fecundity is to be used in a model. For example, the particular case of estimating fecundity for the Euler–Lotka equation is discussed in Chapter 5.

In general, fecundity is one of the easier demographic parameters to estimate in birds and mammals. Captive breeding data will usually provide estimates of gestation period and clutch or litter size that can be applied to wild populations without too much error. One adjustment that does need to be made, however, is to estimate the proportion of adult females that are breeding in a given population at a particular time. This is not a straightforward exercise (see Clobert *et al*., 1994). In the fisheries literature, much attention is given to estimating fecundity, usually by estimating the number of eggs carried by gravid females. This is valuable for determining the relative contribution of various age or size classes to the overall reproductive output of the population. Without an estimate of larval mortality, however, it sheds limited light on the reproductive output itself.

Fecundity will always be age-dependent. It is frequently also size-dependent, particularly in ectothermic animals. Juveniles, by definition, do not reproduce, and in almost all organisms the pre-reproductive stage is not such a small fraction of the total lifespan that it can be neglected. This is especially so because the intrinsic growth rate of a population is particularly sensitive to early reproduction (see Chapter 5). Nevertheless, some very abstract models (for example, the simple host–parasite models of Anderson and May, 1978) assume a single birth-rate parameter and require it to be separated from the death rate. Attempting to estimate this as simply the rate of production of offspring by mature females will always cause a gross overestimate of the intrinsic growth rate of the population, to an extent that the qualitative behaviour of the system will probably be misrepresented.

There is no entirely satisfactory way to replace an age-specific fecundity estimate with a simple, unstructured rate. Probably the best approach is to use the full age-specific fecundity and mortality schedules to estimate the intrinsic rate of growth $r$ (see Chapter 5), and then to subtract the death rate from this to yield a single birth-rate parameter.

## Survival

Survival is inevitably a much more difficult process to quantify in wild populations than is fecundity. Survival of captive animals will always be a very poor indicator of survival in the wild, except that it may give an indication of the absolute maximum age the animals can reach. Estimating the rate of survival in the field ideally involves following the fate of individuals, preferably of known age. Given restrictive assumptions, some idea of the age-specific death rate can be gained from the age distribution of animals, but this requires a means of ageing individuals, a problem that is often technically difficult.

The most appropriate ways to describe survival for a model and then to estimate parameters to quantify the chosen form of survival depend critically on the type of organism under study and on the objectives of the study. In any but the most general models, the death rate will be a function of age. The death rate at a given age may also be dependent on time and any number of other variables.

Broadly speaking, the pattern of age-dependent survival in a population can be described in two ways. A life table breaks the life history up into a number of discrete ages or stages, and estimates a survival rate through each stage. A survival curve represents the death rate as a function of age. From the perspective of parameter estimation, the difference between the two approaches is similar to the difference between analysis of variance and regression. A life table requires the estimation of as many survival rates as there are stages in the table, but is completely flexible in the way that it allows survival to change with age. A survival curve will generally require the estimation of two or three parameters only, depending on the functional form fitted, but the way in which survival changes with age is constrained. Alternatively, a life table can be viewed as expressing survival as a nonparametric function of age, in contrast to the parametric approach of a survival curve (Cox & Oakes, 1984, p. 48).

Studies of insects generally have few problems with sample size, and often use life-table approaches to advantage. In many studies of vertebrate populations, however, sample size is a major constraint, and estimating many parameters simultaneously is impractical. An approach based on a two- or three-parameter survival curve may thus be preferable to a life table, even if the resulting mortality estimates are subsequently used in a model with discrete age classes.

Particular problems arise with models that are structured by size or developmental stage. For a given death rate per unit of time, the proportion of individuals surviving through a life-history stage will obviously depend on the time spent in that stage. Unless the development time through a stage is fixed, parameterization of stage-specific survival requires simultaneous estimation of the development and mortality rates.

The most appropriate method of estimating survival also depends on a number of practical considerations. Some methods can only be used if animals can be individually marked and then followed reliably until either death or the end of the study. If individuals are marked, but will not necessarily be detected on every sampling occasion, even if alive and in the study area, then other methods must be used. It may be possible to mark or identify members of, or a sample from, a cohort, but not to follow particular individuals through time. Again, different estimation methods must be used. Finally, in many situations, it is not possible to mark animals in any way, and attempts must be made to estimate survival from the age or size distribution of the animals.

Faced with this diversity of problems and possible approaches, I will first develop some general principles and concepts, and then deal with the actual estimation problems in a series of subsequent sections. Table 4.1 provides a guide to the possible approaches that can be taken to estimate survival or mortality rates.

### Survival curves, failure times and hazard functions

A survival curve or survivor function is simply the proportion of individuals surviving as a function of age or time from commencement of the study. In the general ecological literature, it has become conventional to divide survival curves into three types, I, II and III (Krebs, 1985; Begon *et al.*, 1996a) (Fig. 4.1), depending on whether the relationship between log survivors and age is concave, straight or convex. These types correspond to death rates that are decreasing, constant or increasing with age. Almost all survival curves for real organisms, however, will include segments with very different survival rates, corresponding to different ontogenetic stages.

It is tempting to estimate a survivor function by fitting a line or curve through a plot of the proportion of individuals surviving versus time. This approach is invalid statistically because successive error terms are not independent: the survivors at any time have also survived to each previous time step. A better approach is to use the time until death of each individual (its 'failure time' in the statistical literature) as the response variable. A problem which then arises is that it is likely that some individuals will either survive to the end of the study or otherwise be removed. A minimum failure time can be assigned to these individuals, but the actual failure time is unknown. Such observations are 'censored'.

Survival of individuals throughout the course of a study is the most obvious way that censoring can occur, and will result in observations censored at a value equal to the total study duration. Censoring may, however, occur at other values. In many studies, subjects may enter the study at different times. For example, animals may be caught and radio-collared at times other than the start of the study. This is known as 'staggered entry' (Pollock *et al.*, 1989), and individuals surviving at the conclusion of the study period will have different censoring times, depending on their time of entry (Fig. 4.2). Alternatively, individuals may be removed from the study before its conclusion, due to factors other than natural mortality. Possibilities include trap mortality, permanent emigration, or radio-collar failure.

Statisticians usually describe survival data in terms of the hazard function, or the probability density of death as a function of time, conditional on survival to that time (Cox & Oakes, 1984). To ecologists, this concept will be more familiar as the instantaneous death rate of individuals, as a function

**Table 4.1** A key for choosing a method of survival analysis. Use this like a standard dichotomous key for identifying organisms. Start at 1, on the left-hand side, and decide which of the alternatives best applies to your problem. Then go to the number associated with that alternative, and repeat the process. Continue until you reach a suggested method, or you reach a ×, which indicates that there is no satisfactory method

| | | |
|---|---|---|
| 1 | Do you want a continuous survival function or a life table (survival in discrete stages or intervals)? | |
| | Survival function | 2 |
| | Life table | 7 |
| 2 | Can particular individuals be followed through time? | |
| | Yes | 3 |
| | No | 6 |
| 3 | Can all individuals being followed in the population be detected at each sampling occasion, if they are present, or may some be missed? | |
| | All detected if present | 4 |
| | Some may be missed | 5 |
| 4 | **Parametric survival analysis** (see p. 108) | |
| 5 | **Mark–resight and mark–recapture methods** (see p. 119) | |
| 6 | Can the number of survivors from an initial cohort of known size be determined? | |
| | Yes | 4 |
| | No | × |
| 7 | Can you follow an initial cohort through time? | |
| | Yes | 8 |
| | No | 13 |
| 8 | Are the intervals of the life table of fixed (but not necessarily equal) duration, and can individuals be assigned to an interval unequivocally? | |
| | Yes | 9 |
| | No | 12 |
| 9 | Do you need to use statistical inference, or is simple description sufficient? | |
| | Description only | 10 |
| | Statistical inference required | 11 |
| 10 | **Cohort life-table analysis** (see p. 116) | |
| 11 | **Kaplan–Meier method** (see p. 118) | |
| 12 | **Stage–frequency analysis** (see p. 126) | |
| 13 | Can individuals be aged? | |
| | Yes | 14 |
| | No | 18 |
| 14 | Can the proportions of each age class surviving one time step be determined? | |
| | Yes | 15 |
| | No | 16 |
| 15 | **Standard life-table techniques possible, but yields a *current* or *time-specific* life table** (see p. 116) | |
| 16 | Is the rate of population increase known (or can it be assumed to be 0)? | |
| | Yes | 17 |
| | No | × |
| 17 | **Use 'methods of last resort'** (see p. 130) | |
| 18 | **Try stage–frequency analysis** (see p. 126) | |

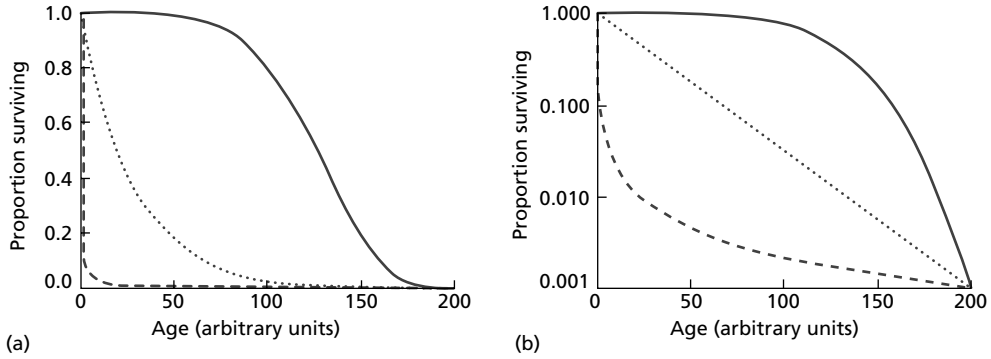


**Fig. 4.1** Survival curves of types I, II, and III, with survival: (a) on a linear scale; (b) on a log scale. The solid line is a type I survival curve, the dotted line is a type II curve, and the dashed line is a type III curve. These were generated using a Weibull survival function (see Table 4.3), with: $\kappa = 5$ and $\rho = 0.007\ 36$ (type I), $\kappa = 1$ and $\rho = 0.034\ 54$ (type II); and $\kappa = 0.2$ and $\rho = 78.642$ (type III). I chose values of $\rho$ so that survival to age 200 was standardized at 0.001 for each value of $\kappa$.

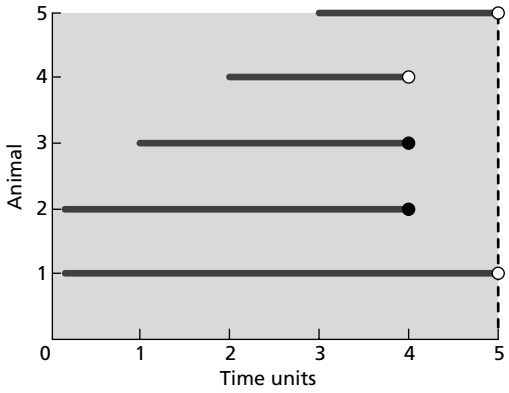


**Fig. 4.2** Staggered entry and censoring in a survival study. This diagram represents the fates of five individuals in a hypothetical survival experiment running from times 0 to 5. The horizontal lines represent times over which animals were part of the study. Deaths are marked by solid dots, and open dots indicate *right-censored* observations: it is known that the animal survived to at least the time shown, but its actual time of death is unknown. Animal 1 was followed from the beginning of the study, and survived through to the end, so its survival time is right-censored at 5 time units. Animal 2 similarly was followed from the beginning, but was dead at time 4. Animal 3 did not enter the study until time 1, and was dead at time 4. Animal 4 entered the study at time 2, but at time 4 its radio-collar had failed, so it was not known if it was alive or dead. Its survival time is therefore censored at 2 time units. Finally, animal 5 entered the study at time 3, and survived until the end. Its survival time is thus also censored at 2 time units.

**Table 4.2** Ways of representing the probability of survival

| Function | Name | Meaning | Relationship to other representations |
|---|---|---|---|
| $S(t)$ | Survivorship function | Proportion of cohort surviving to time $t$ | $S(t) = \exp\left(-\int_0^t \mu(u)\mathrm{d}u\right)$ |
| $\mu(t)$ | Age-specific death rate, or hazard function | Probability density of failure at time $t$, conditional on survival to time $t$ | $\mu(t) = f(t)/S(t)$ |
| $f(t)$ | | Probability density of failure at time $t$ | $f(t) = -\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}$ |

of age. Survival can also be described as an unconditional probability density of failure as a function of age. This is different from an age-specific death rate, because it is the probability density of death at a given age of an individual at the start of the study. The unconditional probability density of failure is important, as it is this quantity that is necessary for maximum likelihood estimation of survival rate parameters. Table 4.2 summarizes these ways of describing age-specific survival, and the relationships between them.

### Parametric survival analysis

The approaches described in this section produce an estimate of survival as a continuous function of age or time from the entry of the subject into the study. Survival may also depend on other variables, either continuous or categorical. These approaches can be applied in two situations: if animals are marked individually, and their continued survival at each sampling occasion can be determined unequivocally; or if animals are not followed individually, but the numbers from an initial cohort surviving to each sampling occasion can be counted. If animals are marked individually, but may not necessarily be detected on each sampling occasion, the mark–recapture methods discussed later in the chapter should be used instead.

It is first necessary to select an explicit functional form for the hazard function (the death rate, as a function of age), and the objective is then to estimate the parameters of that function. There are many forms of the hazard function in the statistical literature (see Cox & Oakes, 1984, p. 17). For most ecological purposes, two are adequate: a constant hazard or death rate, which yields an exponential survival function; and the Weibull distribution, which allows the hazard to either increase or decrease with time depending on the value of an adjustable parameter. Table 4.3 gives the hazard, probability density and

**Table 4.3** Exponential and Weibull survival curves

| Name | Hazard function | Density function | Survivorship function |
|---|---|---|---|
| Exponential | $\rho$ | $\rho e^{-\rho t}$ | $e^{-\rho t}$ |
| Weibull | $\kappa\rho(\rho t)^{\kappa-1}$ | $\kappa\rho(\rho t)^{\kappa-1}\exp[-(\rho t)^{\kappa}]$ | $\exp[-(\rho t)^{\kappa}]$ |

survival functions of these two models. The exponential distribution has a single parameter $\rho$, which is the instantaneous death rate per unit of time. The Weibull distribution also has a rate parameter $\rho$, and, in addition, a dimensionless parameter $\kappa$, which determines whether the hazard increases or decreases with time. From Table 4.3, it can be seen that $\kappa > 1$ generates a death rate increasing with time (a type III survival curve) and that if $\kappa < 1$, the death rate decreases with time (a type I curve). If $\kappa = 1$, the Weibull reduces to the exponential distribution. Figure 4.1 shows examples of Weibull and exponential survival functions.

In many cases, it may be necessary to describe survival as a function of explanatory variables. These might be continuous, such as parasite burden, or discrete, such as sex. It is relatively straightforward to make $\rho$ a function of such explanatory variables. For example, if a continuous variable $X$ affects survival, $\rho$ can be represented as

$$\rho(X) = \exp(\alpha + \beta X), \tag{4.2}$$

where $\alpha$ and $\beta$ are parameters to be estimated. The exponential ensures that the death rate remains positive for all values of $\alpha$, $\beta$ and $X$. It is usual to assume that the index $\kappa$ remains constant and does not depend on the value of the explanatory variable, although there is no good biological, as distinct from mathematical, reason why this should be so.

Most ecologists will use a statistical package to fit a survival function, although it is not difficult to do using a spreadsheet from the basic principles of maximum likelihood. Some statistical packages (for example, SAS) use a rather different parameterization from the one I have described here, which follows Cox and Oakes (1984). It is important to plot out the observed and predicted survival, both to assess the fit of the model visually, and to ensure that the parameter estimates have been interpreted correctly.

The survival rate of black mollie fish infected with *Ichthyophthirius multifiliis* (Table 4.4), an example first discussed in Chapter 2, is modelled using the Weibull distribution in Box 4.1, with the results presented graphically in Fig. 4.3. It is clear from the figure that the Weibull curve, using parasite burden as a continuous linear covariate, does not completely capture all the differences in survival between the different levels of parasite infection. In particular, the predicted mortality rate is too high for the lowest infection

class. The fit can be improved by estimating a separate $\rho$ for each infection class. This example emphasizes that procedures such as checking for linearity and lack of fit, which are standard practice for linear models, are equally important for survival analysis.

If very few data are available, the simplest and crudest way to estimate a death rate is as the inverse of life expectancy. If the death rate $\mu$ is constant through time, then the following differential equation will describe $N(t)$, the number of survivors at time $t$:

$$\frac{dN}{dt} = -\mu N(t). \tag{4.3}$$

This has a solution

$$N(t) = N(0)\ \exp(-\mu t). \tag{4.4}$$

**Table 4.4** Black mollie survival data. The following table shows, as a function of parasite burden, the number of fish $N$ that had died at the given times (in days) after the initial assessment of infection. The experiment was terminated at day 8, so that it is not known when fish alive on that day would have died. The survival time of those fish is thus censored at day 8

| Burden | fail time | $N$ | Burden | fail time | $N$ |
|---|---|---|---|---|---|
| 10 | 6 | 1 | 90 | 4 | 1 |
| 10 | 8* | 10 | 90 | 5 | 2 |
| 30 | 3 | 1 | 90 | 6 | 1 |
| 30 | 7 | 2 | 90 | 7 | 1 |
| 30 | 8 | 4 | 90 | 8* | 2 |
| 30 | 8* | 3 | 125 | 3 | 1 |
| 50 | 3 | 2 | 125 | 4 | 4 |
| 50 | 4 | 1 | 125 | 5 | 1 |
| 50 | 5 | 3 | 125 | 6 | 2 |
| 50 | 6 | 1 | 125 | 8 | 1 |
| 50 | 7 | 3 | 175 | 2 | 2 |
| 50 | 8 | 5 | 175 | 3 | 4 |
| 50 | 8* | 3 | 175 | 4 | 3 |
| 70 | 4 | 2 | 175 | 5 | 3 |
| 70 | 5 | 2 | 175 | 8 | 1 |
| 70 | 6 | 1 | 225 | 2 | 4 |
| 70 | 7 | 1 | 225 | 3 | 4 |
| 70 | 8 | 4 | 225 | 8 | 1 |
| 70 | 8* | 4 | 300 | 2 | 1 |
| | | | 300 | 3 | 1 |
| | | | 300 | 4 | 4 |
| | | | 300 | 6 | 2 |
| | | | 300 | 7 | 1 |

*Censored observations.

**Box 4.1 Fitting a Weibull survival function using maximum likelihood**

The data used in this example are from an experiment investigating the survival of black mollie fish infected by the protozoan *Ichthyophthirius*. The experimental procedure is outlined in the legend to Table 2.1, and the data are presented in Table 4.4. The time of death of individual fish, up to a maximum of eight days after assessment of the infection level (which in turn was two days after infection) was recorded as a function of intensity of parasite infection. The objective is to use maximum likelihood to fit a Weibull survival curve with parasite infection intensity $X$ as a covariate, and the rate parameter of the Weibull distribution given by

$$\rho = \exp(\alpha + \beta X). \tag{1}$$

There are two types of observation in the data set: uncensored observations, for which the actual time until death is known; and censored observations, for which it is known that the fish survived at least eight days, but the actual time of death is unknown. The probability density of death at time $t$, where $t \leq 8$, is given in Table 4.3:

$$f(t) = \kappa\rho(\rho t)^{\kappa-1} \exp[-(\rho t)^{\kappa}]. \tag{2}$$

The probability that a fish survives beyond eight days is given by the survivorship function, evaluated at $t = 8$:

$$S(8) = \exp[-(8\rho)^{\kappa}]. \tag{3}$$

The log-likelihood function has two components, one for uncensored observations, and a second for the censored observations. Following the usual approach of maximum likelihood estimation, the log-likelihood function of the three unknown parameters $\alpha$, $\beta$ and $\kappa$ can then be written as a function of the observed data:

$$l(\alpha,\beta,\kappa) = \sum_{u} \ln(f(t)) + \sum_{c} \ln(S(t)). \tag{4}$$

Here, the $u$ represents a sum over the uncensored observations, and the $c$ a sum over the censored observations. (Note that in this case, the time at censoring entered into $S(t)$ is 8 for every censored observation.) The maximum likelihood estimates of the parameters are those which maximize the value of eqn (4). Most ecologists would probably use a statistical package to perform the calculation. It is, however, quite straightforward to do using a spreadsheet. Substituting eqns (2) and (3) into eqn (4), the problem is to maximize

$$l(\rho,\kappa) = \sum_{u} (\ln\kappa + \kappa\ln\rho + (\kappa-1)\ln(t) - (\rho t)^{\kappa}) - \sum_{c} (\rho t)^{\kappa}. \tag{5}$$

*continued on p. 112*

**Box 4.1** *contd*

Substituting the expression for $\rho$ from eqn (1), and further simplifying, the objective is to find estimates of $\alpha$, $\beta$ and $\kappa$ to maximize

$$l(\alpha,\beta,\kappa) = \sum_{u}[\ln \kappa + \kappa(\alpha + \beta X) + (\kappa - 1)\ln(t)] - \sum_{u,c}(t \exp(\alpha+\beta X))^{\kappa}. \quad (6)$$

This can be done using the Solver facility in Excel, as is shown for a simpler example in Box 2.3. In this case, all that is necessary is to enter each component of the sums in eqn (6) into separate rows of a spreadsheet, and to maximize their sum. The solutions are:

| $\alpha$ | $\beta$ | $\kappa$ |
|---|---|---|
| −2.211 56 | 0.002 789 | 2.892 263 |

As would be expected, $\beta$ is positive, meaning that the death rate increases with parasite burden. As $\kappa > 1$, the death rate for a given parasite burden also increases strongly with time since infection. This is also to be expected, given that this is a parasite which grows rapidly in size over its 10-day lifetime on the host.

The predicted survival curves are shown in Fig. 4.3. Whilst the fit seems fairly good for intermediate parasite burdens, there is evidence of systematic differences between observed and predicted survivals, particularly at the lowest infection intensity. This subjective impression is confirmed if separate values of $\rho$ are fitted for each value of the parasite burden. Again, this can be done quite easily on a spreadsheet. The estimated parameters are shown below.

| Burden | 10 | 30 | 50 | 70 | 90 | 125 | 175 | 225 | 300 |
|---|---|---|---|---|---|---|---|---|---|
| $\ln(\rho)$ | −2.860 88 | −2.139 94 | −1.987 8 | −2.073 74 | −1.979 54 | −1.670 69 | −1.509 17 | −1.427 54 | −1.59 146 |

The estimated value of $\kappa$ is 3.00.

The change in deviance (twice the difference in the log-likelihood between the two models) is 23.58 at a cost of 7 additional parameters, a highly significant difference. The predicted survival using this categorical model is also shown in Fig. 4.3.

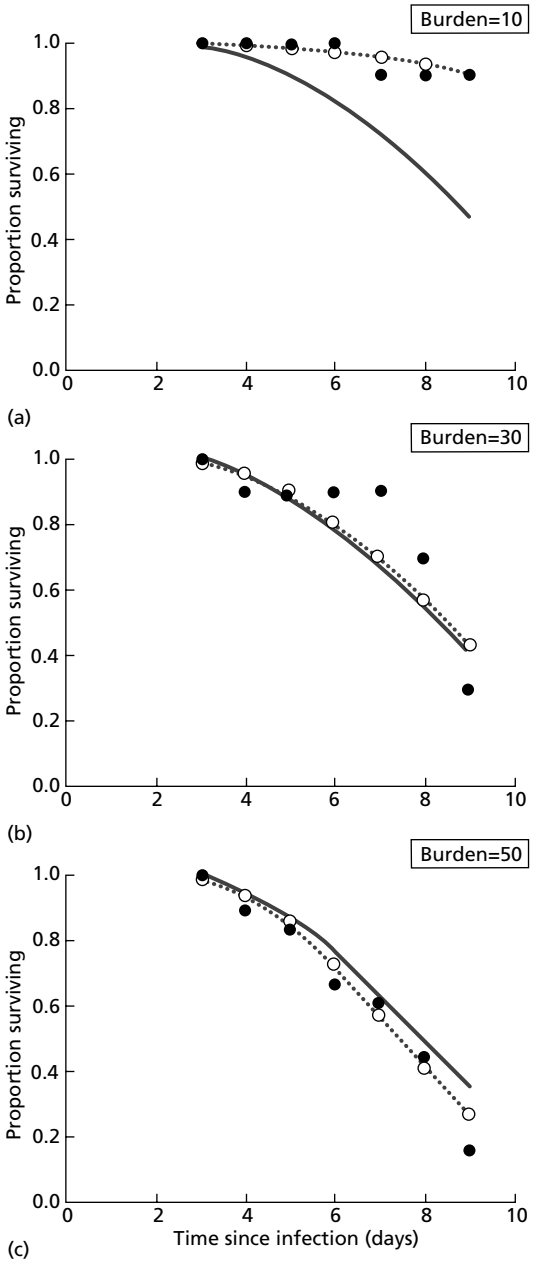


**Fig. 4.3** Survival of black mollies infected with *Ichthyophthirius multifiliis*. Black dots show the observed proportions of fish surviving as a function of time after infection. The proportion surviving as predicted by a Weibull survival function, using burden as a continuous linear predictor, is shown as a solid line. The dotted line and open circles show the predicted proportion surviving if a Weibull with burden treated as a categorical variable is used.

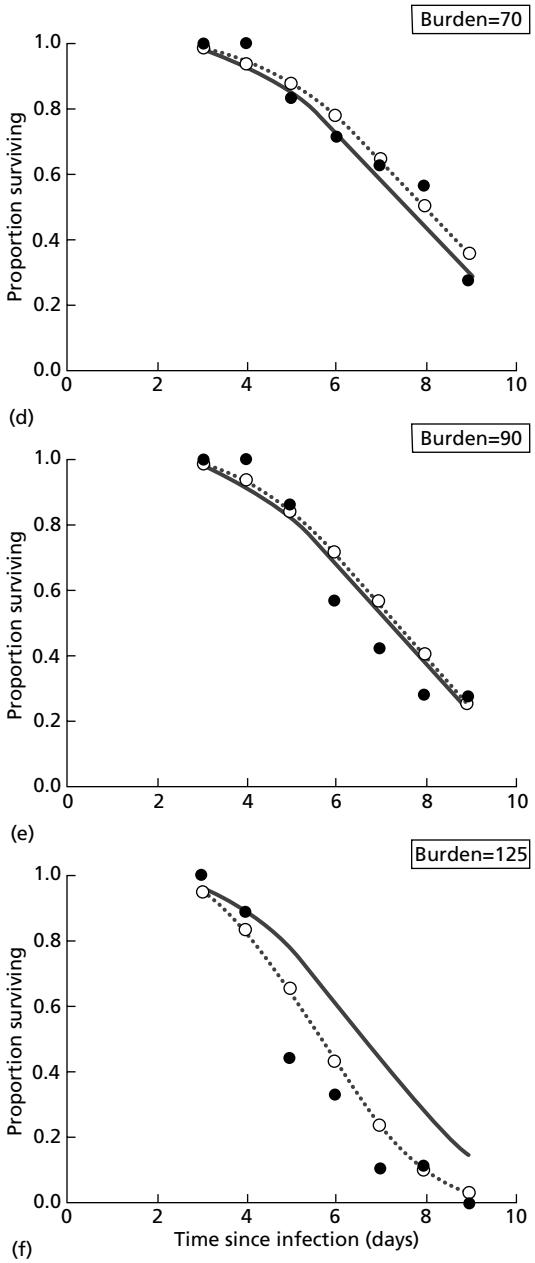
Burden=70
Proportion surviving
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
(d)
Burden=90
Proportion surviving
(e)
Burden=125
Proportion surviving
Time since infection (days)
(f)

**Fig. 4.3** *contd*

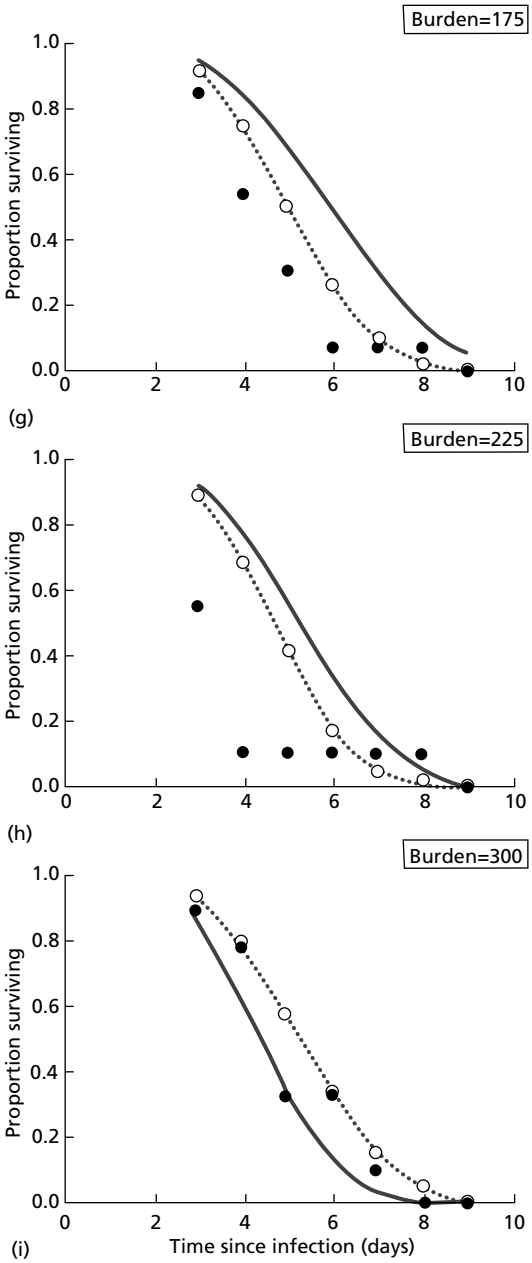
Burden=175
Proportion surviving
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
(g)
Burden=225
Proportion surviving
(h)
Burden=300
Proportion surviving
(i)
Time since infection (days)

**Fig. 4.3** *contd*

The definition of life expectancy, $E(t)$, is the average age each individual survives from time 0, and hence

$$E(t) = \frac{\int_{t=0}^{\infty} t \exp(-\mu t) \mathrm{d}t}{\int_{t=0}^{\infty} \exp(-\mu t) \mathrm{d}t}. \tag{4.5}$$

Integrating by parts, the solution to this is simply $1/\mu$.

This straightforward relationship is useful where a single death-rate parameter needs to be plugged into a model, as long as you remember that the life expectancy of an organism or life-history stage is usually much shorter than the maximum lifespan.

### Life tables

A life table is simply a record showing the number of survivors remaining in successive age classes, together with a number of additional statistics derived from these basic data (Table 4.5). The method has its basis in human demography, and in particular, the life insurance industry (see Keyfitz, 1977). It is the first stage in constructing many age- or stage-structured models. The simplest life tables are single-decrement life tables, in which individuals dying in each age class or stage are recorded, without any attempt to identify the cause of death. A multiple-decrement life table (Carey, 1989) seeks further to attribute deaths to a range of possible causes, permitting the impact of various mortality factors on the population to be determined.

In developed countries, a full register of births and deaths for many years makes obtaining the data for human populations a straightforward exercise. Many animal and plant populations are a very different matter. With long-lived organisms, in particular, it is usually impossible to follow an entire cohort from birth to death. A wide variety of methods has been developed to deal, as far as is possible, with problems of limited sample size, limited study duration, organisms with unknown fate, difficulties with ageing and other sampling difficulties.

There are two basic forms of life table, whether single- or multiple-decrement tables are considered. The standard form is the cohort life table, in which a number of organisms, born at the same time, are followed from birth, through specified stage or age classes, until the last individual dies. The age- or stage-specific survival rates obtained are therefore for a number of different times, as the cohort ages. The second form of life table, a current life table, takes a snapshot of the population at a particular time, and then determines the survival over one time period for each age or stage class in the population.