19

MODELLING SURVIVAL DATA

OBJECTIVES

After reading this chapter, you should be able to:

1. Distinguish between non-parametric, semi-parametric and parametric analyses of survival time data.

2. Carry out non-parametric analyses using either actuarial or Kaplan-Meier lifetables and compare the survival experiences of groups of animals using a variety of statistical tests.

3. Generate survivor and cumulative hazard function graphs to display survival data.

4. Understand the relationships among survivor functions $S(t)$, failure functions $F(t)$, probability density functions $f(t)$, hazard functions $h(t)$ and cumulative hazard functions $H(t)$.

5. Carry out a semi-parametric analysis of survival data using a Cox proportional hazards model.
   (a) Evaluate the model to:
      i. assess the validity of the assumption of proportional hazards,
      ii. assess the validity of the assumption of independent censoring,
      iii. evaluate other aspects of the model such as its overall fit, the functional form of the predictors in the model, and check for outliers and influential points.
   (b) Incorporate time-varying effects into the model to evaluate or account for non-proportional hazards.

6. Carry out a parametric analysis of survival data based on an assumption that the survival times have an exponential, Weibull or log-normal distribution.

7. Incorporate frailty effects into a model to account for unmeasured covariates at the individual or group level.

8. Analyse multiple failure-type (recurrence) data.

9. Fit discrete time survival models when appropriate.

19.1 INTRODUCTION

In previous chapters, we have looked at statistical models for evaluating how much of an outcome occurred (linear regression), whether or not an event occurred (logistic regression), which category of event occurred (multinomial models) and the number of events that occurred (or the rate of event occurrence) (Poisson regression). However, we are often interested in how long it took for an event to occur (time-to-event data). These data are often referred to as 'survival' data because the outcome of interest is often the time until death (*eg* time to death in *Cryptococcus gattii* infected dogs and cats (Duncan *et al*, 2006)). However, the analytical approaches discussed in this chapter apply equally to any time-to-event data (*eg* interval from calving to conception in dairy cows (Meadows *et al*, 2006; Meadows *et al*, 2007), or time to recurrence of *Mycoplasma* infections in swine barns after an eradication program). As these examples suggest, the unit of analysis could be an animal or a group of animals (litter, pen, herd) although, in general, we will present the discussion in terms of animals. The occurrence of the event of interest is often referred to as a 'failure' even though in some cases the outcome is desirable (*eg* time to conception after calving in dairy cows). Some relatively recently published texts which cover the analysis of survival data include (Cleves *et al*, 2008; Collett, 2003; Hosmer & Lemeshow, 2008; Therneau & Grambsch, 2000).

There are specific issues that affect how we quantify and express time to occurrence of an event and how we evaluate the effects of factors (predictors) on that time. However, before discussing these issues, let's look at a simple hypothetical example (Example 19.1).

19.1.1 Features of survival data

Later in this chapter, a dataset derived from a clinical trial of prostaglandin treatment in dairy cows is used for many examples. The outcome of interest in that dataset is 'days to conception' which is the time from the 'end of the voluntary wait period' (*ie* the time at which the farmer will start breeding the cow if she is seen in heat) to conception. A distribution of those time intervals is shown in Fig. 19.1.

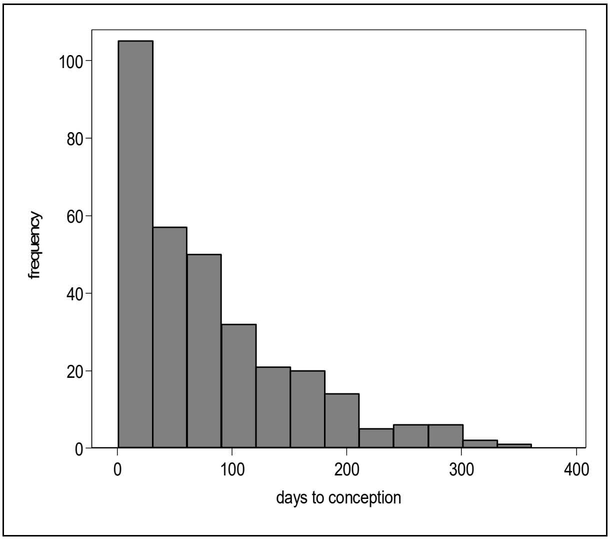


Fig. 19.1 Distribution of survival times in prostaglandin dataset

Three common features of survival data follow.

1. There is **strict left truncation** which means that there are no values <0.
2. Survival data often have a highly **right-skewed distribution** with many individuals 'failing' early and a small number having long times to 'failure'.
3. Survival data are often **censored** (*ie* the animal is lost to follow-up before the event of interest (failure) is observed (see Example 19.1).

Example 19.1 Hypothetical survival data
data = lympho_hyp

Fig. 19.2 shows the time from first diagnosis of lymphosarcoma to the recurrence of the cancer in 12 dogs. The study was carried out over a 5.5 year period with dogs entering the study as they were diagnosed and treated for the first occurrence. Once enrolled, not all dogs were followed for the rest of the study period because some died (from other diseases) or the owner moved away from the study location. For convenience, all dogs were assumed to have had the initial diagnosis and treatment at the start of a year and events (recurrence or loss to study) occurred at the mid-point of a year. In reality, this would not normally be the case.

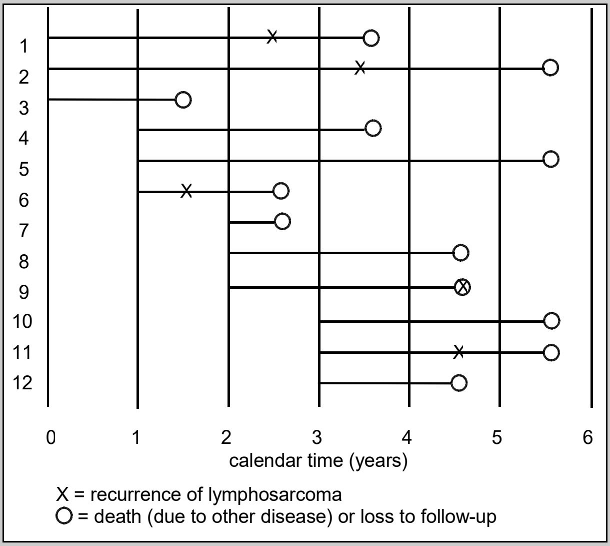


Fig. 19.2 Time from first diagnosis to recurrence of cancer

One way to simplify the graphic representation of these 12 dogs would be to express all times as being relative to the time of first diagnosis (Fig. 19.3).

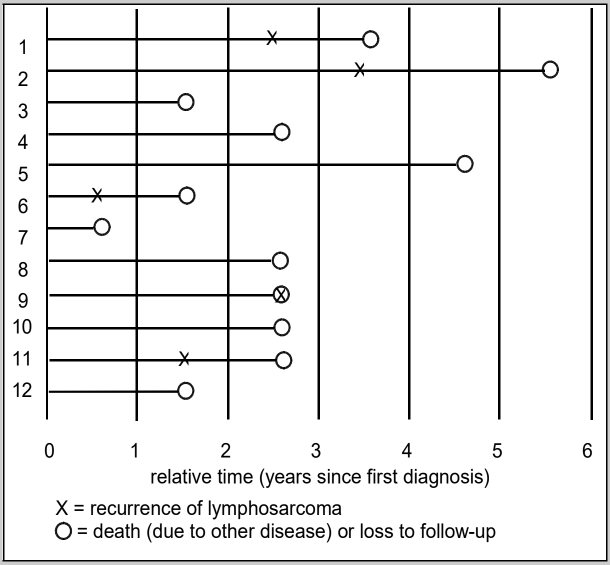


Fig. 19.3 Time to recurrence relative to time of first diagnosis

19.1.2 Quantifying survival time

How should the time to recurrence (*ie* time after initial diagnosis) of lymphosarcoma in dogs that have been treated for lymphosarcoma be quantified and expressed (Example 19.1)? For many dogs, we do not know what the time to recurrence was. All we know is that the disease did not occur in the time period for which the dogs were followed. These 'non-failures' are called censored observations and are a unique feature of time-to-event data.

Some possible ways of quantifying and expressing the time to recurrence follow (using data from Example 19.1).

1. **Mean time to recurrence** The mean time to recurrence can only be computed using data from the dogs in which recurrence has been observed. Consequently, we can only use data from 5 dogs (mean survival=2.1 years). The estimate will have a downward bias because recurrence in dogs which had a long time to recurrence are less likely to be observed. On the other hand, if the follow-up observation period is long, the mean suffers from the fact that it might be heavily influenced by a few animals with long survival times. Time-to-event data often have an asymmetrical distribution with a long right tail (*ie* right skew).

2. **Median time to recurrence** This can only be computed directly if at least 50% of the animals are observed to have the event of interest and if none of the censored observations were censored before the failure of the 50th percentile individual (*ie* if they were going to fail, they had a failure time at least as large as the median). It could not be computed for the data in Example 19.1. However, if it can be computed, the median is not influenced by a few animals having long times to recurrence in the same way that the mean is.

3. **Overall probability of recurrence** The proportion of dogs having a recurrence of the tumour could be computed, but it is not at all clear what constitutes a 'negative' dog (*ie* one which does not have a recurrence). Should the dog be required to have some minimum number of years of follow-up to be considered eligible to contribute to the denominator of the proportion?

4. **n-year survival risk** This expresses the number of dogs which have not had a recurrence by the nth year. For each year (*eg* first, second) it can be computed based on the dogs that were observed for that number of years. This approach is often used in human epidemiology to quantify survival of people diagnosed with various forms of cancer (*eg* 5-year survival of breast cancer patients). The 2-year 'survival' for dogs in Example 19.1 is 0.78 (2 recurrences among 9 dogs that had either 2 complete years of follow-up or a failure at <2 years).

5. **Incidence rate** The number of recurrences relative to the accumulated dog-years at risk would be one way to use all of the available data. In some cases, the average time to recurrence could be estimated from the incidence rate (see Section 4.5). However, this approach assumes that the incidence rate of recurrence remains constant throughout the follow-up period and this is often not the case with time-to-event data. The incidence rate in Example 19.1 is 0.19 cases per dog-year (5 cases in 26 dog-years of follow-up time—dogs no longer contribute to the pool of dog-years once they have experienced a recurrence).

The approaches outlined above identify 2 key problems to be considered when analysing time-to-event data. First, many observations are censored; that is, the individual is not followed for a length of time sufficient to observe the event of interest if it were to occur. Second, the distribution of survival times is often not symmetrical, and might not even be unimodal. For example, tumour recurrences might be common in the first year after first diagnosis and then relatively rare for several years before becoming more common again as the dog ages. These issues are also important when evaluating the effect of predictors on the time-to-event occurrence.

19.1.3 Censoring

Censoring is defined as the occurrence (or possible occurrence) of a failure when the animal is not under observation. Censoring can arise in a variety of ways and these are summarised in Fig. 19.4.

Right censoring occurs when an animal is lost to a study, before the outcome of interest has occurred. This might arise because the study ends before the event occurs or because it is lost to follow-up during the study (*eg* the owner moves to another city). Right censoring is the most common form of censoring that needs to be dealt with in survival analyses.

Interval censoring might arise when an animal is only observed periodically throughout a study period. If examinations are conducted approximately every 6 months and at one examination (t_4 in Fig. 19.4) it is determined that the event had happened in the preceding 6 months, all that is known is that the event occurred sometime between t_3 and t_4. The precise time the event occurred is not known.

Left censoring is similar to interval censoring except that the 'interval' occurs at the start of the study (*ie* the event occurred in the animal before the animal was observed). Consequently, the animal is not put in the study. Left censoring usually arises if the onset of risk occurs before the start of the study. For example, if a study of calving to conception intervals started following cows at 45 days post-partum, a cow which conceived to a breeding at 42 days would be left censored. (**Note** If multiple failures are possible, the animal might be put in the study and the left censoring then becomes left truncation (see below)).

A related concept is that of **truncation**. While censoring relates to the possible occurrence of events during periods when the animal was not observed, truncation refers to periods of time in which nothing is known about the animal in terms of whether or not the event occurred or what the predictors were. These periods of time might be referred to as **gaps**. In cases where multiple events are possible (*eg* cases of mastitis), you have no knowledge of how many cases occurred during the gap. For outbreaks which can only occur once (*eg* death), all that is known is that the event did not occur during the gap (or the animal would not have come back into the study). Truncation can occur throughout a study (**interval truncation**) or at the start of a study (**left truncation**—also known as delayed entry). Right truncation is the same as right censoring.

As noted above, the most common problem is with right censoring and it will be the only type of censoring or truncation that we deal with in examples in this chapter. A more complete discussion of censoring and how the various forms are dealt with can be found in Chapter 4 of Cleves *et al*, (2008).

19.1.4 Evaluating the effect of factors on survival times

Because time-to-event data are continuous, it would be tempting to evaluate the effects of factors on the time to the occurrence of an event using linear regression models. In some cases,

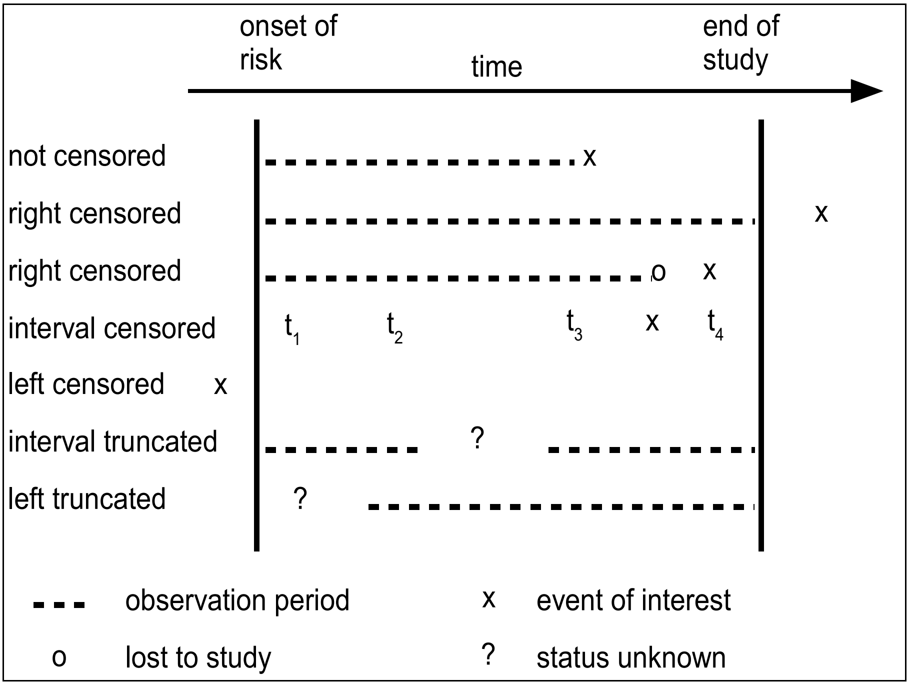


Fig. 19.4 Summary of censoring

this would be appropriate. However, as noted above, time-to-event distributions are often not symmetrical and might not even be unimodal. The assumption of normally distributed errors required for a linear regression model would often be violated in these cases. (In extreme cases, a linear model might predict negative survival times which are impossible). However, linear models have been successfully used to analyse time-to-event data. Calving-to-conception intervals in dairy cattle have been analysed using this approach (Dohoo *et al*, 2001). The data were either log or Box-Cox transformed to deal with the distribution of errors being skewed to the right.

Even if the distribution of the errors is (or can be made) approximately normal, the problem of censored observations remains. In the case of calving-to-conception interval data, because most cows are not culled until the end of the lactation, the follow-up period for most cows is adequate. However, many cows are bred unsuccessfully several times and then the producer stops trying. The data from these cows are lost to the analyses so the effects of factors which reduce conception might be underestimated.

19.1.5 General approaches to analysing survival data

There are 3 general approaches to analysing survival data:

1. non-parametric analyses
2. semi-parametric models
3. parametric models.

These are discussed in much more detail later, but the essential features of each approach are summarised here.

In a **non-parametric** analysis, we make no assumptions about the distribution of survival times, nor about the functional form of the relationship between a factor (predictor) and the survival time. Consequently, they are only appropriate for evaluating the effect of qualitative (categorical) predictors.

In a **semi-parametric** analysis, we make no assumption about the distribution of the survival time, but merely use the survival time to order the observations in terms of time of occurrence of the event. We then evaluate the probability of the event occurring at each of those time points as a function of the predictors of interest. Because the time variable is only used to order the observations, it makes no difference if there was a large time interval or a small time interval between successive events.

In a **parametric** analysis, we replace the distributional assumption that the errors are normally distributed (as required in a linear model) by a more appropriate distribution that reflects the pattern of survival times. Because we specify a distribution for the survival times, the length of the interval between events is relevant for the analysis. Consequently, if the assumed distribution is correct, a parametric model may be more efficient than a semi-parametric model (*ie* it makes better use of the available data).

19.2 NON-PARAMETRIC ANALYSES

As noted above, in a non-parametric analysis of survival data, we make no assumption about either the distribution of survival times or the functional form of the relationship between a predictor and survival. Hence, they can be used to compare survival experiences of groups of animals, but not to evaluate the effect of a continuous predictor on survival times. We will look at 3 non-parametric methods for analysing survival data:

- actuarial life tables
- Kaplan-Meier estimator of the survivor function
- Nelson-Aalen estimator of the cumulative hazard function.

In the following section, we introduce the concepts of survivor and hazard functions. These will be described more formally in Section 19.7.

19.3 ACTUARIAL LIFE TABLES

Life tables were originally developed to summarise long-term human-survival data by dividing the lifespan into short intervals in which the probability of dying was reasonably constant over the time interval. (It certainly is not constant over an entire lifespan.)

The requirements to create an actuarial life table are as follows.

- A clearly demarcated starting point to the period of risk (*eg* birth, calving, first diagnosis, first exposure *etc*)
- A well-defined study outcome (death, seroconversion, pregnancy diagnosis, calving)
- Only one episode or event per individual animal (not multiple remissions or relapses)
- Losses to follow-up should be independent of the study outcome (another way of saying this is that the animals lost from the study should have the same future experience as those that remain under observation)
- The risk of the outcome remains constant over calendar time (no secular (long-term) changes in risk). This does not imply that risk stays the same in an individual over time.

Secular changes in survival rates for cancers (*eg* due to better therapies), might for example, affect validity of studies of survivorship

- The risk of outcome must remain constant within the intervals used for constructing a life table. Intervals of any length could be calculated to meet this requirement. In fact, the intervals need not be of the same length.

19.3.1 Steps in constructing the actuarial life table

Table 19.1 shows the columns required to build an actuarial life table, based on the data from Example 19.1.

Table 19.1 Actuarial life table

j	t_{j-1}, t_j	l_j	w_j	r_j	d_j	q_j	p_j	S_j
1	0 < 1	12	1	11.5	1	0.087	0.913	0.913
2	1 < 2	10	2	9.0	1	0.110	0.890	0.812
3	2 < 3	7	3	5.5	2	0.364	0.636	0.516
4	3 < 4	2	0	2.0	1	0.500	0.500	0.258
5	4 < 5	1	1	0.5	0	0.000	1.000	0.258

where ...

j listing of time intervals (time intervals should be established *a priori*)

t_{j-1}, t_j time span covered in the interval

l_j subjects at risk of failure at the start of the time interval

$l_j = l_{j-1} - (w_{j-1} + d_{j-1})$

w_j subjects withdrawn during interval (censored observations)

These are animals who died of causes other than the condition under study or were otherwise lost to follow up during that interval. Animals who were still free of the outcome when the study ended are counted as withdrawals in the last interval r_j

average number of subjects at risk during the current time interval

$r_j = l_j - (w_j /2)$

This calculation is based on the assumption that the censored observations were withdrawn, on average, at the midpoint of the interval.

d_j outcomes (failures) during the interval

This is the number experiencing the outcome during the time interval (death, seroconversion, relapse *etc*).

q_j risk of event during interval

$q_j = d_j/r_j$

This is the probability that the subject will develop the study outcome during the given interval, conditional upon surviving without the outcome up to the beginning of the time interval.

p_j probability of surviving the interval

$p_j = 1 - q_j$

The conditional probability of surviving the time interval, given survival to the beginning of the interval.

S_j cumulative survival probability to the end of the interval

$S_j = (p_1)(p_2)(p_3)....(p_j)$

The probability of surviving without experiencing the event of interest from the start of follow-up through the end of the current interval in the life table.

The risk of an animal experiencing the event of interest during the interval (q_j) divided by the length of the interval is also known as the **hazard**. The cumulative survival probability (S_j) is also known as the **survivor function**. These 2 quantities are key elements of all survival analyses.

19.4 Kaplan-Meier estimate of survivor function

19.4.1 Overview and comparison to actuarial method

The Kaplan-Meier (K-M) (Kaplan & Meier, 1958) estimate of the survivor function is also known as the **product-limit estimate**. It has 2 important differences from the actuarial estimate described above.

1. The K-M method does not depend on discrete time intervals constructed by the investigator. Each row in the table (hence, each time interval) is defined by the time at which the next subject (or subjects, in the case of 2 events happening at the same time) experiences the event of interest.
2. Censored observations (losses to follow up *etc*) between 2 events are counted as animals at risk only up to the time of the earlier of the 2 events.

The K-M method has the advantage that it avoids the assumption that withdrawals occurred uniformly throughout the interval (*ie* the actuarial assumption) and that the risk is constant over the arbitrarily selected interval. (The only remaining assumption about withdrawals is that they have the same future experiences as those remaining under observation).

19.4.2 Construction of the K-M life table

An ordered list of the event times is constructed from the sample, with subjects ranked in ascending order of the time of the event of interest. Based on these, Table 19.2 can be filled out (using the data from Example 19.1)

Table 19.2 Kaplan-Meier life table

j	t_j	r_j	d_j	w_j	q_j	p_j	S_j
1	0.5	12	1	1	0.083	0.917	0.917
2	1.5	10	1	2	0.100	0.900	0.825
3	2.5	7	2	3	0.286	0.714	0.589
4	3.5	2	1	0	0.500	0.500	0.295
5	4.5	1	0	1	0.000	1.000	0.295

where:

j	listing of time points
t_j	time of event
r_j	subjects at risk of event at time t_j

$r_j = r_{j\text{-}1} - (d_{j\text{-}1} + w_{j\text{-}1})$

Includes all subjects known to be alive and in the study (not censored) at the time of the event at time *t*, plus the number experiencing the event at time *t*. When censored times are tied with event times, the event is usually assumed to have occurred first

d_j number of events at time t_j

w_j number of censored observations at time t_j

Censoring between time t_j and t_{j+1} is assumed to have happened at t_j so the animals will not be considered at risk at time t_{j+1}

q_j risk of event at time t_j

$q_j = d_j/r_j$

Also known as the instantaneous hazard, this is the individual probability of the event at time t_j, conditional upon survival to time t_j

p_j probability of survival at time t_j

$p_j = 1 - q_j$

S_j cumulative probability of surviving up to and including time t_j

$S_j = (p_1)(p_2)(p_3)....(p_j)$

Survivor functions are usually presented graphically as step functions of the cumulative survival over time. They start at one and monotonically descend (*ie* they never go up) as time proceeds. Fig. 19.5 shows a Kaplan-Meier survivor function (and its 95% confidence intervals) based on some published data on calf pneumonia (see Example 19.2 for actuarial life table and Kaplan-Meier estimates of the survivor function (Thysen, 1988)). Some issues related to the presentation of survival plots have been presented (Pocock *et al*, 2002), including a suggestion that plots of failure functions (see Section 19.7) might be more useful.

19.4.3 The Kaplan-Meier function and estimator

The Kaplan-Meier estimator plays an important role in many procedures used for the analysis of survival data. This section describes the estimator and resulting function in slightly more technical detail.

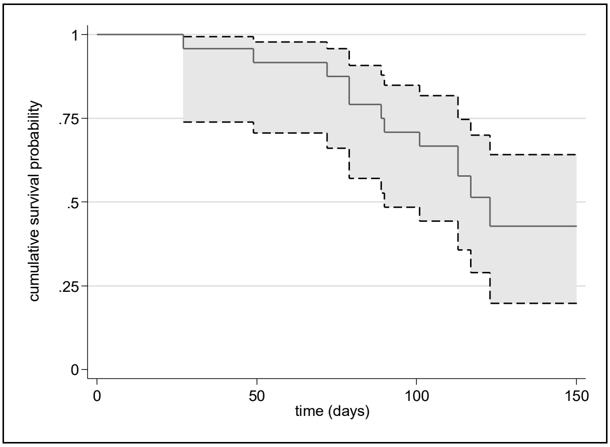


Fig. 19.5 Kaplan-Meier survivor function (95% CI)

Example 19.2 Actuarial and Kaplan-Meier estimates of survivor functions

data = calf_pneu

Data on the occurrence of calf pneumonia in calves raised in 2 different housing systems were published (Thysen, 1988). Calves surviving to 150 days without experiencing pneumonia were considered censored at that time. The table below presents an actuarial life table estimate of the cumulative survivor function.

Actuarial life table

Interval		Beg. total	Deaths	Lost	Cum. Survival	SE	95% CI	
15	30	24	1	0	0.958	0.041	0.739	0.994
45	60	23	1	0	0.917	0.056	0.706	0.979
60	75	22	1	0	0.875	0.068	0.661	0.958
75	90	21	3	0	0.750	0.088	0.526	0.879
90	105	18	2	1	0.664	0.097	0.439	0.816
105	120	15	3	6	0.498	0.110	0.273	0.688
120	135	6	1	0	0.415	0.119	0.189	0.629
150	165	5	0	5	0.415	0.119	0.189	0.629

Note that survival estimates are only presented for intervals in which at least one event or censoring occurred. Thus the cumulative survival at the end of the 30-45 day interval would be exactly the same as at the end of the 15-30 day interval (0.958)

Kaplan-Meier survivor function

Time	Beg. total	Fail	Lost	Survivor function	SE	95% CI	
27	24	1	0	0.958	0.041	0.739	0.994
49	23	1	0	0.917	0.056	0.706	0.979
72	22	1	0	0.875	0.068	0.661	0.958
79	21	2	0	0.792	0.083	0.570	0.908
89	19	1	0	0.750	0.088	0.526	0.879
90	18	1	0	0.708	0.093	0.484	0.849
101	17	1	1	0.667	0.096	0.443	0.817
113	15	2	4	0.578	0.102	0.357	0.747
117	9	1	2	0.514	0.109	0.288	0.700
123	6	1	0	0.428	0.120	0.198	0.641
150	5	0	5	0.428	0.120	0.198	0.641

The 2 estimates of the probability of survival up to day 150 are very close (41.5% and 42.8%).

Assume the following notation and assumptions:

t_j	j=1,...,n are failure times
t^*	is the final failure time = max(t_j)
d_j	the number of failures at time t_j
r_j	the number of subjects at risk at time t_j
I_k	time (0, t^*) is divided into many small intervals (I_k)
p_k	the probability of surviving through I_k if alive at the start of I_k

As I_k gets very small, then p_k=1 if there is no failure during the interval and p_k=(r_j-d_j)/r_j if t_j falls in the interval I_k. If there are ties between failures and censored observations, it is assumed that the failures occurred first (*ie* the censored observations are included in the 'at risk' group).

The Kaplan-Meier estimator of survival $S(t)$ at time t is defined as:

$$S(t)=\prod_{j:t_j\le t}(r_j-d_j)/r_j \qquad \text{for } 0\le t\le t^* \qquad \textit{Eq 19.1}$$

The Kaplan-Meier function is therefore a **piecewise constant** (*ie* remains constant over time intervals), **non-increasing** (*ie* it can be flat or go down, but never up) and **right-continuous** (*ie* after an event, it remains constant up until, but not including, the next event) function on the interval (0, t^*). It only changes value at failure times (t_j).

The most commonly used standard error (SE) of $S(t)$ is attributed to Greenwood (reported in (Collett, 2003)). Because survival probabilities often have a very skewed distribution, it is not usual to compute a confidence interval as an estimate $\pm$ 1.96(SE). Consequently, confidence intervals are computed by estimating $S(t)$ and its SE on either a natural log (ln) scale or on a ln(-ln) scale and then back-transforming the estimates to the original time scale. (**Note** the ln(-ln) transformation maps probabilities (0,1) onto (-∞, ∞).

19.5 Nelson-Aalen estimate of cumulative hazard

In the above 2 sections, we introduced the concept of 'hazard', being the probability of failure at a point in time, given that the animal had survived up to that time point. This is discussed more formally in Section 19.7, but for now, we note that a **cumulative hazard** (Nelson-Aalen estimate) can also be computed. The cumulative hazard is the expected number of outcomes for one subject occurring up to a point in time (assuming that the outcome could occur multiple times in an individual). For example, in the calf pneumonia data, the cumulative hazard at day 60 would be the sum of all the individual hazards (computed at failure times), up to day 60.

The cumulative hazard can range from 0 to infinity (as the time period gets longer, the expected number of outcomes keeps going up with no upper bound). A graph of the cumulative hazard is, like a graph of the survivor function, a way of expressing the overall failure (survival) experience of the population. Fig. 19.6 shows the cumulative hazard (and 95% CI) for the calf-pneumonia data.

Using the notation from Section 19.4.3, the Nelson-Aalen estimator of the cumulative hazard $H(t)$ at time t is computed as:

$$H(t)=\sum_{j:t_j\le t} d_j/r_j \qquad \textit{Eq 19.2}$$

for $0\le t\le t^*$.

As with the Kaplan-Meier estimator of *S*(*t*), SE can be determined and confidence intervals are computed on a ln scale and back transformed.

19.6 STATISTICAL INFERENCE IN NON-PARAMETRIC ANALYSES

19.6.1 Confidence intervals and 'point-in-time' comparisons

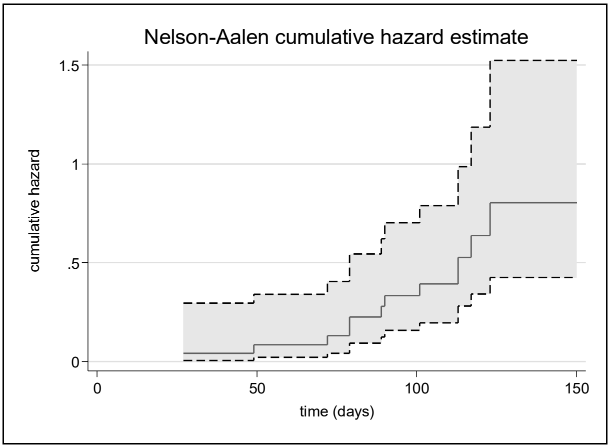


Fig. 19.6 Nelson-Aalen cumulative hazard function (with 95% confidence interval)

Although the formulae have not been shown, SEs of the cumulative survival estimates can be computed from actuarial or Kaplan-Meier survivor functions at any point in time. These SE can be used to test the difference between survivor functions (usually on a log scale) for 2 (or more) populations at any point in time using a standard normal *Z*-test. However, there are potentially, an infinite number of points at which the cumulative survival probabilities could be computed. This could lead to a serious problem of 'data snooping' or multiple comparisons and consequently, 'point-in-time' comparisons are only valid if it is possible to identify specific times at which the comparison of survival probabilities is warranted. These should be specified *a priori* (*ie* before the data are collected) and if multiple time points are evaluated, some adjustment for multiple comparisons must be made.

19.6.2 Tests of the overall survival curve

There are several tests that can be used to test whether the overall survivor functions in 2 (or more) groups are equal. They are all based on a series of contingency tables of observed and expected events for each group at each time point at which an event occurred (assuming the test is based on a Kaplan-Meier survivor function). The observed number of events at each time point is compared to the expected number and a χ^2 test computed. (Under the H_0 that there is no difference between the 2 groups, the expected number of events is a function of the amount of follow-up time in each group.) Consequently, the tests can be viewed as the survival analysis equivalent of the Mantel-Haenszel test for stratified data.

All of the tests assume that the ratio of risks of the event of interest for the 2 groups is constant across all strata (equivalent to the no-interaction assumption in a Mantel-Haenszel test). This assumption is known as the 'proportional hazards' assumption (you will see more of this later). If the survivor functions cross over, then it is clear that this assumption is violated. The differences among the tests depend on the weights used to combine the estimates derived at each point in time.

Log-rank test

The log-rank test is the simplest test as it assigns equal weight to each point estimate (weights $w(t_j)$=1). Consequently, it is equivalent to doing a standard Mantel-Haenszel procedure to combine the estimates. This equivalence is shown in Example 19.3.

Example 19.3 Equivalence of log-rank test and Mantel-Haenszel procedure

data = calf_pneu

Log-rank test for equality of survivor functions. The resultant P-value was 0.084.

	Events Observed	Events Expected
batch	4.00	6.89
continuous	8.00	5.11
Total	12.00	12.00

The layout for Mantel-Haenszel stratified analysis is shown below. The Mantel-Haenszel χ^2 statistic was computed using Eq 13.7

	day=27		day=49		day=79	
	surv	fail	surv	fail	surv	fail
batch	12	0	12	0	etc	
continuous	11	1	10	1		

Mantel-Haenszel OR = 3.09, χ^2 = 2.99, P = 0.084

Wilcoxon test

This test weights the intervals according to the sample size ($w(t_j)=n_j$) . Consequently, it is more sensitive to differences early in the time period when the sample size is larger. Some people advocate using both Wilcoxon and the log-rank test to see if differences in the survival curves occur early or late in the time period studied. The Wilcoxon test is less sensitive than the log-rank test to the assumption of proportional hazards, but will be unreliable if the censoring patterns vary across the groups being compared.

Other tests

Other non-parametric tests include the Cox test, the Tarone-Ware test, and the Peto-Peto-Prentice test. The first is based on a Cox regression procedure (see Section 19.8) while the Tarone-Ware weights the stratum-specific estimates by the square root of the population at risk at each time point. The Peto-Peto-Prentice test weights the stratum-specific estimates by the overall survival experience (an estimate of $S(t)$ just before the time point of interest), and consequently, reduces the influence of different censoring patterns between the groups.

Example 19.4 shows separate survivor functions for 'batch' and 'continuous' stocked calves and the results from several of the tests for the overall equality of the survivor functions.

19.7 Survivor, failure and hazard functions

The concepts of survivor, and hazard functions were introduced when we looked at non-parametric methods of analysis of survival data. Before proceeding with semi-parametric and parametric analyses, we need to develop a more complete understanding of these and related functions.

Example 19.4 Comparing survivor functions

data = calf_pneu

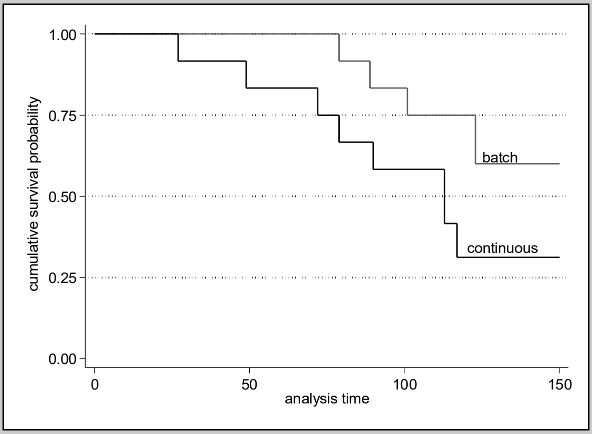


Fig. 19.7 K-M survival curves, by stocking type

Fig. 19.7 shows the Kaplan-Meier survivor functions for batch and continuous-stocked calves.

Continuous-stocked calves appeared to be at greater risk of having pneumonia than batched stock calves. The statistical significance of the test results for the difference between these 2 survivor functions are shown below. All statistical tests provide comparable results (borderline significance).

Test	P-value
log-rank	0.084
Wilcoxon	0.083
Cox	0.088
Tarone-Ware	0.081
Peto-Peto-Prentice	0.078

19.7.1 Survivor function

The survivor function ($S(t)$) is the probability that an individual's survival time (T) (or more generally, the time-to-event occurrence) will exceed some specified time t. It can be written as:

$$S(t) = p(T \geqslant t) \qquad \textbf{\textit{Eq 19.3}}$$

As noted, survivor functions are non-increasing. They start at 1 and drop to 0 if all individuals ultimately experience the event of interest. **Note** By convention, cumulative functions will be designated by upper-case letters and density functions by lower-case letters. The survivor function is a cumulative function in that it represents the cumulative probability of surviving up to a point in time t.

19.7.2 Failure function

The failure function ($F(t)$) is the probability of not surviving past time t. Consequently, it is:

$$F(t) = 1 - S(t) \qquad \textbf{\textit{Eq 19.4}}$$

19.7.3 Probability density function

The probability density function (*f*(*t*)) describes the distribution of survival times and is the slope (derivative) of the failure function. Consequently, it represents the instantaneous rate at which failures are occurring in the study population at a point in time. It is estimated by taking the derivative of a smoothed estimate of the failure function with respect to time (See Fig. 19.8).

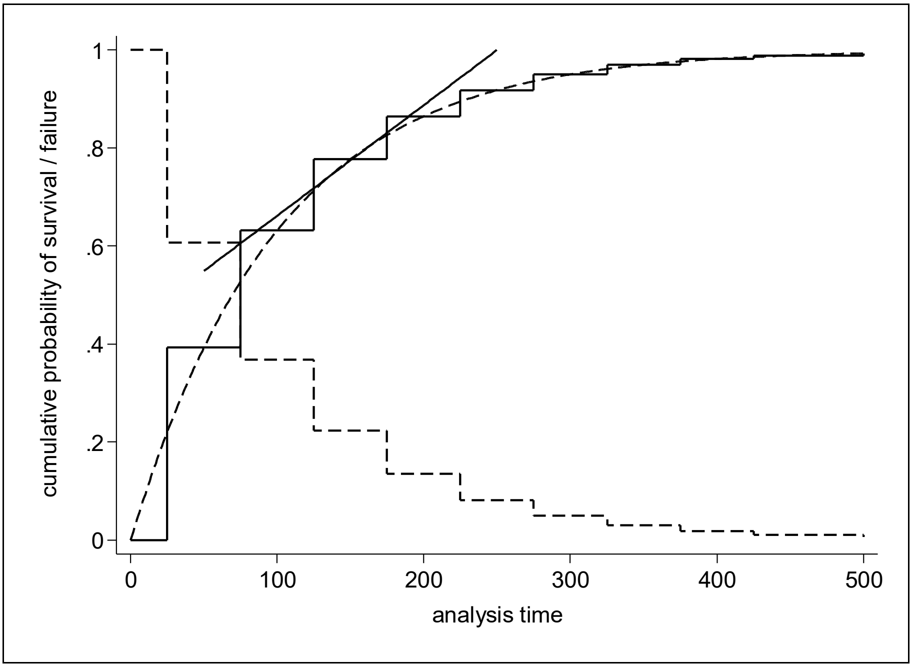


Fig. 19.8 Survivor function (dashed stepped line). Failure function (solid stepped line). Smoothed failure function (dashed curved line) and tangent of smoothed failure function (short solid line) giving the slope at a single point

19.7.4 Hazard function

The hazard function (*h*(*t*)) is the probability of an event occurring at time *t* given that it had not occurred up to time *t*. With time divided into discrete intervals (as in a life table), it can be expressed as:

$$h(t) = p(T = t | T \geq t) \qquad \textbf{\textit{Eq 19.5}}$$

With time on a continuous scale, the hazard function describes the instantaneous probability of an event occurring at a point in time given that it did not occur previously. The hazard function is:

$$h(t) = \lim_{\Delta t \to 0} \frac{p(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \qquad \textbf{\textit{Eq 19.6}}$$

The hazard function can also be computed as the ratio of the probability density function (which represents the rate at which failures are occurring at a point in time) and the survivor function (which represents the probability of surviving up to that point in time). It can further be expressed as:

$$h(t)=\frac{f(t)}{S(t)}=\left[\frac{-\frac{dS(t)}{dt}}{S(t)}\right]=-\left[\frac{d}{dt}(\ln S(t))\right]$$ ***Eq 19.7***

Hazard functions are always non-negative (*ie* greater than or equal to zero) and have no upper bound (their value will change with the time scale used).

19.7.5 Cumulative hazard function

The cumulative hazard (*H*(*t*)), also known as the **integrated hazard**, represents the accumulation of hazard over time. It can be computed as the integral of the hazard function but is more conveniently found using the following equation.

$$H(t)=-\ln S(t)$$ ***Eq 19.8***

As noted, the cumulative hazard represents the expected number of outcomes of interest that would occur in an individual (assuming that repeat occurrences were possible). For example, if you were studying the survival of cats following infection with the feline infectious peritonitis virus and at 3 years you find that the cumulative hazard = 4 [$H(t_3) = 4$], then that would suggest that in 3 years after infection, we would expect to see 4 deaths. Obviously, only one death is possible, but it provides an indication that the probability of the cat surviving to 3 years post-infection is very low.

19.7.6 Relationships among survivor, failure and hazard functions

Some of the relationships between the survivor, failure and hazard functions have already been shown in previous sections. As each of these functions determine the survival time distribution, if one of them is known, the others can all be computed.

$$f(t)=\frac{dF(t)}{dt} \quad h(t)=\frac{dH(t)}{dt} \quad h(t)=\frac{f(t)}{S(t)}$$ ***Eq 19.9***

Note *f*(*t*) and *h*(*t*) are derivatives of *F*(*t*) and *H*(*t*) which are step functions. Consequently, the step functions are smoothed before the derivative is taken.

$$F(t)=1-S(t) \quad H(t)=-\ln S(t) \quad S(t)=e^{-H(t)}$$ ***Eq 19.10***

Note The last expression for *S*(*t*) (above) gives the Flemming-Harrington estimate of the survivor function when the Nelson-Aalen estimate is used for *H*(*t*). This estimate will be larger than the Kaplan-Meier estimate of *S*(*t*) computed directly, but will be close if the number of failures is small relative to the number of individuals at risk.

While survival experiences for groups of animals are usually shown by plotting the survivor function, the hazard function plays a key role in semi-parametric and parametric analyses.

19.7.7 Examples of hazard functions

A wide variety of hazard functions have been studied, but constant and Weibull functions are the 2 most commonly encountered in survival analyses. Other forms used include the log-normal, log-logistic, gamma and Gompertz. The names of these functions refer to the corresponding survival time distributions (see Section 19.9).

Constant hazard

A constant hazard is one which does not change over time. With a constant hazard (λ), the survivor function drops exponentially and survival times will have an exponential distribution. The hazard *h*(*t*), density *f*(*t*) and survivor *S*(*t*) functions are:

$$h(t)=\lambda \qquad f(t)=\lambda e^{-\lambda t} \qquad S(t)=e^{-\lambda t} \qquad \textbf{\textit{Eq 19.11}}$$

The appropriateness of an exponential model can be assessed by plotting the cumulative hazard *H*(*t*) (or equivalently -ln*S*(*t*)) against *t*. If the exponential model is appropriate, the line will be straight. Fig. 19.9 shows a survivor function derived from a constant hazard of λ=0.01 per day.

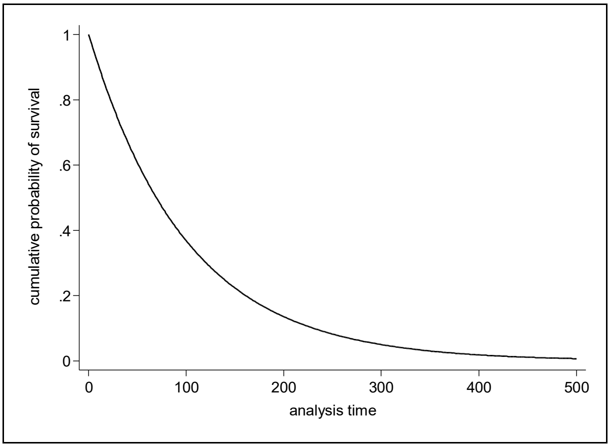


Fig. 19.9 Survivor function from a constant hazard (*h*(*t*)=0.01)

Weibull hazard

A Weibull hazard function depends on 2 non-negative parameters: a scale parameter (λ) and a shape parameter (*p*). If *p*=1, the resulting survival time distribution is the exponential distribution. If *p*<1 then the hazard function decreases monotonically. If *p*>1, then the function is monotonically increasing with a value between 1 and 2 producing a curve that increases at a decreasing rate, *p*=2 produces a hazard function that increases linearly with time and *p*>2 produces a function that increases at an ever-increasing rate. The hazard and survivor functions are:

$$h(t)=\lambda p t^{(p-1)} \qquad S(t)=\exp(-\lambda t^{p}) \qquad \textbf{\textit{Eq 19.12}}$$

Fig. 19.10 shows Weibull hazard functions for several values of *p*. An increasing Weibull hazard function (1<*p*<2) might be appropriate for dairy cow conception data if the fertility of the cow increases with time after parturition, but does so at a decreasing rate. A decreasing Weibull hazard function (*p*<1) might be appropriate for the survival of animals after surgery when the hazard is highest right after surgery and then decreases.

The suitability of the Weibull distribution or hazard can be assessed by evaluating the log-cumulative hazard plot [ln(*H*(*t*)) versus ln(*t*)]. If the data fit a Weibull distribution, the line on the graph should be approximately straight. The intercept and the slope of the line will be ln(λ) and *p*, respectively. Parametric survival models based on exponential and Weibull hazard functions are described in Section 19.9.

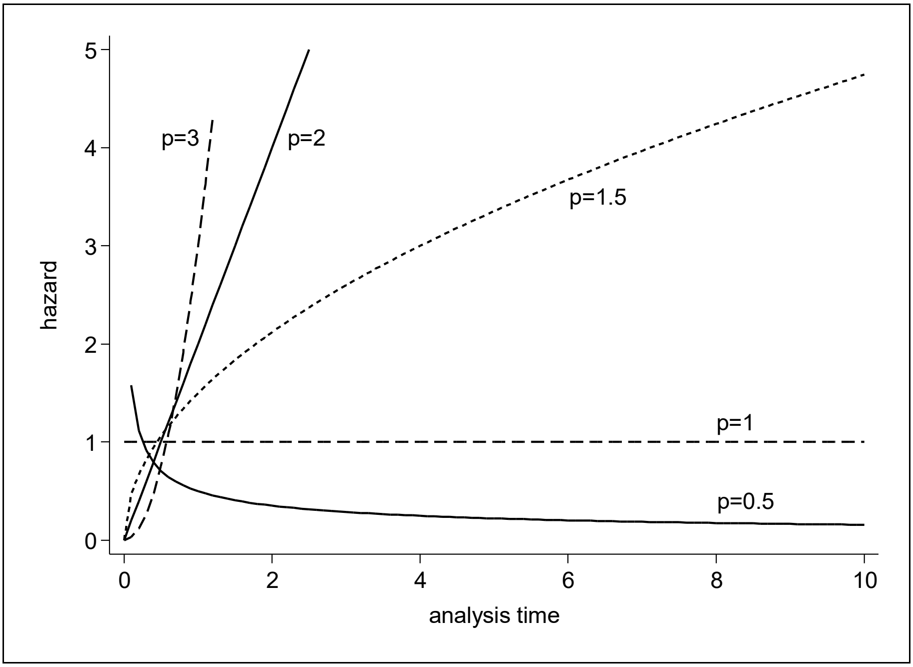


Fig. 19.10 Weibull hazard functions for various values of shape parameter (p)

Other hazard functions

One of the limitations of the Weibull hazard function is that the hazard can only increase or decrease over time. **Gamma**, **log-normal** and **log-logistic** hazards can be used to deal with the situation in which the risk first increases and then decreases (or vice versa). Such a function would be appropriate in a situation where the risk of death was high early in an illness, drops to a lower level and then increases again over time. For example, a new intramammary infection with *Staph. aureus* in a dairy cow might produce a high risk of culling early (if acute clinical mastitis developed), followed by a sharp reduction in the risk and then a gradually increasing risk as the level of chronic udder damage increased over time. Detailed descriptions of these functions can be found in survival analysis texts (Cleves *et al*, 2008; Collett, 2003; Hosmer & Lemeshow, 2008; Therneau & Grambsch, 2000).

19.8 SEMI-PARAMETRIC ANALYSES

Non-parametric analyses are limited to evaluating the effect of one, or a small number of, qualitative variable(s) on survival times. However, we often want to simultaneously evaluate the effects of multiple continuous or categorical explanatory variables. This requires that we model the survival data using a multivariable technique. The most commonly used form of multivariable analysis for survival data is the **proportional hazards model** (also known as the **Cox regression model**) (Cox, 1972). It is a semi-parametric model in that we do not have to assume any specific functional form for the hazard, but we do model the ratio of hazards as a linear function of the predictors.

19.8.1 Cox proportional hazards model

The proportional hazards model is based on the assumption that the hazard for an individual is a product of a baseline hazard (h_0) and an exponential function of a series of explanatory variables.

$$h(t)=h_0(t)e^{\beta X} \qquad \textbf{\textit{Eq 19.13}}$$

where $\beta X= \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$. Equivalently, it can be expressed as:

$$HR=\frac{h(t)}{h_0(t)}=e^{\beta X} \qquad \textbf{\textit{Eq 19.14}}$$

where *HR* is the hazard ratio. The first formulation emphasises that the hazard for an individual is always a multiple ($e^{\beta X}$) of a baseline hazard (see Fig. 19.11—left panel), while the second formulation shows that it is the ratio of the hazards which is assumed to be constant over time.

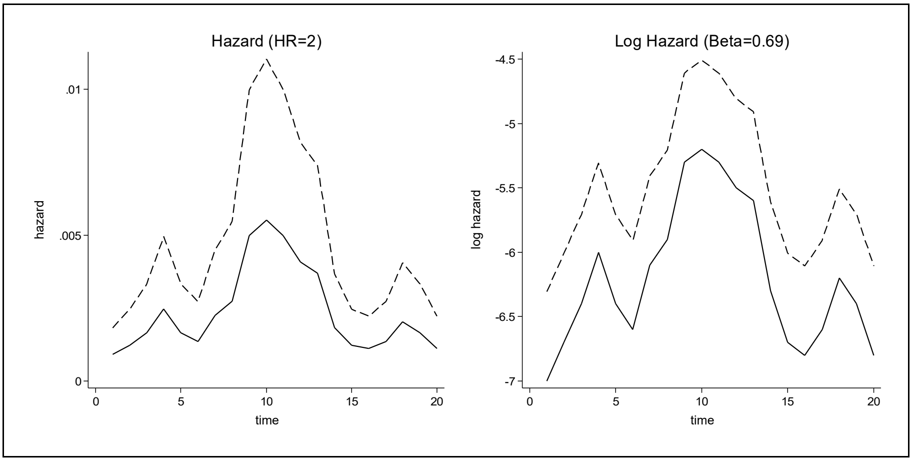


Fig. 19.11 Effect of a hypothetical factor on a baseline hazard shown on 2 scales. Effect on hazard scale in left panel and on log hazard scale in right

On the log scale, the log hazard is a constant (βX) above or below the baseline log hazard as shown below and in Fig. 19.11 (right panel).

$$\ln h(t)=\ln h_0(t)+\beta X \qquad \textbf{\textit{Eq 19.15}}$$

Two important features of this model are that no assumption is made about the shape of the baseline hazard (h_0) and that the model has no intercept. In fact, the intercept (which in most regression models reflects the value of the outcome when all covariates (predictors) are zero) is subsumed into the baseline hazard which represents the hazard when all covariates are zero.

19.8.2 Hazard ratios

Based on Eq 19.15, the lnHR=βX. Consequently, exponentiating the coefficient from a proportional hazards model produces a hazard ratio. Hazard ratios have interpretations similar to odds ratios and risk ratios. They represent the effect of a unit change in the predictor on the frequency of the outcome (which in this case is measured as a hazard). **Note** You will sometimes encounter hazard ratios referred to as relative risks (or risk ratios), but this is not a correct use of the term and should be avoided. For example, if factor X_1 has an HR=2, then a unit change in X_1 will double the hazard of the outcome. If X_1 is a dichotomous variable and, because we are assuming that this HR is constant (over time), this means that, at any point during the risk period, 'failures' will be occurring at twice the rate in animals with X_1=1 than in animals with X_1=0. This is not equivalent to a doubling of the risk over the full study period.

Example 19.5 provides some examples of *HR*s derived from a dataset from a clinical trial of prostaglandin use in dairy cattle. A total of 319 cows in 3 herds were assigned randomly to receive prostaglandin (or not) at the time that the producer had indicated was the beginning of the breeding period (*ie* the number of days after calving that he would start breeding a cow that came into heat). The time from the onset of the breeding period to conception was the outcome of interest. The dataset (-pgtrial-) is described more fully in Chapter 31. The variables in Table 19.3 are those that we will use in examples in this chapter.

Table 19.3 Variables in -pgtrial- dataset

Variable	Description
herd	herd (1,2,3)
tx	treatment (1=yes, 0=no)
lact	age (lactation number – a continuous variable)
thin	body condition score at time of treatment (1=thin, 0=normal or fat)
dar	days at risk (number of days from the start of the breeding period to either conception or censoring); this is the outcome of interest
preg	status of animal at end of -dar- (1=pregnant, 0=censored)

19.8.3 Fitting the Cox proportional hazards model

Obtaining partial maximum likelihood estimates of the parameters in a Cox proportional hazards model requires an iterative estimation procedure (the Newton-Raphson procedure is most commonly used). As with a non-parametric Kaplan-Meier estimation procedure, a Cox model is only evaluated at the times at which failures occur. In fact, fitting a Cox model with no predictors produces exactly the same survival curve as a Kaplan-Meier estimation does. In both procedures, it is not the actual times at which failures occur which is important, it is only the order in which they occur that matters.

The estimation is based on the partial (profile, conditional) likelihood function, which has a different interpretation than the usual likelihood function (as described in Example 19.6), but is used in the same way for statistical inference.

19.8.4 Handling of ties

Because the order in which failures occur is critical for conducting the analysis, there must be a way of handling the problem of 2 (or more) failures being recorded at the same time. Details of various methods of dealing with ties can be found in texts on survival analysis but they fall into 2 general approaches. The first is called a **marginal calculation** or continuous-time calculation and is based on the assumption that the timing of the events was not really tied, but simply due

Example 19.5 Cox proportional hazards model

data = pgtrial

A Cox proportional hazards model was fit to the prostaglandin trial data with herd, treatment, lactation number, and body condition (thin) as predictors. The first table presents the model in terms of coefficients.

No. of subjects = 319
No. of failures = 264 Number of obs = 319
Time at risk = 25018 LR $\chi^2_{(5)}$ = 9.50
Log likelihood = -1307.7329 Prob > χ^2 = 0.0908

Predictor	Coef	SE	Z	P	95% CI	
herd=2	-0.284	0.170	-1.68	0.094	-0.617	0.048
herd=3	0.037	0.174	0.21	0.833	-0.305	0.378
tx	0.184	0.125	1.46	0.143	-0.062	0.429
lact	-0.043	0.041	-1.04	0.297	-0.123	0.038
thin	-0.146	0.138	-1.06	0.291	-0.416	0.125

Although not statistically significant, treatment appears to increase the lnHR by 0.18 units. As we rarely think in terms of lnHRs, it is more common to present the results as HRs

Predictor	HR	SE	Z	P	95% CI	
herd=2	0.752	0.128	-1.68	0.094	0.539	1.050
herd=3	1.037	0.181	0.21	0.833	0.737	1.460
tx	1.202	0.151	1.46	0.143	0.940	1.536
lact	0.958	0.039	-1.04	0.297	0.884	1.038
thin	0.865	0.119	-1.06	0.291	0.660	1.133

Here it appears that treatment increases the hazard of conception 1.2 times. If this effect is real (which appears questionable at this stage given the P-value of -tx-), it means that at any point in time after the onset of the breeding period, conceptions were happening at a 20% higher rate in treated cows than in the non-treated ones. Similarly, for each additional lactation the cows had experienced, the rate of conception dropped by approximately 4% (but this predictor has an even larger P-value).

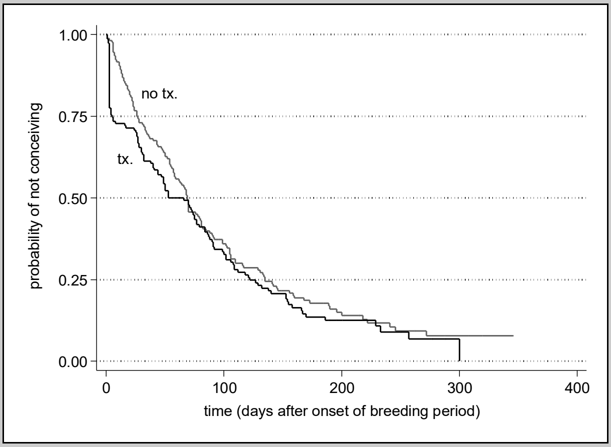


Fig. 19.12 Kaplan-Meier survival estimates, by treatment (prostaglandin data)

A pair of Kaplan-Meier survivor functions (one for each treatment group) provides some additional insight into the possible effect of the treatment (Fig. 19.12). It appeared that treated cows conceived slightly more quickly, although the difference was most pronounced early in the breeding period.

Example 19.6 Partial likelihoods for a Cox model

data = hypothetical time to death following diagnosis of lymphosarcoma for 20 dogs

Assume the following data, sorted by time to death

Dog	Time to death (mo)	Age at diagnosis (yrs)
1	3	9.6
2	8	8.1
...	...	...
20	63	5.7

For the first dog, a maximum likelihood procedure would ask the question ‘What was the probability of this dog dying at 3 months, given that it was 9.6 years old at diagnosis?’ In contrast, a partial likelihood procedure asks the question ‘Given that a death occurred at 3 months, what was the probability that it was dog #1 (given the age of the dog)?’ This likelihood can be written as follows.

$$L_1=\frac{h_1(3)}{h_1(3)+h_2(3)+...+h_{20}(3)}=\frac{h_0(3)\,e^{\beta*9.6}}{h_0(3)\,e^{\beta*9.6}+...+h_0(3)\,e^{\beta*5.7}}=\frac{e^{\beta*9.6}}{e^{\beta*9.6}+...+e^{\beta*5.7}}$$

The partial likelihood of the first failure being dog 1 is that dog’s likelihood relative to the sum of all of the likelihoods. For the second dog, the partial likelihood is

$$L_2=\frac{h_2(8)}{h_2(8)+h_3(8)+...+h_{20}(8)}=\frac{e^{\beta*8.1}}{e^{\beta*8.1}+...+e^{\beta*5.7}}$$

The product of the partial likelihoods is the likelihood of the model. **Note** The analysis only depends on the sequence of events (not the actual time) and that the baseline hazard has no effect because it is common to all dogs.

to the fact that the timing of the failure was not recorded with sufficient precision to differentiate among ‘tied’ observations. The second is called the **partial calculation** and is based on the assumption that the events were actually tied and treats the problem as a multinomial problem.

Exact calculation of the likelihood function under either assumption is computationally demanding and may be slow in large datasets with many ties. Two approximate methods have been developed for marginal calculations. The **Breslow** method is simplest and is adequate if there are not a lot of ties. The **Efron** method provides a closer approximation to the exact calculation. An approximation attributed to **Cox** can be used for partial calculations. However, for a small dataset such as -pgtrial-, exact calculation methods (marginal or partial) are feasible. In this case, the exact methods and the Breslow and Efron approximations all produce very similar results (data not shown).

19.8.5 Baseline hazard

Although, as noted above, no assumption is made about the baseline hazard (h_0) and the Cox model does not estimate it directly, an estimate of it can be derived conditional on the set of coefficients in the estimated model. This baseline hazard represents the hazard in an individual for whom all predictors equal zero. For it to be meaningful, it is important that X=0 is reasonable for all predictors. If computed directly from the -pgtrial- data using the model shown in Example 19.5, this would represent the hazard of conception in a non-treated, normal body condition cow in herd 1 in her 0th lactation. To avoid this nonsensical value for lactation, lactation should be modified so that a cow with a value of 0 is possible (*eg* rescale it by subtracting 1 so a cow in lactation 1 now has a value of 0).

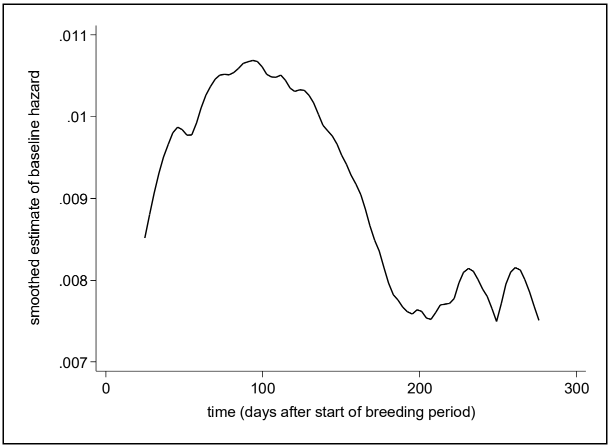


Fig. 19.13 Smoothed estimate of baseline hazard

The baseline hazard can only be estimated on days on which failures occur, and the estimate will bounce around quite a lot from day to day (particularly once the surviving population at risk becomes small. Consequently, it is necessary to smooth the estimate of the baseline hazard and this is shown in Fig. 19.13. The daily hazard of conception in non-treated, normal-weight, first-lactation cows in herd 1 rises from about .006 (0.6% per day) to a peak of about 0.011 (1.1% per day) and then gradually declines over time (up to 300 days—the end of the follow-up period). It is important to note that this reflects the probability of conception among the pool of cows remaining non-pregnant at any specific time. It does not necessarily indicate that the hazard for an individual cow declines after day 100. The decline may be a function of the fact that the population of cows remaining non-pregnant consists increasingly of cows that are very difficult to get pregnant. This issue of the nature of the population changing is discussed further in Section 19.11 (frailty models).

19.8.6 Model-building

In general, model-building procedures for Cox models are similar to those used for other regression-type models. Wald tests and likelihood ratio tests can be used to evaluate the significance of individual predictors or groups of predictors. Confounding and interaction can be assessed using methods presented for other regression-type models. Because the explanatory variables are related to the logarithm of the hazard ratio, it follows that interaction will be assessed on a multiplicative scale. There are, however, 2 issues that are specific to survival models: **stratified analysis** to allow for different baseline hazards in different groups of animals in the study, and the possibility of including **time-varying covariates**.

19.8.7 Stratified analysis

Although we made no assumption about the shape of the baseline hazard, we have assumed that it is appropriate for an animal with all X_j=0. Let's consider the effect of being 'thin' on the hazard of conception in the prostaglandin data. If we obtained a significant *HR* for -thin-, we would assume that it multiplies the h_0 by the *HR* and that this effect was constant over time. If we had reason to believe that the shape of the hazard was different in thin cows than in normal-weight cows, we could stratify the analysis on -thin- and allow for separate estimates of the baseline hazard in each group.

In a stratified Cox model, different baseline hazards ($h_{0j}(t)$) are assumed across groups of animals to yield the following hazard function for the j^{th} group.

$$h_j(t)=h_{0j}(t)\,e^{\beta X}$$ ***Eq. 19.16***

The difference from the unstratified model (Eq. 19.14) is only in the baseline hazards whereas the regression term $e^{\beta X}$ is unchanged. Thus, the effects of predictors on *HR*s relative to the baseline hazard are assumed equal across all strata. Stratum-level predictors can not be assessed in a stratified model because their effects will be absorbed in the baseline hazards. However, you can include interactions between a covariate (*eg* -tx-) and a stratifying variable (*eg* herd). Example 19.7 shows a stratified (by herd) analysis of the -pgtrial- data with a treatment by herd interaction included. (**Note** Stratified analyses provide one means of dealing with clustered data —by stratifying on the clustering variable. Dealing with clustered data is discussed further in Section 19.11).

19.8.8 Time-varying covariates

Up to now, we have focused on exposure factors that do not change their value over time and we have assumed that the effect of a factor was constant over time (proportional hazards assumption). However, survival analysis gives us the opportunity to relax both of these conditions. The terminology used with time-varying covariates may be confusing so we will distinguish between **time-varying predictors** and **time-varying effects**.

Given the long-term nature of many survival studies, it is conceivable that the values of some of those predictors might change over time. These are time-varying predictors. For example, in the prostaglandin trial, if the body condition of the cows had been assessed periodically, rather than just once, some cows that were initially thin could have gained enough weight to be classified as normal or vice versa.

Example 19.7 Stratified Cox proportional hazards model
data = pgtrial

A stratified (by herd) model was fit with a treatment by herd interaction term included.

No. of subjects = 319
No. of failures = 264 Number of obs = 319
Time at risk = 25018 LR $\chi^2_{(7)}$ = 10.32
Log likelihood = -1025.1181 Prob > χ^2 = 0.1710

Predictor	Coef	SE	Z	P	95% CI	
tx	-0.022	0.255	-0.08	0.933	-0.522	0.479
herd_2 * tx	-0.057	0.336	-0.17	0.866	-0.715	0.601
herd_3 * tx	0.545	0.318	1.71	0.087	-0.079	1.169
lact	-0.046	0.041	-1.13	0.258	-0.126	0.034
thin	-0.136	0.138	-0.98	0.326	-0.407	0.135

The main effect of -tx-, which is now the effect of -tx- in herd=1 is completely non-significant. An overall test of the significance of the interaction terms produces a P-value of 0.072 (not shown), which suggests that there is some evidence that the treatment may have different effects in different herds.

On the other hand, a predictor may remain constant, but its effect may change over time. These are time-varying effects. For example, prostaglandin treatment may have a more pronounced effect in the days or weeks immediately after administration than many weeks later. If this is true the assumption of proportional hazards is violated.

Time-varying predictors

Because there were no time-varying predictors in the prostaglandin trial data, we will shift our attention to a study that evaluated the effects of a number of risk factors on the time to occurrence of infectious salmon anemia (ISA) outbreaks in salmon being reared in net-pens in ocean-based aquaculture operations. These data (isa_risk) are from 182 net-pens on 18 sites and are described more fully in Chapter 31. For this example, we will focus on a single predictor: whether or not there was (or had been) another outbreak at the site. At a site with no outbreaks, all records were censored at the end of the study period and there was a single record for each net-pen. For sites where an outbreak occurred, each net-pen would have 2 records. The first would describe the period up to the date of the first outbreak and would end in a censoring for all net-pens except for the one that had the first outbreak. The second record would span the period from the date of the first outbreak until the cage either had an outbreak or was censored. Example 19.8 shows how the data must be modified to account for a time-varying predictor.

Time-varying effects

A time-varying effect represents an interaction between a predictor and time (the effect of the predictor depends on what time point you are looking at). Effects may change at discrete points in time or may change continuously over time. A continually changing effect may change in a linear manner with time (*eg* effect drops by a given amount every 10 days), with ln time (*eg* effect drops by a given amount for every one ln unit increase in time (equivalent to every 2.72 fold increase in time)), or with any other function of time. Evaluating how effects may, or may not, change over time is an important part of validating a Cox proportional hazards model and is discussed further in Section 19.8.10.

One approach to evaluating how effects of predictors change over time is to fit an Aalen's linear hazards model (Hosmer & Royston, 2002). This model plots a cumulative regression coefficient for a predictor against time. If the effect of the predictor remains constant, the cumulative predictor will be expected to increase (or decrease) in a straight line over time. In general this is true, although some curvature to this line has been observed even when hazards are proportional. Example 19.9 shows Aalen's linear hazards model applied to the prostaglandin data.

19.8.9 Validating the model

Validation of a Cox proportional hazards model will be covered in the following 6 sections. The components in the validation process include:

- evaluating the proportional hazards assumption (Section 19.8.10)
- evaluating the assumption of independent censoring (Section 19.8.11)
- evaluating the overall fit of the model (Section 19.8.12)
- evaluating the functional form of predictors (Section 19.8.13)
- checking for outliers (Section 19.8.14)
- detecting influential points (Section 19.8.15).

Example 19.8 Time-varying predictor

data = isa_risk

Data were collected on a number of risk factors for outbreaks of ISA in 182 net-pens of salmon at sea-cage sites. The period of risk was considered to start on 1 April 1997 (day=0) and carried on until the fish were harvested in the fall of 1997. Data from 3 net-pens at site 19 are:

		time		
site	**net-pen**	**start**	**end**	**outcome**
19	39	0	86	1 = outbreak
19	46	0	211	0 = censored
19	56	0	79	1 = outbreak

Net-pen 46 did not have an outbreak and was censored on day 211. Net-pens 39 and 56 had outbreaks on days 86 and 79, respectively with the outbreak in net-pen 56 being the first outbreak at the site. In order to allow for a time-varying predictor to indicate whether or not there had been another net-pen with an outbreak on the site, multiple records for each net-pen need to be created (this data format is referred to as counting process style data). The resulting data follow.

		time			site
site	**net-pen**	**start**	**end**	**outcome**	**positive**
19	39	0	79	0	0
19	39	79	86	1	1
19	46	0	79	0	0
19	46	79	211	0	1
19	56	0	79	1	0

Net-pen 39 now has one record for the period of days 0 to 79 during which the covariate (predictor) for the site being positive was 0 and which ended in censoring. It has a second record for the period from day 79 to 86 when the site was positive and which ended in an outbreak. Similarly, net-pen 46 has 2 records (representing the period before and after the site became positive), but both end in censorings because the net-pen did not have an outbreak. Net-pen 56 still only has one record because it was the first outbreak.

A Cox model fit to these data with the single predictor -pos- (*ie* site was positive) produces:

No. of subjects = 182
No. of failures = 83 Number of obs = 312
Time at risk = 28353 LR $\chi^2_{(1)}$ = 15.24
Log likelihood = -392.91 Prob > χ^2 = 0.0001

Predictor	HR	SE	Z	P	95% CI	
pos	2.610	0.676	3.70	0.000	1.571	4.335

Although it appears that there were 312 observations, the number of subjects is correctly identified as 182. Once a site has become positive, the rate of outbreaks in other cages at the site was 2.6 times higher than prior to the site becoming positive.

Example 19.9 Aalen's linear hazards model

data = pgtrial

Aalen's linear hazards model was fit to the prostaglandin data and the cumulative coefficient for treatment was plotted against time and the cumulative coefficient plotted for the first 50 days.

There was initially a very strong positive effect of treatment (primarily on day 3, some effect up to day 6) followed by a strong negative effect that lasted until about day 23. This was followed by approximately a 9-day period of positive effect, after which there was no consistent evidence of any effect. This fits well with the expected effect of prostaglandin: a strong positive effect as synchronised cows come into heat, followed by a negative effect as those cows go through a heat cycle before returning to oestrus about 21 days later.

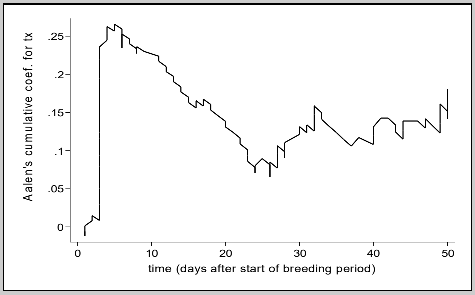


Fig. 19.14 Aalen's linear hazards model of prostaglandin data

19.8.10 Evaluating the assumption of proportional hazards

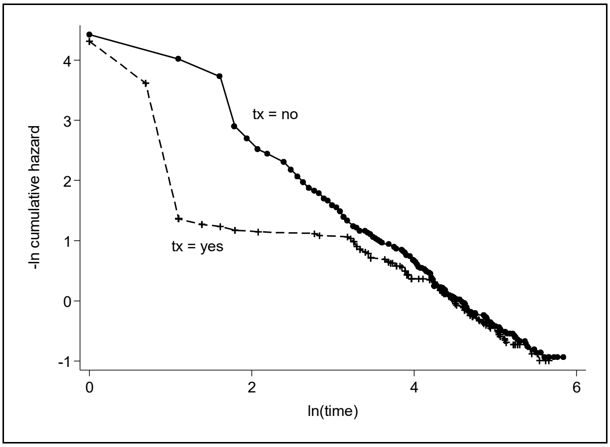
There are 3 general ways of evaluating the assumption of proportional hazards:

- graphical assessment
- the use of time-varying effects
- statistical assessment using Schoenfeld residuals.

Graphical assessment

For a categorical predictor, the assumption of proportional hazards can be tested by examining the log-cumulative hazard plot (ln*H*(*t*) vs ln*t*) to check if the curves for the 2 (or more) study groups are parallel. If they are not parallel, then the assumption has been violated. Fig. 19.15 shows a log cumulative hazard plot for the prostaglandin data. (**Note** it is actually the -log of the cumulative hazard that has been plotted which explains why the curves slope down (instead of up) as the cumulative hazard rises. It is clear that the curves are not parallel, at least up to ln(time) ≈3.5 (33 days), suggesting that the proportional hazards assumption has been violated. This seems reasonable because we would expect prostaglandin treatment to have a more pronounced effect shortly after administration than many weeks later.

Fig. 19.15 Log cumulative hazard plot.



Note It is actually the -log of the cumulative hazard that has been plotted

Another approach to graphical assessment is to compare plots of predicted survival times from a Cox model (which assumes proportional hazards) to Kaplan-Meier survivor function plots (which make no such assumption). If the 2 sets of curves are close together, it suggests that the proportional hazards assumption has not been violated. Fig. 19.16 shows such a plot. Clearly, the predicted values from the Cox model (the 2 curves in the centre of the plot are the predicted values for treated (lower curve) and not treated (upper curve) cows) are not at all close to the observed values prior to day 24. A limitation to graphical assessment is that it is limited to evaluating unconditional associations or situations in which the predictor being evaluated is clearly the strongest predictor in a multivariable setting.

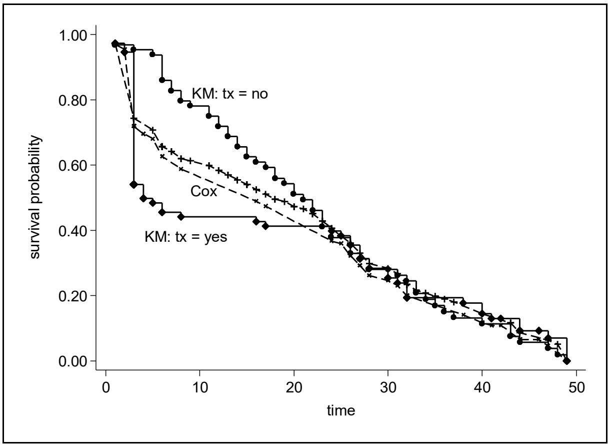


Fig. 19.16 Kaplan-Meier Cox plot

Time-varying effects

A term for the interaction between the treatment and time (or the log of the survival time) can be added to the model. The effect of treatment can be allowed to interact with time in a linear fashion or with ln(time) (or any other function of time for that matter). The advantage of adding a predictor*time interaction term is that if the assumption of proportional hazards is violated, the addition of the interaction term can solve the problem (provided the change in effect over time can be appropriately modelled).

In Example 19.10, a Cox model has been fit in which the effect of treatment is allowed to vary with ln(time). The positive effect of treatment disappears by day 55 and the effect then becomes negative (*HR*<1). No negative effect of treatment was expected beyond day 55 so allowing the effect to decay linearly with ln(time) may not be adequate. Although details are not shown, if interaction terms between both ln(time) and ln(time)^2 are added to the model, the log likelihood rises from -1307 to -1300 suggesting that the latter model improves the model fit substantially. Some special procedures for integrating the use of fractional polynomials into the fitting of time-varying effects are discussed in (Royston & Sauerbrei, 2008).

Schoenfeld residuals

Schoenfeld and **scaled Schoenfeld** residuals are based on the contribution that an observation makes to the partial derivative of the log partial likelihood. Hence they are also sometimes called ‘partial residuals’. There is a separate set of residuals for each regression coefficient in the model, each set corresponding to the partial derivative for that parameter. These residuals are only computed at observed survival times. Scaled Schoenfeld residuals are adjusted using an estimate of the variance of the residual and these are better for detecting departures from the assumed model.

A graph of the scaled Schoenfeld residuals for a given predictor, when plotted against time (or ln(time)) can provide a graphical assessment of the proportional hazards assumption. This is particularly useful for continuous predictors because the log cumulative hazard plot is not

Example 19.10 Assessing proportional hazards assumption—time-varying covariates

data = pgtrial

A Cox model with a single predictor (treatment) was fit but the effect of treatment was allowed to interact with time on a natural log scale. This was chosen because it was assumed that the effect of treatment would drop off rapidly after administration and then more slowly as time went on (instead of a linear, or straight-line, decay in effect).

No. of subjects = 319
No. of failures = 264 Number of obs = 319
Time at risk = 25018 LR $\chi^2_{(2)}$ = 0.51
Log likelihood = -1307.22 Prob > χ^2 = 0.005

Predictor	HR	SE	Z	P	95% CI	
main effect						
tx	3.085	1.102	3.15	0.002	1.532	6.211
ln(time) interaction effect						
tx	0.759	0.072	-2.92	0.003	0.631	0.913

Treatment is now a significant predictor of time to conception. The treatment*ln(time) interaction term is also significant, confirming that the effect of treatment does vary with time (*ie* the proportional hazards assumption does not hold). In the presence of interaction, the effect of treatment can be better understood by computing the *HR* at a number of time points. The *HR* at time *t* is $3.08*0.759^{\ln(t)}$.

Time (days)	ln(time)	HR
1.0	0	3.08
2.7	1	2.34
7.4	2	1.77
20.1	3	1.35
54.6	4	1.02
148.4	5	0.78

The effect of treatment drops off until by day 55, it has completely disappeared.

useful for those variables. This graphical assessment can be enhanced by adding a smoothing line to indicate the overall trend. The residuals should hover around the ‘zero’ line, indicating no trend in the residuals over time. If the residuals trend up or down, it suggests that the effect of the predictor is varying over time. Fig. 19.17 (in Example 19.11) shows a plot of the scaled Schoenfeld residuals for lactation against ln(time). The assumption of proportional hazards appears to be reasonable for this predictor.

Schoenfeld residuals also form the basis of a statistical test of the assumption of proportional hazards. The test checks for a non-zero slope of the scaled Schoenfeld residuals against time (or a function of time) using a generalised linear regression. It provides an overall assessment and a test for each predictor separately. Results of this test for the prostaglandin data are presented in Example 19.11. These suggest that a treatment*time interaction term does need to be added to the model.

Example 19.11 Assessing the proportional hazards assumption—Schoenfeld residuals
data = pgtrial

A Cox model with herd, treatment, lactation and body condition (-thin-) as predictors was fit to the prostaglandin data (without any time-varying covariates). Schoenfeld and scaled Schoenfeld residuals were obtained. Fig. 19.17 shows a smoothed plot of scaled Schoenfeld residuals for lactation plotted against time on a log scale.

The statistical test for non-zero slope for each of the predictors (against ln(time)) resulted in the following.

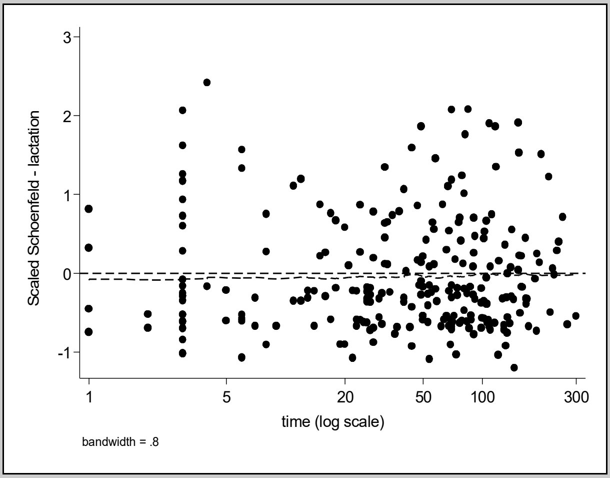


Fig. 19.17 Schoenfeld residuals for lactation

	χ^2	df	prob> χ^2
herd=2	0.34	1	0.559
herd=3	0.09	1	0.760
tx	7.65	1	0.006
lact	0.28	1	0.594
thin	1.81	1	0.179
global test	10.43	5	0.064

While the global test was borderline significant, it is clear that the assumption of proportional hazards was violated for treatment.

19.8.11 Evaluating the assumption of independent censoring

One of the fundamental assumptions of survival models is that censoring is independent of the outcome of interest. This means that censored animals should have the same future survival expectation as non-censored animals (*ie* if the animals were not censored, they would have the same survival distribution as the non-censored animals). There are no specific tests to evaluate the independence of censoring and the event of interest. However, sensitivity analyses can be used to look at the extreme situations of complete positive or negative correlations between censoring and the event of interest.

Complete **positive correlation** would mean that every animal that was censored would have experienced the event of interest immediately if it had not been censored. This could be evaluated by refitting the model after recoding all of the censored observations so that they had the event of interest instead of being censored (at the time of censoring).

Complete **negative correlation** would mean that every animal that was censored would be guaranteed a long 'event-free' existence if it had not been censored. This could be evaluated by refitting the model after changing each censored animal's time at risk to a large, but plausible, value.

The above 2 analyses would provide the possible range of values that the coefficients of the factors of interest could possibly take if the assumption of independent censoring was badly violated. If gross violation of this assumption does not drastically alter the estimates of the parameters of interest, you can be confident that the actual bias in the parameter estimates will be small.

Example 19.12 presents the results of a sensitivity analysis designed to evaluate this assumption in the prostaglandin data.

19.8.12 Evaluating the overall fit of the model

Four approaches to evaluating the overall fit and predictive ability of the model are:

- to evaluate graphically the distribution of the Cox-Snell residuals,
- to use a goodness-of-fit test similar to the Hosmer-Lemeshow test used for logistic regression
- to evaluate concordance between the predicted and observed sequence of pairs of events, and
- to compute an overall r^2 statistic

Example 19.12 Evaluating the assumption of independence of censoring
data = pgtrial

A Cox model with herd, treatment, lactation number, and body condition (-thin-) as predictors was fit to the prostaglandin trial data (with treatment as a time-varying effect on the ln(time) scale). The model was then refit assuming complete positive and complete negative correlations between censoring and conception (see text for description of method). Negative correlation was based on assigning -dar- of 400 to all censored cows. The results are summarised in the following table.

Variable	Original estimate	Assuming complete positive correlation	Assuming complete negative correlation
herd=2	-0.260	-0.199	-0.228
herd=3	0.023	-0.007	0.008
tx	1.089	0.983	0.927
lact	-0.043	-0.006	-0.061
thin	-0.145	-0.141	-0.050
tx*ln(time)	-0.259	-0.209	-0.215

Both sensitivity analyses resulted in a small reduction in the coefficient for treatment, but the changes were not large and the treatment effect remained highly significant (P-values not shown).

Cox-Snell residuals are the estimated cumulative hazards for individuals at their failure (or censoring) times. If the model is appropriate, these residuals are a censored sample from a unit exponential distribution (*ie* an exponential distribution with a mean of one and variance of 1). Consequently, the range of these residuals is zero to +∞. Cox-Snell (CS) residuals can be used to assess the overall fit of a proportional hazards model by graphically assessing how close these residuals are to having a unit exponential distribution. To do this, you:

- compute the CS residual
- fit a new proportional hazards model with the CS residuals used as the 'time' variable (along with the original censoring variable)
- derive an estimate of the cumulative hazard function ($H(t)$) from this new model
- plot $H(t)$ against the CS residuals

If the residuals have a unit exponential distribution, the cumulative hazard should be a straight line with an intercept of 0 and a slope of 1. In practise, these graphs have been of limited value. Assessment of the linearity of the graph is a very subjective procedure and substantial apparent departures from the 45° line can result from a few observations with long survival times (when most of the observations are clustered at the lower left end of the line) (see Example 19.13).

For censored observations, the estimated cumulative hazard is an underestimate of the true

Example 19.13 Evaluating overall fit of a model

data = pgtrial

A Cox proportional hazards model was fit to the data with fixed effects for herd, treatment, lactation number and body condition (-thin-) (treatment effect was not time-varying).

Cox-Snell residuals were computed and plotted as described in the text (Fig. 19.18). It appears that there is relatively good agreement between the plotted values and the expected (45°) line.

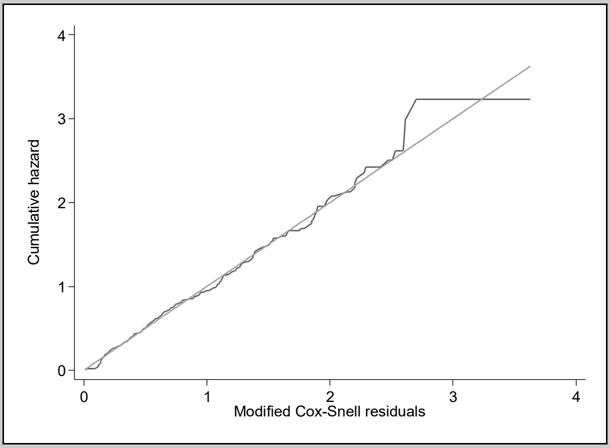


Fig. 19.18 Plot of Cox-Snell residuals

Goodness-of-fit tests

The Grønnesby and Borgan omnibus goodness-of-fit test produces a P-value of 0.34 (no evidence of lack of fit), while the Moreau, O'Quigley and Mesbah test, designed specifically for detecting non-proportional hazards, generates a P-value of 0.004 (significant lack of fit).

Concordance

Harrell's C statistic was 0.56 indicating that the model only correctly predicts the sequence of 2 observed failures 56% of the time (*ie* very limited predictive ability).

r^2

The r^2 for the Cox model with -herd-, -tx-, -lact-, and -thin- as predictors produced an estimated r^2 of 0.022 (2.2%) with a bootstrapped 95% confidence interval of (0.009, 0.065). Clearly, collectively these predictors have relatively little ability to predict exactly when a cow is going to conceive.

cumulative hazard for an individual (by virtue of the fact that we don't observe them for the full period until they have the outcome of interest). Consequently, Cox-Snell residuals are sometimes modified by the addition of a constant (either 1 or ln(2)=0.693) for censored observations. There is no evident rationale for choosing one adjustment over the other, but this is only important if a substantial proportion of the observations are censored.

Several **goodness-of-fit** tests similar to a Hosmer-Lemeshow test for logistic regression models can be computed (May & Hosmer, 2004b). These tests all divide the data into groups and add indicator variables for these groups to the models and assess the overall significance of the indicator variables, with significance indicating evidence of lack-of-fit. An omnibus test designed to detect all causes of lack of fit was proposed by Grønnesby and Borgan (1996). The observed number of failures in groups defined by quantiles of risk from the fitted model are compared to the expected number of failures which are based on martingale residuals (see Section 19.8.13). However, the validity of the test depends on choosing an appropriate number of groups (May & Hosmer, 2004a). The number of groups should roughly equal the number of failure events in the data divided by 40, with a minimum of 2 and a maximum of 10. Using this strategy the test has reasonable power provided the sample size is greater than 200 with no more than 50% censoring (in smaller samples, the power is low). However, this test fails to identify the problem of non-proportional hazards in the prostaglandin trial data (Example 19.13). An alternative test designed specifically to evaluate the proportional hazards assumption was proposed by Moreau *et al* (1985). It requires the computation of time-dependent indicator variables and it successfully detects the problem of non-proportional hazards (Example 19.13). These tests should not be used in situations in which there are time-varying covariates in the model.

Closely related to the issue of evaluating overall fit is the question of evaluating overall predictive ability. **Harrell's C concordance statistic** computes the proportion of all pairs of subjects in which the model correctly predicts the sequence of events (*ie* which one would have come first). It ranges from 0 to 1 with a value of 0.5 indicating no predictive ability at all (you would expect to get 50% correct by chance alone).

For a linear regression model, we would use $\boldsymbol{r^2}$ as a measure of predictive ability. Recently, Royston (2006) described several possible measures of explained variation for survival models and proposed an r^2 statistic for proportional hazard models. Comparable to the adjusted r^2 from linear regression, it is also possible to adjust the proposed r^2 for the number of predictors in a survival model. The r^2 compares a fitted model with a null model and provides an estimate of the amount of variation in survival times that is explained by the predictors. However, it can not be used to compare models with different hazard structures (*eg* a semi-parametric Cox model with a parametric Weibull model—see Section 19.9 for hazard structures in parametric models) because the null models are different. An estimate of the r^2 for the prostaglandin data, and bootstrap 95% confidence intervals are shown in Example 19.13.

19.8.13 Evaluating the functional form of predictors

Martingale residuals can be used to evaluate the functional form of the relationship between a continuous predictor and the survival expectation for individuals. These residuals represent the difference between the observed final outcome for an individual and the cumulative hazard for that individual at the final point in time. (As such, they are more like typical residuals which represent a difference between an observed and a predicted value). Because they are based on

Example 19.14 Evaluating functional form of predictors
data = pgtrial

Fig. 19.19 shows a lowess smoothed graph of martingale residuals against lactation number. It appears that a linear relationship may not be appropriate. To further evaluate this possibility, a model was fit with lactation included as both a linear and a quadratic term. Both the linear (*β*=-0.124) and quadratic (*β*=0.046) terms were significant at P=0.03. This confirms that the effect of lactation number on time to conception is not linear.

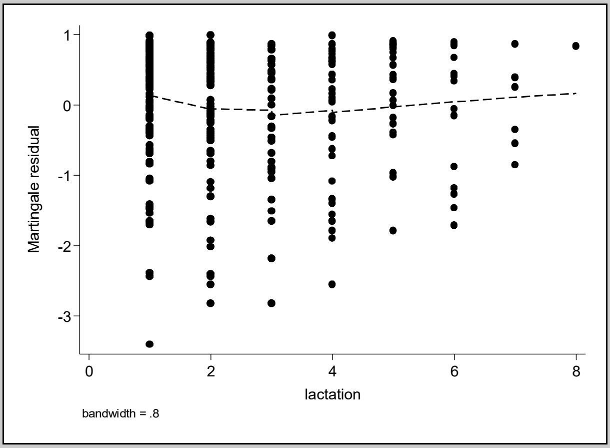


Fig. 19.19 Plot of martingale residuals vs lactation number

the estimated cumulative hazard, these residuals are similar to Cox-Snell residuals except their range is from -∞ to 1. The values of these martingale residuals are:

- uncensored observation *i*: $1-\hat{H}_i(t_i)$
- censored observation *i* : $0-\hat{H}_i(t_i)$

Consequently, residuals will be negative for all censored observations and for observations in which $H(t_i)>1$ (equivalent to $S(t_i)<0.37$).

To check for the functional form of continuous predictors, martingale residuals should be computed from a model which does not include the continuous predictor of interest. These residuals are then plotted against the predictor. A smoothing function (*eg* kernel smoothing) can be used to better visualise the relationship. If the relationship is linear, the smoothed martingale residual line should be approximately straight. Fig. 19.19 (Example 19.14) shows a kernel smoothed graph of martingale residuals against lactation number.

19.8.14 Checking for outliers

Deviance residuals can be used to identify outliers (*ie* points that are not well fit by the model). Deviance residuals are martingale residuals that have been rescaled so they are symmetric around 0 (if the fitted model is appropriate). The sum of the squared deviance residuals is the deviance (*D*) of the model.

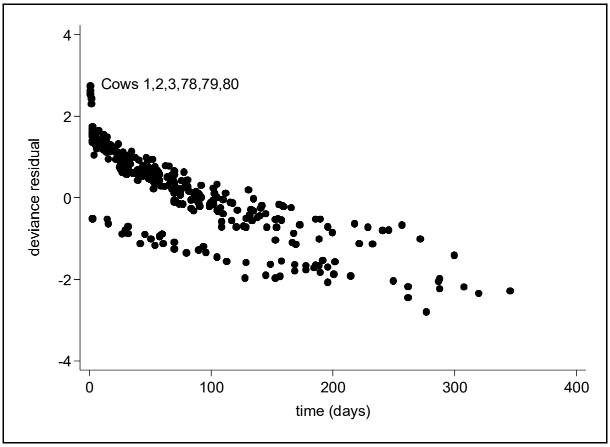


Fig. 19.20 Deviance residuals

If plotted with an observation number as the plotting symbol, they can be used to identify outlying observations. Fig. 19.20 is a plot of deviance residuals from the model without -tx- as a time-varying effect. The cluster of large positive residuals at the top left are residuals from 6 cows (1,2,3,78,79,80) that conceived on day 1 or day 2 (before the large block of cows that conceived on day 3). The cumulative hazard was low on days 1 and 2 because relatively few cows conceived on those days (relative to the large pool of cows 'at risk' of conception). Hence, for any cow that did conceive, the martingale and deviance residuals were 'large'.

19.8.15 Detecting influential points

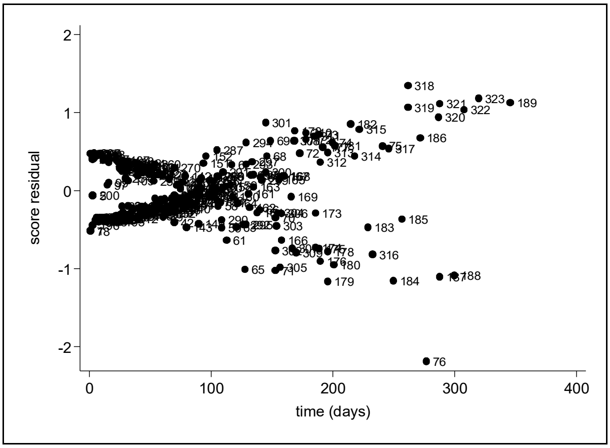


Fig. 19.21 Score residuals

Score residuals and scaled score residuals can be used to identify influential observations. The former have a 'leverage like' property while that latter measure the impact of an observations on coefficients in the model.

Score residuals are a variation of martingale residuals but are computed for each predictor (covariate) in the model. They have a 'leverage-like' property in that observations that are far from the mean of the predictor have larger (positive or negative) residuals. When plotted against time, they typically form a 'fan-shaped' pattern (with the centre of the fan at the mean of the time variable) and observations lying outside this 'fan' should be considered as potentially influential. Fig. 19.21 shows score residuals for treatment—cow 76 was a treated, 1st lactation cow in herd 1 that was censored at 277 days. (She was the only cow left in herd 1 at this time and hence had high leverage).

Score residuals can be modified to compute a **delta-beta** like parameter for coefficients in the model. This modification involves multiplying the score residual by the estimated variance of the coefficient (from the variance-covariance matrix of the coefficients) and produces what is called a **scaled score residual**. Fig. 19.22 shows a plot of the scaled score residuals for treatment against time. Once again, cow 76 warrants some further investigation. The main effect of this cow is to reduce the estimated treatment effect (determined by refitting the model without cow 76—results not shown).

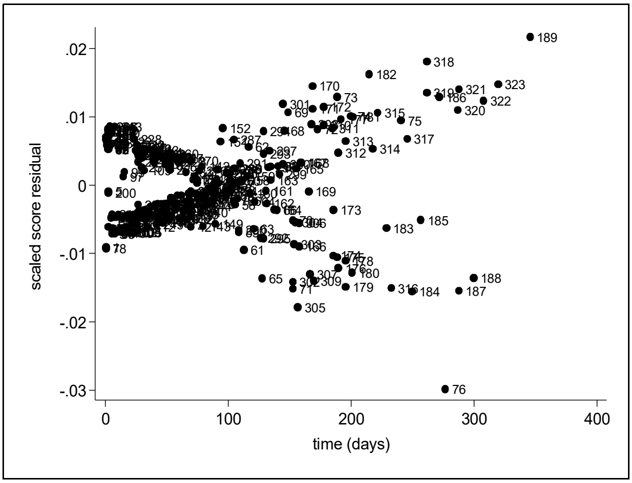


Fig. 19.22 Scaled score residuals (delta-beta)

19.9 Parametric models

As noted previously, Cox proportional hazards models make no assumption about the shape of the baseline hazard, which can be a real advantage if you have no idea what that shape might be, or if it has a very irregular form. However, these models achieve this flexibility at a price. Because they only use information about the observations at times at which one or more of the subjects fail, they do not efficiently use all of the information you have about the observations. For example, because the Cox model is based solely on the rank ordering of the observations, it makes no difference if 2 successive failures are one day apart or one year apart. The length of the interval, which provides some valuable information in terms of survival times, is ignored. Consequently, if you can correctly specify the form of the baseline hazard, a parametric model will be more efficient (*ie* use more of the available information).

A parametric model satisfying the proportional hazards assumption could be written in the same way as a semi-parametric model:

$$h(t)=h_0(t)\mathrm{e}^{\beta X}$$

but $h_0(t)$ is assumed to have a specified functional form. The major difference is that βX now includes an intercept term (β_0). (An alternative method of writing these models is described in Section 19.10).

Not all parametric models are proportional hazards models. (Models which are not are discussed in Section 19.10). Three parametric models which are proportional hazards models are the exponential, Weibull and Gompertz. Each of these will be discussed briefly. When using any of these models, it must be kept in mind that in addition to specifying a correct function for the baseline hazard, the assumption of proportional hazards must also be evaluated and met.

19.9.1 Exponential model

An exponential model is the simplest form of parametric model in that it assumes that $h_0(t)$ is constant over time (*ie* in the baseline group, the rate at which failures are occurring remains constant). Consequently

$$h(t)=\lambda=c(\mathrm{e}^{\beta X}) \qquad \textit{Eq 19.17}$$

where c is the constant baseline hazard and λ is the time-constant value of $h(t)$ for any given set of predictor values. The density and survivor functions of the exponential distribution were given in Eq 19.11. As noted previously, the survival times will have a decreasing exponential distribution.

Interpretation of coefficients

Coefficients for predictors in an exponential model can be interpreted the same way as coefficients from a Cox model. The exponentiated coefficient is the **hazard ratio** (Section 19.8.2). The intercept in the model is the estimate of the log of the (constant) baseline hazard. In Example 19.15, an exponential model is fit to the prostaglandin data. If this model was appropriate (which it isn't—but more on that later), the baseline hazard would be estimated to be $e^{-4.41}$=0.012. That is, on any given day, a cow in the baseline group which had not previously conceived had a 1.2% chance of conceiving on that day.

Evaluating the assumption of constant hazard

The assumption that the baseline hazard is constant over time can be evaluated in several ways. The first is to generate an estimate of the baseline hazard from a Cox model and graph it to see if it approximately follows a straight, horizontal line. Fig. 19.13 showed that the baseline hazard rose up to day 100 and then fell gradually over time. A second approach is to fit a model with a piecewise-constant baseline hazard (Dohoo *et al*, 2003). In this case, the baseline hazard is allowed to vary across time intervals by including indicator variables for each of the time intervals in the model. The baseline hazard is assumed to be constant within each time period, but can vary between time periods. This produces the results and step graph shown in Example 19.17. In general it appears that the hazard falls over time, but the pattern is not clear in early lactation (prior to day 50). However, a model which assumed that the hazard declined in a curved manner, might be a reasonable approximation. A third approach to evaluating the assumption of constant hazard is to evaluate the shape parameter from a Weibull model (see Section 19.9.2).

Example 19.15 Exponential regression

data = pgtrial

An exponential survival model was fit to the prostaglandin data after rescaling -lact- (by subtracting one) so that first lactation animals had a value of 0.

No. of subjects = 319
No. of failures =264 Number of obs = 319
Time at risk = 25018 LR $\chi^2_{(5)}$ = 11.42
Log likelihood = -528.4 Prob > χ^2 = 0.0437

Predictor	Coef	SE	Z	P	95% CI	
herd=2	-0.315	0.169	-1.86	0.063	-0.647	0.017
herd=3	0.038	0.175	0.21	0.830	-0.306	0.381
tx	0.218	0.125	1.74	0.083	-0.028	0.464
lact	-0.042	0.041	-1.01	0.314	-0.123	0.039
thin	-0.157	0.138	-1.14	0.255	-0.428	0.114
constant	-4.405	0.161	-27.28	0.000	-4.721	-4.088

The *HR* for treatment would be $e^{0.218}$ = 1.24, suggesting that, at any given point in time, a treated cow was 1.24 times more likely to conceive than a non-treated cow (if this model were correct).

19.9.2 Weibull model

In a Weibull model, it is assumed that the baseline hazard function has a shape which gives rise to a Weibull distribution of survival times. The Weibull hazard was discussed in Section 19.7.7 and shown graphically in Fig. 19.10. In addition, Eq 19.13 gives the formulae for the hazard and survivor functions.

If a vector of covariates (predictors) is added to a Weibull model, the formula for the hazard function becomes:

$$h(t)=\lambda p t^{p-1} e^{\beta X} \qquad \textbf{\textit{Eq 19.18}}$$

where βX does not include an intercept term (β_0). Example 19.16 shows a Weibull model fit to the prostaglandin data. The estimate of the shape parameter (p) is 0.867 (95% CI: 0.784, 0.958) suggesting that the hazard is decreasing over time, but at a relatively slow rate (p close to 1).

Evaluating the Weibull distribution
As was noted previously, the suitability of the assumption that the survival times follow a Weibull distribution can be assessed by generating a log-cumulative hazard plot. If the distribution is Weibull, this graph will show as a straight line. A rough evaluation can be obtained by generating a simple plot of ln$H(t)$ vs ln(t) for all of the data. Fig. 19.15 shows a plot of -ln$H(t)$ vs ln(t) for each of the 2 treatment groups in the prostaglandin data. The baseline hazard will be included in the non-treated group and that line was approximately straight

Example 19.16 Weibull model
data = pgtrial

A Weibull model was fit to the prostaglandin data.

```
No. of subjects = 319
No. of failures = 264                                   Number of obs   319
Time at risk = 25018                                    LR chi2 (5) = 9.96
Log likelihood = -524.2                               Prob > chi2 = 0.0764
```

Predictor	Coef	SE	Z	P	95% CI	
herd=2	-0.289	0.169	-1.71	0.088	-0.621	0.043
herd=3	0.039	0.175	0.22	0.825	-0.304	0.381
tx	0.205	0.125	1.63	0.102	-0.041	0.450
lact=2+	-0.041	0.041	-1.01	0.315	-0.122	0.039
thin	-0.136	0.138	-0.99	0.324	-0.406	0.134
constant	-3.790	0.259	-14.64	0.000	-4.297	-3.282
/ln_p	-0.143	0.051	-2.80	0.005	-0.243	-0.043
p	0.867	0.044			0.784	0.958
1/p	1.154	0.059			1.044	1.275

The treatment effect is similar to that seen in the exponential, piecewise exponential, and Cox models and is similar in terms of statistical significance. The shape parameter (p) from the Weibull distribution indicates that the hazard is falling with time (*ie* p<1).

Example 19.17 Piecewise constant exponential regression model

data = pgtrial

A model which allows the baseline hazard to vary between time periods, but forces it to remain constant within time periods is called a piecewise constant exponential model. Results from such a model, and a graph of the resulting baseline hazard are shown below.

```
No. of subjects = 319
No. of failures = 264                                  Number of obs = 1725
Time at risk = 25018                                    LR chi2(9) = 16.74
Log likelihood = -525.7                               Prob > chi2 = 0.0529
```

Predictor	Coef	SE	Z	P	95% CI	
day21_40	-0.377	0.195	-1.940	0.053	-0.759	0.005
day41_80	-0.310	0.171	-1.820	0.069	-0.645	0.025
day81_120	-0.238	0.195	-1.220	0.223	-0.619	0.144
day121+	-0.416	0.192	-2.170	0.030	-0.792	-0.041
herd=2	-0.295	0.170	-1.730	0.083	-0.628	0.038
herd=3	0.040	0.175	0.230	0.820	-0.303	0.383
tx	0.211	0.125	1.680	0.092	-0.035	0.457
lact	-0.041	0.041	-1.000	0.318	-0.122	0.040
thin	-0.145	0.138	-1.050	0.294	-0.416	0.126
constant	-4.164	0.188	-22.180	0.000	-4.532	-3.796

The coefficients for predictors day21_40 through day121+ show how the log hazard changes relative to the value for the baseline time period (days 1 to 20).

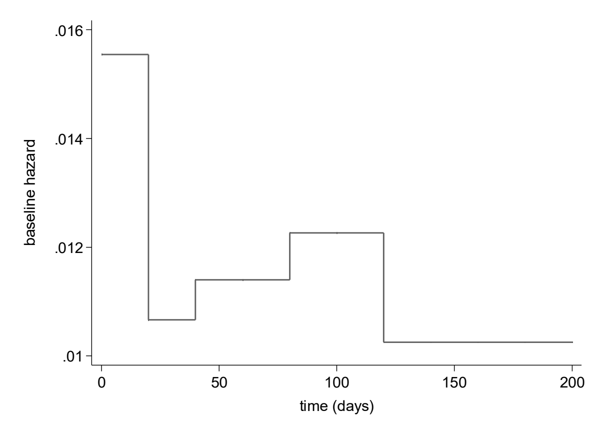


Fig. 19.23 Piecewise constant estimate of baseline hazard

suggesting that the Weibull model might be appropriate. The step graph of the baseline hazard (Fig. 19.23 in Example 19.17) however, suggests that the Weibull model, although preferable to the exponential, might not be ideal because the hazard appears to initially fall, then rise until about day 120 and then fall again. A Weibull model with a shape parameter of 0.87 assumes the hazard monotonically decreases over time.

19.9.3 Gompertz model

The Gompertz model is used less frequently than the exponential and Weibull models but has been used to model mortality data. In a Gompertz model, the log of the baseline hazard varies linearly with time so the baseline hazard is as follows.

$$h_0(t)=\lambda e^{pt} \qquad \textit{Eq 19.19}$$

The baseline hazard increases exponentially if p>0 and decreases exponentially if p<0. If p=0, the hazard is constant at λ (exponential model). A Gompertz model fit to the prostaglandin data (results not shown) produces an estimate of p of -0.002 (95% CI: -0.004, 0) which also suggests that the hazard is falling with time, but slowly.

19.10 ACCELERATED FAILURE TIME MODELS

As noted above, not all parametric models are proportional hazards models. However, those that are can be written in one of 2 ways: as a proportional hazards model (which is what has been presented thus far) or as an accelerated failure time model (AFT). Other parametric models (discussed below) can only be written in the AFT metric, because the predictors in these models do not necessarily multiply the baseline hazard by a constant amount.

The general form of an AFT model is:

$$\ln t=\beta X+\ln\tau \qquad \text{or} \qquad t=e^{\beta X}\tau \qquad \textit{Eq 19.20}$$

where ln$t$ is the natural log of the time to the failure event, βX is a linear combination of explanatory variables and lnτ is an error term with an appropriate distribution. **Note** The values of the βs in this representation will not be the same as the βs in a proportional hazards representation.

From Eq 19.20 it can be seen that τ is the distribution of survival times when βX=0 (*ie* $e^{\beta X}$=1). τ is assumed to have a specific distribution (*eg* Weibull, log-normal). If τ has a log-normal distribution, then the log of survival times will have a normal distribution which is equivalent to fitting a linear model to ln(survival times) (assuming you can ignore the problem of dealing with censored observations). Three specific distributions of survival times (log-logistic, log-normal and generalised gamma) are discussed in Section 19.10.2.

Eq 19. 20 can be rearranged as follows:

$$\tau=e^{-\beta X}t \qquad \text{or} \qquad \ln(\tau)=-\beta X+\ln(t) \qquad \textit{Eq 19.21}$$

The linear combination of predictors in the model (βX) act additively on log(time) or multiplicatively on time (*ie* they accelerate or decelerate the passage of time by a multiplicative factor) where $e^{-\beta X}$ is called the **acceleration parameter** because if:

- $e^{-\beta X}$>1, then t< τ so time passes more quickly (*ie* failures expected sooner)
- $e^{-\beta X}$ =1, then t= τ so time passes at a 'normal' rate (*ie* no effect of predictors)
- $e^{-\beta X}$<1, then t> τ so time passes more slowly (*ie* failures expected later)

As indicated above, the exponential and Weibull models can be written either as proportional hazards models or as AFT models. The relationship between the coefficients from a proportional hazards expression (β_{ph}) of a Weibull model and an AFT expression (β_{aft}) is:

$$\beta_{aft}=\frac{-\beta_{ph}}{p}$$ ***Eq 19.22***

where p is the shape parameter from the Weibull model.

19.10.1 Coefficients in AFT models

A coefficient in an AFT model represents the expected change in the ln(survival time) for a 1-unit change in the predictor. For example, assume you have a dichotomous predictor (X with a coefficient of 2). If, in the absence of X, a study subject is expected to fail at t=5 days (ln(t)=1.61), the presence of X would increase the expected ln(survival time) to 1.61+2=3.61 or the survival time to 37 days. The presence of X in a subject which was expected to survive 30 days would result in an increase expected survival time from 30 to 222 days. As you can see, in absolute time, factors have a greater impact at longer expected survival times.

An alternative interpretation is to exponentiate the coefficient to compute a **time ratio** (*TR*). A coefficient of 2 produces a *TR* of e^2=7.4 which means that the presence of X increases the expected survival time by a factor of approximately 7 times.

19.10.2 Specific survival time distributions

Log-logistic model

A log-logistic model assumes that survival times follow a log-logistic distribution, or alternatively, log survival times follow a logistic distribution (a symmetric distribution similar to a normal distribution). The hazard function for a log-logistic distribution is as follows.

$$h(t)=\frac{e^{\theta}}{\gamma t(t^{-1/\gamma}+e^{\theta})}$$ ***Eq 19.23***

where $\gamma > 0$ is a scale parameter. $h(t)$ decreases as a function of t if $\gamma >1$, otherwise it is increasing and then decreasing with a peak at:

$$t=\left(\frac{(1/\gamma)-1}{e^{\theta}}\right)^{\gamma}=e^{-\theta\gamma}((1/\gamma)-1)^{\gamma}$$ ***Eq 19.24***

In the log-logistic model $-\theta\gamma$ is modelled as a function of the predictors (*ie* $-\theta\gamma=\beta X$). The p^{th} percentile (and median) of a log-logistic distribution are the following.

$$t_p=\left(\frac{p}{100-p}\right)^{\gamma}e^{-\theta\gamma} \qquad t_{50}=e^{-\theta\gamma}$$ ***Eq 19.25***

Fig. 19.24 shows hazard functions for various values of γ (left panel) and a histogram of log-logistic distributed survival times (based on 2,000 simulated observations) when γ=0.25. (In all cases the median survival time is set to 20 days.)

An example of a log-logistic survival model expressed in AFT terms is shown in Example 19.18. Because log-normal or log-logistic models can rise and then fall, but not the opposite, neither may be appropriate for the apparent shape observed in Fig. 19.23 (falling and then rising). However, the example is provided for pedagogic purposes.

Example 19.18 Log-logistic model of prostaglandin data
data = pgtrial

A log-logistic model was fit to the prostaglandin data and produced the following.

Predictor	Coef	SE	Z	P	95% CI	
herd=2	0.254	0.236	1.08	0.281	-0.208	0.715
herd=3	-0.102	0.244	-0.42	0.676	-0.579	0.376
tx	-0.387	0.177	-2.19	0.029	-0.733	-0.040
lact	0.061	0.055	1.11	0.266	-0.047	0.169
thin	0.040	0.189	0.21	0.833	-0.331	0.411
constant	4.016	0.225	17.87	0.000	3.575	4.456
/ln_gam	-0.126	0.052	-2.45	0.014	-0.227	-0.025
gamma	0.882	0.045			0.797	0.975

Gamma equals 0.882 which would correspond to a distribution with a peak at 9.4 days in the baseline group (calculated as $e^{4.016}((1/0.882)-1)^{0.882}$) or 6.4 days in the treated group.

The time ratio for -tx- was $e^{-0.387}$=0.68 which suggests that, on average, time to conception in treated cows was 68% of what it would be in non-treated cows. Using Eq 19.25, the median survival time of the baseline group is:

$$t_{50} = e^{\text{constant}} = e^{4.016} = 55.5 \text{ days.}$$

Log-normal model
In a log-normal model the survival times are distributed normally on a log time scale, or alternatively, log times are distributed normally. The survivor function is:

$$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \qquad \textit{Eq 19.26}$$

where *Φ* is the cumulative distribution function of a standard normal (Gaussian) distribution and *μ* and *σ* are the mean and standard deviation of log survival times. (The formulae for the hazard functions for the log-normal and generalised gamma distributions can be derived from *f*(*t*) and *S*(*t*), but are complex and beyond the scope of this text. See Cleves *et al* (2008); Collett, 2003) for details.

Generalised gamma model
A generalised gamma distribution is a 3-parameter (*μ, κ, σ*) distribution for which the hazard function can take a wide variety of shapes which include the Weibull, log-normal and gamma distributions. Consequently, it is particularly useful for evaluating the shape of the hazard function (see Section 19.10.3).

19.10.3 Choosing a parametric model

Selecting an appropriate parametric model involves both biological and statistical procedures. The selection should be guided by knowledge of how failures arise and insights into what we would be expected in terms of a hazard function.

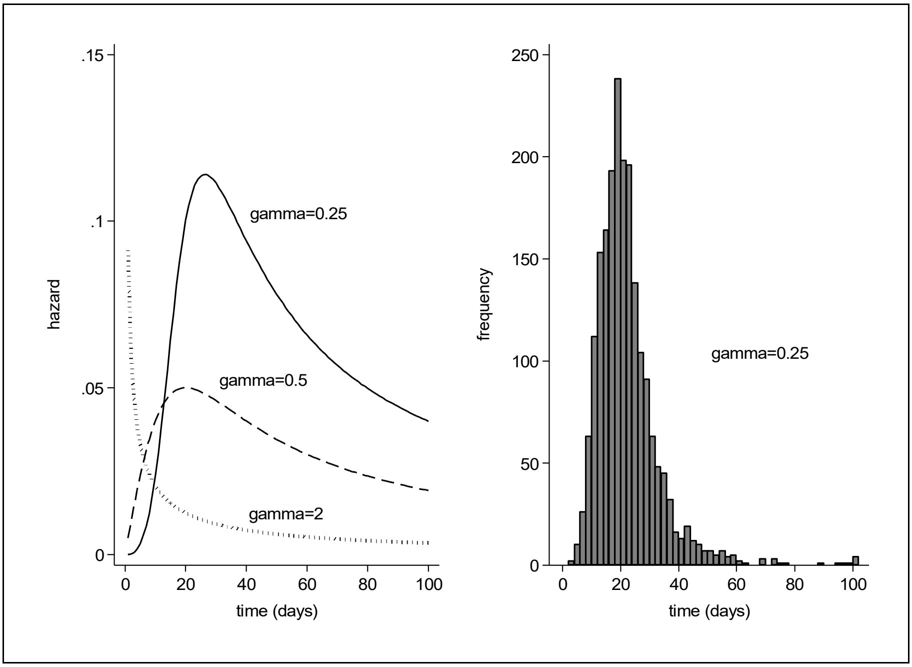


Fig. 19.24 Hazard functions (left) and survival times (right) for a log-logistic distribution

As noted, the generalised gamma distribution provides some insight into what might be an appropriate distribution.

- if κ=1, the distribution is Weibull and $\sigma = 1/p$ is the inverse shape parameter
- if κ=1 and σ =1, the distribution is exponential
- if κ=0, the distribution is log-normal.

For the prostaglandin data, a generalised gamma model produces estimates of κ=1.12 (95% CI: 0.61, 1.64) and σ=1.10 (95% CI: 0.90, 1.36). Both are numerically different from one, but their confidence intervals include one. This suggests that an exponential model may be adequate. Example 19.19 shows the log-likelihood for each of the 5 parametric models, along with the number of distribution parameters and the point estimate of the effect of -tx-.

19.11 Frailty models and clustering

As noted in previous sections, predictors in survival models (semi-parametric and parametric) act multiplicatively on the baseline hazard (*ie* the hazard for an individual is a multiple of the baseline function). In a frailty model, an additional latent (unobserved) effect (*ie* the frailty) acts multiplicatively on the hazard. The frailty is not measured directly, but is assumed to have a specified distribution and the variance of the distribution is estimated from the data.

There are 2 general types of frailty model: individual frailty and shared frailty (Gutierrez, 2002). In an individual frailty model, the additional variance is unique to individuals and serves to account for additional variability in the hazard among individuals in much the same way that the negative binomial model accounts for more variability than a Poisson model. Shared frailty

Example 19.19 Comparison of parametric models
data = pgtrial

Five parametric models were fit to the prostaglandin data and compared.

		# Parameters			
Model	Log L	Distribution	Predictors	AIC	Time ratio for -tx-
exponential	-528.4	1	5	1068.7	0.80
Weibull	-524.2	2	5	1062.3	0.79
log-logistic	-535.7	2	5	1085.5	0.70
log-normal	-533.5	2	5	1081.0	0.58
generalised gamma	-524.1	3	5	1064.1	0.80

The while the generalised gamma model fits the best (largest log L) the AIC suggests that the Weibull is a suitable alternative. The shape parameter from the Weibull model is 0.87 (roughly equivalent to the corresponding 1/κ= 0.89 from the generalised gamma) and it has a 95% CI of 0.78, 0.96 which suggests that it is different from 1. The Gompertz model is not shown because a time ratio cannot be computed from this model, but it had a log L of -526.8 and an AIC of 1067.5, suggesting that it is inferior to the Weibull model.

models constitute one approach to dealing with clustered data and are discussed starting in Section 19.11.3

19.11.1 Individual frailty models

Within a population described by an average hazard *h*(*t*), some individuals fail early and some late. This variation in survival time may be attributed to 3 components. Part may be due to differences among individuals in terms of measured covariates and this variability will be removed by including those covariates in the model. Part may be due to unmeasured covariates which make some individuals more prone to fail early (*ie* 'frail' individuals). The final part is that attributable to random variation and is explained by the survival time distribution that is selected. The effect of frailty (unmeasured covariates) can be thought of as overdispersion—more variability in the survival times than would be expected based on the chosen distribution.

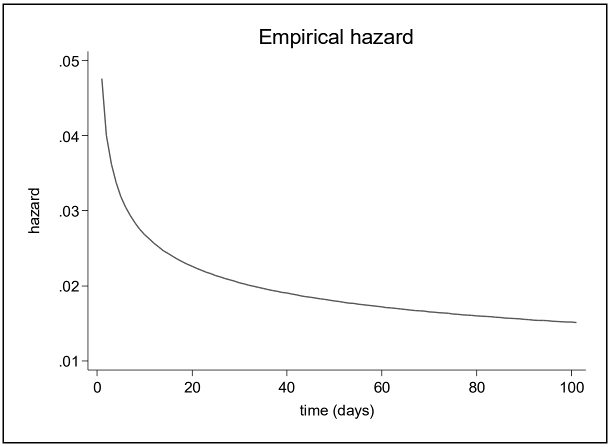


Fig. 19.25 Effects of individual frailties (see text for explanation)

The effect of individual frailty can be seen in Fig. 19.25 which shows the empirical hazard for a population of 2,000 individuals. All individuals in this population had a constant hazard set to 0.05 (exponential model), but individuals were assigned individual frailties (gamma distribution, μ=1, σ=1) which made some individuals more prone to fail than others. The graph shows the estimated hazard from a Weibull regression model (shape parameter p=0.75). Although every individual had a constant hazard, the average hazard for the population clearly falls as the frail individuals fail and the remaining population increasingly consists of more robust individuals.

An individual frailty model can be written as follows:

$$h(t|\alpha)=\alpha h(t) \qquad \textit{Eq 19.27}$$

Conditional on the frailty, the hazard at any point is multiplied by a factor (variable) α, which is assumed to have a distribution with a mean of 1 and a variance of θ. Two commonly assumed distributions of α are the gamma and the inverse Gaussian.

A frailty effect can account for apparent changes in the hazard in a population over time. A Weibull model fit to the data used to create Fig. 19.25 has a shape parameter of 0.76 suggesting that the hazard is falling over time. If the model is refit with a gamma frailty added, the shape parameter changes to 1.3, suggesting that, for individuals with the same frailty, the hazard is actually rising. It is impossible to separate individual frailty effects from the distributional assumptions of the model, so in practice, individual frailties have limited applicability unless the expected distribution of survival times is known with certainty (O'Quigley & Stare, 2002).

Example 19.20 shows the addition of a gamma frailty to the Weibull model of the prostaglandin data (with no time-varying predictors).

The concept of individual frailty does not apply to Cox (semi-parametric) models because the frailty effect represents variation in survival times in excess of what would be expected from the assumed distribution of survival times. However, in a Cox model, there is no assumed distribution of survival times. Any 'overdispersion' would be incorporated into the baseline hazard (h_0) which has no specified form.

19.11.2 Clustering in survival data

Animals within a group or cluster (*eg* cows within a herd) have features in common (*eg* housing, feed) that lead to a lack of independence among animals within a cluster and could result in more similar survival times (*eg* cows on one herd may have a longer lifespan than cows in another herd). The general problem of clustering is covered in Chapters 20 through 24. However, in terms of survival models, there are several approaches that can be used to deal with clustered data. If the number of clusters is limited, fixed effects representing the clusters can be included in the model. Stratified models (Section 19.8.7), in which the strata are the clusters, can also be used to address the issue of clustering, but like the use of fixed effects models, preclude the evaluation of cluster level predictors. Robust standard errors (Chapter 20) are a general approach that can be used to address the problem of lack of independence in many types of model, but have some limitations (Lin & Wei, 1989). Shared frailty models are based on the assumption that groups of animals within a cluster have a common frailty and model that frailty so they are analogous to random effects models (see Chapters 21 and 22).

Example 19.20 Individual frailty model—prostaglandin trial data
data = pgtrial

A Weibull model with a gamma individual frailty was fit to the prostaglandin trial data.

No. of subjects = 319	
No. of failures = 264	Number of obs = 319
Time at risk = 25018	LR χ^2 (5) = 9.96
Log likelihood = -524.2	Prob > χ^2 = 0.0764

Predictor	Coef	SE	Z	P	95% CI	
herd=2	-0.289	0.169	-1.710	0.088	-0.621	0.043
herd=3	0.039	0.175	0.220	0.825	-0.304	0.381
tx	0.205	0.125	1.630	0.102	-0.041	0.450
lact2	-0.041	0.041	-1.010	0.315	-0.122	0.039
thin	-0.136	0.138	-0.990	0.324	-0.406	0.134
_cons	-3.790	0.259	-14.640	0.000	-4.297	-3.282
ln p	-0.143	0.051	-2.800	0.005	-0.243	-0.043
ln theta	-14.870	756.631	-0.020	0.984	-1497.840	1468.099
p	0.867	0.044			0.784	0.958
1/p	1.154	0.059			1.044	1.275
theta	0.000	0.000			0.000	.

The variance of the gamma frailty (theta) was estimated to be zero (*ie* no frailty effect at all), suggesting that the Weibull hazard might be appropriate for these data.

19.11.3 Shared frailty models—introduction

Just as individual frailties can be considered to represent the effects of unmeasured covariates, shared frailties represent the effects of unmeasured covariates that a group of individuals have in common. These can represent the random effect of a grouping variable such as herd. (See Chapters 20-24 for more discussion of random effects). A shared frailty would be an appropriate way of dealing with the lack of independence observed when we have multiple failure times in an individual. (The frailty would represent the common characteristics of the individual that affect time to each event occurrence.)

A shared frailty model can be written as follows:

$$h_i(t|\alpha_i)=\alpha_i h(t)$$ ***Eq 19.28***

where α_i represents the frailty for the i^{th} group (and $h_i(t)$ and $h(t)$ incorporate the effects of the predictors). The survival probability, conditional upon the frailty is written:

$$S_i(t|\alpha_i)=S(t)^{\alpha_i}$$ ***Eq 19.29***

Frailties can take on a variety of distributions, but the most commonly used ones are gamma, inverse Gaussian and positive stable distributions. The statistical significance of a frailty can be assessed with a likelihood ratio test, but the usual χ^2 reference statistic is not correct because variances cannot be less than 0, so the P-value should be cut in half.

With only 3 herds represented in the prostaglandin data, it would not be logical to fit a shared (herd) frailty model to those data. Consequently, we will shift our attention to the -culling- dataset. This dataset consists of records of 721 cows that were tested for Johne's disease (*Mycobacterium avium* subspecies *paratuberculosis*) using an ELISA and then followed for a period of approximately 3.5 years to determine if Johne's status had an impact on survival. Only 13 cows were Johne's positive and during the follow-up period 466 of the cows were culled. The effect of the parity of the cow (3 categories, 1st, 2nd 3rd+) was also evaluated. (See Chapter 31 for details of dataset). Example 19.21 shows the results of fitting a Weibull model with a gamma-distributed shared frailty for herd to the culling data.

19.11.4 Shared frailty models—Cox models

A Cox model with a frailty term added can be written either as:

$$h_i(t|\alpha_i)=h_0(t)e^{\beta X}\alpha_i \qquad \textbf{\textit{Eq 19.30}}$$

with the α_i being the frailty on the hazard scale (frailties on the hazard scale are often assumed to have a gamma distribution), or as:.

Example 19.21 Shared frailty Weibull model—culling data
data = culling

A shared frailty model (Weibull distribution with a gamma distributed frailty common to all cows in a herd) was fit to the culling data.

No. of subjects = 721	Obs per group min = 4
No. of groups = 30	avg = 24.03333
No. of failures = 466	max = 31
Time at risk = 606875	LR chi2 (3) = 52.17
Log likelihood = -963.8	Prob > chi2 = 0.0000

Predictor	Coef	SE	Z	P	95% CI	
lact=2	0.252	0.145	1.74	0.081	-0.031	0.535
lact=3+	0.764	0.122	6.28	0.000	0.525	1.002
johnes	0.591	0.305	1.94	0.052	-0.006	1.188
_cons	-8.590	0.348	-24.65	0.000	-9.273	-7.907
/ln_p	0.144	0.041	3.51	0.000	0.064	0.224
/ln_the	-1.857	0.410	-4.53	0.000	-2.660	-1.053
p	1.155	0.047			1.066	1.252
1/p	0.866	0.036			0.799	0.938
theta	0.156	0.064			0.070	0.349

The coefficient for -johnes-=0.59 (*HR*=1.8) which suggests that a Johne's positive cow has almost twice the hazard of being culled as a Johne's negative cow from the same herd. The estimated variance of the gamma frailty distribution is 0.16 and is highly significant (*LRT* χ^2=27.7, P<0.001) which suggests that some herds have higher culling hazards than other herds.

$$h_i(t|\delta_i)=h_0(t)e^{\beta X+\delta_i} \qquad \textbf{\textit{Eq 19.31}}$$

with the δ_i (the shared frailty for the i^{th} group) on the log-hazard scale.

Estimating shared frailties in a Cox model is not straightforward. Four possible approaches are available: using a penalised likelihood function (see Example 19.22), using an EM (expectation maximisation) algorithm, fitting a random effects Poisson model (see below) or using Bayesian methods. With the exception of the Poisson model approach, these methods will not be discussed further except to state that the penalised likelihood approach is the computationally simplest and most commonly used method.

Shared frailty Cox model—Poisson regression

Poisson regression methods can be used to fit a standard Cox proportional hazards model and it produces exactly the same results. While this is not necessary (or practical) for fitting a standard Cox model, it has an advantage for shared frailty models in that random effects (equivalent to frailties) can be added to the Poisson model. This allows for the possibility of having more than one level of random effect and those effects can take on either gamma or log-normal distributions.

The procedure for fitting a Poisson model to survival data is as follows.

- Split each observation into multiple records according to the complete set of failure times in the dataset (*ie* each record will represent the time interval between the times of 2 consecutive failures). (**Note** This may create a very long dataset and cause numerical problems).
- Compute the length of time represented by each record (*ie* the interval between the 2 failure times) and log transform it.
- Fit a Poisson model which includes fixed effects for each time interval represented in the dataset and the log of the interval length as an offset

Example 19.22 Shared frailty Cox model—culling data

data = culling

A shared frailty Cox model (with a gamma distributed frailty common to all cows in a herd) was fit to the culling data.

```
No. of subjects = 721                          Obs per group min = 4
No. of groups = 30                                           avg =  24.03
No. of failures = 466                                        max = 31
Time at risk = 606875                                Wald chi2 (3) = 52.53
Log likelihood = -2830.6                             Prob > chi2 = 0.0000
```

Predictor	Coef	SE	Z	P	95% CI	
lact=2	0.249	0.144	1.730	0.084	-0.034	0.532
lact=3+	0.769	0.122	6.320	0.000	0.531	1.008
johnes	0.578	0.304	1.900	0.058	-0.019	1.174
theta	0.155	0.063				

The results are very similar to those obtained from the Weibull model (Example 19.21).

To avoid fitting the large number of fixed effects for the time periods, you can create a set of orthogonal polynomials (see Section 15.6.3) for time and use them instead of the set of fixed effects.

To fit a shared frailty model, include time as the set of polynomials (as described above) and add a random effect for the group variable (*eg* herd). Example 19.23 compares the results from fitting a proportional hazards model using standard Cox regression and Poisson regression procedures.

19.11.5 Frailty models—interpretation of coefficients

In a frailty model, the effects of predictors on the hazard or survival are *'conditional'* on the frailty, that is, they represent the effect of the predictor compared to an individual without the factor, but from the same group.

As noted above, the effects of predictors (*eg HR*) are the effects *'conditional'* on the frailty. For proportional hazards models (*eg* Weibull), the *HR* at any time *t* represents the shift in the hazard due to a unit change in the predictor, conditional on the frailty (*ie* assuming a comparable frailty). For a dichotomous predictor, it represents the effect of the factor being present compared with an individual with exactly the same frailty but with the predictor absent. This is analogous to a 'subject-specific' effect—see Section 22.4.1.

Example 19.23 Cox model fit by Poisson regression
data = culling

Several Cox proportional hazards models were fit to the culling data with lactation number (-lact_c3-) and Johne's status (-johnes-) as predictors. The models were:

- standard Cox proportional hazards model
- Cox model fit by Poisson regression (time intervals as fixed effects)
- Cox model fit by Poisson regression with time as a 4th order polynomial
- Cox model with shared frailty (gamma distribution) and with time as a 4th order polynomial
- Cox model fit by Poisson regression with random effect (gamma distribution) and time as a 4th order polynomial

The coefficient for -johnes-, its Wald test P-value and the estimate of the variance of the gamma distributions are presented.

Model	Coef	SE	P	Variance
standard Cox	0.648	0.293	0.027	
Cox – Poisson (time as fixed effects)	0.648	0.293	0.027	
Cox – Poisson (time as polynomial)	0.644	0.294	0.028	
Cox – with shared frailty (gamma distn.)	0.578	0.304	0.058	0.155
Cox – Poisson (gamma dist. random effect) and time as a polynomial	0.572	0.305	0.061	0.157

As can be seen, the standard Cox model and Poisson models produce identical results. Expressing time as a 4th order polynomial instead of a set of fixed effects produces results which are quite close. The 2 approaches to fitting the shared frailty model produce slightly different, but very close results.

In gamma frailty models, the population hazards (analogous to marginal effects—see Section 22.4.1) are not proportional over time and the hazard ratio only represents the population effect of the predictor at time 0. In general, the effect of the predictor on the population hazard will diminish over time in favour of the frailty effect. In simple terms, the frailty of the individual (or group) accounts for the fact that, over time, the population is increasingly 'robust' and the predictor has less and less influence on the hazard. With gamma frailties, the population *HR* tends to 1 as time approaches infinity, while, for an inverse Gaussian frailty, the *HR* tends toward the square root of the *HR*. This problem in interpreting the marginal effects of predictors is not present if the model is expressed as an AFT model—time ratios remain the same.

19.12 Multiple outcome event data

In all of the material presented in this chapter so far, we have assumed that there was only one possible occurrence of the outcome of interest (*eg* onset of pneumonia in calves, conception in dairy cows). However, in some instances, multiple outcome events are possible, and these fall into 3 general classes.

- Multiple different failure events—These arise in situations where you want to evaluate the effect of a predictor on multiple possible outcomes such as an evaluation of the use of a nutritional supplement in dairy cows after calving on the time to first service, the time to achieving positive energy balance and the time to peak milk production. These are sometimes referred to as competing risks data.
- Multiple 'same' endpoints (not ordered)—These arise in situations where multiple possible outcomes of the same event are possible, but there is not necessarily any ordering to them (*eg* time to onset of clinical mastitis in each of the quarters of a cow). One way of dealing with these is to change the unit of observation to the quarter, but in many cases, most of the risk factors will be at the cow level.
- Multiple 'same' endpoints (ordered)—These are also called **recurrence data**. They arise when it is possible for the outcome event to occur multiple times in the same animal (*eg* breedings, cases of clinical mastitis). The key feature to these is that there is a natural ordering to them (*ie* the second case cannot happen before the first case). The lack of independence among episodes must be accounted for (see below). This type of data is the focus of this section.

19.12.1 Models for recurrence data

Event times within an individual are often correlated for 2 reasons. First, there is likely to be heterogeneity among individuals, with some individuals more likely to experience the outcome than others, leading to clustering of events within the individual. As a result, observations within an individual are not independent. The second is that the probability of occurrence of one event may increase or decrease the probability of subsequent events (called event dependence).

There are 2 general approaches to dealing with the problem of heterogeneity. One is to adjust the variance estimate using robust standard errors (see Sections 19.11.2 and 20..5.4). An alternative is to fit a shared frailty model with the frailty representing the intrinsic susceptibility of the individuals (Therneau & Grambsch, 2000). The former approach produces population averaged estimates of effect while the latter generates subject-specific estimates (Cain & Cole, 2006; Kelly & Lim, 2000).

The problem of event dependence can be dealt with either by including a covariate for the number of previous events in the model (see Anderson-Gill model below) or by stratifying the data according to the number of events (see Prentice-William-Peterson model below). The former approach assumes there is a common baseline hazard function for all events. The latter allows for the baseline hazard to vary for each event (*ie* a different baseline hazard for first events compared to second *etc*). Models for repeated events data have been reviewed recently along with a proposal for a conditional frailty model which addresses both the issues of heterogeneity and event dependence (Box-Steffensmeier & De Boef, 2006) (beyond the scope of this text).

Three approaches to modelling recurrence data have been reviewed (Wei & Glidden, 1997). Two of these will be summarised below, but the third (a marginal model (Wei *et al*, 1989)) is no longer recommended (Hosmer & Lemeshow, 2008) and will not be described. Details of structuring data appropriately for these analyses is presented in Cleves (1999). The 2 approaches are shown in Example 19.24 using some clinical mastitis data. The data (-clin_mast-) are from 4595 cows in 105 herds. Each cow was followed for a minimum of 100 days in one lactation, starting at calving. The number of cows with 0, 1, 2, 3 and 4 cases of clinical mastitis was 3987, 497, 90, 18 and 3 respectively. Factors of interest that were investigated were parity and number of previous cases of mastitis (in that lactation).

Anderson-Gill model

This model is a generalised proportional hazard model and is the simplest approach to analysing recurrence data. The risk of recurrence is assumed to be independent of previous events, although the assumption of independence can be relaxed by including a time-varying predictor for the number of previous occurrences. The model is fit by assuming each subject's 'at-risk' time starts over again after each outcome is observed. If an animal is not considered to be at risk for a defined period after the occurrence of a case, then the time not at risk can be excluded (interval censored or gap). For example, it is common when defining cases of clinical mastitis that 7-14 days elapse between cases for the second occurrence to be considered a new case.

Prentice-William-Peterson model—conditional risk sets model

This model is a proportional hazards model that is conditional on previous occurrences. It is equivalent to carrying out a stratified analysis with the strata defined by the number of previous outcome events. All first occurrences would be in the first stratum, the second stratum would consist of second cases, but only animals that had experienced a first case would be at risk *etc.* Time at risk for each outcome can be measured either from the start of the study period or from the time of the previous event. The choice of approach depends on whether you feel that there is reason to 'reset the clock' each time an event occurs. An example of the former approach is shown Example 19.24. As noted above, this approach allows for a different baseline hazard in each risk set (stratum).

19.13 DISCRETE-TIME SURVIVAL ANALYSIS

Up to this point, we have assumed that failure times were recorded on a continuous basis, that is, we knew exactly when each failure time occurred (at least to the unit of time measurement—which was 'days' in both the prostaglandin and culling datasets). However, we are often faced with the situation in which failures are known to occur in an interval, but the exact time is not available. These are called interval-censored data. Such a situation would arise if we measured time to seroconversion in animals and they were only tested every 6 months. At some point, we

Example 19.24 Multiple failure event models
data = clin_mast

The structure of the clinical mastitis data for the Anderson-Gill and Prentice-William-Peterson models is shown in the table below. Cow 5 had 2 cases of mastitis and hence, has 3 records. Cows 7 and 15 had no cases.

					Anderson-Gill		Prentice-Williams-Peterson		
Herd	Cow	Parity	Prev. cases	Mast.	Start	End	Start	End	Risk set
125	5	4	0	1	0	14	0	14	1
125	5	4	1	1	14	108	0	94	2
125	5	4	2	0	108	359	0	265	3
125	7	4	0	0	0	336	0	336	1
125	15	3	0	0	0	357	0	357	1

The results from fitting a variety of models to the data are shown below.

	Parity		Previous cases	
	Coef	SE	Coef	SE
Anderson-Gill model				
Cox – robust SE	0.102	0.021	0.782	0.047
Weibull – robust SE	0.109	0.021	0.725	0.045
Weibull – gamma frailty	0.137	0.030	-0.452	0.170
Prentice-William-Peterson (PWP) model				
Cox – robust SE	0.075	0.024	na	na
Weibull – robust SE	0.089	0.024	na	na

All of the models which used robust SE to deal with heterogeneity produce lower estimates for the effect of parity because they produce population averaged estimates instead of subject-specific estimates. There is a marked difference in the apparent effect of the number of previous cases between robust SE models and the frailty model. The population average models (robust SE) indicate that in general, previous cases increase the risk of additional cases. The subject-specific model (frailty) suggests that for a given individual, having a case reduces the hazard of another case. This may be, in part, artificial in that there is a period of time after a case during which the hazard is zero —until enough time has lapsed to classify a new case as such (instead of as a relapse).

would observe a serologic response and would know that seroconversion had occurred some time during the preceding 6 months. Discrete-time survival analysis can be used to analyse such data. In some cases, failure times may have been recorded on a continuous basis, but actual failure times are uncertain and grouping them into intervals may improve data quality.

Discrete-time models may also be used for continuous time data if:

- the dataset is very large and not amenable to standard survival analysis methods, or
- there are many time-varying predictors, or
- there are time-varying effects which are not easily modelled as some function of time.

This last situation was evident in the prostaglandin data in which treatment had positive effects in time periods 3-4 days and 24-27 days, a negative effect in between and very little effect after 34 days (see Example 19.9).

Discrete time—basis for analysis

Time is divided into intervals, denoted I_j. In each interval, the number of subjects at risk is n_j and the number of failures is d_j. The probability of failure during the interval (or discrete time hazard) is then

$$h_j = d_j / n_j \qquad \textbf{\textit{Eq 19.32}}$$

Intervals can either be chosen to reflect the underlying biology of the situation or at convenient points which balance the width of the interval and the number of failures in each interval. (**Note** A rule of thumb is to have a minimum of 5 failures in each interval and it is important to avoid choosing intervals based on the observed data—*ie* 'data snooping'). Observations which are censored during the interval may be considered to have been censored at the start of the interval (*ie* not included in n_j), censored at the end (*ie* included in n_j) or counted for 1/2 of the time interval (as was done in an actuarial life table analysis). Table 19.4 shows the prostaglandin data divided into 15-day intervals and intervals based on expected biological effect of the treatment.

Table 19.4 Prostaglandin data divided into intervals

t_{j-1}	t_j	n_j[1]	d_j	h_j
regularly spaced intervals (15, 30 or 60 days)				
0	15	319	63	0.198
15	30	251	34	0.136
30	45	213	24	0.113
etc...				
intervals based on expected effect of prostaglandin				
0	2	319	6	0.019
3	4	311	34	0.109
5	23	275	40	0.146
24	27	234	13	0.056
28	60	212	54	0.255
60	90	153	43	0.281
etc...				

[1] observations censored during the interval considered censored at the start of the interval

If the data are truly discrete-time data (*ie* collected only at specific times), the time periods are defined by the data collection periods, so intervals do not need to be created. If there are many time periods (intervals), you may want to replace the fixed effects for each time period with some form of polynomial model of time (*eg* orthogonal polynomials as was done in the Poisson model in Section 19.11.4). See (Singer & Willett, 1993; Singer & Willett, 2003) for a review of discrete-time methods.

Discrete-time—logistic regression

Once the data have been structured as described above, they can be analysed using logistic regression according to the following model.

$$\text{logit}(h_j) = \beta_0 + \alpha_j + \beta X \qquad \textit{Eq 19.33}$$

where logit(h_j) is the probability (hazard) of failing in interval I_j given being present at the start of the interval and β_0 is the logit(hazard) in the baseline time period for a baseline individual, α_j is the effect of the j^{th} time period (compared to the baseline period) and βX represents the predictors in the model. This model assumes additivity on the logit(hazard) scale or proportional odds for the hazard probabilities. (**Note** This corresponds to a continuation-ratio model (see Section 17.2.4) for the multinomial probabilities across all intervals.)

Discrete-time logistic regression models can easily be extended to include one or more random effects (shared frailties) using procedures for modelling multilevel data (see Chapter 22). Specialised software is required to fit individual frailty models to discrete time data (Jenkins, 1995). Example 19.25 shows the results of such a model with time intervals as shown in the lower half of Table 19.4 and treatment by time interaction included.

Discrete-time—complementary-log-log regression

As noted above, the logistic model assumes that the log-odds of the outcome are additive, or alternatively that the odds are proportional. This is the same as saying that the *OR* for a predictor is constant across all time intervals (although this assumption can be relaxed by including interaction terms with the predictor). An alternative to logistic regression is to use a complementary log-log model which is based on the assumption of proportional hazards (not proportional odds) and consequently is a more natural fit with models such as the Cox proportional hazards model.

The complementary log-log function transforms a probability according to the following formula.

$$\text{cloglog}(p) = \ln[-\ln(1-p)] \qquad \textit{Eq 19.34}$$

Fig. 19.26 shows the relationship between probability and both the complementary log-log and logit functions. At p<0.2 the 2 functions are very close, but become substantially different at large values of p (and may produce substantially different results in binary regression analyses). As noted, the main advantage of a complementary log-log model is that it is based on the proportional hazards assumption and consequently, exponentiated coefficients can be interpreted as hazard ratios (as opposed to odds ratios).

Example 19.26 shows the results from a complementary log-log model of the prostaglandin data.

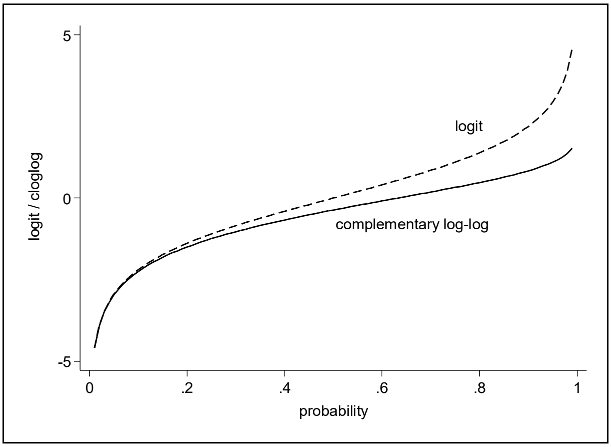


Fig. 19.26 Complementary log-log and logit functions

19.14 SAMPLE SIZES FOR SURVIVAL ANALYSES

Computation of sample sizes for studies with survival time as the outcome can be a complex process. For studies where the primary focus is the comparison of survival times across 2 (or more) groups, as it often is in controlled trials, one approach is to compute the sample size required to have a desired power in an analysis based on an unweighted log-rank test. If an assumption of proportional hazards is likely not valid, basing the sample size on that required for a weighted version of the test (*eg* Tarone-Ware or Harrington-Flemming tests) might be more appropriate.

However, there are many factors which will influence the required sample size. Some of the following have been discussed under sample size estimation in Chapter 2 and some are unique to studies of survival time.

1. Sample size might need to be increased to account for multiple predictors in the analysis, and/or to adjust for clustering of the data (*ie* non-independence among observations) (see Chapter 2).

Example 19.25 Discrete-time analysis—logistic regression
data = pgtrial

A discrete-time analysis using logistic regression was carried out on the prostaglandin data. Treatment by time period interaction terms were included. Not all model coefficients are shown.

Logistic regression — Number of obs = 1705
Log likelihood = -605.6 — LR chi2 (23) = 255.13
Prob > chi2 = 0.0000

Predictor	Coef	SE	Z	P	95% CI	
period 3-4	-0.687	1.230	-0.56	0.576	-3.098	1.723
period 5-23	3.093	0.737	4.19	0.000	1.648	4.538
period 24-27	...					
tx	0.814	0.874	0.93	0.352	-0.899	2.526
per 3-4 * tx	3.077	1.345	2.29	0.022	0.441	5.713
per 5-23 * tx	-2.355	0.989	-2.38	0.017	-4.292	-0.417
per 24-27 * tx	...					
herd=2	-0.323	0.199	-1.63	0.103	-0.713	0.066
herd=3	0.044	0.204	0.21	0.830	-0.356	0.444
lact	-0.035	0.047	-0.75	0.451	-0.127	0.057
thin	-0.201	0.161	-1.25	0.212	-0.517	0.115
constant	-4.143	0.735	-5.64	0.000	-5.583	-2.704

The coefficient for treatment (β = 0.81), represents the effect of treatment in the first time period (0-2 days), but it was not significantly different from 0. Subsequently there was a very strong positive effect of treatment in the period 3-4 days (0.814+3.077=3.89) and a strong negative effect (β =0.814-2.355=-1.54) in the period 5-23 days.

Example 19.26 Discrete-time analysis—complementary log-log regression

data = pgtrial

A complementary log-log model of the prostaglandin data that were used in Example 19.25 produced the following results.

Complementary log-log regression
Zero outcomes = 1442 Number of obs = 1705
Non-zero outcomes = 263 LR chi2 (23) = 255.33
Log likelihood = -605.5 Prob >chi2 = 0.0000

Predictor	Coef	SE	Z	P	95% CI	
period 3-4	-0.683	1.225	-0.56	0.577	-3.084	1.717
period 5-23	2.972	0.728	4.08	0.000	1.546	4.398
period 24-27	...					
tx	0.806	0.866	0.93	0.352	-0.891	2.504
per 3-4 * tx	2.950	1.334	2.21	0.027	0.334	5.565
per 5-23 * tx	-2.248	0.973	-2.31	0.021	-4.155	-0.341
per 24-27 * tx	...					
herd=2	-0.290	0.172	-1.68	0.093	-0.628	0.048
herd=3	0.040	0.176	0.22	0.822	-0.306	0.386
lact	-0.027	0.041	-0.65	0.518	-0.108	0.054
thin	-0.176	0.140	-1.26	0.207	-0.450	0.098
constant	-4.191	0.725	-5.78	0.000	-5.612	-2.770

The results are very close to those obtained from the logistic regression analysis. This was to be expected as the hazards of failure in all intervals were generally <0.2.

2. As pointed out in Chapter 11, multiple comparisons (often arising from interim analyses), losses in the follow-up process and subgroup analyses are common features of controlled trials which require adjustment to the sample size.

3. The shape of the baseline hazard function might not be known in advance of the study so a sample size estimate based on a non-parametric test (*eg* log-rank) would be appropriate.

4. The possibility of non-proportional hazards needs to be considered.

5. In controlled trials, crossover might occur in which animals could move from one treatment group to another (*eg* treated to non-treated if the owner fails to comply with treatment instructions).

6. Recruitment of animals into the study could take place over time which might affect the length of follow-up period for animals recruited.

7. Survival analyses are often used in randomised controlled trials. In non-randomised studies of therapeutic interventions, subjects with the new treatment are often matched

to those receiving a standard treatment within strata defined by covariates of interest. Sample size issues related to this study design have recently been discussed (Mazumdar *et al*, 2006).

A general discussion of sample size issues can be found in (Hosmer & Lemeshow, 2008). A review of some of the issues identified above and a description of a software program for computing samples sizes for survival analysis studies has recently been published (Barthel *et al*, 2006; Royston & Babiker, 2002) (see Example 19.27).

Example 19.27 Sample size calculations for a randomised controlled trial

data = hypothetical

Assume that you are about to start a randomised controlled trial of 2 drugs designed to prevent recurrence of a certain type of cancer in dogs, following initial treatment of the condition. Past experience has shown that as the risk of recurrence goes down, the longer the dog remains in remission. In the absence of treatment, you expect the cumulative probabilities of recurrence in each of 4 time periods to be as follows:

• end of year 1	30%
• end of year 2	50%
• end of year 4	60%
• later in life	65%

Relative to untreated controls, you expect treatment A to have a *HR* of 0.75 and treatment B to have a *HR* of 0.5. You consider the following 5 scenarios

(a) no loss to follow, no cross-over of treatments, equal allocation of subjects to the 3 groups

(b) same as (a) except cumulative loss to follow up of 5%, 15%, 30% and 40% in the 4 time periods

(c) same as (a) except cumulative loss to follow up of 10%, 30%, 60% and 80% in the 4 time periods

(d) same as (c) except 20% of control dogs and 10% of treatment A dogs cross over into treatment B

(e) same as (d) except you initially allocate dogs in the following ratio: control=1, treatment A=1, treatment B=2.

For each scenario, you want to determine the sample size required to have an overall power of 80% for detecting a difference among treatment groups. The required sample sizes and expected number of recurrences are:

	Scenario				
	(a)	(b)	(c)	(d)	(e)
Total sample size	742	837	965	1379	1347
Expected number of cases	353	353	353	502	500

As expected, the sample size goes up with increasing loss to follow up (with no increase in the expected number of cases). Subjects switching treatments (d) increases the required sample size, while changing the allocation of subjects was able to reduce the total sample size required.

REFERENCES

Barthel FM, Babiker A, Royston P, Parmar MKB. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over Stat Med. 2006; 25: 2521-42.

Box-Steffensmeier JM, De Boef S. Repeated events survival models: the conditional frailty model Stat Med. 2006; 25: 3518-33.

Cain LE, Cole SR. Survival analysis for recurrent event data: an application to childhood infectious diseases Stat Med. 2006; 25: 1431-3; author reply 1433.

Cleves M. Analysis of multiple failure-time data with Stata Stata Tech Bull. 1999; 49: 30-9.

Cleves M, Gould W, Gutierrez R. An Introduction to Survival Analysis Using Stata. 2nd Ed. Stata Press; College Station (TX). 2008.

Collett D. Modelling Survival Data in Medical Research. 2nd Ed. Chapman and Hall; New York. 2003.

Cox D. Regression models and life-tables (with discussion) Journal of the Royal Statistical Society, B. 1972; 34: 187-220.

Dohoo I, Stryhn H, Sanchez J. Beyond Cox : 1. hazard functions and time varying effects in parametric survival models. In: Proceedings of the 10th Symposium of the International Society for Veterinary Epidemiology and Economics, Vina Del Mar, Chile. 2003. p. 200.

Dohoo I, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island Preventive Veterinary Medicine. 2001; 50: 127-44.

Duncan C, Stephen C, Campbell J. Clinical characteristics and predictors of mortality for Cryptococcus gattii infection in dogs and cats of southwestern British Columbia Can Vet J. 2006; 47: 993-8.

Grønnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score Lifetime Data Anal. 1996; 2: 315-28.

Gutierrez R. Parametric frailty and shared frailty survival models The Stata Journal. 2002; 2: 22-44.

Hosmer D, Lemeshow S. Applied Survival Analysis. Regression modelling of time to event data. 2nd Ed. John Wiley & Sons; New York. 2008.

Hosmer D, Royston P. Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model The Stata Journal. 2002; 2: 331-50.

Jenkins S. Easy estimation methods for discrete-time duration models. Oxford Bulletin of Economics and Statistics. 1995; 57: 129-38.

Kaplan E, Meier P. Nonparametric estimation from incomplete observations Journal of the American Statistical Association. 1958; 53: 457-81.

Kelly PJ, Lim LL. Survival analysis for recurrent event data: an application to childhood infectious diseases Stat Med. 2000; 19: 13-33.

Lin D, Wei L. The robust inference for the Cox proportional hazards model Journal of the American Statistical Association. 1989; 84: 1074-8.

May S, Hosmer DW. A cautionary note on the use of the Grønnesby and Borgan goodness-of-fit test for the Cox proportional hazards model Lifetime Data Anal. 2004a; 10: 283-91.

May S, Hosmer DW. Hosmer and Lemeshow type Goodness-of-Fit statistics for the Cox proportional hazards model Handbook of Statistics. 2004b; 23: 383-94.

Mazumdar M, Tu D, Zhou XK. Some design issues of strata-matched non-randomized studies with survival outcomes Stat Med. 2006; 25: 3949-59.

Meadows C, Rajala-Schultz P, Frazer G, Meiring R, Hoblet KH. Evaluation of a contract breeding management program in selected Ohio dairy herds with event-time analysis I. Cox proportional hazards models Prev Vet Med. 2006; 77: 145-60.

Meadows C, Rajala-Schultz P, Frazer G, Phillips G, Meiring R, Hoblet K. Evaluation of a contract breeding management program in selected Ohio dairy herds with event-time analysis II. Parametric frailty models Prev Vet Med. 2007; 80: 89-102.

Moreau T, O'Quigley J, Mesbah M. A global goodness-of-fit statistic for the proportional hazards model Applied Statistics. 1985; 34: 212-8.

O'Quigley J, Stare J. Proportional hazards models with frailties and random effects Stat Med. 2002; 21: 3219-33.

Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls Lancet. 2002; 359: 1686-9.

Royston P. Explained variation for survival models The Stata Journal. 2006; 6: 83-96.

Royston P, Babiker A. A menu driven facility for complex sample size calculation in randomized controlled trials with a survival or binary outcome The Stata Journal. 2002; 2: 151-63.

Royston P, Sauerbrei W. Multivariable model-building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. John Wiley & Sons, Ltd; Chichester, England. 2008.

Singer J, Willett J. It's about time: using discrete-time survival analysis to study duration and the timing of events J Educational Stat. 1993; 18: 155-95.

Singer J, Willett J. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence Oxford University Press; Oxford. 2003.

Therneau TM, Grambsch PM. Modelling survival data: extending the Cox model. Springer-Verlag; New York. 2000.

Thysen I. Application of event time analysis to replacement, health and reproduction data in dairy cattle research Preventive Veterinary Medicine. 1988; 5: 239-50.

Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials Stat Med. 1997; 16: 833-9; discussion 841-51.

Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modelling marginal distributions J Am Stat Assoc. 1989; 84: 1065-73.