7
Competing risks and selection ★

7.1 Censoring in follow-up studies

Up to this point we have lumped all the different reasons for censoring together. In this chapter we look at this practice more carefully and make a distinction between censoring due to practical difficulties in maintaining follow-up (such as migration, refusal to participate further and so on), and censoring due to competing causes of failure.

The first class of events causes removal of a subject from observation, but after censoring the subject is still at risk of failure – a subject does not cease to run the risk of a myocardial infarction simply because he or she has ceased to participate in a follow-up study. Such observations are censored in the sense that this later experience is removed from our view. The second class of censoring events also causes removal of a subject from observation, but this time the subject is no longer at risk from the failure of interest. This is obviously true when a subject dies from a competing cause, but onset of a non-fatal competing disease can also remove a subject from the risk under study. For example, in a study of myocardial infarction in previously healthy subjects, a subject who suffers the onset of lung cancer would be considered as no longer at risk — although patients with lung cancer suffer myocardial infarctions quite frequently, the aetiology is so different as to be regarded as a different type of event.

7.2 Competing causes

The termination of follow-up by a competing cause is not due to imperfection of any one study, but is intrinsic to all imaginable studies. The binary model which underlies the measurement of disease frequency by rates and risks assumes only one type of failure. To allow for more than one type, the model must be extended. Fig. 7.1 illustrates a model with two causes of failure over a single study period of fixed duration. There are now three possible outcomes, labelled F1 and F2 for the two types of failure and S for survival. The probabilities of F1 and F2 are referred to as π_1 and π_2, so the probability of survival is $1 - \pi_1 - \pi_2$. In incidence studies, π_1 and π_2 represent *cause-specific* failure probabilities or risks.

It is easy to use likelihood to estimate the parameters π_1 and π_2. If N

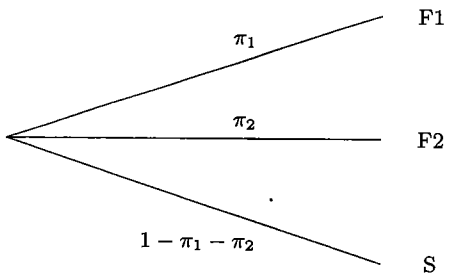

Fig. 7.1. Two causes of failure

subjects are studied and we observe D_1 failures of the first type and D_2 failures of the second type, the likelihood is

$$(\pi_1)^{D_1}(\pi_2)^{D_2}(1-\pi_1-\pi_2)^{N-D_1-D_2},$$

and the log likelihood is

$$D_1\log(\pi_1)+D_2\log(\pi_2)+(N-D_1-D_2)\log(1-\pi_1-\pi_2).$$

This takes its maximum value when $\pi_1 = D_1/N$ and $\pi_2 = D_2/N$ so that the most likely values correspond with the intuitive measures — the *proportions* of subjects failing due to each cause.

Exercise 7.1. In a 5-year follow-up study of 1000 subjects, 27 suffered myocardial infarctions during the study period while 8 suffered strokes. (If any subject suffered both events, only the first was counted.) Estimate the cause-specific risks for these conditions. If myocardial infarctions and strokes are grouped together as 'cardiovascular events', what is the estimated risk of a cardiovascular event?

Fig. 7.2 illustrates the extension of this model to describe observation of a subject through several consecutive bands. Superscripts denote band and subscripts continue to indicate the type of failure. As in the case of a single cause, the π parameters are defined as *conditional probabilities*. For example, π_1^3 represents the probability of failure F1 during the third band, *conditional upon survival* through all preceding bands. The log likelihood behaves as if the time bands form separate studies involving different groups of subjects, so for each band the cause-specific failure probabilities are estimated by the proportion of those subjects at risk during the band, failing from the specified cause.

Exercise 7.2. The conditional probabilities of F1 and F2 remain constant at 0.1 and 0.2 respectively over three bands. List the 7 possible outcomes and calculate their probabilities.

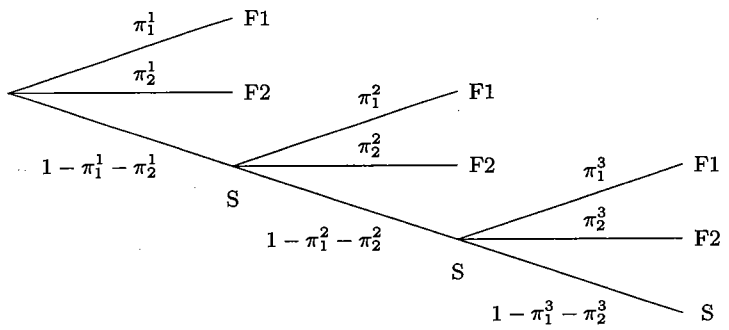

Fig. 7.2. Consecutive time bands

Table 7.1. Log likelihood contributions for a subject during one click

Outcome	Log likelihood
F1	$\log\lambda_1+\log h$
F2	$\log\lambda_2+\log h$
S	$-(\lambda_1+\lambda_2)h$

7.3 Cause-specific rates

The same argument can be extended to rates by dividing the time scale into clicks. Fig. 7.1 now represents the possible outcomes for one subject during a single click. The conditional failure probabilities are

$$\pi_1=\lambda_1 h, \qquad \pi_2=\lambda_2 h,$$

where h is the duration of a click and λ_1 and λ_2 are cause-specific *rates* — conditional probabilities per unit time. Because the probabilities of failure are very small, we can make the approximation

$$\log(1-\pi_1-\pi_2)\approx -\pi_1-\pi_2=-(\lambda_1+\lambda_2)h,$$

and the contributions to the log likelihood of a single subject during a single click are then those shown in Table 7.1. The total log likelihood is obtained by summing such terms over subjects and over clicks. There are D_1 clicks which result in failure of type F1 and these contribute a total of

$$D_1\log(\lambda_1)+D_1\log(h)$$

to the log likelihood. Since the second term does not depend upon parameters it can be ignored. Similarly the D_2 failures of type F2 contribute $D_2 \log(\lambda_2)$. Because every subject, regardless of eventual outcome, survives all the clicks save the last, the sum of all of these log likelihood contributions over both subjects and clicks is

$$\sum -(\lambda_1 + \lambda_2)h = -(\lambda_1 + \lambda_2)Y,$$

where Y is the total person-time of observation of the cohort. The grand total of all these contributions to the log likelihood is

$$D_1 \log(\lambda_1) + D_2 \log(\lambda_2) - (\lambda_1 + \lambda_2)Y.$$

A minor rearrangement of this expression leads to

$$D_1 \log(\lambda_1) - \lambda_1 Y \quad + \quad D_2 \log(\lambda_2) - \lambda_2 Y$$

so that the log likelihood is the sum of two parts, both Poisson in form, the first referring to F1 and the second to F2. The fact that the log likelihood falls into two distinct parts, one for each cause, justifies the standard practice of analyzing each cause separately, allowing for competing causes only in that they curtail further observation. The argument is easily generalized to allow for more than two causes.

7.4 Interpreting cause-specific rates

There has been some controversy as to whether the practice of estimating cause-specific rates in this way requires us to assume *independence* of causes – an assumption which might often not be justified. In fact, the split of the log likelihood into a sum of separate parts, one for each cause-specific rate, does not arise as a result of any assumption of independence of causes, but out of the way cause-specific rate parameters are defined. The rate for cause 1 is defined as the probability per unit time of failure due to cause 1, conditional upon the subject having previously survived *all* causes of failure. This quantity is not truly specific to one cause. Influences which directly influence one cause can, because of this, have an indirect affect on rates for another cause. The term *cause-specific* is misleading. For example, it is likely that myocardial infarction and stroke compete for the same high risk subgroup of the population: those with advanced atherosclerosis. A preventive measure which reduced the incidence rate of myocardial infarction without reducing the prevalence of atherosclerosis would result in an *increase* in the rate of stroke, since more of the atherosclerotic group would survive to be at risk from stroke.

It is a common practice to apply the formula

$$\log(\text{Cumulative survival probability}) = -\text{Cumulative rate}.$$

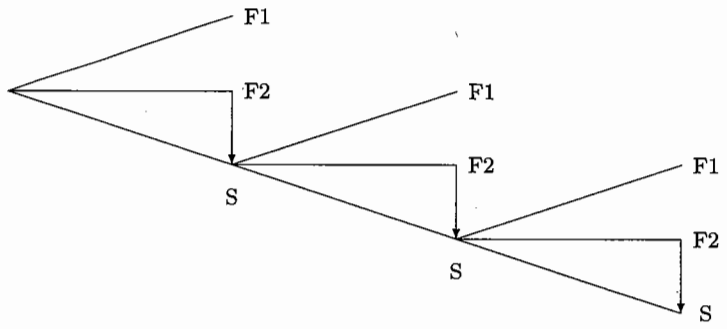

Fig. 7.3. Elimination of cause F2

to the cumulative *cause-specific* rate to calculate a *cause-specific* survival probability, interpreted as the probability of survival which would be observed if all other causes of failure were eliminated. However, this interpretation *does* depend on the assumption that the different causes of failure are independent. This is illustrated in Fig. 7.3. If the causes are independent, subjects who would have failed failed due to F2 have exactly the same conditional probabilities of failure due to F1 as those who would not. Under these circumstances, elimination of cause F2 will have no effect on the subsequent rate for F1, and the exponential function of minus the cumulative cause-specific rate for F1 can be interpretated as a survival probability when cause F2 is eliminated. More generally we might expect elimination of other causes to have an effect on the rate for the remaining one and the cumulative cause-specific rate will then have no such interpretation. Since the independence of different causes is usually untestable, it is best to avoid such interpretations and to leave estimates of cumulative cause-specific failure rates, calculated by the modified life table or Aalen–Nelson method, without converting them to cumulative probabilities of survival. Conversely, if the actuarial life table and Kaplan–Meier methods of Chapter 4 are applied to cause-specific failure probabilities, the resulting 'survival probability' should be transformed to a cumulative rate by taking minus its logarithm.

7.5 Selection bias

We now turn to the other reasons for censoring in follow-up studies. The statistical theory is exactly the same as for competing causes – we simply relabel the two causes as failure and loss to follow-up (Fig. 7.4). However, the question of dependence between failure and censoring takes on a new significance, because censoring arises as a result of the imperfection of

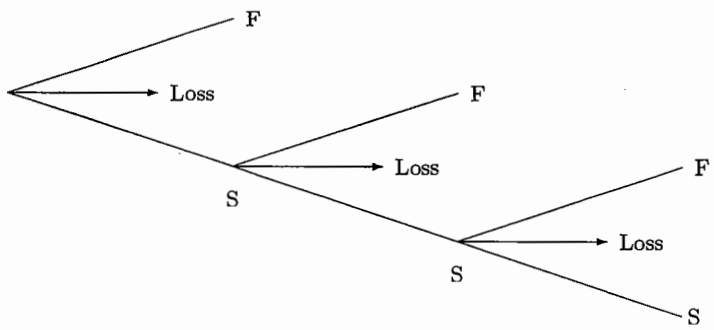

Fig. 7.4. Loss to follow-up

real studies, rather than unavoidable biological realities. We would like to estimate what would have happened in an ideal study in which no censoring occurred, but in order to do this we need to assume that censoring and failure are independent. More precisely, this means that that those lost to follow-up due to censoring must have the same probabilities of failure in later bands as those remaining under observation. If this is the case, censoring is said to be *non-informative*. If not, the study results will be subject to *selection bias*.

A well known example of selection bias due to censoring arises in clinical trials when patients become so ill that their clinicians are unable ethically to maintain them on a randomized double blind protocol. The randomization code is then broken and the clinician is free to modify the treatment as necessary. If observation of such patients is regarded as censored at this point, the analysis is seriously biased, because these patients have a worse prognosis than those remaining in the trial. It will almost always be preferable to continue the follow-up of these patients and to analyze the data according to the initial treatment assigned. This is known as analysis by *intention to treat*.

Similar considerations apply when there is *late entry* to follow-up studies. Ideally, subjects should be recruited at the starting point for the failure process under study. This is usually the case in clinical epidemiology, where patients are recruited into the study at diagnosis, the natural starting point for a prognostic study. In many epidemiological studies, however, subjects are recruited some time after the natural starting point (see Fig. 7.5) This is known as *late entry*; it can introduce further selection bias if the new subjects have different subsequent probabilities of failure from the survivors they join. For example, clinical follow-up studies are frequently carried out in cohorts initially recruited from patients under treatment in a group

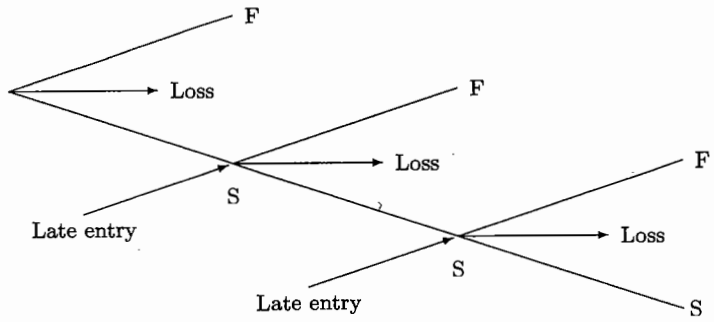

Fig. 7.5. Selection due to loss and late entry

of participating hospitals when the study starts. These cohorts are then extended by addition of new patients as they are diagnosed. When carrying out analyses of survival times from diagnosis, the initial members of such a cohort are late entries, because their diagnosis preceded their entry into the study. This introduces possibilities for selection bias, because the initial cohort could include patients diagnosed elsewhere, but would omit patients diagnosed in the participating hospitals and referred elsewhere for treatment.

In epidemiological studies of the causes of disease, late entry is almost universal and we must be careful that it does not introduce bias. An example of bias arising this way is the *healthy worker* effect, so called because of the widespread empirical finding that occupationally recruited cohorts have lower mortality than general population rates would suggest. This arises partly because of selective recruitment into occupations but mainly because early retirement and job changes in response to ill health act to prevent entry of 'unhealthy' workers into the cohort. The ideal study would recruit subjects on entry to the occupation, but in practice the subjects actually recruited are those in employment on a particular date. Follow-up then starts on that date. Factors such as early retirement, and job changes in response to ill health, can operate in the period between joining the occupation and recruitment to prevent entry of unhealthy workers into the cohort.

The phenomenon of late entry is closely related to the distinction sometimes drawn between 'closed' cohorts (in which only failures and censoring can occur) and 'dynamic' cohorts, which can be refreshed by new entrants. This distinction implies that being closed or dynamic is a property of the cohort, but this is incorrect. The occurrence or non-occurrence of late entry is not a property of the cohort, but depends on the time scale chosen for

analysis. If survival is analyzed by time in study there are no late entries, but in an analysis of the same study by age, or by time since entering an occupation, there will be late entries.

Solutions to the exercises

7.1 The estimated 5-year risk of myocardial infarction is 27/1000 while that for stroke is 8/1000. The risk of a cardiovascular event is 35/1000.

7.2 The outcomes and their probabilities are listed below.

Outcome	Probability
Band 1	
F1	0.1
F2	0.2
Band 2	
F1	$0.7 \times 0.1 = 0.07$
F2	$0.7 \times 0.2 = 0.14$
Band 3	
F1	$0.7 \times 0.7 \times 0.1 = 0.049$
F2	$0.7 \times 0.7 \times 0.2 = 0.098$
S	$0.7 \times 0.7 \times 0.7 = 0.343$

7.1 Censoring in follow-up studies

JH would use a different word for the so-called 'censoring' that removes a subject from pool of candidates for the transition of interest.

7.2 Competing causes

A good way to visualize how things play out over time is to use a table with rows for time intervals, and 3 columns (one can also show columns of cumulative totals for the transitions):

interval	no. candidates	transitions, type 1	transitions, type 2
interval	no. candidates	transitions, type 1	transitions, type 2
...	...	...	...
interval	no. candidates	transitions, type 1	transitions, type 2

The numbers of transitions of type i, when the transition rates are λ_1 and λ_2, in an interval of length w,[1] are:

$$\text{no. of candidates} \times (1 - \exp\{-\lambda_i \times w\})$$

If λ_i is a smooth function of age or time, then we replace the $\lambda_i \times w$ product by the integral $\int \lambda_i(t)dt$ over the interval.

The number of candidates for the next interval is the number for the previous one minus the sum of the numbers of transitions in that previous interval.

The number of candidates at the beginning of an interval can also calculated as

$$\text{initial no. of candidates} \times \exp\{-\Sigma(\lambda_1 + \lambda_2) \times w\}$$

where the summation (or integral, if the If λ's are smooth functions of age or time) is over the time span already elapsed.

[1] C&H use h for the duration of a 'click.'

Supplementary Exercise 7.1. The BIOS601 website has, under the tab 'Competing Risks', (and inside the R code) data on age-specific breast cancer incidence and mortality, as well as all cause mortality.

1. Use these (and if you wish, the `R` code provided) to calculate the 'lifetime'[2] risk of being diagnosed with breast cancer.

2. Compare your answers with those given in **the report (now on the 601 site) 'Google Canadian Cancer Statistics 2013' (which JH obtained by 'Google-ing' Canadian Cancer Statistics 2013). The relevant Table 1.1 is at the end of Chapter 1.**

   Incidentally, do you agree with the wording 'Lifetime probability (%) of developing cancer in next 10 years by age group' further to the right in the Table? If not, suggest a better wording.

3. Calculate the lifetime risk in the absence of competing causes of removal from the candidate pool. **You can do this by repeating the calculation in Q1, but setting the mortality function to zero.**

4. Calculate how many more years of life, on average, women could expect to live if all deaths from breast cancer mortality could be averted. Either carry the calculations out to age 105, using a sensible extrapolation of the breast cancer mortality curve, **or stop where the breast cancer mortality curve ends**.

   Note added 2014.11.06: *JH realizes that he gave you misleading advice today – he had not read part 4 of the question carefully, and thought the question referred to <u>just</u> those women who died of breast cancer. In fact, as worded, it refers to an average of <u>all</u> women, under 2 scenarios.*

   The question asks what is the average life expectancy at birth if we use (i) the all-cause death rate function generated in the R code and (ii) if we subtract from this rate function the breast cancer death rate function, also generated in the code.

   For (ii) one can subtract the breast cancer death rates from the overall rates.

   And to get the life expectancy at birth, one can indeed – as some of you had surmised – use the area under the survival curve.

   One caveat is that by removing the breast cancer deaths, one is not removing a cause of some other disease(s) that kill(s) women. C & H write about this issue when considering stroke and MI.

[2] Since the available breast cancer rates only go as far as age 92 or so, and are negligible before age 20, consider the lifespan from 22 to 92.

2014.11.06: *Here is what JH was thinking about when he spoke with some of you after class: he thought the question was 'For those who die of breast cancer, how many years of life do they lose, on average?'*

For this, one could use the breast cancer death rate function and the other cause death rate function to create a 3-ply table, with the columns (a) alive at this age, (b) died at this age of breast cancer, (c) died at this age of other causes. The second column would give you the age distribution of the ages at which women die of breast cancer. And from this, one could indeed then use these as weights to get a weighted average of the remaining life expectancies at these ages. Some of you saw me find a Canadian life table with such life expectancies in it.

This distribution is somewhat synthetic or artificial. Another more real way is to use data from an actual population, such as those provided by the WHO – see the beginning of the R code, which extracts the distribution of ages at breast cancer death from recent WHO cancer statistics from Canada. These can then be used as as weights to get a weighted average of the remaining life expectancies at these ages.

Supplementary Exercise 7.2. The website also has, under the same tab, data on the 767 men reported on in Albertsen et al. 2005 article on prostate cancer mortality in men who had their prostate cancer managed conservatively. See also the statistical methods in the 1998 and 2005 articles.

1. Use Poisson regression to fit a model for the rate (incidence density, hazard) function governing all-Other-cause mortality – ie fit a smooth $\lambda_O(age)$ function Hint: you will probably find that Gompertz' law (log[rate] is linear in age) gives a reasonably acceptable fit. The one other variable, besides age, that would matter is how many concurrent diseases (e.g, diabetes, heart disease, etc..) the man has – we tried to capture this information by using the Charlson score, but you can ignore it in this exercise. *You can also ignore (i.e., pool the experience over) Gleason categories.*

   This is the 'blue' function in the lower right panel. Inspect the R code and explain how it was fitted. Think of a rate function as an 'intensity' function, i.e. how the intensity of the blue dots (deaths from other causes) in the population-time space increases with age. This way of thinking about it should drive home the point that these rates are NOT proportions: a rate of 0.03 deaths per man-year does not mean that there is a 3% probability that death will occur within the year. It means that it is $1 - \exp[0.03 deaths/year \times 1 year] = 1 \exp[-(\mu =)\ 0.03\ deaths]$. *This is close to a 0.03 probability in this instance, but it would not be if we were dealing with a death rate of 2 deaths per person year if we are at age 110! With this high rate, an average of 2 110-year olds are needed to create a continuous 1-person chain from age 110 to 111.*
   Another example, related to a fast-killing disease, is cancer of the pancreas. Suppose one were running a 5-bed hospice for such patients, that – because of the few beds – was always full, and that the average longevity from admission to death was 0.5 years. Then in a calendar year, one would accumulate 10 deaths in 5 bed-years, so the death rate would be 2 deaths/bed-year. Now the chance of a patient living a whole year is $\exp[-2\ deaths/year \times 1\ year = 2 deaths] = 14\%$ *and the chance of dying within the year is therefore 86%.*

2. Use Poisson regression to fit **separate** (**Gleason-category-specific**) models for the rate (incidence density, hazard) function governing prostate Cancer mortality – i.e., for each category, fit a separate smooth-in-time $\lambda_C(t)$ function where t is the number of years elapsed since the date of diagnosis. Assume the function is independent of age.

   This is the 'red' function in each of the 5 panels. Inspect the R code and explain how it was fitted. The rate function shows how the intensity of the red dots (deaths) in the population-time space varies with follow-up time. The code assumes a log-linear model, but of course, if there were more data, more complex models could be used. In the JAMA paper, for parsimony, some parameters were 'shared' across the Gleason categories.

3. Compute '3-ply' curves, of the types contained in the articles, for a few selected combinations of age-at diagnosis, and Gleason score categories – they won't match exactly those in the JAMA article, where a slightly more extensive model was used for the prostate cancer mortality rate function. You can use integrals or sums.

   E.g.: men diagnosed at age 65, with a Gleason 7 cancer. To reach 85, they must avoid 2 (red and the blue) ongoing threats. Some textbooks use the image of people being shot at by red and blue guns, with different lethalities, as they walk down that street. The proportion reaching the end alive is $\exp[-\int_0^{20}[\lambda_{red}(u) + \lambda_{blue}(u)]\ \times du\ ]$.

   The proportion alive at any t is $S[t] = \exp[-\int_0^u [\lambda_{red}(u)+\lambda_{blue}(u)]\ \times du\ ]$. *This forms the upper boundary of the 'white' area in the JAMA figure.*

   The numbers of red and blue casualties in the interval $(t, t + dt)$ *are* $S(t)dt\ \times \lambda_{red}(t)$ *and* $S(t)dt\ \times \lambda_{blue}(t)$ *respectively. Cumulated over time, they form the darker and lighter bands in the JAMA figure.*

The concept of population-time as a 2-D surface

The vertical dimension is the population size, and the horizontal dimension is time.

The schematic on the right illustrates the 'population time' segments that form the denominators of incidence densities. Shown are when, and for how long 100 different motor vehicle drivers drove while using or not using cellular telephones, during a specific time-window on a particular morning. The raw data are depicted in two formats

(1) in detail, driver by driver, with the driving time shown as thin horizontal lines, and the time driving while using a cellular telephone in darker and thicker horizontal lines and

(2) collectively, i.e., de-personalized. The height of the upper curve indicates how many were driving at the indicated instant, and the lower curve how many of them were at that instant using the phone while driving. The area under the lower curve represents the total amount of driver-time 'on-the-phone', and that between the two curves the total driver-time 'off-the-phone.'

We could imagine dots in these 'population-time surfaces' representing traffic accidents in the 'on the phone' and 'off the phone' experience.

These lead naturally to an *intensity function*.