Statistics in Question

SHEILA M GORE

ASSESSING METHODS—SURVIVAL

The organisation and graphical representation of survival data are emphasised in this article. The comparison of lifetables using the logrank test¹ is recommended because it is more efficient than the comparison of survival rates at one point in time. Moreover, comparison of *n*-year survival rates can be misleading if investigators decide to make the comparison at a given time just because the observed difference is greatest then. Bad practice such as this does not escape notice if lifetables are necessarily reported.

Looking at lifetables readily suggests which factors are important for prognosis; plotting the rates of mortality year by year in addition to lifetables gives insight to the disease process and is the first step in assessing statistical models for describing survival. Diversity of times to peak hazard and the unimportance in later follow-up of prognostic factors that were relevant at the time of diagnosis may describe the pattern of mortality when there are long-term survivors, as in breast cancer.² Plotting the lifetable on a logarithmic scale is one method of assessing graphically whether survival is exponential. Other methods may be easier to interpret, however.

Organising survival data

(17) Summarise the survival data at the time of analysis for patients A to D who were entered serially in a clinical trial comparing treatments for advanced cancer of the pancreas.

—patient A entered the trial on 14 January 1980 and was

still alive at the date of analysis 30 June 1981, giving survival time of at least 533 days; the information that the patient died on 24 August 1981 is not available at the time of preparing the trial report

—patient B entered the trial on 31 January 1980 and died on 9 October 1980; exact survival time is 252 days

—patient C entered the trial on 27 February 1980 and was lost to follow-up on 1 April 1980, the latest date of contact with the patient; survival time for patient C is (right) censored at 34 days

—patient D was admitted to the trial on 16 August 1980 and was still alive at the date of analysis, though lost to follow-up later; being alive on 30 June 1981 means that the survival of this patient is (right) censored at 318 days, another way of saying that the patient's survival time is at least 318 days

—at the time of analysis survival is (right) censored for patients A, D, and C because patients A and D are still alive and patient C was lost to follow-up; survival time is exact for patient B, who has died

COMMENT

The following discussion is about time to death, but the event that is most important could as well be tumour regression, development of metastasis, rehabilitation after a stroke, reinfarction, discharge from hospital, or regaining birth weight for a preterm infant. Several of these events are less easily defined

than death, so that in practice the study may be more difficult but in principle the same type of information is being recorded—namely, the time from entering the study to the occurrence of the particular event.

In most clinical trials patients enter serially—in the order in which they are referred. The first step in organising survival data, after fixing the date for analysis, is to update the follow-up on all patients so that each case history is summarised by only one of the following three descriptions. Description (1): patient died before analysis on — (give the date, and cause of death if inferences about competing risks are to be made). Description (2): patient is “known” to be alive at the time of analysis (presumption is not good enough, because of the possibility of bias). Description (3): latest date before analysis on which the patient was known to be alive is — (give date and reason for information being incomplete, such as the patient is abroad or no reply has been received to letters requesting follow-up information or the case notes have been mislaid, etc). The next step is to check that the sequence of dates is correct for a given patient and also that trial numbers and entry dates correspond. Any inversion—such as patient 12 was entered on 2 May 1981 while patient 13 was entered earlier on 16 April 1981—suggests an error in transcribing dates or in the trial procedures. Having checked the orderliness of the data, the third step is to convert dates into survival time in days. Survival time is exact for patients who have died and is right-censored for patients who satisfy descriptions (2) and (3), these patients having survived for a time of *at least so many days*. The information is now in a form that makes it straightforward to estimate the lifetable for different treatment groups as they were randomised and to compare the duration of survival, using, for example, the logrank test. An excellent guide to the comparison of lifetables has been published by Peto *et al.*¹

Design and analysis of randomised clinical trials requiring prolonged observation of each patient:

Introduction and design *Br J Cancer* 1976; 34: 585-612

Analysis and example *Br J Cancer* 1977; 35: 1-39

R Peto *et al*

Censored survival times are dealt with easily when computing lifetables (experimental survival curves) for different treatment groups. We assume, however, that the reasons for censoring are independent of treatment. If this is not so the comparison of lifetables could be biased. In particular, ensure that the follow-up of patients who have been withdrawn from treatment—because of toxicity, for example—is as intense as for patients who continue on treatment. The former group may be disadvantaged also in terms of survival, so that if toxic reactions occur more often in one treatment group failure to follow-up such patients shows that group in a relatively better light than is warranted. Notice also that the detection of a local recurrence, say, or metastasis tends to be earlier when the interval between follow-up visits is short, so that routine appointments should be arranged similarly for all treatment groups. The reasons for censoring are important even when estimating a single lifetable as the following example shows. In the Western General Hospital breast cancer series,² patients were dismissed from follow-up at the 20th anniversary of first treatment if there was no recurrent disease. Patients who continued to be followed up after the 20th anniversary had an unfavourable prognosis therefore, and the duration of their survival would be a poor reflection of the outcome for all 20-year survivors. Langlands *et al*² avoided potential bias by censoring survival at the 20th anniversary for *all* patients in the Western General Hospital breast cancer series.

Certified cause of death is often unreliable in cancer series

because deaths from other causes are underreported,³ a proportion of them being described as deaths from cancer. When analysing overall survival all deaths are accounted for irrespective of the cause. One justification for this type of analysis is that what is important to the patient is life. To assess curability, however, it is useful to have some measure²—such as the excess death rate or the ratio of observed to expected deaths—which allows for the normal mortality in the general population. In one sense, therefore, death from cancer and deaths from other causes are competing risks. Analyses of duration of survival corrected for age and the study of curability and of competing risks are problems that should be referred to a statistician.

Lifetables and annual rate of mortality

(18) *Figure 1 shows the lifetable for breast cancer patients referred to the Western General Hospital, Edinburgh in 1956. Estimate the probability that a patient referred for treatment of breast cancer survives for 10 years or more from date of diagnosis.*

—for any time t , the lifetable gives the estimated probability of surviving for at least time t

—always plot survival data as a lifetable (another name for the lifetable is Kaplan-Meier experimental survival curve)

—plot also the annual proportion dying, which is related to what statisticians call the hazard function

—28% of patients with breast cancer survive for 10 years or more from date of diagnosis

COMMENT

A simple identity gives the principle on which calculation of lifetables is based. The identity is that the probability of surviving for $(t+1)$ days from diagnosis equals the probability of surviving for t days, multiplied by the probability of surviving the next day. The last term, the probability of surviving day $(t+1)$, is got by counting (a) how many patients are at risk $(t+1)$ days after diagnosis—only patients whose recorded survival time is $(t+1)$ days or more are in the group at risk; (b) how many patients died on day $(t+1)$; and then computing (c) the probability of surviving that day as *one minus the proportion who died*—the proportion who died being the number of deaths on day $(t+1)$ divided by the number of patients in the group at risk. Notice that the lifetable is a step function that estimates

Lifetable for 347 patients with breast cancer referred to the Western General Hospital, Edinburgh.

the true survival curve in an unbiased way and that steps occur only at times of death. Readers are referred for a worked example and further explanation to the expository paper by Peto *et al.*¹ This gives sound advice about the design⁴ and analysis¹ of randomised clinical trials that require prolonged observation of each patient.

By plotting the lifetable authors report survival data more completely than by summarising 5-year and 10-year survival rates. The lifetable gives for any time t in the range of the observations the estimated probability of surviving for at least time t . In addition, comparison of lifetables gives a preliminary impression of which factors are important for prognosis. Figure 2 shows that international stage separates patients with

breast cancer into four groups with progressively poorer survival—69% and 57% of patients with stage 1 and 2 breast cancer survive for five years or more compared with 35% of patients with stage 3 and 7% of patients with stage 4.

Lifetables are not the only graphical representation of survival data and, because they show cumulative information, questions about how the rate of mortality changes year by year cannot be answered directly—although the answers are implicit in figure 2. I have in mind questions such as this: is the relative disadvantage of stage 3 breast cancer as great 15 years after diagnosis as it was at five years? The solution is to plot for each year of follow-up or other interval covering a sufficient number of deaths the estimated proportion of those alive at the beginning of a year who die within the year. To ensure proper handling of censored survival times, derive these proportions from the lifetable—for example, 41% of stage 3 patients survived for four years and 35% for five years, making the estimated proportion who died in the fifth year of follow-up equal to $6/41=0.15$, or 15% (see figure 3). Obvious from figure 3 but concealed in figure 2 is

that the rate of mortality decreases for stage 4 patients: more than 60% die within one year of diagnosis. For patients with international stage 1, 2, or 3 breast cancer a different pattern emerges. Hazard increases during the first one to four years after diagnosis and then slowly declines. Peak hazard is not only greater but also occurs earlier in stage 3 and stage 2 disease. By the 10th year after diagnosis the annual percentage dying is similar for survivors from all three stages—remembering that a standard error qualifies the estimators. In particular, survivors from stage 3 breast cancer seem to experience the same rate of mortality at 15 years after diagnosis as do survivors from stages 1 and 2.

Reference points for choosing a description of survival are exponential survival (constant annual proportion dying), the Weibull family (risk of dying is monotone—it increases (or decreases) steadily from time of diagnosis), and the proportional hazards model (irrespective of the elapsed time since diagnosis the ratio of the risks of dying in given prognostic groups remains constant, in particular the time when the force of mortality is greatest is the same for all prognostic groups). The problem of fitting a parametric distribution to survival needs to be referred to a statistician, but an initial assessment can be made from figure 3. The annual proportion dying is neither constant nor monotone and so the exponential model and the Weibull are not in the running. Diversity of times to peak hazard and the relative unimportance in later follow-up of description of the tumour (international stage) that was relevant initially are a denial of the proportional hazards model. These features suggest that a log-logistic distribution might fit the data. Careful statistical analysis is called for to substantiate these impressions.

When analysing long-term follow-up of patients investigators should plot mortality year by year in addition to the cumulative proportion of patients who survive. They should be on the look out for prognostic factors which, although relevant at diagnosis, are less important when it comes to describing later follow-up. Variable time to peak hazard indicates accelerated failure in some groups compared to others.

Look out for prognostic factors which, although relevant at diagnosis, have less importance when it comes to describing later follow-up

(19) If the graph of $\log p_t$ (where p_t is the estimated probability of surviving for t years or more) plotted against time is a straight line then the survival distribution is exponential. What do you infer from the figure below about the survival of patients with breast cancer?

—survival is not exponential

—rate of dying for breast cancer patients as a whole is rapid during the first four to five years after diagnosis; thereafter the annual proportion dying is less, but moderately constant

—the hazard pattern for important subgroups of patients (according to international stage, for example) is different from the pattern for the series as a whole

COMMENT

The figure shows the lifetable from figure 1 in question 18 plotted on a logarithmic scale. I have mentioned already that the exponential distribution is one reference standard for survival data—a particularly simple one, the time-specific risk of dying

Distribution of survival in 347 patients with breast cancer referred to the Western General Hospital, Edinburgh.

or hazard being constant irrespective of the elapsed time since diagnosis. This means that the rate of mortality neither intensifies nor mitigates at any time during follow-up. Fairly simple statistical theory tells us that if survival is exponential, then the graph of $\log p_t$ against time looks like a straight line. For any other survival distribution a similar theoretical exercise indicates the appropriate transformation of p_t , which induces a straight-line relationship and so gives a visual test of goodness of fit. The exponential example is most familiar, however. The slope or gradient is identified with the rate of mortality, which is most intense when the gradient is steepest.

The difficulty of translating from gradient to hazard is avoided in figure 3 in question 18 by plotting annual mortality directly. The logarithmic plot is the more commonly reported, however, and so I include this example dealing with survival in the 1956 cohort as a whole. I prefer the method described in question 18.

From the figure we see that the rate of dying is rapid and probably constant during the first four to five years after diagnosis—by placing a ruler along the line of the early part of the slope observe the more gentle gradient in the later part of the graph. This lies above the ruler and approximates to a different straight line. Our conclusions are (a) that survival is not exponential, and (b) that the rate of dying does not increase for the series as a whole. Recall, however, the initially increasing mortality associated with international stage 1, 2, and 3 breast cancer (see figure 3 under question 18). It is not unusual for the pattern of mortality in subgroups to be different from the overall impression. Investigation of prognostic factors should include looking at graphs of mortality plotted year by year.

Inefficient or misleading comparison

(20) *Comment on the comparison of survival at the fixed times shown in the figure below.*

—comparison of three-year survival rates misses the emergent difference between treatments A and B

—comparison of two-year survival rates favours treatment A, whereas by seven years after diagnosis the advantage is with patients randomised to treatment B; the logrank test would identify neither treatment as superior

COMMENT

Comparison of n -year survival rates is inefficient because it ignores the structure of the lifetable and is open to the criticism that if the time of analysis had been chosen differently the results might be slanted differently. Peto *et al*¹ list comparison

Undesirable comparisons of survival rates.

of survival rates at one point in time as the first of 13 bad methods of analysis. Avoid comparison of n -year survival rates.

The inefficiency of this method is illustrated in the figure on the left. By comparing three-year survival rates a difference between treatments that becomes apparent in time is missed. Comparison of lifetables over the entire follow-up period by the logrank test makes efficient use of the data and is likely to identify the superiority of treatment B.

In the second example, on the right of the figure, the treatment that is singled out as advantageous is different at two years and seven years. Treatment B is associated with higher initial mortality. In this case the logrank test is likely to report no significant difference between treatments. Only by plotting lifetables will the investigator discover the cross-over of the survival curves, an unusual contingency in practice—one that the logrank test does not deal with happily, nor does any other simple test statistic.

A third example, not illustrated, is when an initial treatment advantage (in the first year, for instance) is not sustained thereafter. This means that the rates of mortality in the two treatment groups are likely to cross-over, being first higher in one treatment group and then in the other before approaching a common level. Significant difference in disease-free interval but not in overall survival is the guise in which this problem commonly presents.⁵ Statistical advice should be sought.

Comparison of n -year survival rates is at best inefficient, at worst misleading. Unless lifetables are plotted the two extremes cannot be identified separately. Organising survival data to compute lifetables and calculation of the logrank statistic are so much the same exercise that there is little justification for presenting an inferior analysis—and lifetables should of course be reported always.

References

- 1 Peto R, Pike MC, Armitage P, *et al*. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II Analysis and examples. *Br J Cancer* 1977;**35**:1-39.
- 2 Langlands AO, Pocock SJ, Kerr GR, Gore SM. Long-term survival of patients with breast cancer: a study of the curability of the disease. *Br Med J* 1979;iii:1247-51.
- 3 Cutler SJ, Axtell LM, Schottenfeld D. Adjustment of long-term survival rates for deaths due to intercurrent disease. *J Chron Dis* 1969;**22**:485-91.
- 4 Peto R, Pike MC, Armitage P, *et al*. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I Introduction and design. *Br J Cancer* 1976;**34**:585-612.
- 5 Henk JM, Kunkler PB, Smith CW. Radiotherapy and hyperbaric oxygen in head and neck cancer. Final report of first controlled clinical trial. *Lancet* 1977;ii:101-3.

Sheila M Gore, MA, is a statistician in the MRC Biostatistics Unit, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH.

No reprints will be available from the author.