 1-2 per cent units in patients diagnosed in 1977–1985 (ref?). With male lung cancer patients the converse is true, since lower social classes (and smokers) are overrepresented among the patients, the general population mortality overestimates the expected survival proportion, and the relative survival ratio is slightly underestimated.

Inaccuracy of death certificates will lead to underestimation of the cause-specific survival of the patients when the person signing the death certificate in unaware of the underlying cancer, particularly its true location. Conversely, deaths caused by competing risks may be erroneously attributed to cancer if it is known that the deceased had been diagnosed cancer. In Finland these problems have almost disappeared but care must be taken with cause-specific survival analyses based on older materials [43]. This is further discussed in Section 5.4.

If the target is net survival, the choice of cause-specific or relative survival analysis depends on which problems are foreseen to be the largest. If information on cause of death is unavailable or unreliable, relative survival analysis would be the method of choice. On the other hand, if general population mortality data are unavailable or non-representative, cause-specific survival analysis may be preferable.

When the social-class specific mortality tables were not available in Finland, the net survival comparisons between social classes were made using cause-specific survival, as there was good reason to believe that the accuracy of cause of death information in Finland was comparable between social classes [44]. When social-class specific mortality tables were later constructed, it was shown that the use of social-class specific general mortality data was very important for relative survival comparisons between social classes [36]. On the other hand, a social class selection of the cancer patients by site was not substantial enough to affect the relative survival ratios for all social classes combined, even when the general population mortality was not stratified by social class in the calculations.

Similar results, but of a smaller magnitude were found when relative survival ratios were estimated for geographic regions [45]. A region-specific stratification of general mortality was modestly important for relative survival comparisons between regions but did not practically affect estimated relative survival ratios for the entire country. These findings, however, may not necessarily be generalizable to countries other than Finland. Methods for estimating social class and region-specific survival proportions are further discussed in Chapter 5.15.

If the aim of the analysis is other than the estimation of net survival, it is quite natural and acceptable that the estimated cause-specific and relative survival ratios differ. For example, prostate-cancer patients with a previously diagnosed primary cancer at another site often have much lower relative survival ratios than cause-specific survival proportions [46]. The excess mortality caused by the previous cancer is attributed to the prostate cancer when estimating relative survival ratios, whereas the estimated cause-specific survival proportions are unaffected. In this application, the cause-specific and relative survival ratios describe different aspects of prostatic-cancer patients' survival by providing summary measures of cause-specific and excess mortality, respectively.

Note that the standard error of the CSR is smaller than the standard error of the RSR, even though they both use essentially the same data to estimate the same quantity (Table 2.6). This is partly because the estimation of the CSR assumes the survival times of the patients who died of causes other than cancer to be fixed, when they are in fact random. This is discussed further in Section 2.20.

2.18 ESTIMATING EXPECTED SURVIVAL

In order to estimate the relative survival ratio (Section 2.13) for a group of cancer patients, we require an estimate of the expected survival proportion

for a comparable group from the general population who are assumed to be practically free of the cancer of interest, matched by age, sex, and calendar time. The estimates of the expected survival proportion are based on tables of annual probabilities of death in the general population. A subsection of such a table for Finland is shown in Table 2.7, where the annual probabilities of death ($\times 10^5$) are tabulated for single year age groups between 0 and 99, stratified by calendar period (5 year periods) and sex. The table of annual probabilities of death is sometimes further stratified by other important variables, such as race, region of residence, or social class (Chapter 5.15). We will use the abbreviation 'popmort' (for population mortality) tables to refer to such tables. Although these tables include deaths due to the cancer being studied, Ederer et al. [22] showed that this does not, in practice, affect the estimate of the expected survival proportion. Oksanen [46] showed that the effect was very small for prostate cancer patients, who are relatively old and subject to comparatively high general population mortality. Note that it is necessary only to assume that the comparison group is free of mortality due to the specific cancer of interest, and not all cancers.

The estimation of the expected survival proportion for a group of cancer patients first involves calculating, for each individual cancer patient, the expected survival probability for a person in the general population, similar to the patient with respect to age, sex, and calendar period.

The first 100 numbers (10 rows) of Table 2.7 represent the age specific annual probabilities of death ($\times 10^5$) for Finnish males during the calendar period 1986–1990. For example, the probability of death during the first year of life for a male born in Finland during 1986–1990 was 0.00653. Similarly, the probability of death while aged 79, of a Finnish male who had his 79th birthday during 1986–1990, is 0.09384. The second block of 100 numbers gives similar figures for females for the same period. The next two blocks give the same information for the calendar period 1991–1995. Examples of expected survival calculations for a 79 year old man, alive during 1989, are shown in Table 2.8. Note that the first two annual probabilities are calculated from the first block in the popmort file (males 1986–1990), while the next four probabilities are calculated from the third block in the popmort file (males 1991–1995).

It could be claimed that a person diagnosed with cancer at an age of 79 years was actually, on average, aged 79.5 years. Thus, slightly lower expected survival probabilities should be applied, for example, the arithmetic averages of two consecutive expected probabilities [21]. A simple refinement would therefore be the use of $\frac{1}{2}\left[p_i^*(h) + p_{i+1}^*(h)\right]$ instead of $p_i^*(h)$.

A geometric average, $\left[p_i^*(h)p_{i+1}^*(h)\right]^{\frac{1}{2}}$, would be theoretically more appealing as it can be motivated by an assumption of equal conditional expected survival probabilities during half-years within a year [28], of which the latter half of age i and the first half of age $i+1$ would be applied. This argument is

Table 2.7 Section of a general population mortality file for Finland showing annual probabilities of death ($\times 10^5$) at each age from 0 to 99 for males and females during 1986–90 and 1991–95.

653	54	23	25	18	25	16	24	29	21	Males
24	16	20	31	37	52	90	87	130	122	86-90
135	135	135	138	133	136	143	132	166	159	
167	184	195	221	235	234	259	266	293	313	
368	367	378	422	466	446	510	562	631	679	
795	806	844	930	1005	1120	1271	1358	1543	1703	
1814	2002	2210	2433	2646	2934	3043	3426	3703	3896	
4432	4761	5100	5662	6324	6714	7265	8045	8748	9384	
10461	11213	11947	13462	13936	15042	17443	17831	19606	20932	
21668	22325	26020	27423	30204	34281	30391	32045	41261	37603	
535	42	25	19	24	14	11	15	13	15	Females
13	9	15	12	17	25	40	34	40	50	86-90
43	41	42	39	41	37	45	48	46	48	
55	52	63	70	74	74	86	86	107	118	
123	127	136	144	167	179	214	224	253	275	
274	312	330	370	371	394	473	469	551	580	
671	752	840	900	1058	1133	1357	1461	1581	1883	
2006	2351	2692	2982	3202	3701	4262	4803	5243	5983	
6869	7477	8492	9746	10689	11782	13027	14312	15809	17361	
18852	21035	22274	23716	26315	28156	30113	31547	33039	35680	
533	38	27	23	15	14	22	20	18	17	Males
12	18	15	23	31	44	64	76	118	119	91-95
129	117	110	128	128	142	122	134	135	154	
156	166	169	172	204	212	244	248	280	296	
312	344	391	386	420	457	497	514	556	612	
611	695	789	807	830	937	1046	1125	1209	1372	
1553	1739	1807	2041	2321	2499	2643	3021	3243	3530	
4037	4343	4738	5071	5584	6233	6743	7502	7886	8969	
9662	10640	11372	12814	13615	15141	16046	17385	18971	19509	
21760	23398	25708	26981	29281	29334	32270	35056	37685	41015	
425	39	21	19	14	16	14	18	17	11	Females
16	11	16	14	18	19	25	22	33	36	91-95
38	36	40	34	34	36	43	47	48	44	
42	52	54	65	58	86	89	94	95	117	
116	125	131	149	165	179	205	207	224	246	
249	297	334	307	392	376	423	437	453	531	
603	668	731	775	876	971	1120	1304	1408	1547	
1762	2057	2362	2556	2876	3393	3758	4286	4945	5528	
6093	6916	7986	8945	10107	11452	12525	14104	15424	17298	
18495	19321	22269	23631	26955	28519	30349	31283	33855	33497	

Table 2.8 Annual probabilities of death ($q_i^*(h)$), taken from POPMORT.DAT (Table 2.7), and interval-specific ($p_i^*(h)$) and cumulative ($_1p_i^*(h)$) expected survival probabilities for a 79 year-old man in Finland in 1989.

age (i)	79	80	81	82	83	84
year	1989	1990	1991	1992	1993	1994
$q_i^*(h)$ ($\times 10^5$)	9384	10461	10640	11372	12814	13615
$p_i^*(h)$	0.9062	0.8954	0.8936	0.8863	0.8719	0.8639
$_1p_i^*(h)$	0.9062	0.8114	0.7250	0.6426	0.5602	0.4840

also easy to generalise to dates with one month's accuracy by assigning an expected survival probability of $[p_i^*(h)]^{\frac{1}{12}}$ for each month within age i. These refinements are not considered further in the present text since the size of the improvement they bring is likely to be rather small.

2.18.1 The Ederer I method for estimating expected survival

The simplest method of estimating the corresponding expected survival proportion for a group of cancer patients is the so-called Ederer I method [22]. Under this method, the cumulative expected survival proportion from the date of diagnosis to the end of the ith interval is given by

$$_1p_i^* = \sum_{h=1}^{l_1} {}_1p_i^*(h)/l_1, \tag{2.15}$$

where l_1 is the total number of patients alive at the start of follow-up and $_1p_i^*(h)$ is the expected probability of surviving to the end of the ith interval for a person in the general population, similar to the hth patient alive at the beginning of follow-up with respect to age, sex, and calendar time, given by

$$_1p_i^*(h) = \prod_{j=1}^{i} p_j^*(h). \tag{2.16}$$

That is, the expected 5-year survival proportion is estimated as the average of the expected 5-year survival probabilities for every individual in the life table, even those individuals who are censored in the first interval. Estimates of expected survival proportions using the Ederer I method for the 35 patients in Table 1.2 are shown in Table 2.10. Note that the estimated expected survival proportions for each interval use information from all of the 35 patients, even though not all 35 patients are under follow-up at the start of every interval.

Although the Ederer I method provides unbiased estimates of the expected survival proportion, its application,together with a potentially biased observed

survival proportion, results in biased estimates (usually overestimates) of the relative survival ratio [47] due to the fact that it does not allow for the fact that the potential follow-up times of the patients are of unequal length (Section 2.18.3).

2.18.2 The Ederer II method for estimating expected survival

Ederer and Heise [48] suggested an alternative method of estimating ${}_1p_i^*$, which allows for heterogeneous actual (rather than potential) follow-up times. This method is commonly known as the Ederer II method and is given by

$$ {}_1p_i^* = \prod_{j=1}^{i} p_{j2}^*, \tag{2.17} $$

where

$$ p_{j2}^* = \sum_{h=1}^{l_j} p_j^*(h)/l_j \tag{2.18} $$

is the average of the annual expected survival probabilities $p_j^*(h)$ of the patients alive at the start of the jth interval.

That is, interval-specific expected survival proportions are estimated for each interval, based on only those patients alive at the start of the interval. The cumulative expected survival proportion is then estimated as the product of the interval-specific survival proportions. Estimates of expected survival proportions using the Ederer II method for the 35 patients in Table 1.2 are shown in Table 2.11.

This method allows for heterogeneous potential and observed follow-up times. However, the expected survival proportion is dependent on the observed mortality, leading to biased estimates (usually underestimates) of the relative survival ratio [47]. The observed mortality in a given interval determines the patients who will form the basis for the calculation of p_{j2}^* in the next interval. The cumulative expected survival rate of the patients is therefore dependent on the fatality of the disease in question during preceding intervals. Although the Ederer II method is not recommended for estimating cumulative expected survival proportions, p_{j2}^* is, however, a good estimator for the interval-specific expected survival proportion (see Section 2.19).

Table 2.9 Interval-specific expected survival probabilities for the first six years of follow-up for the 35 patients in Table 1.2. These probabilities are the complements of the annual probabilities of death shown in Table 2.7 (i.e. the expected survival proportion = 1 – the probability of death). ID number, sex, age at diagnosis, year of diagnosis, and survival time in years (t) are also shown.

					Interval-specific expected survival probabilities					
ID	sex	age	year	t	1	2	3	4	5	6
1	m	72	89	0	0.9490	0.9434	0.9442	0.9377	0.9326	0.9250
2	f	82	91	0	0.9201	0.9105	0.8989	0.8855	0.8748	0.8590
3	m	73	93	0	0.9493	0.9442	0.9377	0.9326	0.9250	0.9211
4	m	63	88	0	0.9757	0.9735	0.9707	0.9736	0.9698	0.9676
5	m	67	89	0	0.9657	0.9630	0.9647	0.9596	0.9566	0.9526
6	m	74	92	0	0.9442	0.9377	0.9326	0.9250	0.9211	0.9103
7	f	56	86	0	0.9953	0.9953	0.9945	0.9942	0.9933	0.9933
8	f	52	86	0	0.9967	0.9963	0.9963	0.9961	0.9953	0.9956
9	m	64	94	1	0.9768	0.9750	0.9736	0.9698	0.9676	0.9647
10	f	70	94	1	0.9824	0.9794	0.9764	0.9744	0.9712	0.9661
11	f	83	90	1	0.9025	0.8989	0.8855	0.8748	0.8590	0.8458
12	m	64	89	1	0.9735	0.9707	0.9736	0.9698	0.9676	0.9647
13	f	79	93	2	0.9447	0.9391	0.9308	0.9201	0.9105	0.8989
14	m	70	93	2	0.9799	0.9765	0.9731	0.9744	0.9712	0.9661
15	f	70	88	2	0.9596	0.9566	0.9526	0.9493	0.9442	0.9377
16	f	68	91	2	0.9859	0.9845	0.9824	0.9794	0.9764	0.9744
17	m	58	90	2	0.9846	0.9863	0.9845	0.9826	0.9819	0.9796
18	f	86	93	2	0.9900	0.9906	0.9895	0.9887	0.9879	0.9863
19	m	54	90	2	0.8748	0.8590	0.8458	0.8270	0.8151	0.8068
20	m	31	90	2	0.9982	0.9983	0.9983	0.9980	0.9979	0.9976
21	f	75	93	2	0.9661	0.9624	0.9571	0.9506	0.9447	0.9391
22	f	85	92	3	0.8855	0.8748	0.8590	0.8458	0.8270	0.8151
23	f	68	86	3	0.9842	0.9812	0.9799	0.9765	0.9731	0.9744
24	m	54	85	3	0.9881	0.9888	0.9873	0.9864	0.9846	0.9830
25	m	80	91	4	0.9034	0.8936	0.8863	0.8719	0.8638	0.8486
26	f	52	89	6	0.9967	0.9963	0.9961	0.9962	0.9958	0.9956
27	m	52	89	6	0.9916	0.9907	0.9917	0.9906	0.9895	0.9887
28	m	65	89	6	0.9707	0.9696	0.9698	0.9676	0.9647	0.9596
29	m	60	88	7	0.9819	0.9800	0.9779	0.9796	0.9768	0.9750
30	f	71	87	8	0.9765	0.9731	0.9702	0.9680	0.9661	0.9624
31	m	58	87	8	0.9846	0.9830	0.9819	0.9800	0.9819	0.9796
32	f	80	87	8	0.9313	0.9252	0.9151	0.9025	0.8989	0.8855
33	m	66	86	8	0.9696	0.9657	0.9630	0.9610	0.9557	0.9566
34	m	67	87	8	0.9657	0.9630	0.9610	0.9557	0.9566	0.9526
35	f	56	86	9	0.9953	0.9953	0.9945	0.9942	0.9933	0.9933

Table 2.10 Cumulative expected survival probabilities for the first six years of follow-up for each of the 35 patients in Table 1.2 (calculated as the products of the interval-specific expected survival probabilities given in Table 2.9) and Ederer I estimates of the cumulative expected survival probabilities for the group ($_1p_i^*$), which are equal to the average of the 35 individual cumulative expected survival probabilities. Patient ID number, sex, age at diagnosis, year of diagnosis, and survival time in years (t) are also shown. A life table containing estimates for all 10 intervals is shown in Table 2.12.

					Cumulative expected survival probabilities					
ID	sex	age	year	t	1	2	3	4	5	6
1	m	72	89	0	0.9490	0.8953	0.8453	0.7927	0.7392	0.6838
2	f	82	91	0	0.9201	0.8378	0.7531	0.6668	0.5833	0.5011
3	m	73	93	0	0.9493	0.8963	0.8405	0.7838	0.7251	0.6678
4	m	63	88	0	0.9757	0.9498	0.9220	0.8977	0.8706	0.8424
5	m	67	89	0	0.9657	0.9300	0.8971	0.8609	0.8235	0.7845
6	m	74	92	0	0.9442	0.8854	0.8257	0.7638	0.7035	0.6404
7	f	56	86	0	0.9953	0.9906	0.9852	0.9795	0.9729	0.9664
8	f	52	86	0	0.9967	0.9930	0.9893	0.9855	0.9808	0.9765
9	m	64	94	1	0.9768	0.9524	0.9272	0.8992	0.8701	0.8394
10	f	70	94	1	0.9824	0.9622	0.9395	0.9154	0.8890	0.8589
11	f	83	90	1	0.9025	0.8113	0.7184	0.6284	0.5398	0.4566
12	m	64	89	1	0.9735	0.9450	0.9200	0.8922	0.8633	0.8329
13	f	79	93	2	0.9447	0.8872	0.8258	0.7598	0.6918	0.6219
14	m	70	93	2	0.9799	0.9569	0.9311	0.9073	0.8812	0.8513
15	f	70	88	2	0.9596	0.9180	0.8744	0.8301	0.7838	0.7350
16	f	68	91	2	0.9859	0.9706	0.9535	0.9339	0.9119	0.8885
17	m	58	90	2	0.9846	0.9711	0.9561	0.9394	0.9224	0.9036
18	f	86	93	2	0.9900	0.9807	0.9704	0.9594	0.9478	0.9348
19	m	54	90	2	0.8748	0.7515	0.6356	0.5256	0.4284	0.3457
20	m	31	90	2	0.9982	0.9965	0.9948	0.9928	0.9907	0.9884
21	f	75	93	2	0.9661	0.9298	0.8899	0.8459	0.7991	0.7505
22	f	85	92	3	0.8855	0.7746	0.6654	0.5628	0.4654	0.3794
23	f	68	86	3	0.9842	0.9657	0.9463	0.9240	0.8992	0.8762
24	m	54	85	3	0.9881	0.9770	0.9646	0.9515	0.9369	0.9209
25	m	80	91	4	0.9034	0.8073	0.7155	0.6238	0.5389	0.4573
26	f	52	89	6	0.9967	0.9930	0.9891	0.9854	0.9812	0.9769
27	m	52	89	6	0.9916	0.9824	0.9742	0.9651	0.9549	0.9441
28	m	65	89	6	0.9707	0.9412	0.9128	0.8832	0.8520	0.8176
29	m	60	88	7	0.9819	0.9623	0.9410	0.9218	0.9004	0.8779
30	f	71	87	8	0.9765	0.9502	0.9219	0.8924	0.8622	0.8297
31	m	58	87	8	0.9846	0.9679	0.9503	0.9313	0.9145	0.8958
32	f	80	87	8	0.9313	0.8616	0.7885	0.7116	0.6397	0.5664
33	m	66	86	8	0.9696	0.9363	0.9017	0.8665	0.8281	0.7922
34	m	67	87	8	0.9657	0.9300	0.8937	0.8541	0.8170	0.7783
35	f	56	86	9	0.9953	0.9906	0.9852	0.9795	0.9729	0.9664
$_1p_i^*$ (Equation 2.15)					0.9640	0.9272	0.8899	0.8518	0.8138	0.7757

Table 2.11 Interval-specific expected survival probabilities for the first six years of follow-up (extracted from Table 2.9) for the patients who were alive at the beginning of the interval. Interval-specific expected survival probabilities for the group (p^*_{j2}) are estimated as the average of these probabilities, and the Ederer II estimate of the cumulative expected survival proportion for the group ($_1p^*_i$) is given by the product of the p^*_{j2}'s up to and including the given interval. A life table containing estimates for all 10 intervals is shown in Table 2.12.

					Interval-specific expected survival probabilities					
ID	sex	age	year	t	1	2	3	4	5	6
1	m	72	89	0	0.9490					
2	f	82	91	0	0.9201					
3	m	73	93	0	0.9493					
4	m	63	88	0	0.9757					
5	m	67	89	0	0.9657					
6	m	74	92	0	0.9442					
7	f	56	86	0	0.9953					
8	f	52	86	0	0.9967					
9	m	64	94	1	0.9768	0.9750				
10	f	70	94	1	0.9824	0.9794				
11	f	83	90	1	0.9025	0.8989				
12	m	64	89	1	0.9735	0.9707				
13	f	79	93	2	0.9447	0.9391	0.9308			
14	m	70	93	2	0.9799	0.9765	0.9731			
15	f	70	88	2	0.9596	0.9566	0.9526			
16	f	68	91	2	0.9859	0.9845	0.9824			
17	m	58	90	2	0.9846	0.9863	0.9845			
18	f	86	93	2	0.9900	0.9906	0.9895			
19	m	54	90	2	0.8748	0.8590	0.8458			
20	m	31	90	2	0.9982	0.9983	0.9983			
21	f	75	93	2	0.9661	0.9624	0.9571			
22	f	85	92	3	0.8855	0.8748	0.8590	0.8458		
23	f	68	86	3	0.9842	0.9812	0.9799	0.9765		
24	m	54	85	3	0.9881	0.9888	0.9873	0.9864		
25	m	80	91	4	0.9034	0.8936	0.8863	0.8719	0.8638	
26	f	52	89	6	0.9967	0.9963	0.9961	0.9962	0.9958	0.9956
27	m	52	89	6	0.9916	0.9907	0.9917	0.9906	0.9895	0.9887
28	m	65	89	6	0.9707	0.9696	0.9698	0.9676	0.9647	0.9596
29	m	60	88	7	0.9819	0.9800	0.9779	0.9796	0.9768	0.9750
30	f	71	87	8	0.9765	0.9731	0.9702	0.9680	0.9661	0.9624
31	m	58	87	8	0.9846	0.9830	0.9819	0.9800	0.9819	0.9796
32	f	80	87	8	0.9313	0.9252	0.9151	0.9025	0.8989	0.8855
33	m	66	86	8	0.9696	0.9657	0.9630	0.9610	0.9557	0.9566
34	m	67	87	8	0.9657	0.9630	0.9610	0.9557	0.9566	0.9526
35	f	56	86	9	0.9953	0.9953	0.9945	0.9942	0.9933	0.9933
p^*_{j2} (Equation 2.18)					0.9640	0.9614	0.9586	0.9554	0.9585	0.9649
$_1p^*_i$ (Equation 2.17)					0.9640	0.9268	0.8884	0.8488	0.8136	0.7850

Table 2.12 Estimated interval-specific expected survival proportions (p_i^*) and cumulative expected survival rates using the Ederer I, Ederer II, and Hakulinen methods for the 35 patients shown in Table 1.2.

i	l_i	d_i	w_i	l'_i	p_i^*	Ederer I	Ederer II	Hakulinen
1	35	8	0	35.0	0.9640	0.9640	0.9640	0.9640
2	27	2	2	26.0	0.9614	0.9272	0.9268	0.9267
3	23	5	4	21.0	0.9586	0.8898	0.8884	0.8907
4	14	2	1	13.5	0.9554	0.8518	0.8488	0.8587
5	11	0	1	10.5	0.9585	0.8137	0.8136	0.8308
6	10	0	0	10.0	0.9649	0.7757	0.7850	0.8050
7	10	0	3	8.5	0.9617	0.7377	0.7549	0.7793
8	7	0	1	6.5	0.9486	0.6995	0.7161	0.7530
9	6	2	3	4.5	0.9395	0.6620	0.6728	0.7307
10	1	0	1	0.5	0.9903	0.6249	0.6663	0.7142

2.18.3 The Hakulinen method for estimating expected survival

Hakulinen [47] proposed an alternative method of estimating the expected survival, and is the method we recommend for the estimation of cumulative expected survival proportions for the purpose of estimating relative survival ratios. Expected survival proportions estimated using the Hakulinen method are adjusted for potentially heterogeneous follow-up times among the patients and are independent of the observed mortality of the patients.

The problem with using the Ederer I method in relative survival analysis does not as such lie in Equation 2.15. An average of the cumulative expected survival probabilities provides a proper estimate for an overall expected survival probability. The problem arises from the fact that the relative survival ratio is a ratio between the observed and expected survival proportions and it is the observed survival proportion that has problems when the potential withdrawal patterns are heterogeneous, for example, when older patients have shorter potential follow-up times than younger patients (Section 2.14). In that case, the observed survival proportion is overestimated as the earlier withdrawn older patients are assumed to have the same survival as the younger patients continuing under follow-up after the withdrawal dates of the old patients.

When an overestimated observed survival proportion is divided by an unbiased estimate of the expected survival proportion then the ratio, the relative survival ratio, is also overestimated. The idea behind the Hakulinen method is to introduce a biased estimator for the expected survival proportion, with bias similar to that present in the observed survival proportion, so that the ratio of the biased observed and expected survival proportions will produce

an unbiased estimate of the relative survival ratio. It must be stressed that the Hakulinen expected survival proportion has been constructed solely to yield an unbiased estimate of the relative survival ratio. If an estimate of the expected survival proportion is required for other purposes, then this is best obtained using the Ederer I method, which is not affected by the potential withdrawal patterns of the patients. The Hakulinen method accounts only for potential withdrawal times in calculating the expected survival proportion, unlike the Ederer II method where the observed survival in an interval affects the estimates of expected survival in subsequent intervals.

A drawback of the Hakulinen method is that information on potential follow-up times are required for all patients. This is not a problem if the follow-up is terminated concurrently for all patients under follow-up, which is usually the case in population-based cancer registries. However, potential follow-up times are more difficult to determine for patients with individual last days of contact, particularly for deceased patients. It may even not sound very sensible to try to consider the time for which the patient would have been followed if death had not occurred. Moreover, for cases lost to follow-up, the original potential follow-up time has to be redefined, as no deaths can be recorded for the patients after they become lost to follow-up. The recommendation for the lost cases is to set the potential follow-up time equal to the actual follow-up time. Mathematical details of the Hakulinen method follow, which may be skipped by readers who find the mathematical detail tedious.

The expected survival proportion using the Hakulinen method is derived as follows. Let k_j be the number of patients with a potential follow-up time which extends beyond the beginning of the jth interval. Let the first k_{ja} of these k_j patients have a potential follow-up time which extends past the end of the jth interval and the last k_{jb} be potential withdrawals during the jth interval. It follows that $k_1 = l_1$, $k_{j+1} = k_{ja}$, and $k_j = k_{ja} + k_{jb}$. We will use the notation K_{ja} to refer to the set of k_{ja} patients etc. and h to index the k_{ja} patients in the set K_{ja}.

The expected number of patients alive and under observation at the beginning of the jth interval is given by:

$$l_j^* = \begin{cases} \sum_{h \in K_j} {}_1p_{j-1}^*(h) & \text{for } j \geq 2 \\ l_1 & \text{for } j = 1 \end{cases} \tag{2.19}$$

For the k_{jb} patients with potential follow-up times ending during the jth interval, it is assumed that each patient is at risk for half of the interval, so the expected probability of dying during the interval is given by $1 - \sqrt{p_j^*}$. The expected number of patients withdrawing alive during the jth interval is therefore given by:

$$w_j^* = \begin{cases} \sum_{h \in K_{jb}} {}_1p_{j-1}^*(h)\sqrt{p_j^*(h)} & \text{for } j \geq 2 \\ \sum_{h \in K_{1b}} \sqrt{p_1^*(h)} & \text{for } j = 1 \end{cases} \tag{2.20}$$

The expected number of patients dying during the jth interval, among the k_{jb} patients with potential follow-up time ending during the same interval is given by:

$$\delta_j^* = \begin{cases} \sum_{h \in K_{jb}} {}_1p_{j-1}^*(h)[1 - \sqrt{p_j^*(h)}] & \text{for } j \geq 2 \\ \sum_{h \in K_{1b}} [1 - \sqrt{p_1^*(h)}] & \text{for } j = 1 \end{cases} \tag{2.21}$$

and the expected total number of patients dying during the jth interval is given by:

$$d_j^* = \begin{cases} \left\{ \sum_{h \in K_{ja}} {}_1p_{j-1}^*(h)[1 - p_j^*(h)] \right\} + \delta_j^* & \text{for } j \geq 2 \\ \left\{ \sum_{h \in K_{1a}} [1 - p_1^*(h)] \right\} + \delta_1^* & \text{for } j = 1 \end{cases} \tag{2.22}$$

The expected interval-specific survival proportion is then written as:

$$g_j^* = 1 - d_j^*/(l_j^* - w_j^*/2), \tag{2.23}$$

and, finally, the expected survival proportion from the beginning of follow-up (usually diagnosis) to the end of the ith interval is obtained by calculating:

$${}_1p_i^* = \prod_{j=1}^{i} g_j^*. \tag{2.24}$$

It is possible, following Chiang [28], to provide an improved estimator for g_j^* by making an explicit distinction between deaths among the δ_j^* patients due to be censored and deaths among the $d_j - \delta_j^*$ patients not due to be censored.

$$g_j^* = \tfrac{1}{4}(l_j^* - \tfrac{1}{2}f_j^*)^{-2} \left\{ -\tfrac{1}{2}\delta_j^* + \tfrac{1}{4}\left[(\delta_j^*)^2 + 4(l_j^* - \tfrac{1}{2}f_j^*)(l_{i+1}^* + \tfrac{1}{2}w_j^*) \right]^{\frac{1}{2}} \right\}^2, \tag{2.25}$$

where $f_j^* = w_j^* + \delta_j^*$. In practice this refinement has a very small, almost negligible, impact on the results.

The Ederer I and II methods and the Hakulinen method give very similar results for follow-up times up to ten years [47]. The methods give similar results even for longer periods of follow-up provided that the estimates are made separately for age groups. Four age groups (e.g. 0-44, 45-64, 65-74, and 75-years) are usually sufficient for practical applications, with a slight reservation for the oldest group that may still, occasionally, be overly heterogeneous. The age-specific relative survival ratios may then be combined by taking a weighted average with weights proportional to the numbers of patients at the beginning of follow-up. This may be problematic for the highest age group, as estimates may not be available over extended periods of time (Section 2.19).

In summary, the simple Ederer I and II methods work well until ten years of follow-up, after which the Hakulinen method becomes important. A stratification by age helps to extend this time limit but becomes problematic if overall

results are required for a patient group including very old patients. Statistical modelling may be helpful by providing more stable model-based estimates of the cumulative relative survival ratios which can be used in weighted averages in place of the empirical cumulative relative survival ratios presented here (Chapters 3 and 4).

2.18.4 The Hakulinen method for interval-specific rates

Although the Hakulinen method for deriving expected survival proportions was entirely motivated by a desire to obtain unbiased estimates of cumulative relative survival ratios, it may also be applied to obtain similarly improved interval-specific relative survival ratios. This involves the application of Equations 2.19–2.23 for a single interval i and, in particular, for the l_i patients under observation at the beginning of the interval.

The l_i patients under observation at the beginning of interval i can be categorised into two groups. The first group, denoted by K'_{ia}, contains the k'_{ia} patients with a potential follow-up time extending beyond the endpoint of interval i and the second group, denoted by K'_{ib}, contains the $k'_{ib} = l_i - k'_{ia}$ patients whose potential follow-up time ends during the interval. The expected number of patients withdrawing alive during the interval is then

$$w'_i = \sum_{h \in K'_{ib}} \sqrt{p^*_i(h)}. \tag{2.26}$$

The expected number of patients dying among those with potential follow-up time ending during the interval is

$$\delta'_i = \sum_{h \in K'_{ib}} \left[1 - \sqrt{p^*_i(h)}\right] = k'_{ib} - w'_i. \tag{2.27}$$

Finally, the total expected number of deaths is

$$d'_i = \sum_{h \in K'_{ia}} [1 - p^*_i(h)] + \delta'_i. \tag{2.28}$$

The expected interval-specific survival proportion is then

$$p^*_{i3} = 1 - d'_i/(l_j - w'_j/2). \tag{2.29}$$

The estimator can be refined in a similar manner to Equation 2.25, although such a refinement will have little impact in practice.

In practice, even with annual intervals, p^*_{i2} (Equation 2.18) and p^*_{i3} and give extremely similar results. The quantity g^*_i (Equation 2.23), on the other hand, gives very different results than p^*_{i2} and p^*_{i3}. The latter two give an expected

survival proportion for the l_i patients under follow-up at the beginning of interval i. The l_i patients have their potential follow-up ending during i or beyond i and have also avoided the death before i. They have been thus selected also by the observed mortality. Each g_i^* is a theoretical conditional proportion of survivors in interval i defined solely by the potential withdrawal patterns and the expected mortality for persons to be followed up, l_1. Thus, the g_i^* should not be used in practice as an interval-specific expected survival proportion. Its correct role is a building block of a cumulative expected survival proportion (Equation 2.24), which in turn has a role only as a denominator of the cumulative relative survival ratio ${}_1r_i = {}_1p_i/{}_1p_i^*$. For annual relative survival ratios $r_i = p_i/p_i^*$ either p_{i2}^* or p_{i3}^* should be employed as the method for estimating the expected survival proportion p_i^*.

2.19 COMPARISON OF METHODS FOR ESTIMATING EXPECTED SURVIVAL

Even if the patients did not have their cancer but belonged to a comparable general population, their expected survival would be heavily age-dependent. Figure 2.10 shows cumulative expected survival proportions by age, calculated using the Ederer I method, for females diagnosed with localised melanoma of the skin in Finland during 1975–84. The expected survival proportions would be the same if these same women were instead diagnosed at the same time with, for example, non-localised melanoma of the skin or colon carcinoma, since the Ederer I expected survival proportions depend solely on the characteristics of the patients at the time of diagnosis. If we were to plot similar curves showing the age-specific expected survival proportions for the women who were actually diagnosed with non-localised melanoma of the skin in Finland during 1975–84, then we would observe similar curves to those shown in Figure 2.10, although with very small differences due to differences in the actual age distributions within the age groups and to a negligible extent on the distribution of the actual calendar years of diagnosis. The Ederer I method does not account for censoring (neglects it or eliminates its effect) and is to be preferred if expected survival were desired for its own sake. When estimating expected survival proportions for individual age groups, there is very little difference between the three methods, and essentially no differences during the first ten years of follow-up (Figures 2.11 and 2.12).

There is, however, one conceptual problem with the Ederer I method. In order to produce the curves in Figure 2.10, Equation 2.16 requires that expected survival probabilities be available for each of the patients for twenty years following diagnosis. The patients were diagnosed during the years 1975–84, implying the need for expected survival probabilities for the years 1975–2004. At the time this text was written (1999), expected survival proportions were

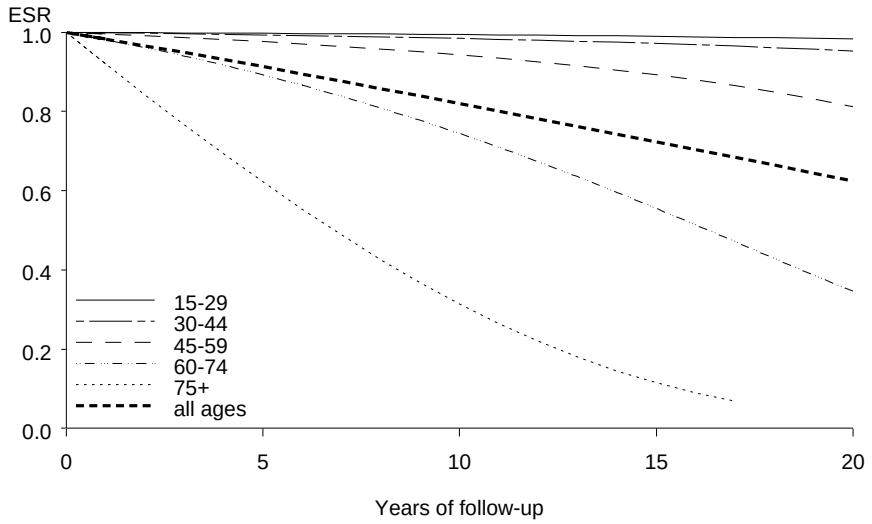

Fig. 2.10 Cumulative expected survival proportions (ESR) by age using the Ederer I method. Females diagnosed with localised melanoma of the skin in Finland 1975–84.

available only up to 1995 and many of the years for which they were required were in the future, so the rates for 1996–2004 had to be predicted. These expected probabilities are used in Equation 2.15 to produce an expected survival proportion for the patient group, which is then compared with the observed survival proportion to produce a relative survival ratio. This creates a problem, because the observed survival proportion is based solely on patients who we have been able to observe in the past, but we are comparing this to an expected survival proportion which involves predicted expected survival proportions for time periods in the future. Of course, the observed survival, because of censoring, also contains predictions of the withdrawn patients' fate, that being the same as the fate of the patients not withdrawn from follow-up. In practice, the expected survival probabilities for the future years are often, as also in Figure 2.10, assumed to be equal with corresponding rates from the last time period from which expected survival probabilities are empirically available.

The use of the Ederer II or Hakulinen methods obviates the need for expected survival probabilities for future periods. The Hakulinen method gives the same results as the Ederer I method, provided that no expected probabilities are needed for future years (e.g. with shorter periods of follow-up) when the potential follow-up time does not depend on age (or calendar time). This is particularly true when there is no censoring, for example, when all potential follow-up times exceed the length of follow-up of interest, usually 5 or 10 years.

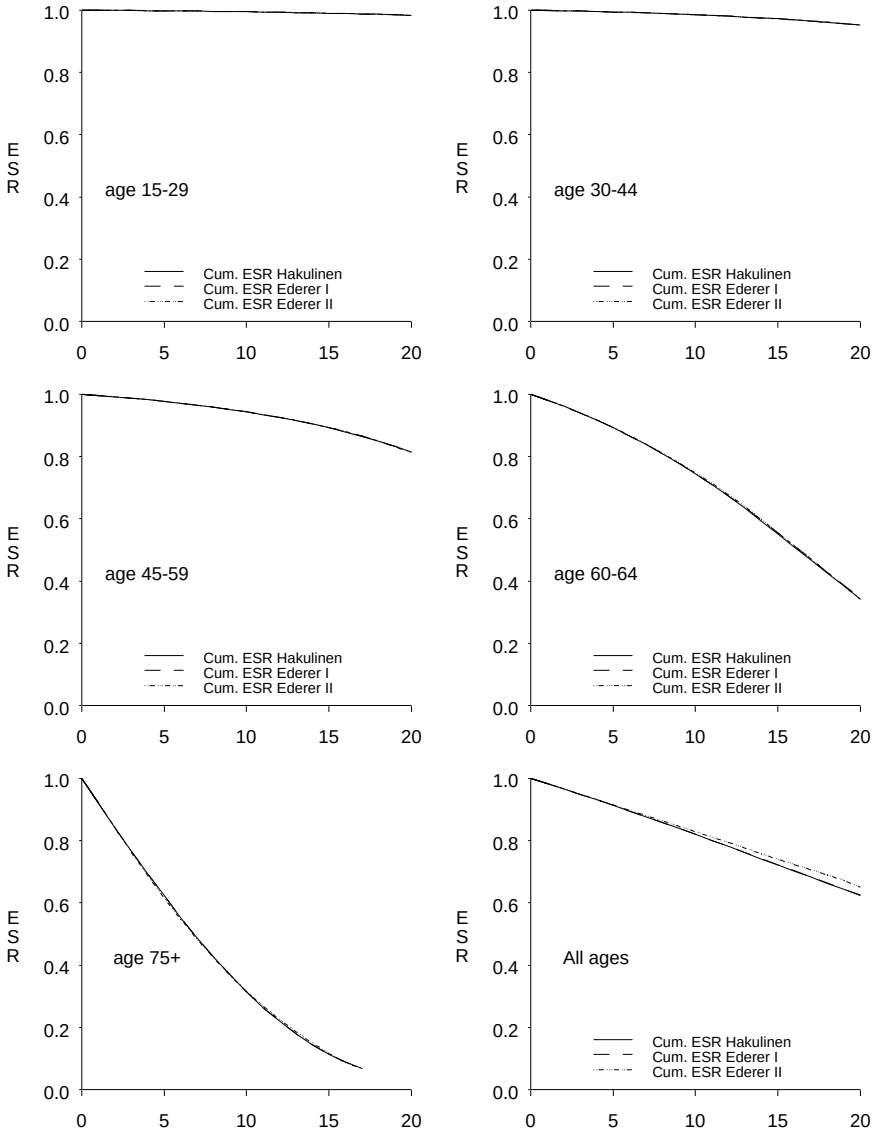

Fig. 2.11 Cumulative expected survival proportions (ESR) using 3 different methods. Females diagnosed with localised melanoma of the skin in Finland 1975–84.

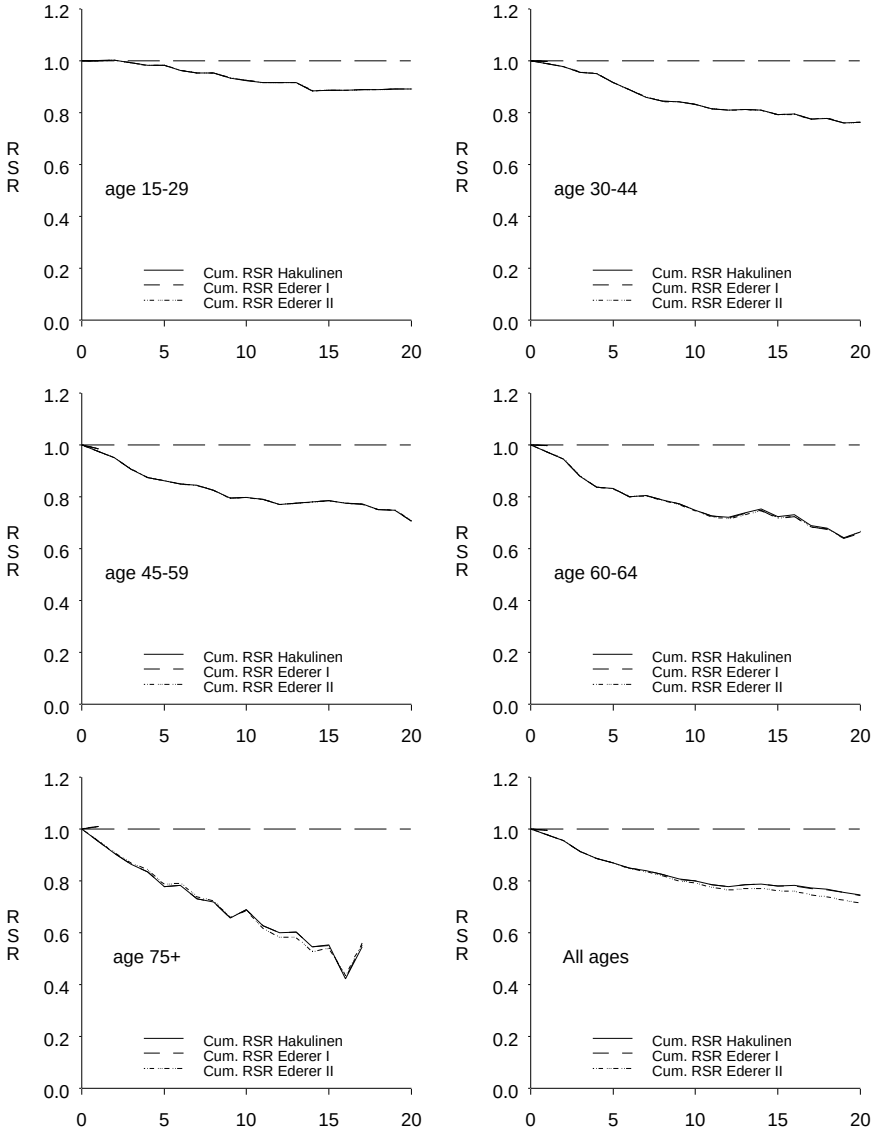

Fig. 2.12 Cumulative relative survival ratios (RSR) based on 3 different methods of estimating the expected survival rate. Females diagnosed with localised melanoma of the skin in Finland 1975–84.

As outlined in Section 2.18, the Hakulinen method has been constructed to produce unbiased cumulative relative survival ratios when the potential follow-up times depend on age (and calendar time) and when thereby the observed survival proportions are biased. Thus, the Hakulinen method in general yields similarly biased cumulative expected survival proportions that depend on the potential follow-up times of the patients. As the latter do not represent any biological entity, in principle the expected survival proportions produced by the Hakulinen method should not be used for purposes other than estimating the relative survival ratio.

The cancer of the patients does not have any effect on the cumulative expected survival when the Ederer I or the Hakulinen method is used. This makes sense: groups of patients with localised melanoma and non-localised melanoma should have the same expected survival proportions if their demographic background is exactly the same. However, this is not true for the Ederer II method, as the cumulative expected survival is based on the patients actually under observation (Equation 2.18) and patients diagnosed with high-fatality cancers are more likely to die (and therefore be removed from observation) than patients diagnosed with less fatal cancers. Figure 2.13 shows that the Ederer II method gives a superior cumulative expected survival for the non-localised patients. This is because differences in the age-specific hazard ratios between the two groups result in changes in the age distribution of the groups over time. Although the age distributions of the two groups are similar at diagnosis, the oldest patients diagnosed with non-localised disease are rapidly killed by the disease, leaving only the relatively young patients to determine the expected survival. Note that the difference in overall survival proportions between the groups does not contribute to the differences in the expected survival proportions. It is the difference in the age-specific hazard ratios that causes the difference. The hazard ratio for the oldest patients compared to the youngest patients is higher in the non-localised than the localised group, which leads to differences between the two groups with respect to the age distribution of the patients under follow-up and, consequently, differences in the Ederer II estimate of the expected survival proportion.

Actually, during the last years of the follow-up, only young patients are being followed. Thus, it may be argued that the use of Ederer II method implies a better comparison between like and like. This is true locally in a given follow-up interval although not necessarily so in an overall cumulative fashion. The Ederer II method produces good interval-specific expected survival proportions but these depend on observed survival proportions in the previous intervals. Thus, the cumulative expected survival proportion using the Ederer II method becomes dependent on the cumulative observed survival proportion and the comparison between these two quantities by dividing the observed rate by the expected becomes problematic. The problem of method choice is, however, very often purely theoretical. The follow-up is often less

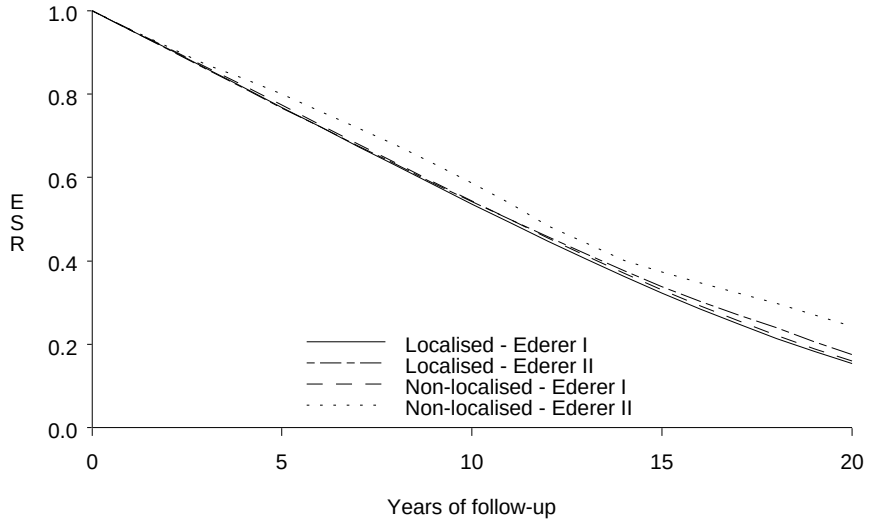

Fig. 2.13 Cumulative expected survival proportions (ESR) for localised and non-localised skin melanoma using the Ederer I and Ederer II methods. Females diagnosed in Finland 1975–84 aged 65 years or more. The distribution of age at diagnosis is similar for the two groups, with a mean age of 73.5 years for localised and 73.3 years for non-localised.

than ten years or the relative survival ratios are calculated by age group and combined by taking a weighted average.

A very simple method of improving the estimate of the relative survival ratio is to calculate a weighted average of the observed survival proportions by age groups in order to obtain an unbiased observed survival proportion and to divide this with the expected survival proportion produced by the Ederer I method. However, the cumulative relative survival ratio for all ages combined should, nevertheless, be combined as a weighted average of age-specific cumulative relative survival ratios [34], when the age-specific cumulative relative survival ratios differ. The Ederer I and II methods and the Hakulinen method have been designed to estimate a common cumulative relative survival ratio. However, in practice it is more a rule than an exception that the relative survival ratios differ by age (and by other variables, too, although not so strongly) [19].

If there are differences between the cumulative relative survival ratios by age, the overall cumulative relative survival ratio converges towards that of the young patients when the follow-up time becomes longer [34]. After, say, a thirty-year follow-up the patients that are observed to be alive are those who were young at the beginning of follow-up. On the other hand in a general population, also only the young comparable persons are expected to be alive. Thus the relative survival ratio, the observed survival proportion divided by the expected becomes close to the same ratio in young patients. Because the young patients often have the highest cumulative relative survival ratio the overall cumulative relative survival curve gives a false impression of an increase by time.

It is possible to avoid this problem by estimating a cumulative relative survival ratio for each age group separately and then taking a weighted average using the numbers of patients at the beginning of the follow-up as weights. Even this is problematic if the patients include a very high age group, e.g., those aged 75 years a more at diagnosis. After a 30-year follow-up this group would consist of patients aged 105 years or more. A 30-year relative survival ratio for this age would then become rather shaky to estimate, virtually an observed survival proportion of zero divided by an expected survival proportion of zero. This is reflected in cumulative relative survival curves for the oldest ages for fairly non-lethal tumours, cf. the oldest age group in Figure 2.14. Including such an inaccurate relative survival ratio in a weighted average makes the weighted average inaccurate, as well. Statistical modelling may be helpful for a ‘stabilisation’ of the weighted rates, cf. Sections 3 and 4.

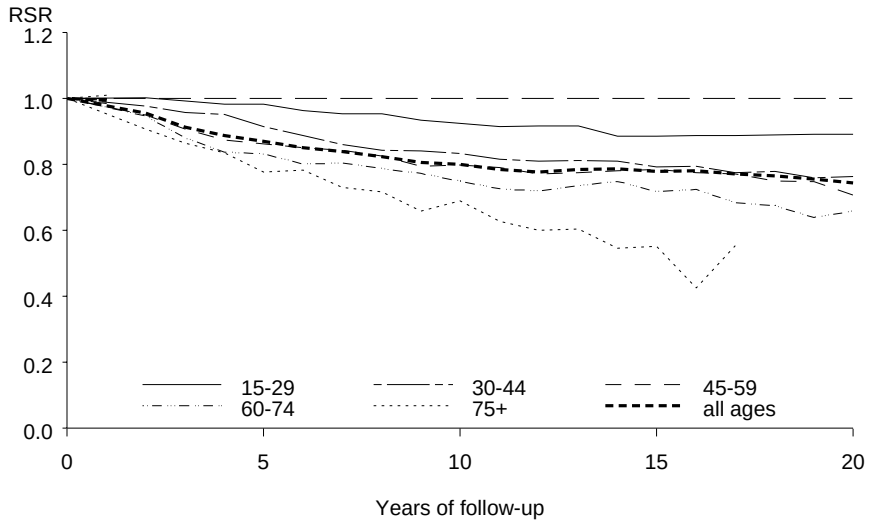

Fig. 2.14 Cumulative relative survival ratios (RSR) based on the Ederer I method of estimating the expected survival rate. Females diagnosed with localised melanoma of the skin in Finland 1975–84.

2.20 STANDARD ERROR OF THE ESTIMATED SURVIVAL PROPORTION

Patient survival proportions are generally estimated for one of two reasons:

1. To describe the survival experience of the patients who have been reported to the registry; or
2. To generalise the estimates to other patients (including those yet to be diagnosed).

In order to generalise the results to other patients, it is assumed that the patients reported to the registry are a random sample from some hypothetical ('super') population. The process of using estimates based on a finite sample to draw inference about the properties of some other population is known as statistical inference. Irrespective of whether or not we wish to view the patients as a sample from some super population and generalise the estimates, the frequencies and estimated rates based on a finite material are always subject to random error and it is desirable to quantify the size of the random error.

It is first necessary to define what is meant by the term 'random error'. Assume that 10 women of the same age are diagnosed with a certain type of

cancer (at the same stage) on the same day at the same hospital and each undergoes the same treatment. The survival times of these 10 women will be similar, although not identical. We say that the differences in the survival times of these women are due to random error. Some synonyms for 'random error' are 'chance', 'chance variation', and 'random variation'. There may actually be many thousands of factors, which, if measured, might explain the differences in the survival times of these women, although we lump these together under the heading 'random error'.

A common framework (known as the frequentist framework) for statistical inference is to assume that the patients reported to the registry are a random sample of all individuals who could possibly be reported to the registry. This includes all individuals in the population at risk of developing cancer. It is also assumed that, if we could study all possible individuals, then we could calculate the true survival proportion. Our population of 'all possible individuals' includes 'all individuals with cancer', which includes, 'all individuals diagnosed with cancer', which includes 'all individuals reported to the cancer registry'. We are only able to study all individuals reported to the cancer registry, which we hope is equivalent to 'all individuals diagnosed with cancer' (see the discussion of ascertainment bias in Section 5.3).

If a different set of individuals (with identical age, sex, time of diagnosis, and tumour characteristics) were reported to the registry then we would obtain slightly different estimates of the survival proportion. The differences in the survival proportions estimated from these different samples will depend on the size of the sample and the amount of variation in the survival times of individuals. If we were able to estimate the survival proportion for all possible samples of a given size, n, we would have a very good picture of the amount of random error in the estimated survival proportion. The collective distribution of these estimates is known as the sampling distribution of the survival proportion for samples of size n and will follow a normal distribution with mean equal to the true population mean. The standard deviation of the distribution of estimates is known as the standard error of the estimated survival proportion and is the most widely used measure of the amount of random error in the estimate. Fortunately, it is possible to calculate the standard error of an estimate based upon the single sample at hand (Section 2.20.1), rather than requiring estimates for all possible samples from the population.

Another method for presenting the magnitude of random error in an estimate is with a confidence interval (Section 2.21). The most common method for calculating a confidence interval is to assume that the estimated survival proportion is normally distributed around the true (but unknown) survival proportion for the population, with variance given by the square of the standard error. A 95% confidence interval for the survival proportion p can then be constructed as $p \pm 1.96 \times \mathrm{SE}(p)$ where $\mathrm{SE}(p)$ is the standard error of p. In the aforementioned hypothetical long repetition of studies, this confidence

interval will include the true (but unknown) survival proportion with a 95% relative frequency. Various methods for constructing confidence intervals are discussed in Section 2.21. The standard method of constructing the confidence interval as $p \pm 1.96 \times \mathrm{SE}(p)$ is not suitable when the material is small or the survival proportions are close to the extreme values (0 or 1 for the observed and cause-specific survival proportion, 0 or $1/p^*$ for the relative survival ratio). As a guide, this method is valid when both np and $n(1-p)$ are both greater than 5 [49, pp. 153].

2.20.1 Standard error of the observed survival proportion

The most widely used method for estimating the standard error of the observed survival proportion (and, equivalently, the cause-specific survival proportion) is the method described by Greenwood (1926) [22, 50], which is known as Greenwood's method or Greenwood's formula.

The formula,

$$\mathrm{SE}({}_1p_i) = {}_1p_i \left[\sum_{j=1}^{i} \frac{d_j}{l'_j(l'_j - d_j)} \right]^{\frac{1}{2}}, \tag{2.30}$$

is slightly laborious for hand calculation but readily available in many computer programs. For a single interval, Equation 2.30 reduces to

$$\mathrm{SE}(p_i) = p_i \left\{ \frac{d_i}{l'_i(l'_i - d_i)} \right\}^{\frac{1}{2}} = \sqrt{p_i(1-p_i)/l'_i}, \tag{2.31}$$

which is the familiar binomial formula for the standard error of the observed interval-specific survival proportion based on l'_i trials. Non-integer values for l'_i, e.g. $l'_i = 20.5$, do not cause any problems in practical use. Equation 2.30 also reduces to the familiar binomial standard error in the absence of censoring (see Section 2.20.1.1).

Equation 2.30 can also be used to estimate the standard error of the cause-specific survival proportion, where deaths due to competing risks are treated as censored observations instead of deaths. The standard error of the cause-specific survival proportion will generally be smaller than the standard error of the observed survival proportion in absolute terms, provided some deaths due to competing risks are observed. This may give the impression that the CSR is estimated with greater efficiency than the OSR, an impression which may be somewhat misleading, since the survival times of the censored individuals are considered fixed when they are, in fact, random.

In the absence of censoring, a standard error for an estimated cumulative observed survival proportion ${}_1p_i$ based on l_1 individuals alive at the start of follow-up can be estimated based on the binomial distribution, $\mathrm{SE}({}_1p_i) =$

$\sqrt{{}_1p_i(1-{}_1p_i)/l_1'}$, but an estimate of ${}_1p_i$ based on the same number of individuals where some survival times are censored is based on less information, so we would expect the standard error to be larger, which is the case when using Greenwood's formula. It may be considered that the censoring has a diminishing effect on the sample size. The survival proportion and its standard error actually correspond to ones obtained from a complete follow-up of a smaller number of patients, called the effective sample size. This quantity, denoted by ${}_1l_i^*$, is obtained as

$$ {}_1l_i^* = \frac{{}_1p_i(1-{}_1p_i)}{[\mathrm{SE}({}_1p_i)]^2}. \tag{2.32} $$

For example, in Table 2.2 (page 35) $l_i = 5285$ but a large number of survival times are censored. The ten-year observed survival proportion and its standard error presented in the table correspond to a complete follow-up of ${}_1l_i^*$=953 persons, a figure much smaller than 5285.

[Comment on the statistical properties of SEs estimated using Greenwood's method. I understand that large sample sizes are required and that the method tends to underestimate the SE, especially in the tails. Timo says that Greenwood's formula is not so bad. It's at least as good as the binomial.]

Other methods for estimating the standard error are also available, such as that described by Peto (1984) [51].

[Give an example of a KM curve with error bars and number at risk shown at the bottom.]

[CIs are more important than SEs.]

2.20.1.1 Greenwood's method in the absence of censoring In the absence of censoring, Greenwood's method (Equation 2.30) reduces to the familiar binomial standard error. First note that $l_j' = l_j$ and

$$ \begin{aligned} \sum_{j=1}^{i} \frac{d_j}{l_j'(l_j'-d_j)} &= \sum_{j=1}^{i} \frac{l_j - l_{j+1}}{l_j l_{j+1}} \\ &= \sum_{j=1}^{i} \left(\frac{1}{l_{j+1}} - \frac{1}{l_j} \right) \\ &= \frac{1}{l_{i+1}} - \frac{1}{l_1} \\ &= \frac{l_1 - l_{i+1}}{l_1 l_{i+1}}. \end{aligned} $$

Consequently

$$\begin{aligned}
\mathrm{SE}({}_1p_i)^2 &= {}_1p_i^2\left(\frac{l_1 - l_{i+1}}{l_1 l_{i+1}}\right) \\
&= {}_1p_i\frac{l_{i+1}}{l_i}\left(\frac{l_1 - l_{i+1}}{l_1 l_{i+1}}\right) \\
&= {}_1p_i\left(\frac{l_1 - l_{i+1}}{l_1}\right)/l_1 \\
&= {}_1p_i(1 - {}_1p_i)/l_1.
\end{aligned}$$

2.20.2 Standard error of the relative survival ratio

The variance of the expected survival proportion is very small in comparison to the variance of the observed survival proportion so, in practice, it is assumed that the expected survival proportion is a fixed constant. The variance of the relative survival ratio (both interval-specific and cumulative) is then given by

$$\begin{aligned}
\mathrm{var}(r) &= \mathrm{var}(p/p^*) \\
&= \mathrm{var}(p)/(p^*)^2. \qquad (2.33)
\end{aligned}$$

That is, the variance of the relative survival ratio is given by the variance of the observed survival proportion, divided by the square of the expected survival proportion. In Equation 2.33, we have made use of the result that for a random variable X and a constant a, $\mathrm{var}(aX) = a^2\mathrm{var}(X)$. The variance of the observed survival proportion ($\mathrm{var}(p) = \mathrm{SE}(p)^2$) is calculated using Greenwood's formula (Section 2.20.1). The standard error (SE) of the relative survival ratio is given by $\mathrm{SE}(r) = \mathrm{SE}(p)/p^*$.

2.21 CONFIDENCE INTERVALS FOR ESTIMATED SURVIVAL PROPORTIONS

Since we (assume that we) have registered every person with a diagnosis of cancer in the population (all cancer cases in Finland, for example), the idea of sampling from a population which has a 'true mean' survival time is not obvious. We actually assume that among the entire population of Finland, the survival time for every individual, given a diagnosis of cancer, follows some theoretical distribution with a 'true mean' survival time, and the people actually diagnosed make up our sample.

Confidence intervals can be calculated for any estimated survival proportion in order to provide a measure of uncertainty associated with the point estimate. A 95% confidence interval (CI) is an interval, i.e. a range of values, such that under repeated sampling, the true survival proportion will be contained in the

interval 95% of the time. The CI is often called an interval estimate for the true survival proportion, while the estimated survival rate is called the point estimate.

Estimated confidence intervals provide an indication of the level of statistical uncertainty in the estimated survival proportions. They do not represent the range of possible prognoses for an individual patient.

2.21.1 Estimating confidence intervals

A confidence interval for the true survival proportion can be obtained by assuming that the estimated survival proportion is normally distributed around the true value with estimated variance given by the square of the standard error (Section 2.20). A two-sided $100(1-\alpha)\%$ confidence interval ranges from $p - z_{\alpha/2}\text{SE}(p)$ to $p + z_{\alpha/2}\text{SE}(p)$, where p is the estimated survival proportion (which can be an interval-specific or cumulative OSR, CSR, or RSR), SE(p) the associated standard error, and $z_{\alpha/2}$ the upper $\alpha/2$ percentage point of the standard normal distribution. For a 95% confidence interval, $z_{\alpha/2} = 1.96$, and for a 99% confidence interval, $z_{\alpha/2} = 2.58$.

As a rule of thumb, the normal approximation for a single interval i is usually appropriate when both $l'_i p_i$ and $l'_i(1 - p_i)$ are greater than or equal to 5 [52]. Confidence intervals obtained in this way are symmetric about the point estimate and can sometimes contain implausible values for the survival proportion, i.e., values less than zero or greater than one. The theoretical upper bound for the relative survival ratio, which can be greater than one, is $1/p^*$, where p^* is the expected survival proportion. When the CI is out-of-bounds, we recommend using one of the transformations described in the following section. Examples are shown in Table 2.13.

2.21.2 Transformation to avoid implausible values

One method of obtaining confidence intervals for the observed survival proportion in the range [0,1] is to transform the estimate to a value in the range $[-\infty, \infty]$, obtain a confidence interval for the transformed value, and then back-transform the confidence interval to [0,1]. Suitable transformations are the logistic transformation, $\log[p/(1-p)]$, or the complementary log-log transformation, $\log[-\log(p)]$.

We recommend using the complementary log-log transformation, which involves constructing the confidence intervals on the log-hazard scale. To estimate confidence intervals for the RSR using this method, we first transform the estimated cumulative OSR. We will write this transformation as $g(\text{OSR}) = \log[-\log(\text{OSR})]$, where g is the complementary log-log transfor-

Table 2.13 95% confidence intervals for the cumulative relative survival ratios given in Table 2.6 (page 61) using three different methods, the naive method (estimate $\pm 1.96SE$), the EUROCARE method (Section 2.21.3) and the complementary log-log transformation (Section 2.21.2).

i	RSR	Naive	EUROCARE	Transformation
1	0.6943	0.6806–0.7079	0.6808–0.7075	0.6804–0.7077
2	0.5980	0.5827–0.6132	0.5830–0.6129	0.5826–0.6131
3	0.5516	0.5352–0.5680	0.5356–0.5677	0.5351–0.5680
4	0.5284	0.5108–0.5460	0.5112–0.5457	0.5107–0.5459
5	0.5121	0.4931–0.5310	0.4936–0.5308	0.4931–0.5310
6	0.4930	0.4724–0.5136	0.4730–0.5134	0.4724–0.5136
7	0.4874	0.4649–0.5099	0.4656–0.5097	0.4649–0.5099
8	0.4801	0.4550–0.5052	0.4558–0.5050	0.4551–0.5052
9	0.4725	0.4438–0.5013	0.4449–0.5012	0.4440–0.5014
10	0.4575	0.4219–0.4930	0.4235–0.4932	0.4223–0.4934
11	0.4480	0.3954–0.5005	0.3986–0.5014	0.3965–0.5014

mation, which transforms the cumulative observed survival proportion to the log cumulative hazard scale. We also require an estimate of the variance of the OSR on the log hazard scale. Using a Taylor series approximation, the variance of a differentiable function, g, of a random variable, X, can be approximated by

$$\text{var}\{g(X)\} \approx \left\{\frac{\text{d}g(X)}{\text{d}X}\right\}^2 \text{var}(X). \tag{2.34}$$

If we denote the cumulative observed survival proportion by X then, noting that

$$\frac{\text{d}\log[f(X)]}{\text{d}X} = \frac{1}{f(X)}\frac{\text{d}f(X)}{\text{d}X}, \tag{2.35}$$

we have

$$\text{var}\{g(X)\} = \text{var}\{\log[-\log(X)]\} \approx \frac{1}{[X\log(X)]^2}\text{var}(X). \tag{2.36}$$

An estimated 95% confidence interval for the OSR on the log hazard scale is therefore given by $g(\text{OSR}) \pm 1.96\sqrt{\text{var}\{g(X)\}}$, which is then back-transformed to give a 95% confidence interval for the OSR. To obtain a CI for the RSR, the upper and lower confidence limits for the OSR are simply divided by the ESR. This method is easier to implement then the EUROCARE method, which is described in the following section. An additional advantage of this method is that the Normal approximation is more appropriate when applied to data on the transformed than the untransformed scale (give references).

Table 2.14 Males aged 75–84 diagnosed with localised skin melanoma in Finland during 1975–84. 95% confidence intervals for the cumulative relative survival ratios using three different methods, the naive method (estimate $\pm 1.96SE$), the EUROCARE method (Section 2.21.3) and the complementary log-log transformation (Section 2.21.2).

I	L	D	W	OSR	2SE	RSR	2SE	ESR	Naive	EUROCARE	Transformation
1	65	8	0	0.8769	0.0815	0.9852	0.0916	0.8901	0.8936–1.0767	0.8712–1.0519	0.8606–1.0530
2	57	6	0	0.7846	0.1020	0.9969	0.1296	0.7870	0.8674–1.1265	0.8517–1.1018	0.8394–1.1025
3	51	15	0	0.5539	0.1233	0.8020	0.1786	0.6906	0.6234–0.9806	0.6276–0.9677	0.6120–0.9655
4	36	6	0	0.4615	0.1237	0.7680	0.2058	0.6010	0.5622–0.9738	0.5756–0.9676	0.5575–0.9632
5	30	6	0	0.3692	0.1197	0.7120	0.2309	0.5186	0.4812–0.9429	0.5058–0.9464	0.4854–0.9391
6	24	4	0	0.3077	0.1145	0.6951	0.2587	0.4427	0.4365–0.9538	0.4719–0.9668	0.4488–0.9564
7	20	3	0	0.2615	0.1090	0.6985	0.2912	0.3744	0.4074–0.9897	0.4545–1.0136	0.4284–0.9995
8	17	4	0	0.2000	0.0992	0.6384	0.3167	0.3133	0.3217–0.9551	0.3855–0.9981	0.3570–0.9784
9	13	3	0	0.1539	0.0895	0.5944	0.3458	0.2588	0.2486–0.9402	0.3314–1.0067	0.3005–0.9800
10	10	2	0	0.1231	0.0815	0.5828	0.3859	0.2112	0.1969–0.9687	0.3016–1.0631	0.2675–1.0282
11	8	3	0	0.0769	0.0661	0.4528	0.3891	0.1699	0.0637–0.8420	0.1960–0.9875	0.1632–0.9398
12	5	2	0	0.0462	0.0521	0.3423	0.3861	0.1348	-0.0437–0.7284	0.1173–0.9430	0.0876–0.8800
13	3	0	1	0.0462	0.0521	0.4367	0.4925	0.1057	-0.0558–0.9292	0.1497–1.2031	0.1118–1.1226
14	2	0	0	0.0462	0.0521	0.5666	0.6390	0.0815	-0.0724–1.2056	0.1942–1.5610	0.1451–1.4566
15	2	0	0	0.0462	0.0521	0.7464	0.8417	0.0618	-0.0953–1.5881	0.2558–2.0560	0.1911–1.9185
16	2	1	0	0.0231	0.0417	0.5033	0.9104	0.0459	-0.4071–1.4138	0.1007–2.3433	0.0494–2.1170
17	1	0	1	0.0231	0.0417	0.6935	1.2544	0.0333	-0.5609–1.9479	0.1387–3.2283	0.0681–2.9166

2.21.3 EUROCARE confidence intervals

Another method for obtaining confidence intervals restricted to [0,1] was described in the EUROCARE publication [53]. Assume the true observed survival proportion is μ and that p is an unbiased estimator for μ, which can be assumed to follow a normal distribution. That is, $p \sim N(\mu, \sigma^2)$.

The limits of a 95% confidence interval for μ are given by the roots of the following second-degree polynomial in ϕ:

$$\left[\frac{(p-\phi)}{\sigma}\right]^2 = 1.96^2 \tag{2.37}$$

For the cumulative observed survival proportion, σ is estimated by $\phi(1-\phi)/{}_1l_i^*$, where ${}_1l_i^*$ is the effective sample size for follow-up until the end of the ith subinterval (Equation 2.32).

For the cumulative relative survival ratio, ${}_1r_i$, the limits for the 95% CI are computed as the roots of the second-degree polynomial in ϕ:

$$({}_1p_i - {}_1p_i^*\phi)^2 = (1.96)^2 {}_1p_i^*\phi(1 - {}_1p_i^*\phi)/{}_1l_i^* \tag{2.38}$$

Confidence intervals for ${}_1r_i$ computed in this way are constrained to the interval $[0,1/{}_1p_i^*]$, as they should be, since ${}_1p_i$ is constrained to the interval [0,1].

Omitting the subscripts and expanding Equation 2.38 we have

$$p^{*2}\phi^2 - 2pp^*\phi + p^2 = \frac{(1.96)^2}{l^*}[p^*\phi - p^{*2}\phi^2]. \tag{2.39}$$

Hence

$$\phi^2 p^{*2}[1 + \frac{(1.96)^2}{l^*}] - \phi[2pp^* + \frac{(1.96)^2}{l^*}p^*] + p^2 = 0 \tag{2.40}$$

This is a quadratic equation in the form $ax^2 + bx + c = 0$ where

$$\begin{aligned} a &= p^{*2}[1 + \frac{(1.96)^2}{l^*}], \\ b &= -[2pp^* + \frac{(1.96)^2}{l^*}p^*], \text{ and} \\ c &= p^2. \end{aligned}$$

The roots of the equation are given by $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$, the standard quadratic formula.

For the cumulative observed survival proportion, ${}_1p_i$, the limits for the 95% CI are computed as the roots of the second-degree polynomial in ϕ:

$$({}_1p_i - \phi)^2 = (1.96)^2\phi(1-\phi)/{}_1l_i^*. \tag{2.41}$$

This is identical to Equation 2.38, except that the cumulative expected survival, ${}_1p_i^*$, rate is substituted by 1. For the interval-specific relative survival ratio, r_i, the limits for the 95% CI are computed as the roots of the second-degree polynomial in ϕ:

$$(p_i - p_i^*\phi)^2 = (1.96)^2 p_i^* \phi(1 - p_i^*\phi)/l_i', \tag{2.42}$$

where l_i' is the effective number of patients at risk during the ith subinterval, estimated as $l_i' = l_i - w_i/2$. For the interval-specific observed survival proportion, p_i, the limits for the 95% CI are computed as the roots of the second-degree polynomial in ϕ:

$$(p_i - \phi)^2 = (1.96)^2 \phi(1 - \phi)/l_i'. \tag{2.43}$$

2.21.4 Comparison of the 3 methods for estimating confidence intervals

Examples of the three different methods for estimating confidence intervals applied to real data are shown in Tables 2.13 and 2.14.

The confidence intervals in Table 2.13 are based on a large number of individuals, with a moderate number of deaths, so the standard errors of the estimated relative survival proportions are small (compared to, for example, those in Table 2.14). As such, the confidence intervals are narrow. Furthermore, none of the estimated confidence intervals extend outside the range of plausible values, since none of the estimated relative survival proportions, which range from 0.69 down to 0.38, are close to 0 or $1/p^*$. In such a situation, we would not expect any problems when using the naive method and this is indeed the case. The estimated confidence intervals for each of the three methods are very close.

The data in Table 2.14, however, are more typical of the kind of data where use of the naive method can be problematic. The estimates are based on relatively few individuals, meaning that the standard errors of the estimated relative survival ratios are relatively large. As such, use of the naive method results in estimated confidence intervals which extend below zero. Note that although the upper bounds for many of the confidence intervals are above one, none of them extend beyond the range of plausible values. In the first interval, for example, the upper limit of the range of plausible values is $1/0.8901 = 1.12$ and for the second interval the upper limit is $1/0.7870 = 1.27$. Problems with the upper bound of the confidence interval exceeding the plausible limit are more likely to occur when the patients are young, where the upper limit of the range of plausible values is much closer to one. When the RSR is close to 0 or 1 the EUROCARE method and the transformation method give more plausible results.

[Write more on this topic.]

2.21.5 Confidence intervals for the entire survival curve

The methods previously described in this section are appropriate for calculating a confidence interval for the survivor function at a fixed point in time. It is not appropriate to construct 'confidence bands' for the survivor function by drawing lines through the upper and lower limits of the confidence intervals estimated at each time point. Methods for calculating confidence intervals for the entire survival curve are described by Harris and Albert (1991) [54, Ch. 2].

[I haven't read the book cited above. Ask Guy about this, he probably has a more appropriate reference.]

2.22 EXPECTATION OF LIFE OF THE PATIENTS

The expectation of life is a single measure summarising the survival of the patients. Technically, expectation of life is the area under the survival curve, obtained by joining the consecutive observed survival proportions (Figure 2.15). Directly calculating the area under the curve is impractical for most cancer sites, however, as we must follow-up all patients until their death in order to obtain a solution. However, a study of annual relative survival ratios may reveal that the patients no longer experience excess mortality compared to the general population after a given point in the follow-up. Thus, after this point, called the point of cure, it may be assumed that the patients have, on average, the same fate as the comparable general population group [55]. It is then possible to estimate expectation of life for the patients by combining two components; the first based on the observed survival of the patients whilst they were under follow-up, and the second component based on the expected survival of the general population.

In Figure 2.16, for example, patients diagnosed with extra-nodal non-Hodgkin's lymphomas in Finland in 1985–94 do not experience excess mortality due to the cancer after 5 years of follow-up. The only problem is that the comparable general population group will survive well into the 2000s, so survival probabilities for the future are required. A practical solution is to apply the most recent survival probabilities available for the general population to all future periods, which is how the expectation of life of the general population is usually calculated in practice.

Breast cancer is conspicuous among cancers in that a point of cure is not reached and patients experience excess mortality for decades after diagnosis [38]. It may well be possible to quantify this excess mortality in a few coefficients and to apply these coefficients in adjusting the general-population mortality for an asymptotic estimation of survival of breast-cancer patients [55].

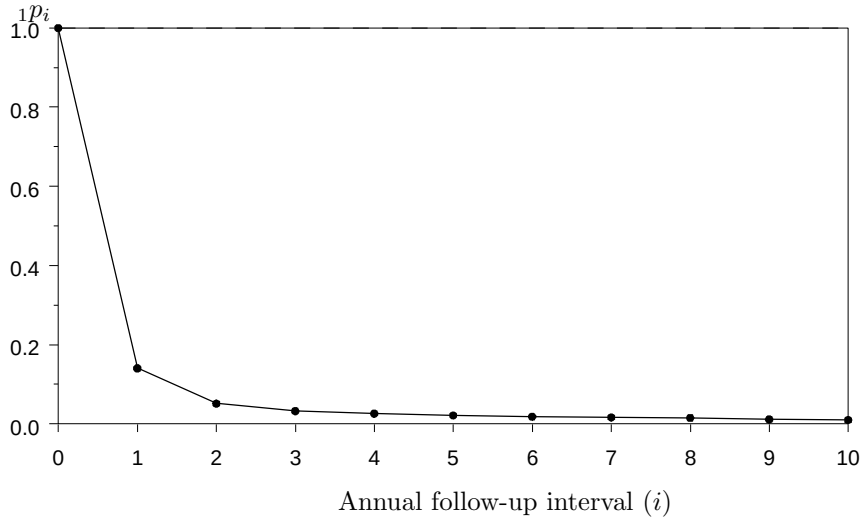

Fig. 2.15 Cumulative observed survival proportions ($_1p_i$) for males diagnosed with pancreatic cancer in Finland 1985–94.

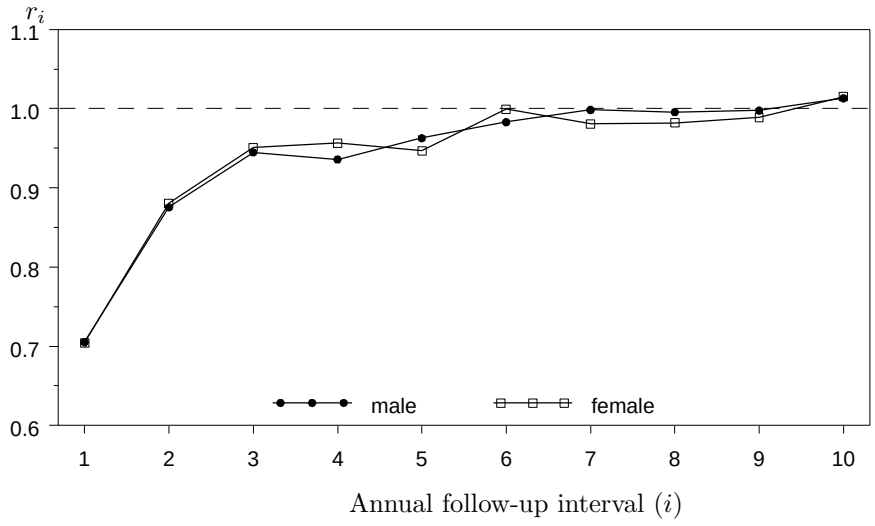

Fig. 2.16 Interval-specific (annual) relative survival ratios (r_i) for males and females diagnosed with extra-nodal non-Hodgkin's lymphomas in Finland 1985–1994 and followed up to the end of 1995.

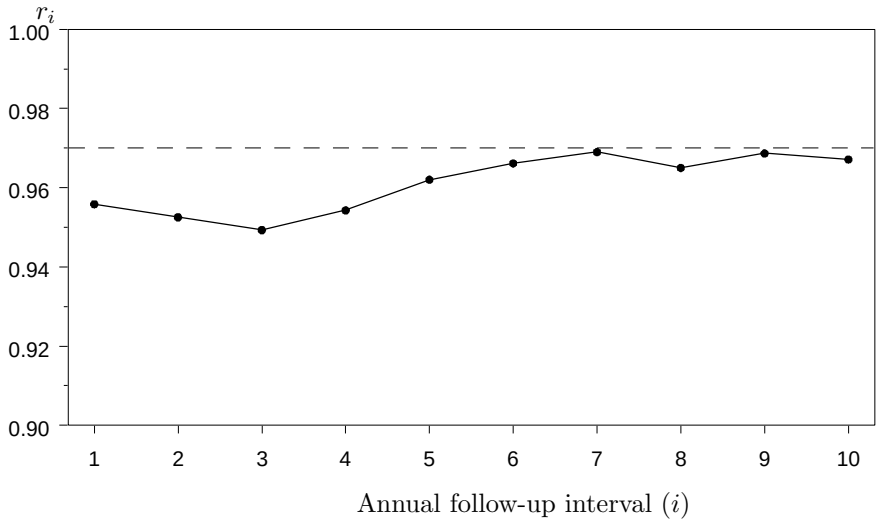

Fig. 2.17 Interval-specific (annual) relative survival ratios (r_i) for females diagnosed with cancer of the breast in Finland 1985–1994 and followed up to the end of 1995.

Most typically, the annual relative survival ratios stabilise asymptotically around a value of slightly less than one.

The most convenient formula for the expectation of life at the time origin (i.e. diagnosis) is

$$e_1 = 0.5 + \sum_{i=1}^{j-1} {}_1p_i + {}_1p_j(0.5 + e_{j+1}), \tag{2.44}$$

in which j is the length of the available life table and e_{j+1} is the expectation of life for patients alive at the beginning of interval $j+1$ (end of interval j). Note that many authors use e_0 to denote the expectation of life at the time origin (e.g. birth or diagnosis of cancer). We have attempted to index all quantities in this text according to life table intervals, where the time origin occurs at the start of interval 1, and have thereby used e_1 to denote the expectation of life at the time origin. The expectation of life at, for example, the start of the mth interval, e_m, can be calculated by replacing the lower limit of the summation in Equation 2.44 with $i = m$ and the cumulative survival probabilities ${}_1p_i$ and ${}_1p_j$ with ${}_mp_i$ and ${}_mp_j$.

If interval j is beyond the point of cure then

$$e_{j+1} = \sum_{h=1}^{l_j} e^*_{j+1}(h)/l_{j+1} = e^*_{j+1}, \tag{2.45}$$

in which $e^*_{j+1}(h)$ is the expectation of life for a person in the general population with the same sex, age and calendar period as the hth patient alive at the beginning of interval $j+1$ (Figure 2.18). If interval j is beyond a point of stabilisation (i.e. the interval specific RSRs have stabilised at a value less than 1) then

$$e_{j+1} = \sum_{h=1}^{l_j} e'_{j+1}(h)/l_{j+1}, \tag{2.46}$$

in which

$$e'_{j+1}(h) = 0.5 + \sum_{s=j+1}^{\infty} r^s \times {}_{j+1}p^*_s(h) \tag{2.47}$$

where r is the asymptotic value of the annual relative survival ratio, that is, $p_s/p^*_s = r$ for $s \geq j+1$. In practice, the sum does not extend to values of age higher than, say 105 years. If ${}_1p_j = 0$ the last term in the formula for e_1 is zero. A problem arises when the length of the life table has not reached the point of cure or stabilisation and the patients with the longest follow-up times have not died. In this situation, the estimation of expectation of life is not possible without making further assumptions.

The formula for e_1 presupposes that deaths occur, on average, at the midpoint of each follow-up interval. This is usually not the case in the first interval if annual intervals are used and the cancer is very fatal. In this situation, shorter intervals can be used or formula for e_1 can be adjusted to account for the uneven distribution of deaths in the first interval. To make the adjustment, we let a_1 be the proportion of the first interval survived, on average, by the individuals who died during the first interval. If $a_i = 0.5$ for $i > 1$ (that is, deaths occur, on average, at the midpoint of the interval for all intervals except the first) an adjusted estimate for e_1 is given by

$$e_1 = a_1 + (1.5 - a_1)p_1 + \sum_{i=2}^{j-1} {}_1p_i + {}_1p_j(0.5 + e_{j+1}). \tag{2.48}$$

In practice, the estimate of the expectation of life changes very little as a result of using Equation 2.48 in place of Equation 2.44 (see the numerical example later in this section). A general formula for the case where $a_i \neq 0.5$ when $i > 1$ is given by Chiang [28]. The formula for e_1 also presupposes that there are patients alive and under observation at the beginning of interval $j+1$, or that all patients with the longest follow-up time have died. If this is not true, i.e., there were patients withdrawing alive in interval j, in addition to possible deaths, but none surviving with a follow-up extending to interval $j+1$, there are two possible courses of action. The simplest alternative is to neglect the information from interval j and to use the formula for e_1 with $j-1$ taking the place of j. However, several alternatives exist which utilise all of the available information, which are described in Appendix 2 of Hakama and Hakulinen [55].

The standard error of the expectation of life, by holding e_{j+1} as a constant for the formula for e_1, is approximately [55]

$$\mathrm{SE}(e_1) = (\sum_{i=1}^{j}[f_i \mathrm{SE}(p_i)]^2)^{\frac{1}{2}} \tag{2.49}$$

in which

$$f_i = {}_1p_{i-1}\{\sum_{s=i+1}^{j-1} {}_{i+1}p_s + {}_{i+1}p_j(0.5 + e_{j+1})\} \text{ when } 1 < i < j, \tag{2.50}$$

$$f_1 = \sum_{s=2}^{j-1} {}_2p_s + {}_2p_j(0.5 + e_{j+1}), \tag{2.51}$$

and

$$f_j = {}_1p_{j-1}(0.5 + e_{j+1}). \tag{2.52}$$

Since the expectation of life is based on observed survival proportions, it is crucial that they are estimated with as little bias as possible. Biased estimation is often reflected in implausible values for e_1. The most important consideration is to ensure that the potential follow-up times and risk of death do not become heavily dependent. For example, if the potential follow-up times for young patients are, on average, longer than those for older patients, the quantity e_{j+1} in the formulae for e_1 will be estimated on the basis of young patients, resulting in an overestimate of e_1. Consequently, it is important to stratify the calculation of e_1 by age, using at least four age groups, when the recruitment period is not short – more than 10 years, for example. With short recruitment periods, there is little room for heterogeneity in potential follow-up times by age or any other variable. When age-specific calculations have been made, the final overall e_1 is calculated from the age-specific expectations as a weighted average using the age-specific numbers of patients at the beginning of follow-up as weights.

The expectation of life summarises the observed survival of the patients and, as with relative survival, it is possible to compare the observed e_1 with an expected value, e_1^*, based on the expected mortality of a comparable general population group. This value may be obtained using published age-specific expectations of life in the general population as

$$e_1^* = \sum_{h=1}^{l_1} e_1^*(h)/l_1 \tag{2.53}$$

where $e_1^*(h)$ is the expectation of life for a person in the general population with the same sex, age, and calendar period as the hth person at the beginning of the follow-up.

As the quantities $e_1^*(h)$ are often published for calendar-time periods and the patients may be followed through several such periods, an estimator which takes better account of calendar-time changes is given by

$$e_1^*(h) = 0.5 + \sum_{i=1}^{\infty} {}_1p_i^*(h) \tag{2.54}$$

where the terms of the product

$${}_1p_i^*(h) = \prod_{j=1}^{i} p_j^*(h) \tag{2.55}$$

are obtained from appropriate time periods if possible. Again the infinite sum should be terminated at high finite age values, e.g., at the age of 105 years.

Finally, the e_1 can be compared with the e_1^* by calculating the difference

$$\Delta_1 = e_1^* - e_1, \tag{2.56}$$

which is the loss in the expectation of life attributable to the excess mortality of the patients. When this difference is further divided by e_1^*, the resulting quantity

$$\gamma_1 = (e_1^* - e_1)/e_1^* \tag{2.57}$$

is the proportion of expected life lost due to the excess mortality of the patients. Both Δ_1 and γ_1 can be negative, indicating that the survival experience of the patients was superior to that expected of a comparable group from the general population.

Let us study the survival of patients diagnosed with extra-nodal non-Hodgkin's lymphomas in Finland 1985-94 as an example (Figure 2.16). These patients do not experience excess mortality after five years of follow-up. Thus, it is reasonable to use Equation 2.44 to estimate the expectation of life with $j = 10$ and $e_{j+1} = e_{j+1}^*$. The expectation of life e_1 of the patients decreases with age, as does the expected value e_1^* (Table 2.15). The difference, Δ_1, also decreases with age, but the proportion of expected life lost, γ_1, increases with age in both sexes. The effect of calculating e_1 as a weighted average is, in this example, only slight: e_1=12.89 for males from a single life table for all ages combined vs. e_1=11.73 obtained as the weighted average of the age-specific expectations of life. Within a 10-year recruitment period there is not a whole lot of room for differences in potential follow-up times.

If we made the (rather extreme) assumption that the patients who died during the first year of follow-up survived, on average, 2.4 months (i.e. $a_1 = 0.2$), rather than the 6 months ($a_1 = 0.5$) assumed by Equation 2.44, the first two terms in Equation 2.44, $0.5 + {}_1p_1$ (noting that ${}_1p_1 = p_1$), would be replaced by the first two terms of Equation 2.48, $0.2 + 1.3p_1$. The estimated expectation of life would thereby be reduced by $0.3(1 - p_1)$. For males aged 0–44 years

at diagnosis, $p_1 = 0.8733$, so the reduction in the estimate of expectation of life, e_1, would be 0.038 years. For males aged 85 years or more at diagnosis, $p_1 = 0.3333$ so the reduction in the estimate of e_1 would be 0.2 years. The estimate of the expectation of life is therefore quite robust to violations (during the first interval) of the assumption that deaths occur, on average, at the midpoint of each interval.

Breast-cancer patients experience a relatively constant excess mortality for at least 20 years after diagnosis [38, 56]. The annual relative survival ratio appears to stabilise at approximately $r = 0.97$ between 5 and 10 years after diagnosis (Figure 2.17). An application of this value asymptotically for $j > 10$ has a marked effect on the expectation of life (Table 2.16). In practice it may take longer than 10 years to find an asymptotic value of r and this value may vary by age and other background factors. Other functional forms for asymptotic excess mortality, other than a constant relative survival, are also possible.

The expectation of life has one feature which is quite unsatisfactory, particularly with highly fatal cancers such as cancer of the pancreas (Figure 2.15). Almost all patients die, leaving very little room for any asymptotic assumptions. Nevertheless the value of e_1 is 0.7 years (based on Equation 2.48 with $a_1 = 0.2$), and it can be seen from Figure 2.15 that only a small proportion of the patients, approximately 30%, live longer than the expectation of life. This is a known phenomenon for skewed distributions with few long survival times and many short ones. Thus, a median length of life might be a more satisfactory measure than the expectation of life which, in essence, is the mean length of life. For quantities such as e_1, Δ_1, and γ_1 it does not matter if ten years are lost by one patient or one year lost by each of ten patients. There are also problems, although of a different kind, with the median (Section 2.23).

Table 2.15 Number of cases l_1, expectation of life for a comparable general population group e_1^*, that for the patients e_1, its standard error, loss in the expectation of life Δ_1, and proportion of expected life lost γ_1, for patients diagnosed with extra-nodal non-Hodgkin's lymphomas in 1985-94 in Finland, by sex and age.

	males						females					
Age	l_1	e_1^*	e_1	SE(e_1)	Δ_1	γ_1	l_1	e_1^*	e_1	SE(e_1)	Δ_1	γ_1
0–44	132	41.49	35.53	3.229	5.96	0.144	100	46.80	35.06	5.558	11.74	0.251
45–54	116	25.36	17.09	2.726	8.27	0.326	92	31.19	22.59	3.480	8.60	0.276
55–64	253	17.43	10.38	1.443	7.05	0.404	190	21.91	14.92	1.959	6.99	0.319
65–74	264	11.35	5.26	1.114	6.09	0.537	315	14.50	7.47	0.965	7.03	0.485
75–84	160	6.77	3.37	0.854	3.40	0.502	293	8.16	3.78	0.704	4.38	0.537
85+	36	3.95	1.25	0.501	2.70	0.684	63	4.55	1.50	0.563	3.05	0.670
All[1]	961	17.74	12.89	1.198	4.85	0.273	1053	18.00	11.40	0.966	6.60	0.367
All[2]	961	17.74	11.73		6.01	0.339	1053	18.00	11.37		6.63	0.368

[1] based on the life table for all ages combined

[2] e_1 is calculated as a weighted average of the stratum-specific values using age-specific numbers of cases l_1 as weights

2.22.1 Example: expectation of life of the general population

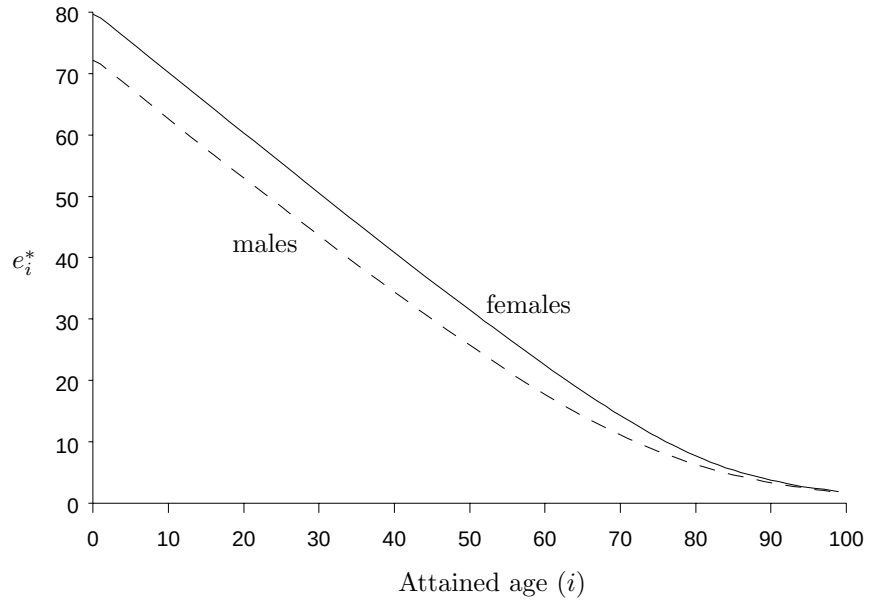

Fig. 2.18 Expectation of life (e_i^*) as a function of attained age for the Finnish general population 1990–1995.

2.23 MEDIAN SURVIVAL TIME

Since the distribution of survival times is often positively skew, that is, there are many short but relatively few long survival times, the median survival time is often preferable to the mean as a summary measure of patient survival. The median survival time is the time beyond which 50% of the individuals in the population are expected to survive. It is estimated from the life table as the time at which the cumulative observed survival proportion falls below 0.5. The median is estimated by extrapolation if the cumulative observed survival proportion does not sink below 0.5 during the period the patients are under follow-up.

For the pancreatic-cancer patients in Figure 2.15 the median falls in the range of observation and is XXX years. But for young patients with localised melanoma the cumulative observed survival proportion stays well above 50 per cent for the entire 10-years of follow-up (Figure 2.19). The best alternative here is to use another quantile than the median to summarise the survival

Table 2.16 Number of cases l_1, expectation of life for a comparable general population group e_1^*, and that for the patients e_1, both without assuming an asymptotic annual relative survival ratio after 10 years of follow-up ($r = 1$) and with an assumption of $r = 0.97$, along with the loss in the expectation of life Δ_1 and proportion of expected life lost γ_1 for females diagnosed with cancer of the breast in 1985-94 in Finland, by age.

			$r = 1$			$r = 0.97$		
Age	l_1	e_1^*	e_1	Δ_1	γ_1	e_1	Δ_1	γ_1
0–44	3481	41.03	28.40	12.63	0.308	20.59	20.44	0.498
45–54	5841	31.50	24.67	6.83	0.217	19.65	11.85	0.376
55–64	5598	22.60	17.30	5.30	0.235	15.18	7.42	0.328
65–74	5156	14.47	10.65	3.82	0.264	10.07	4.40	0.304
75–84	4071	8.22	6.19	2.03	0.247	6.08	2.14	0.260
85+	1055	4.40	3.29	1.11	0.252	3.26	1.14	0.259
All[1]	25202	22.46	16.47	5.99	0.267	13.70	8.76	0.390
All[2]	25202	22.46	16.80	5.66	0.252	13.95	8.51	0.379

[1] based on the life table for all ages combined

[2] e_1 is calculated as a weighted average using age-specific numbers of cases l_1 as weights

experience, e.g., the lower quartile where the survival curve sinks below 75%. In Figure 2.19 this value is XXX years.

If however, an estimate of the median is desired, extrapolation techniques introduced in Section 2.22 may be taken into use. The patients in Figure 2.19 have a point of cure before ten years and it is therefore reasonable to assume that the 36 patients alive at ten years will have a future survival similar to a comparable general population group. Thus, the cumulative observed survival proportion after the last life table interval j ($j = 10$ in Figure 2.19) is extrapolated as

$$ {}_1p_s = {}_1p_j \; {}_jp_s^* \text{ for } s > j, \tag{2.58} $$

where

$$ {}_jp_s^* = \sum_{h=1}^{l_j} {}_jp_s^*(h)/l_j, \tag{2.59} $$

when the cure point is reached before j. If the cure point is not reached before the end of interval j, but the relative survival ratio has stabilised at a value r (where $r < 1$) and we are prepared to assume that $r_s = r$ for $s > j$ then, in place of Equation 2.58, the cumulative observed survival proportion for $s > j$ is estimated by

$$ {}_1p_s = {}_1p_j \; {}_jp_s^* \, r^s. \tag{2.60} $$

Other functional forms are also possible.

[Complete this example!] For patients in Figure 2.19 the result is XXX years, which clearly reflects not only the disease-specific mortality, but also the general mortality during the years to come. This property is also shared by the expectation of life (Section 2.22).

The median survival time may be a rather unstable measure when patients are young and approximately one half die because of the disease. When slightly more than one half survive, the median survival time becomes very long and is found a number of years after the point of cure. When slightly less than one-half survive, the median survival time can be dramatically shorter. The difference between the estimated medians in these two situations clearly exaggerates the difference in the severity of the disease.

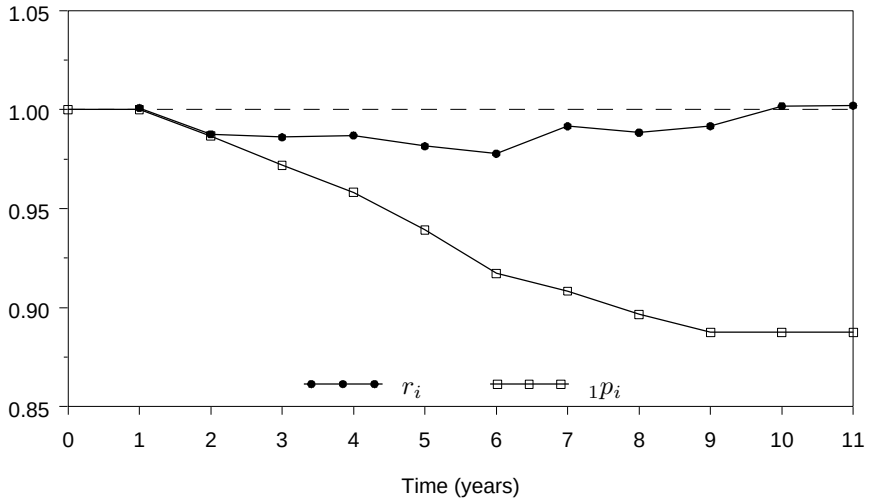

Fig. 2.19 Cumulative observed survival ($_1p_i$) and annual relative survival ratios (r_i) for 469 females diagnosed with localised melanoma in Finland 1985–1994, aged 0-44 years at diagnosis

2.24 AGE ADJUSTED SURVIVAL PROPORTIONS

The survival proportions for males diagnosed with localised colon carcinoma and localised skin melanoma are presented as an illustration of the relative survival ratio (Table 2.17). Considering first the men diagnosed with localised colon carcinoma, we see that the 5-year observed survival proportion (p) decreases with increasing age. The 5-year observed survival proportion for the men aged 75 years and over at diagnosis was $p = 0.411$. For a group of men of

similar age, but without a diagnosis of colon carcinoma, drawn from the same population as the cancer patients, the five-year expected survival proportion was $p^* = 0.520$. That is, we would expect 52.0% of these men to survive 5 years or more if they were free of the colon carcinoma. The relative survival ratio (r) is therefore calculated as $0.411/0.520 = 0.790\%$. Therefore, 79.0% of these men would survive at least five years in the hypothetical situation where colon carcinoma were the only possible cause of death. Although the observed survival proportions decrease with age, the relative survival ratios remain relatively constant at around 0.80. This indicates that the excess mortality due to a diagnosis of localised colon carcinoma is similar for all age groups.

Table 2.17 Number of cases (N) and 5-year observed (p), expected (p^*), and relative (r) survival proportions for males diagnosed with localised colon carcinoma and localised skin melanoma in Finland during 1985–1994.

	localised colon carcinoma				localised skin melanoma			
Age	N	p	p^*	r	N	p	p^*	r
15–29	2	-	0.994	-	67	0.947	0.993	0.954
30–44	84	0.805	0.980	0.821	273	0.856	0.982	0.872
45–59	257	0.795	0.936	0.849	503	0.824	0.943	0.874
60–74	731	0.628	0.793	0.792	449	0.679	0.815	0.833
75+	500	0.411	0.520	0.790	200	0.396	0.505	0.784

The pattern of age-specific RSRs for the men diagnosed with localised colon carcinoma is the exception rather than the rule. For most cancer sites, the excess mortality due to cancer increases with age, as is the case with localised skin melanoma (Table 2.17).

When survival (also cause-specific and relative) depends on age and comparisons between groups are desired, it is good practice to make the comparisons by age group. This, however, means that the number of comparisons increases. But if age-specific comparisons are not done, there is a danger that misleading conclusions will be made.

Let us take Table 2.17 as an example and assume that the female localised skin melanoma patients had a four per-cent unit higher cumulative observed survival ratio than the male patients in every age group, as shown in Table 2.18. Further, let us assume that the age distribution of the female patients were that of male localised colon carcinoma patients. With these assumptions the female patients have a higher relative survival ratio in each age group than the male patients, but nevertheless, have a lower overall relative survival ratio (Table 2.18).

Obviously, the controversial result is due to differences in the age distribution of the patients. The female patients are older than the male patients, and the

Table 2.18 A hypothetical example - observed survival proportions (p) following a diagnosis of localised skin melanoma, by sex and age. [Find a better example]

	females			males		
Age	N	%	p	N	%	p
30–44	84		0.987	273		0.947
45–59	257		0.896	503		0.856
60–74	731		0.864	449		0.824
75+	500		0.719	200		0.679
Total†	1572		0.830	1425		0.839

†Age-adjusted estimates for all patients

lower relative survival ratios at high ages get a larger weight in determining the overall rate in females than in males. This phenomenon would also hold true even if the overall rates had not been obtained as weighted averages of the age-specific rates. But then there might be also problems with censoring patterns, at least when long-term survival is concerned (see Section 2.16).

In order to get comparable results for males and females in Table 2.18 it is clear that the overall relative survival ratios for males and females must be calculated by using the same age distribution, producing so-called age-standardised rates. The age distribution of males and females combined is a natural choice for this distribution (Table 2.18).

[Further discussion of the use and interpretation of age-standardised rates.]

[HERE SOME DISCUSSION OF THE RESULTS]

[Find an example with relative survival, preferably a real-life example. Thyroid maybe?]

2.25 TESTING EQUALITY OF SURVIVAL PATTERNS

[The following points are taken from the course notes and will be rewritten.]

- Comparing survival at a fixed time point (e.g. five years) wastes available information.
- It is invalid to select the time of evaluation after inspecting the data, for example, testing for a difference at the point where the Kaplan-Meier curves show the largest difference.
- Various tests are available (both parametric and non-parametric) for testing equality of survival curves.

- The most common is the log rank test, which is non-parametric (it makes no assumption about the shapes of the underlying survival curves). An example is shown on the next slide.
- Start by tabulating the number at risk in each group and the total number of events (deaths) at every time point when one of more deaths occur.
- Under the null hypothesis (the two survival curves are the same), the expected number of deaths in each group is proportional to the number at risk in each group.
- For example, at $t = 2$ months we observed 2 deaths (one male and one female). Conditional on 2 deaths being observed, we would expect $2 \times 19/35 = 1.086$ deaths among the 19 males at risk and $2 \times 16/35 = 0.914$ deaths among the 16 females at risk.
- Now calculate the totals of the observed and expected number of deaths for each group, calling them O_1, O_2, E_1, and E_2, and calculate the following test statistic:

$$\theta = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \tag{2.61}$$

- Under the null hypothesis, θ will follow a χ^2 distribution with 1 degree of freedom. That is, if θ is greater than 3.84 then we reject the null hypothesis and conclude that there is a statistically significant difference between the two survival curves.
- For k groups, the test statistic is

$$\theta = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{2.62}$$

  which is approximately χ^2_{k-1} under the null hypothesis.
- The log rank test is similar to the Mantel-Haenszel procedure for stratified 2×2 tables. We essentially construct a 2×2 at each time a death occurs and test whether the binary outcome (death) is independent of exposure.
- The log rank test is designed to be sensitive to departures from the null hypothesis in which the two hazards (instantaneous death rates) are proportional over time. It is very insensitive to situations in which the survival curves cross.
- The log rank test puts equal weight on every observation (event time). An alternative test, the generalised Wilcoxon test, is weighted according

Table 2.19 Logrank test for comparing survival of males and females for the 35 patients shown in Table 1.2.

event	males			females		
time	at risk	observed	expected	at risk	observed	expected
2	19	1	1.086	16	1	0.914
3	18	1	0.545	15	0	0.455
5	17	1	0.531	15	0	0.469
7	16	1	0.516	15	0	0.484
8	15	1	0.500	15	0	0.500
9	14	0	0.483	15	1	0.517
11	14	0	0.500	14	1	0.500
19	13	0	0.520	12	1	0.480
22	13	1	0.542	11	0	0.458
27	12	0	0.545	10	1	0.455
28	11	0	0.550	9	1	0.450
32	11	2	1.158	8	0	0.842
33	9	1	0.563	7	0	0.438
43	8	0	0.615	5	1	0.385
46	8	1	0.667	4	0	0.333
102	2	0	0.500	2	1	0.500
103	2	1	0.667	1	0	0.333
		$O_1 = 11$	$E_1 = 10.488$		$O_2 = 8$	$E_2 = 8.512$

to the total number of individuals at risk and is consequently more sensitive to differences early in the follow-up period.

See the example in Table 2.19. The test statistic is given by $\chi^2 = (O_1 - E_1)^2/E_1 + (O_2 - E_2)^2/E_2 = 0.056$ which is compared to a chi-square statistic with 1 degree of freedom.

Describe the Buckley test (reported as Pearson test in the program).

The Hakulinen LR tests are preferred.

Hakulinen *et al.* (1987) [57] described likelihood ratio tests for testing the equality of relative survival ratios based on grouped data between patient groups. The tests are applicable to the situation where the patients are stratified according to one variable, for example, age at diagnosis or year of diagnosis, and a life table constructed for each patient group. We then test the equality of relative survival ratios between the subgroups. Note that there is no limit on the number of patient subgroups, although the subgroups can only be defined on the basis of one variable. It is therefore not possible to, for example, test for the effect of age while controlling for sex, or to test whether

age and sex are both independently associated with survival. It is, of course, possible to construct combinations of variables and test for survival differences across levels of the composite variable.

Because of this limitation, we recommend that testing the equality of survival between patient groups be carried out in a model-based framework. Hakulinen and Tenkanen (1987) [58] described a life table regression model for the relative survival ratio which performs equivalent likelihood ratio tests to the ones described in this section. Advantages of the model-based approach are that more than one variable can be tested simultaneously and standard errors and confidence intervals for estimated hazard ratios can be easily obtained.

The methods are also appropriate for testing equality of observed survival proportions. We will present an overview of the methods, which are then applied to an example where the relative survival ratios for patients with localised melanoma (males and females combined) diagnosed during 1975–1994 are compared between four age groups (30–44, 45–59, 60–74, and 75+).

Let k be the index for the groups being compared (m groups in total) and i the index for follow-up interval (w intervals in total). The relative survival ratio in the ith follow-up interval of the kth patient group is given by

$$r_{ki} = p_{ki}/p^*_{ki} \text{ for all } k \text{ and } i. \tag{2.63}$$

The null hypothesis is the equality of relative survival ratios in each follow-up interval:

$$H_0 : r_{ki} = r_{1i} \text{ for all } k \text{ and } i. \tag{2.64}$$

Alternative hypotheses of interest are the proportional hazards hypothesis:

$$H_1 : r_{ki} = r_{1i}^{c_k} \text{ for all } k \text{ and } i, \tag{2.65}$$

which is clearer when written in terms of the excess hazard (λ_{ki}):

$$H_1 : \lambda_{ki} = \lambda_{1i} c_k \text{ for all } k \text{ and } i, \tag{2.66}$$

and the hypothesis of general inequality,

$$H_2 : r_{ki} \text{ unrestricted.} \tag{2.67}$$

It is not always useful to reject H_0 in favour of H_2, because the order of the groups with respect to the relative survival ratio may vary from interval to interval under H_2.

[Give some graphs showing examples of these hypotheses.]

Hakulinen et al. (1987) [57] described the following likelihood ratio tests:

- H_0 vs H_1, with $m - 1$ degrees of freedom,
- H_0 vs H_2, with $w(m - 1)$ degrees of freedom, and

- H_1 vs H_2, with $(m-1)(w-1)$ degrees of freedom.

In calculating these test statistics, the life tables are condensed by pooling a follow-up interval with the previous one if either the number alive at the start of the interval (l_{ki}) or number of deaths during the interval (d_{ki}) are equal to zero. If d_{ki} is zero in the first interval, the first interval is pooled with the second. When testing the proportional hazards hypothesis (H_1), the life tables are all truncated before the estimates of the relative survival ratio under the null hypothesis exceed unity. If the life tables shortened in this way contain less than two intervals, the truncation is abandoned and the tests of the proportional hazards hypothesis are not performed.

Another hypothesis of interest is the hypothesis of constant odds ratios:

$$H_1' : \frac{r_{ki}}{1-r_{ki}} / \frac{r_{1i}}{1-r_{1i}} = R_k \text{ for all } k \text{ and } i. \tag{2.68}$$

Brown (1983) [59] presented a score statistic for testing H_0 against H_1', although Hakulinen et al. (1987) [57] discussed the relative merits of H_1 and H_1' and gave reasons for preferring H_1. [Should we elaborate on this?]

[Include details of Buckley's test (labelled as Pearson in the program output).]

2.25.1 Application of the tests of equality of the RSR

[Rework this section so that it is not program-driven.]

The Relative Survival Program produced the following output:

```
H0  Null hypothesis
      30-44 45-59 60-74  75+
   1  0.994 0.994 0.994 0.994
   2  0.960 0.960 0.960 0.960
   3  0.959 0.959 0.959 0.959
   4  0.969 0.969 0.969 0.969
   5  0.972 0.972 0.972 0.972

H1  Proportional alternative
      30-44 45-59 60-74  75+
   1  0.995 0.994 0.992 0.988
   2  0.969 0.962 0.950 0.923
   3  0.969 0.962 0.949 0.921
   4  0.975 0.970 0.960 0.937
   5  0.978 0.973 0.964 0.944

H2  General alternative
```

```
     30-44 45-59 60-74  75+
  1  0.994 0.994 0.995 0.995
  2  0.973 0.963 0.947 0.915
  3  0.972 0.959 0.947 0.931
  4  0.974 0.971 0.958 0.938
  5  0.972 0.974 0.974 0.939
```

Statistic	Chi sq.	D.F.	Critical value 5%	Critical value 1%	p-value
H0 - H2	33.97	15	25.00	30.58	0.003
H0 - H1	28.41	3	7.81	11.35	0.000
H1 - H2	5.56	12	21.03	26.22	0.937
Pearson	24.65	3	7.81	11.35	0.000

Hence, the data can be described by a proportional hazards model. [more]

Can estimate the hazard ratios from the RSRs under H_1.

[Table of hazard ratios here]

2.25.2 Equivalent tests in a model-based framework

[These results will eventually be shifted to the section on modelling]

A life table regression model was fitted to the same data. The data contained 20 observations (4 age groups by 5 follow-up intervals). The model containing only the term for follow-up (as a categorical variable with 5 levels) had deviance of 33.97 on 15 degrees of freedom, indicating lack-of-fit. Note that the data contained 20 observations and we fitted a model containing 5 parameters, which left us with 15 residual degrees of freedom.

After adding age to the model (as a categorical variable with 4 levels), the deviance is 5.56 on 12 degrees of freedom, indicating a model which provides a good fit to the data. The likelihood ratio test (changed in deviance) for the effect of age is therefore $33.97 - 5.56 = 28.41$ on $15 - 12 = 3$ degrees of freedom, indicating that age is highly significant in the model. This is the identical test to the test of H_0 against H_1 described in Section 2.25 and leads to the same test statistic as the one shown in Section 2.25.1.

The model-based estimates of the hazard ratios (with 95% CIs) are as follows:

Parameter	Level	Estimated RR	Lower limit 95% CI	Upper limit 95% CI

INTERCEPT		0.005	0.003	0.009
FU	1	.	.	.
FU	2	6.570	3.669	11.767
FU	3	6.686	3.722	12.009
FU	4	5.274	2.893	9.615
FU	5	4.686	2.538	8.654
AGE	30-44	.	.	.
AGE	45-59	1.225	0.950	1.580
AGE	60-74	1.644	1.267	2.133
AGE	75+	2.579	1.835	3.625

Note that these are identical to the estimated hazard ratios in Section 2.25.1, except only point estimates were available there.

Adding an age by follow-up interaction to the model produces the saturated model, where the observed data are fitted exactly by the model. The saturated model corresponds to the general hypothesis, H_2, described in Section 2.25. The age by follow-up interaction term uses 12 degrees of freedom and reduces the deviance by 5.56 when it is added to the model, indicating that the interaction term is not statistically significant in the model and that the data can be described by the proportional hazards model. The model-based likelihood ratio test is equivalent to the test of H_1 against H_2 described in Section 2.25 and leads to the same test statistic as the one shown in Section 2.25.1.

The test statistics obtained using the Hakulinen et al. (1987) [57] likelihood ratio tests for testing the equality of relative survival ratios will be the same as the model-based tests, provided all life table intervals contain at least one death. When no deaths are observed in an interval, the interval is combined with an adjacent interval when the Hakulinen et al. (1987) [57] tests are performed, which leads to slightly different tests to the model-based tests, although the interpretation should not change. For example, if we included a fifth age group, patients aged 15–29, the following test statistics are obtained:

	standard tests			model-based tests		
Test	Chi sq.	df	p-value	Chi sq.	df	p-value
H0 - H2	54.69	16	0.000	60.67	20	0.000
H0 - H1	47.77	4	0.000	51.02	4	0.000
H1 - H2	6.92	12	0.863	9.65	16	0.884

No deaths occurred in the first follow-up interval among the 15–29 year olds, meaning the first and second intervals are combined when the tests are performed.

2.26 DEFINING PATIENT GROUPS

survival proportions are estimated for patient groups defined to be as homogeneous as possible, although, because of the relatively large number of patients required in order to obtain a reasonably precise estimate, the patient groups are often broader than what would be clinically desired. It is, for example, possible to estimate sex-specific cancer incidence rates with reasonable precision for 5-year age categories for individual geographic regions in most countries, but survival proportions estimated for the same patient groups would be essentially meaningless due to the lack of precision.

Due to the relatively large size of the population covered by the Finnish Cancer Registry and the high accuracy of and large number of items collected by the registry (not all population-based registries, for example, have access to accurate information on stage and morphology), it is possible to estimate survival for patient groups with relatively narrow definitions in terms of what can usually be achieved in survival analyses, although the definitions are still broad in a clinical sense. For example, estimated survival proportions for females aged 45–59, diagnosed with localised colon carcinoma during 1985-1994 are published in Dickman et al. (1998) [19], as are stage and sex specific rates for each cancer site according to 15-year age groups. Even though this is a rather strict definition, we know that the individual patients in this group are heterogeneous to some extent; there will be differences, for example, in their general health and receptiveness to treatment, and differences in the tumour characteristics, even though the group is restricted to localised colon carcinomas.

The estimated five-year observed survival proportion estimated for this group represents the best estimate of the average survival proportion for the women who make up this group. Any associated confidence interval also applies to the average survival proportion for the group, and does not, for example, represent the range of possible prognoses for an individual patient. The width of the confidence depends on both the homogeneity of survival times in the group - more homogeneous survival times will result in more certainty in the estimated rate, and hence narrower confidence intervals - and the number of individuals in the group - more individuals will result in narrower confidence intervals.

The group defined above, which is a subset of the total Finnish population of approximately 5 million (both sexes), consisted of 297 individuals. This is sufficient to obtain a reasonably accurate estimate of the five-year survival rate and associated confidence interval. [more]

References

1. Collett D. *Modelling Survival Data in Medical Research*. London: Chapman & Hall, 1994.
2. Parmar MKB, Machin D. *Survival Analysis: A Practical Approach*. John Wiley & Sons, 1995.
3. Marubini E, Valsecchi MG, Emmerson M. *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley & Sons, 1995.
4. Kleinbaum DG. *Survival Analysis: A Self-Learning Text*. Springer, 1996.
5. Le CT. *Applied Survival Analysis*. John Wiley & Sons, 1997.
6. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, 1997.
7. Muir CS. The cancer registry in cancer control: An overview. In: Parkin *et al.* [10], 1985; 13–26.
8. Young JL. The hospital-based cancer registry. In: Jensen *et al.* [87], 1991; 177–184.
9. Bell CMJ, Lawrence G, Pheby DFH, Smith J, Coleman MP. The role of cancer registries. *Clinical Oncology* 1995;**7**:143–144.
10. Parkin DM, Wagner G, Muir CS, eds. *The Role of the Registry in Cancer Control*. IARC Scientific Publications No. 66. Lyon: International Agency for Research on Cancer, 1985.

11. Jensen OM, Storm HH. Purposes and uses of cancer registration. In: Jensen *et al.* [87], 1991; 7–21.

12. Muir CS, Nectoux J. Role of the cancer registry. *National Cancer Institute Monograph* 1977;**47**:3–6.

13. Armstrong BK. The role of the cancer registry in cancer control. *Cancer Causes and Control* 1992;**3**:569–579.

14. Enstrom JE, Austin DF. Interpreting cancer survival rates. *Science* 1977; **195**:847–851.

15. Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J, eds. *Cancer Incidence in Five Continents, Volume VII*. IARC Scientific Publications No. 143. Lyon: International Agency for Research on Cancer, 1997.

16. Brandberg Y. Quality of life in clinical trials: Assessment and utility with special reference to rHuEPO. *Medical Oncology* 1998;**15 (Suppl. 1)**:S8–S12.

17. Olschewski M, Schlugen G, Schumacher M, Altman DG. Quality of life assessment in clinical cancer research. *British Journal of Cancer* 1994; **70**:1–5.

18. Coleman MP, Babb P, Damiecki P, Grosclaude P, Honjo S, Jones J, *et al.*. *Cancer Survival Trends in England and Wales, 1971–1995: Deprivation and NHS Region*. No. 61 in Studies in Medical and Population Subjects. London: The Stationery Office, 1999.

19. Dickman PW, Hakulinen T, Luostarinen T, Pukkala E, Sankila R, Söderman B, Teppo L. Survival of cancer patients in Finland 1955-1994. *Acta Oncologica* 1999;**38 (Suppl. 12)**:1–103.

20. Estève J, Benhamou E, Raymond L. *Statistical Methods in Cancer Research Volume IV: Descriptive Epidemiology*. IARC Scientific Publications No. 128. Lyon: International Agency for Research on Cancer, 1994.

21. Parkin DM, Hakulinen T. Analysis of survival. In: Jensen *et al.* [87], 1991; 159–176.

22. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: A statistical methodology. *National Cancer Institute Monograph* 1961;**6**:101–121.

23. Percy CL, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. *American Journal of Public Health* 1981;**71**:242–250.

24. Berkson J, Gage RP. Calculation of survival rates for cancer. *Proceedings of Staff Meetings of the Mayo Clinic* 1950;**25**:270–286.

25. Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. *Annual Reviews in Public Health* 1997;**18**:83–104.

26. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncologica* 1994;**33**:365–369.

27. Hakulinen T. Health care system, cancer registration and follow-up of cancer patients in Finland. In: Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T, Estève J, eds., *Survival of Cancer Patients in Europe: The EUROCARE Study*, IARC Scientific Publications No. 132. Lyon: International Agency for Research on Cancer, 1995; 53–54.

28. Chiang CL. *Introduction to Stochastic Processes in Biostatistics*. New York: John Wiley and Sons, 1968.

29. Böhmer PE. Theorie der unabhängigen Wahrscheinlichkeiten. *Rapports Memoires et Proces verbaux de Septieme Congres International dActuaires Amsterdam* 1912;**2**:327–343.

30. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958;**53**:457–481.

31. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B* 1972;**34**:187–220.

32. Gail MH. Competing risks. In: Kotz S, Johnson NL, eds., *Encyclopedia of Statistical Sciences*, vol. 2. John Wiley & Sons: New York, 1982; 75–81.

33. Tsiatis AA. Competing risks. In: Armitage P, Colton T, eds., *Encyclopedia of Biostatistics*, vol. 1. John Wiley & Sons: Chichester, 1998; 824–834.

34. Hakulinen T. On long-term relative survival rates. *Journal of Chronic Diseases* 1977;**30**:431–443.

35. Dickman PW, Gibberd RW, Hakulinen T. Estimating potential savings in cancer deaths by eliminating regional and social class variation in cancer survival in the Nordic countries. *Journal of Epidemiology and Community Health* 1997;**51**:289–298.

36. Dickman PW, Auvinen A, Voutilainen ET, Hakulinen T. Measuring social class differences in cancer patient survival: Is it necessary to control for social class differences in general population mortality? A Finnish population-based study. *Journal of Epidemiology and Community Health* 1998;**52**:727–734.

37. Cutler SJ, Axtell LM. Partitioning of a patient population with respect to different mortality patterns. *Journal of the American Statistical Association* 1963;**58**:701–712.

38. Kerr GP, Kunkler IH, Langlands AO, Rodger A. (In)curability of breast cancer: a 30-year report of a series of 3933 patients. *Breast* 1998;**7**:190–194.

39. Elandt-Johnson RC. Definition of rates: Some remarks on their use and misuse. *American Journal of Epidemiology* 1975;**102**:267–271.

40. Altman DG, Bland JM. Time to event (survival) data. *British Medical Journal* 1998;**317**:468–469.

41. Halpern DF, Coren S. Handedness and life span. *New England Journal of Medicine* 1991;**324**:998.

42. Raymond L, Fischer B, Fioretta G, Bouchardy C. Migration bias in cancer survival rates. *Journal of Epidemiology and Biostatistics* 1996;**1**:167–173.

43. Saxén E. Trends: facts or fallacy. In: Magnus K, ed., *Trends in Cancer Incidence. Causes and Practical Implications*. Hemisphere: Washington, 1982; 5–16.

44. Auvinen A, Karjalainen S, Pukkala E. Social class and cancer patient survival in Finland. *American Journal of Epidemiology* 1995;**142**:1089–1102.

45. Dickman PW, Hakulinen T. Adjusting for region of residence in relative survival analysis. *Journal of Epidemiology and Biostatistics* 1996;**1**:213–218.

46. Oksanen H. *Modelling the Survival of Prostate Cancer Patients*. Ph.D. thesis, University of Tampere, Finland, 1998. Acta Universitatis Tamperensis, vol 615.

47. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982;**38**:933–942.

48. Ederer F, Heise H. Instructions to IBM 650 programmers in processing survival computations. Methodological note No. 10, End Results Evaluation Section, National Cancer Institute, Bethesda MD, 1959.

49. Colton T. *Statistics in Medicine*. Boston: Little, Brown, 1974.

50. Greenwood M. *The Errors of Sampling of the Survivorship Table*, vol. 33 of *Reports on Public Health and Medical Subjects*. London: Her Majesty's Stationery Office, 1926.

51. Peto J. The calculation and interpretation of survival curves. In: Buyse ME, Staquet MJ, Sylvester RJ, eds., *Cancer Clinical Trial: Methods and Practice*. Oxford: Oxford University Press, 1984; 361–380.

52. Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.

53. Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T, Estève J, eds. *Survival of Cancer Patients in Europe: The EUROCARE Study*. IARC Scientific Publications No. 132. Lyon: International Agency for Research on Cancer, 1995.

54. Harris EK, Albert A. *Survivorship Analysis for Clinical Studies*. New York: Marcel Dekker, 1991.

55. Hakama M, Hakulinen T. Estimating the expectation of life in cancer survival studies with incomplete follow-up information. *Journal of Chronic Diseases* 1977;**30**:585–597.

56. Hakulinen T, Pukkala E, Hakama M, Lehtonen M, Saxén E, Teppo L. Survival of cancer patients in Finland in 1953–1974. *Annals of Clinical Research* 1981;**13 (Suppl. 31)**:1–101.

57. Hakulinen T, Tenkanen L, Abeywickrama K, Päivärinta L. Testing equality of relative survival patterns based on aggregated data. *Biometrics* 1987;**43**:313–325.

58. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Applied Statistics* 1987;**36**:309–317.

59. Brown CC. The statistical comparison of relative survival rates. *Biometrics* 1983;**39**:941–948.

60. Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: Elements for further discussion. *Statistics in Medicine* 1990;**9**:529–538.

61. Andersen PK, Væth M. Simple parametric and nonparametric models for excess and relative mortality. *Biometrics* 1989;**45**:523–535.

62. Baker RJ, Nelder JA. *The GLIM System Release 3 Manual*. Numerical Algorithms Group, 1978.

63. Francis B, Green M, Payne C, eds. *The GLIM System: Generalized Linear Interactive Modelling*. Oxford: Oxford University Press, 1993. Release 4.

64. McCullagh P, Nelder JA. *Generalized Linear Models*. London: Chapman and Hall, 2nd edn., 1989.

65. Dyba T, Hakulinen T. Confidence intervals for model-based cumulative relative survival rates using GLIM. *GLIM Newsletter* 1994;**23**:29–35.

66. SAS Institute Inc. *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary, NC: SAS Institute Inc., 1997.

67. Hédelin G. *RELSURV 2.0 a program for relative survival analysis*. Department of Epidemiology and Public Health, Faculty of Medicine, Louis Pasteur University, Strasbourg, France, 1997.

68. Monnet E, Boutron MC, Arveux P, Milan C, Faivre J. Different multiple regression models for estimating survival: Use in a population-based series of colorectal cancers. *Journal of Clinical Epidemiology* 1992;**45**:267–273.

69. Hess KR. Graphical methods for assessing vioations of the proportional hazards assumption in Cox regression. *Statistics in Medicine* 1995; **14**:1707–1723.

70. De Angelis R, Capocaccia R, Hakulinen T, Söderman B, Verdecchia A. Mixture models for cancer survival analysis: Application to population-based data with covariates. *Statistics in Medicine* 1999;**18**:441–454.

71. Day NE. The assessment of lead time and length bias in the evaluation of screening programmes. *Maturitas* 1985;**7**:51–58.

72. Seddon DJ, Williams EM. Data quality in population-based cancer registration: An assessment of the Merseyside and Cheshire cancer registry. *British Journal of Cancer* 1997;**76**:667–674.

73. Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL, eds. *Comparability and Quality Control in Cancer Registration*. IARC Technical Report No. 19. Lyon: International Agency for Research on Cancer, 1994.

74. Lee PN. Comparison of autopsy, clinical and death certificate diagnosis with particular reference to lung cancer. A review of the published data. *APMIS* 1994;**102 (Suppl. 45)**:1–42.

75. Hoel DG, Ron E, Carter R, Mabuchi K. Influence of death certificate errors on cancer mortality trends. *Journal of the National Cancer Institute* 1993;**85**:1063–1068.

76. Percy C, Ries LG, Van Holten VD. The accuracy of liver cancer as the underlying cause of death on death certificates. *Public Health Reports* 1990;**105**:361–367.

77. Percy CL, Miller BA, Gloeckler Ries LA. Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality (with discussion). *Annals of the New York Academy of Sciences* 1990;**609**:87–99.

78. Samphier ML, Robertson C, Bloor MJ. A possible artefactual component in specific cause mortality gradients: Social class variations in the clinical accuracy of death certificates. *Journal of Epidemiology and Community Health* 1988;**42**:138–143.

79. Carstensen B. Regression models for interval censored survival data: Application to HIV infection in Danish homosexual men. *Statistics in Medicine* 1996;**15**:2177–2189.

80. Farrington CP. Interval censored survival data: A generalized linear modelling approach. *Statistics in Medicine* 1996;**15**:283–292.

81. 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. *New England Journal of Medicine* 1985;**312**:1604–1608.

82. Farrow DC, Hunt WC, Samet JM. Biased comparisons of lung cancer survival across geographic areas: Effects of stage bias. *Epidemiology* 1995; **6**:558–560.

83. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *British Journal of Cancer* 1995; **72**:511–518.

84. Gamel JW, Weller EA, Wesley MN, Feuer EJ. Parametric cure models of relative and cause-specific survival for grouped survival times. *Computer Methods and Programs in Biomedicine* 2000;**61**:99–110.

85. Teppo L, Dickman PW, Hakulinen T, Luostarinen T, Pukkala E, Sankila R, Söderman B. Cancer patient survival - patterns, comparisons, and trends: A population-based cancer registry study in Finland 1955-1994. *Acta Oncologica* 1999;**38**:283–294.

86. Engeland A, Haldorsen T, Dickman PW, Hakulinen T, Möller TR, Storm HH, Tulinius H. Relative survival of cancer patients. A comparison between Denmark and the other Nordic countries. *Acta Oncologica* 1998; **37**:49–59.

87. Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, eds. *Cancer Registration: Principles and Methods*. IARC Scientific Publications No. 95. Lyon: International Agency for Research on Cancer, 1991.