6. MODEL FITTING

CHAPTER 6

MODEL FITTING

6.1 Introduction

As most animal carcinogenesis experiments aim at determining whether or not a particular treatment increases the risk of cancer at one or more sites, the statistical methods described so far lean heavily towards techniques for hypothesis testing. However, many long-term animal experiments are analysed to provide an appropriate condensation of the information contained in the data and not merely to answer a yes–no question.

Many statistical models fitted to experimental data have fruitfully influenced the thinking on chemical carcinogenesis. The multistage model proposed by Armitage and Doll (1961) has been able to account for various experimental as well as epidemiological observations. Age-specific incidence of tumours induced by continuous exposure to chemical carcinogens has been described successfully by such models (Lee & O'Neill, 1971; Berry & Wagner, 1969), which also predicted increasing incidence with age as being a consequence of prolonged time since first exposure. This prediction was confirmed experimentally (Peto *et al.*, 1975).

On the epidemiological side, a marked regularity of age-incidence curves for most spontaneous human tumours of epithelial origin has been noted (Cook *et al.*, 1969) and detailed dose-response curves, as observed for the dependence of lung cancer risk on daily cigarette consumption, show a high degree of consistency with the multistage theory (Doll, 1971; Peto, 1977; Doll & Peto, 1978). Epidemiological data on the joint effect of two exposures are also interpretable in terms of the multistage model (Wahrendorf, 1984). Day and Brown (1980) have explored the multistage model concerning changes in risk after cessation of exposure. They were able to show both for experimental and epidemiological data two different types of behaviour under the multistage model.

An interesting empirical observation for chemical carcinogenesis data was made by Druckrey *et al.* (1967). They noted that if d is the daily dose rate and t the median time to tumour induction (death with tumour), then the relationship, $d \cdot t^n = \text{constant}$, holds for many carcinogens, especially nitroso compounds. This formula is predicted by the Weibull model (Carlborg, 1981) and has received wide attention in the discussion of thresholds in chemical carcinogenesis (Chand & Hoel, 1974; Port *et al.*, 1976).

The fitting of statistical models has to compromise between specificity and identifiability. On the one hand, one would imagine that inserting all relevant

knowledge about the carcinogenic process into a mathematical model would result in many parameters which could not all be identified by the limited amount of information provided from a long-term animal experiment. On the other hand, models which relate the probability of tumour development throughout life to the dose administered appear very simple.

The degree of specificity which a model might be allowed to have depends on the details of the experimental design, including whether and when animals are inspected for diagnostic purposes and whether and when animals are killed – either by scheduled sacrifice or by normally occurring deaths. The experimental design also specifies the schedule of the dose application. Of special interest in this respect are chronic exposures stopped at a certain time or the application of fractionated doses which can differ in respect to total dose, number of fractions and time between single doses.

One essential problem with the fitting of statistical models should be stated clearly. These models are usually fitted to data sets from single experiments done in one strain, in one sex, by one route of application and by considering tumours at one site. Consequently, the scope of generalization of one such fit is limited, particularly in the absence of consistency in studies done under different conditions.

In the following sections, we will give an overview of some statistical models which have been proposed for the analysis of long-term animal experiments. This will include simple dose-response models, time-to-tumour models and models based on different states in the course of the tumour development.

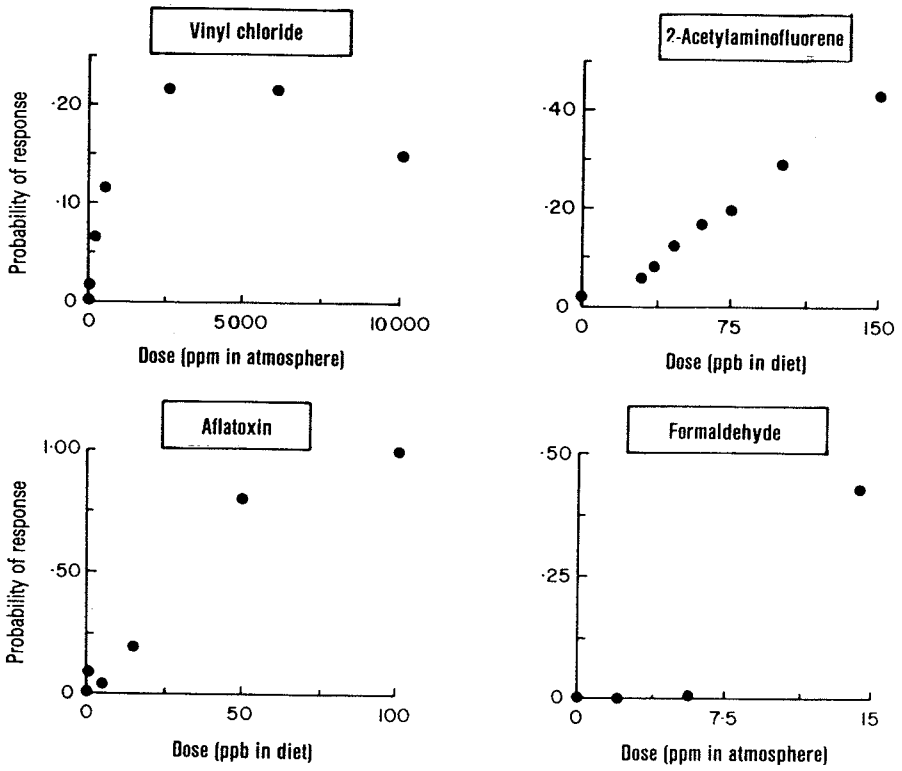
6.2 Dose-response models

As noted in Chapter 3, the probability of tumour occurrence depends on both the dose and the period of exposure. In many cases, however, the dose-response relationship at a fixed point in time may be of interest. In the absence of decreased survival at high doses, for example, the proportion of animals developing tumours during the course of a long-term study generally increases with dose.

The shape of such dose-response curves can vary widely depending on the agent used (Fig. 6.1). While the dose-response curve for liver tumours induced in mice as a result of exposure to 2-acetylaminofluorine (2-AAF) for 24 months is nearly linear (Littlefield *et al.*, 1980a), other curves may be distinctly nonlinear. The dose-response curve for liver tumours induced as a result of exposure to gaseous vinyl chloride (six hours per day for two years) increases somewhat linearly at low doses and then tends to level off at higher doses (Maltoni, 1975). This plateau effect is thought to be due to saturation of the metabolic activation mechanism for vinyl chloride in the liver (Gehring *et al.*, 1978). Conversely, the dose-response curve for squamous-cell carcinomas of the nasal passage induced as a result of exposure to gaseous formaldehyde (also six hours per day for two years) shows a marked increase in response above 5.6 ppm (Swenberg *et al.*, 1983), possibly due to saturation of the mucociliary clearance mechanism. Finally, the dose-response curve for liver tumours resulting from ingestion of aflatoxin in the diet (Wogan *et al.*, 1974) increases at low doses but levels off at high doses as a response rate approaching 100% is reached.

These examples clearly illustrate that the dose-response curves for different chemical

Fig. 6.1 Examples of dose-response relationships for vinyl chloride (from Maltoni, 1975), aflatoxin (from Wogan *et al.*, 1974), 2-acetylaminofluorene (from Littlefield *et al.*, 1980a) and formaldehyde (from Swenberg *et al.*, 1983)



agents can be quite dissimilar. This kind of variation in shape is also encountered in radiation carcinogenesis (Ullrich *et al.*, 1976; Ullrich & Storer, 1979a,b; Ullrich, 1980). In the remainder of this section, we consider the problem of modelling the dose-response relationship for a given compound in order to obtain a more quantitative description of the data. We shall consider also the use of such models in estimating the response rate at doses not included in the experimental protocol.

Some mathematical models

The relationship between the crude proportion of animals developing tumours during the course of a bioassay and the level of exposure may be described by means of a statistical model relating the probability of tumour induction $P(d)$ and the dose d. Statistical or tolerance distribution models are based on the concept that each animal has its own tolerance to the test compound and will develop a lesion only if that tolerance is exceeded. The tolerances are presumed to vary within the population according to some tolerance distribution, $G(t)$, so that the probability that an animal selected at random will respond to dose d is given by

$$P(d) = \Pr\{\text{tolerance} \leq d\} = G(d). \tag{6.1}$$

A general class of tolerance distribution models is defined by

$$G(t) = F(\alpha + \beta \log t), \tag{6.2}$$

where F denotes some suitable cumulative distribution function and α and $\beta > 0$ are parameters (Chand & Hoel, 1974). Three commonly encountered models in this class are the probit, logit and extreme value, defined by

$$F(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-u^2/2)\, du, \tag{6.3}$$

$$F(x) = [1 + \exp(-x)]^{-1}, \tag{6.4}$$

and

$$F(x) = 1 - \exp\{-\exp(x)\} \tag{6.5}$$

respectively. Since under the extreme value model $G(t) = 1 - \exp(-at^b)$, where $a = \exp(\alpha)$ and $b = \beta$, this model is sometimes called the Weibull model (see Section 6.3).

Stochastic or mechanistic models are based on the concept that a toxic response is the result of the random occurrence of one or more fundamental biological events. Under the multi-hit model, for example, a response is assumed to be induced once the target tissue has been 'hit' by $k \geq 1$ biologically effective units of dose within a specified time period. Assuming that the number of hits during this period follows a Poisson process, the probability of a response is given by

$$P(d) = \Pr\{\text{at least } k \text{ hits}\} = 1 - \sum_{j=0}^{k-1} \exp(-\lambda d) \frac{(\lambda d)^j}{j!}, \tag{6.6}$$

where $\lambda d > 0$ denotes the expected number of hits during this period (Rai & Van Ryzin, 1981). When $k = 1$, the multi-hit model reduces to the one-hit model given by

$$P(d) = 1 - e^{-\lambda d}. \tag{6.7}$$

Incorporation of background response

All of the above models imply that the background response rate $P(0)$ is zero. In many cases, however, the response of interest will also occur spontaneously in control animals (Tarone *et al.*, 1981). Spontaneously occurring lesions may be assumed to arise as a result of a variety of biological mechanisms. Two commonly encountered assumptions in this regard are independence and additivity (Hoel, 1980). In the first case, spontaneous and induced lesions are presumed to occur independently of each other so that the probability of either a spontaneous or a treatment-induced response occurring is given by

$$P^*(d) = \pi_0 + (1 - \pi_0)P(d), \tag{6.8}$$

where $0 < \pi_0 < 1$ denotes the background rate of response (Abbott, 1925). Under additivity, spontaneously occurring effects are considered to be due to an effective background dose $\delta > 0$, with

$$P^*(d) = P(d + \delta). \tag{6.9}$$

Note that, with the one-hit model, the independence and additivity assumptions are indistinguishable. A combination of both independent and additive background may be

represented by the model

$$P^*(d) = \pi_0 + (1 - \pi_0)P(d + \delta). \tag{6.10}$$

Other models

A simple class of tolerance distribution models in which background response arises in neither an independent nor additive fashion is defined by

$$P(d) = F(\alpha + \beta d), \tag{6.11}$$

with $\beta > 0$ as in (6.2). When F follows the logistic distribution in (6.4), Cox (1970) has shown that the uniformly most powerful, unbiased test for positive slope β in the proportion of animals responding with increasing dose is the Cochran–Armitage test discussed in Chapter 5. Subsequently, Tarone and Gart (1980) demonstrated the robustness of this test by showing that it is the locally most powerful for any monotone increasing distribution F. However, because the model given by (6.11) involves only two parameters, it is less flexible than those given by (6.8), and may not always provide an adequate description of the observed dose-response curve.

Armitage–Doll multistage model

Perhaps the most widely applied model in the case of carcinogenesis is the Armitage–Doll multistage model (Armitage, 1982). In this case, it is assumed that a cell line progresses through k distinct stages prior to becoming cancerous and that the rate of occurrence of the ith change is of the form $\lambda_i = \alpha_i + \beta_i d$, where $\alpha_i > 0$ and $\beta_i \geq 0$ for $i = 1, \ldots, k$. The parameter α_i represents the spontaneous rate of occurrence of the ith change in the absence of any exposure, and the rate is supposed to be linearly dependent on dose through $\beta_i d$. The probability of a response within a given time period is then approximately

$$P^*(d) = 1 - \exp\left\{-c \prod_{i=1}^{k} (\alpha_i + \beta_i d)\right\}, \tag{6.12}$$

where $c > 0$ (Crump *et al.*, 1976). Noting that the exponent in (6.12) is a polynomial in dose, this model can also be viewed as

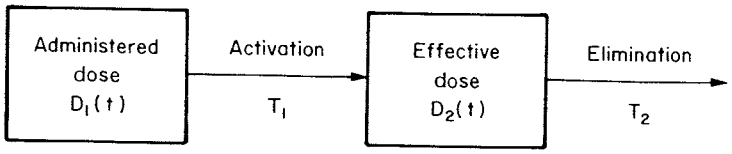
$$P^*(d) = 1 - \exp\left\{-\sum_{i=0}^{k} b_i d^i\right\}, \tag{6.13}$$

where the b_i are subject to certain nonlinear constraints (Krewski & Van Ryzin, 1981). For simplicity, however, the linear constraints $b_i \geq 0$ are often employed in practice, providing a more general model than given by (6.12) (Crump *et al.*, 1977).

Pharmacokinetic model

In many cases, a chemical will require some form of metabolic activation before it may exert its toxic effects (Cornfield, 1977). Rai and Van Ryzin (1983), for example,

Fig. 6.2 A simple pharmacokinetic model for metabolic fate of a compound



consider the simple compartmental model shown in Figure 6.2. Here, the administered dose $D_1(t)$ at time t undergoes a transformation T_1 to the activated form $D_2(t)$ and may then be eliminated *via* a second transformation T_2. Each transformation T_i is assumed to follow saturable Michaelis–Menten kinetics (Karlson, 1965, p. 80), with

$$\text{Rate } (T_i) = \frac{b_i D_i(t)}{c_i + D_i(t)}, \tag{6.14}$$

where $b_i, c_i > 0$ $(i = 1, 2)$. Assuming that the dose is administered at a constant rate k, the system satisfies the nonlinear differential equations

$$\frac{dD_1(t)}{dt} = k - \frac{b_1 D_1(t)}{c_1 + D_1(t)} \tag{6.15}$$

and

$$\frac{dD_2(t)}{dt} = \frac{b_1 D_1(t)}{c_1 + D_1(t)} - \frac{b_2 D_2(t)}{c_2 + D_2(t)}. \tag{6.16}$$

Under the steady state conditions $dD_i(t)/dt = 0$, it follows from (6.16) that

$$D_2^* = \frac{a_1 d}{1 + a_2 d}, \tag{6.17}$$

where $d = D_1(t)$ is constant, $a_1 = b_1 c_2 / b_2 c_1 > 0$ and $a_2 = (b_2 - b_1)/b_2 c_1 > -1/M$, with M being the highest dose $D_1(t)$ such that the rate of T_1 does not exceed the rate of T_2. If both transformations follow linear kinetics, then $a_2 = 0$ in (6.17) and the effective dose D_2^* is directly proportional to the administered dose d.

The probability of a response is assumed to depend only on the steady-state level of the effective dose D_2^* in (6.17), say

$$P(d) = F[D_2^*(d)]. \tag{6.18}$$

Taking F to be of the one-hit form with an additive constant

$$F(x) = 1 - \exp\{-(\alpha + \beta x)\} \tag{6.19}$$

yields the dose-response model

$$P(d) = 1 - \exp\left\{-\left[\theta_1 + \theta_2\left(\frac{d}{1 + \theta_3 d}\right)\right]\right\}, \tag{6.20}$$

where $\theta_1 = \alpha > 0$, $\theta_2 = \beta a_1 > 0$ and $\theta_3 = a_2 > -1/M$. This model can, depending on the rate coefficients governing T_1 and T_2, describe both downward bending curves, such as that noted for vinyl chloride (saturable activation), as well as 'hockey-stick' shaped curves, such as that for formaldehyde (saturable elimination). Thus, even though the

dose-response curve follows a simple one-hit model in terms of the effective dose D_2^*, a variety of curves may still arise as a result of the saturability of the activation and elimination steps.

More generally, Rai and Van Ryzin (1983) also consider F to be of the form

$$F(x) = 1 - \exp\{-(\alpha + \beta x^{\gamma})\}, \tag{6.21}$$

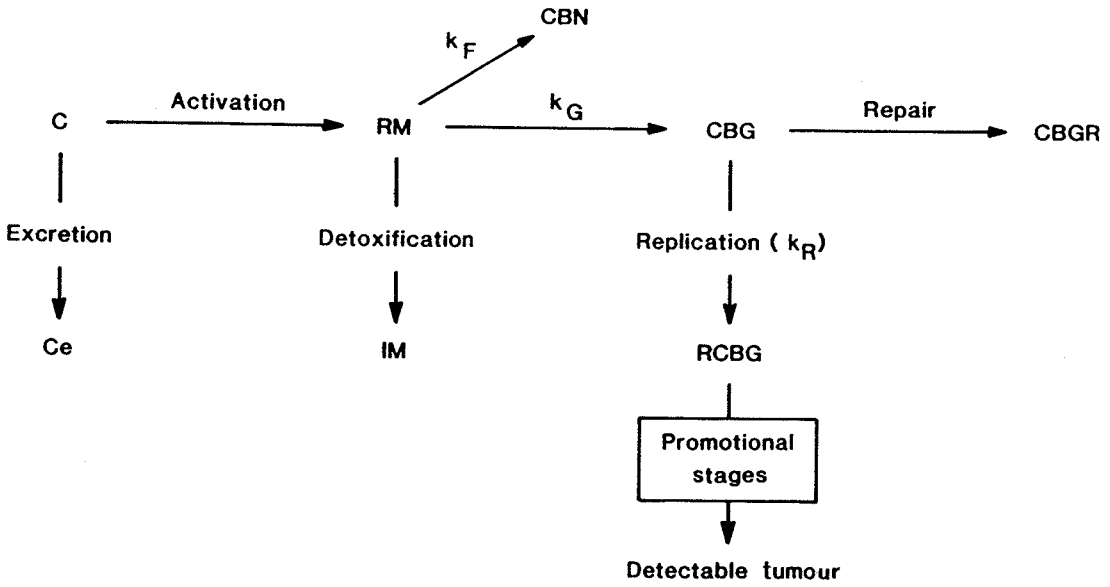
with $\gamma > 0$. In this case, the overall dose-response model in (6.20) becomes

$$P(d) = 1 - \exp\left\{-\left[\theta_1 + \theta_2\left(\frac{d}{1+\theta_3 d}\right)^{\theta_4}\right]\right\}, \tag{6.22}$$

where $\theta_4 = \gamma$. The use of the additional parameter θ_4 allows for additional curvature in the model $P(d)$ which cannot be accommodated by the pharmacokinetic parameters θ_2 and θ_3.

Gehring and Blau (1977) considered the somewhat more complex model shown in Figure 6.3. Once taken up by the body, a chemical C may be either eliminated immediately or activated to form a reactive metabolite RM. This in turn may be detoxified or react with cellular macromolecules to form covalently-bound genetic material (CBG). In this model, it is also possible that the reactive metabolite may be neutralized by nongenetic covalent binding (CBN). The covalently-bound genetic material may then be repaired (CBGR) or replicated (RCBG) resulting in the development of a genetic lesion.

Fig. 6.3 A more complex pharmacokinetic model for metabolic fate of a compound (from Gehring & Blau, 1977). C, chemical; RM, reactive metabolite; Ce, excreted chemical; IM, inactive metabolite; CBN, covalent binding, nongenetic; CBG, covalent binding, genetic; CBGR, repaired covalently bound genetic material; RCBG, retained genetic programme, critical and noncritical



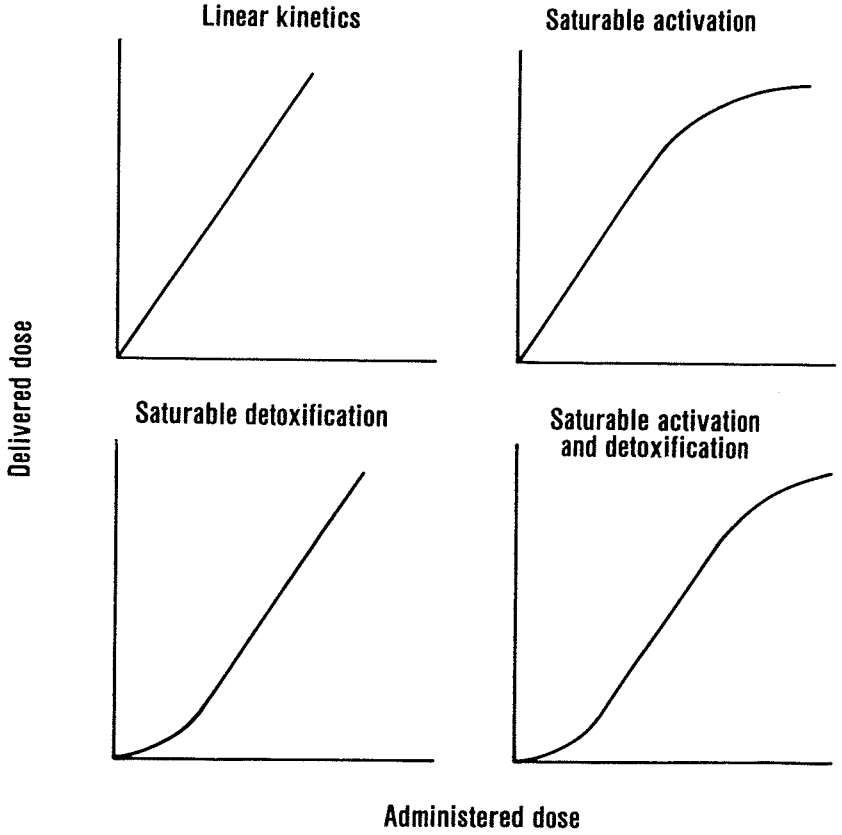
This model was subsequently examined in detail by Hoel *et al.* (1983). They assumed that all reactions are governed by linear first-order kinetics except for activation, detoxification and repair, which are allowed to be saturable in accordance with Michaelis–Menten kinetics. Since replication follows linear kinetics, the amount of damage is proportional to [CBG], the concentration of CBG. Under this model, [CBG] provides a measure of the effective dose.

Assuming that the dosing regimen is such that the concentration of a chemical [C] is proportional to the administered dose, one can use the two nonlinear differential equations describing the system first to solve for the concentration of a reactive metabolite [RM] in terms of [C], in steady state, and then to express [CBG] in terms of [RM].

Considering F to be of the one-hit form (6.18), Hoel *et al.* (1983) (see also Anderson *et al.*, 1980) noted that, under this model, the overall dose-response curve could assume any one of the four shapes shown in Figure 6.4, depending on which of the activation steps are saturable. With all processes being essentially linear, a linear dose-response curve results. If only repair or detoxification is saturable, the dose-response curve will be 'hockey-stick' shaped. If only activation is saturable, the shape is similar in form, although inverted. If more than one process is saturable, a combination of these shapes will occur.

Under this more complex model, the explicit expression for the overall dose-response model $P(d)$, as in (6.20), is very complicated. For the purposes of describing

Fig. 6.4 Relationship between delivered and administered dose and different pharmacokinetic conditions (from Hoel *et al.*, 1983)



actual data, the model in (6.22) may be fitted, however, using standard maximum likelihood procedures, as described below.

Maximum likelihood estimation

Whatever the specific parametric form of a dose-response model, the probability of observing a response at dose level d depends on some unknown parameters. In general, there may be p such parameters $\theta_1, \ldots, \theta_p$, summarized as a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, the probability of response being then denoted as $P^*(d; \boldsymbol{\theta})$. Subsequently, we shall outline how these unknown parameters can be estimated from the observed data by maximum likelihood methods. Suppose that a total of n animals are used in an experiment involving $I+1$ dose levels $0 = d_0 < d_1 < \cdots < d_I$ and that x_i of the n_i animals at dose d_i $(i = 0, 1, \ldots, I)$ develop tumours in the course of the study. Assuming that each animal responds independently of all other animals in the experiment, we find the likelihood of the observed outcome under any dose-response model $P^*(d; \boldsymbol{\theta})$ is given by

$$L(\boldsymbol{\theta}) = \prod_{i=0}^{I} \binom{n_i}{x_i} (P_i^*)^{x_i} (1 - P_i^*)^{n_i - x_i}, \tag{6.23}$$

where $P_i^* = P^*(d_i; \boldsymbol{\theta})$. Those values $\hat{\boldsymbol{\theta}}$ of the parameters $\boldsymbol{\theta}$ which maximize $L(\boldsymbol{\theta})$ are called 'maximum likelihood estimates'. Since maximization of $L(\boldsymbol{\theta})$ using direct analytical procedures is generally not possible, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is usually obtained using iterative numerical procedures. Under mild regularity conditions, it can be shown theoretically that $\hat{\boldsymbol{\theta}}$ is a consistent estimate for $\boldsymbol{\theta}$ as $n \to \infty$ (Krewski & Van Ryzin, 1981). Under these same conditions, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is approximately normally distributed with mean $\mathbf{0}$ and variance/covariance matrix $\mathbf{V} = [(v_{rs})]$. The elements of the inverse of this covariance matrix $\mathbf{V}^{-1} = [(v^{rs})]$ represent the Fisher information and are derived through the second derivatives of the likelihood function $L(\boldsymbol{\theta})$ in (6.23):

$$v^{rs} = \sum_{i=0}^{I} c_i \frac{\partial P_i^*}{\partial \theta_r} \frac{\partial P_i^*}{\partial \theta_s} \Big/ (P_i^* Q_i^*), \qquad (r, s = 1, \ldots, p), \tag{6.24}$$

where $c_i = \lim_{n \to \infty} n_i/n > 0$ and $Q_i^* = 1 - P_i^*$.

These theoretical results provide the basis for computing the estimated dose-response model $\hat{\mathrm{P}}(\mathrm{d}) = \mathrm{P}^*(\mathrm{d}; \hat{\boldsymbol{\theta}})$ and, when using the estimated covariance matrix of the parameter estimates, to calculate corresponding confidence intervals. For this purpose, the matrix $[(v^{rs})]$ is computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, the actual values of the maximum likelihood estimates, and using $c_i = n_i/n$. The usual chi-square statistic

$$\chi^2 = \sum_{i=0}^{I} (x_i - n_i \hat{P}_i^*)^2 / (n_i \hat{P}_i^* \hat{Q}_i^*) \tag{6.25}$$

may be used to assess the goodness-of-fit of $\hat{P}^*(d)$. Provided that the assumed model $P^*(d)$ is correct, the asymptotic distribution of this statistic is chi-square with $(I+1) - t > 0$ degrees of freedom.

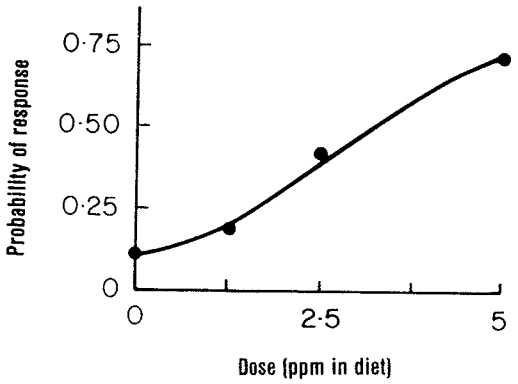
Estimation procedures for the multistage model (6.13) are complicated by the non-negativity constraints on the parameters b_i. Because of these constraints, the asymptotic distribution of the maximum likelihood estimators will generally not be normal (Guess & Crump, 1978). Similarly, the usual chi-square statistic (6.25) will generally be inapplicable. Nonetheless, efficient algorithms for obtaining the restricted maximum likelihood estimators have been developed by Crump (1984) and Hartley and Sielken (1978).

As an example, consider the following data on liver tumours induced in mice in a lifetime study of feeding dieldrin (Walker *et al.*, 1973).

Dose (ppm):	0	1.25	2.50	5.00
Response (x_i/n_i):	17/156	11/60	25/58	44/60

These data, along with the fitted extreme value model, assuming independent background, are shown in Figure 6.5. The maximum likelihood estimates of the parameters (± standard error) are $\hat{\alpha} = -2.46 \pm 0.50$, $\hat{\beta} = 1.66 \pm 0.35$ and $\hat{\pi}_0 = 0.106 \pm 0.024$. As may be expected with most monotonically increasing data sets, the model fits the data reasonably well, with no evidence of lack-of-fit provided by the chi-square statistic (p-value > 0.4). Similar results may be obtained with other data sets (Fig. 6.6 and Table 6.1) and with the other models discussed above. The fitted dose-response curves assuming additive background will also be similar, although in this case the likelihood surface is generally quite flat, and the parameters are thus less well determined.

Fig. 6.5 Dose-response curve for dieldrin-induced liver tumours in mice fitted under the extreme value model



Estimation of quantiles

In some applications, interest centres on the added risk over background which, under the assumption of independence (6.8), would be

$$\Pi(d) = \frac{P^*(d) - P^*(0)}{1 - P^*(0)}. \tag{6.26}$$

Fig. 6.6 Dose-response curves for eight compounds fitted under the extreme value model

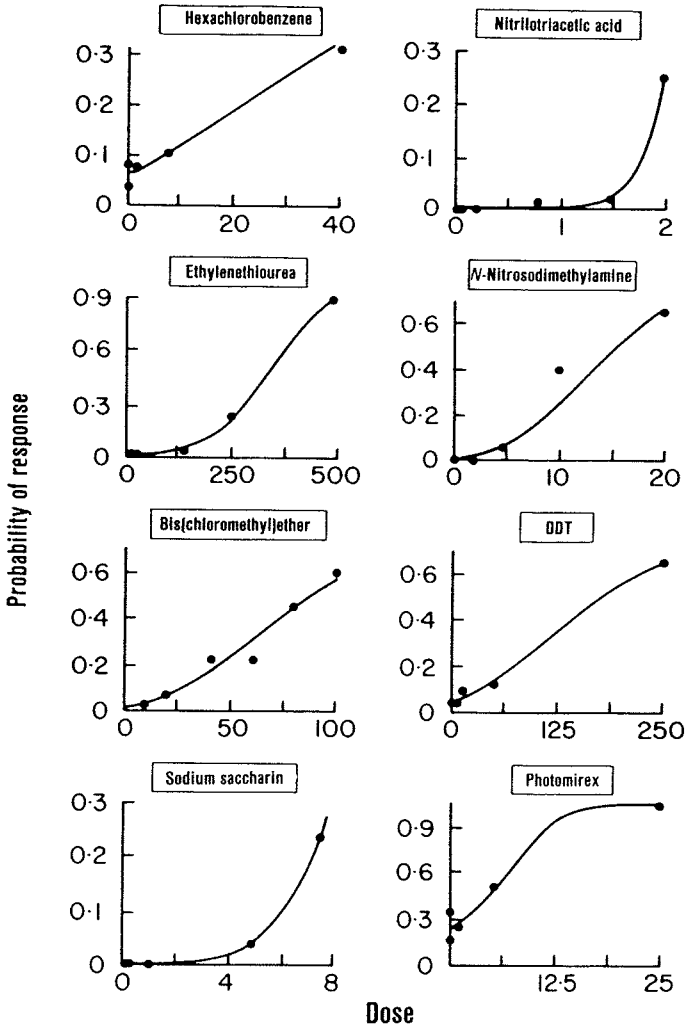


Table 6.1 Lesions induced by eight rodent carcinogens

Compound	Reference	Species	Tumour	Duration	Dose units
Hexachlorobenzene	Arnold *et al.* (1985)	Rat	Phaeochromocytoma	2 years	ppm in diet
Nitrilotriacetic acid	Food Safety Council (1978)	Rat	Kidney	2 years	ppm in diet
Ethylenethiourea	Graham *et al.* (1975)	Rat	Thyroid	2 years	ppm
N-Nitrosodimethylamine	Terracini *et al.* (1967)	Rat	Liver	120 weeks	ppm
Bis(chloromethyl)-ether	Kuschner *et al.* (1975)	Rat	Respiratory	lifetime	no. of exposures
DDT	Tomatis *et al.* (1972)	Mouse	Liver	130 weeks	ppm
Sodium saccharin	Taylor & Friedman (1974)	Rat	Bladder	2 years	% in diet
Photomirex	Chu *et al.* (1981)	Rat	Thyroid	21 months	ppm in diet

In the same way, one may wish to evaluate the dose level corresponding to a certain level of risk over background, q, say $(0<q<1)$, which would be $d_q = \Pi^{-1}(q)$. The maximum likelihood estimator of d_q is defined by $\hat{d}_q = \hat{\Pi}^{-1}(q)$, where $\hat{\Pi}(d) = [\hat{P}^*(d) - \hat{P}^*(0)]/[1 - \hat{P}^*(0)]$. Since $\sqrt{n}\,(\hat{d}_q - d_q)$ is asymptotically normally distributed with mean zero and variance

$$\sigma^2 = \left\{ \frac{\partial \Pi}{\partial d} \Big|_{d=d_q} \right\}^{-2} \sum_{r=1}^{p} \sum_{s=1}^{p} \frac{\partial \Pi}{\partial \theta_r} \frac{\partial \Pi}{\partial \theta_s} v_{rs}, \tag{6.27}$$

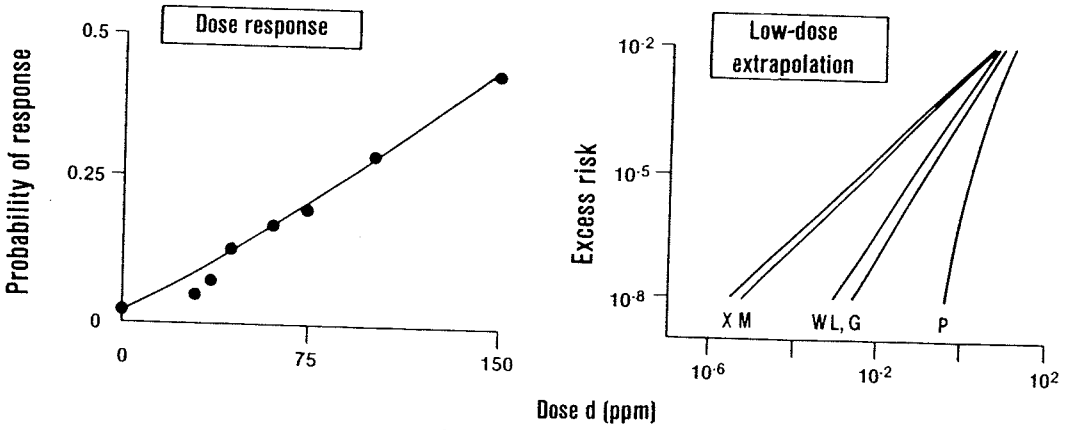
an approximate $100(1-\alpha)\%$ confidence interval for d_q is given by

$$\hat{d}_q \pm z_{\alpha/2} \hat{\sigma} / \sqrt{n}, \tag{6.28}$$

where $z_{\alpha/2}$ denotes the $100(1-\alpha/2)$ percentile of the standard normal distribution, and $\hat{\sigma}$ is an estimate of σ obtained by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ in (6.27). Other possible confidence limit procedures, including those based on the asymptotic distribution of the log-likelihood (Cox & Hinkley, 1974, p. 343), are reviewed by Crump and Howe (1985).

Two applications of doses associated with certain levels of risk over background are of particular interest. First, Mantel and Bryan (1961) proposed the use of some suitably low risks (for example, $q = 10^{-6}$) as a means of defining a 'virtually safe' level of exposure in the absence of a threshold in the dose-response curve. It is now widely recognized that estimation of such extreme risks is subject to considerable model uncertainty. Extrapolation of the data on 2-AAF-induced liver tumours shown in Figure 6.7 (Littlefield *et al.*, 1980a), for example, using the probit, logit, extreme value, multihit and multistage models, yields estimates of virtually safe doses spanning several orders of magnitude. Because of this uncertainty, it has been proposed that some form of linear extrapolation be used to obtain a lower limit on such extreme doses. The assumption of low-dose linearity follows, in fact, immediately from the assumption of additive background, since in that case a simple Taylor expansion shows

Fig. 6.7 Dose-response curve for 2-acetylaminofluorene(2-AAF)-induced liver tumours in mice fitted under the Weibull model and performance of six extrapolation procedures (from Krewski *et al.*, 1984b). *X*, linear extrapolation; *M*, multistage model; *W*, Weibull model; *L*, logit model; *G*, gamma multi-hit model; *P*, probit model



that

$$\Pi(d) \simeq f(0)d \tag{6.29}$$

for small d, where $f(d) = \partial\Pi(d)/\partial d$. One simple procedure which may be used for this purpose is to extrapolate linearly from some higher quantile such as $d_{0.01}$ (Van Ryzin, 1980). (For the 2-AAF data, this form of linear extrapolation yields results close to those predicted by the multistage model.) A similar form of linear extrapolation has also been proposed by the World Health Organization (1984, p. 50).

The second application of the above concept, which can also be termed as estimating certain quantiles of the dose-response curve, is to derive a measure of carcinogenic potency. The measure proposed by Clayson *et al.* (1983) is defined by

$$C_q = K - \log_{10} d_q, \tag{6.30}$$

where the dose d is expressed in μmol/kg body weight/day. The logarithm of dose is used to put the index on an order-of-magnitude basis, with the minus sign associating large values of C_q with low values of d_q. The constant K is set equal to 7 in order that C_q will usually lie in the range 1–10. By choosing a moderate value of q, say $0.10 \leq q \leq 0.50$, the model dependency encountered in estimating lower quantiles is avoided.

Despite its simplicity, the index C_q seems to provide a useful method of ranking animal carcinogens. Values of C_q with $q = 0.25$ for a selection of suspected and well-known animal carcinogens are shown in Table 6.2. Saccharin, the carcinogenicity of which has been widely debated, is assigned a potency index of 1.8, whereas the highly potent 2,3,7,8-tetrachlorodibenzo-*para*-dioxin (TCDD) has a value of 9.1. For a more complex ranking system that takes into account other factors, such as the spectrum of neoplasia induced and genotoxicity, the reader is referred to Squire (1981) and the related discussion by Crump (1983) and Theiss (1983).

Other quantitative measures of carcinogenic potency have also been proposed. Historically, Twort and Twort (1930, 1933) and Iball (1939) proposed several measures of potency in an attempt to summarize the data obtained from their experimental studies. For example, one of their measures was based on the time at which 25% of the

Table 6.2 Potencies $C_{0.25}$ of some selected compounds

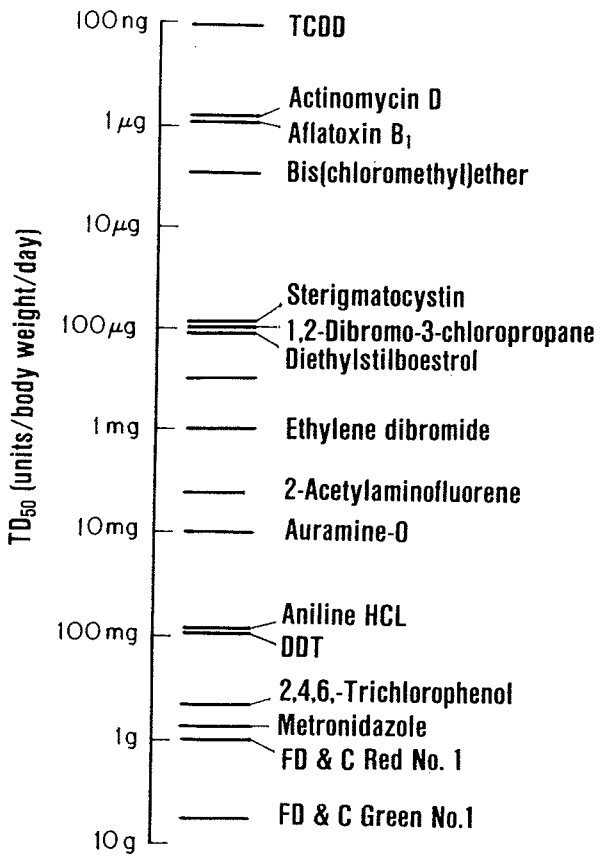
Compound (Reference)	Species	Site	Potency $C_{0.25}$
Saccharin (Scientific Review Panel, 1983)	Rat	Bladder	1.8
2-Acetylaminofluorine (Littlefield *et al.*, 1980a)	Mouse	Bladder	4.3
		Liver	4.4
DDT (Thorpe & Walker, 1973)	Mouse	Liver	5.0
Aflatoxin (Wogan *et al.*, 1974)	Rat	Liver	8.6
Dioxin (National Toxicology Program, 1982)	Rat	Thyroid	9.1

animals would have developed tumours. Irwin and Goodman (1946) subsequently considered other related measures of potency, and noted that the different indices tended to give somewhat similar results. Bryan and Shimkin (1943) suggested the use of the dose required to induce tumours in 50% of the exposed animals, the quantity on which Meselson and Russel's (1977) index is defined.

More recently, Jones *et al.* (1981) considered the use of the time until 50% of the exposed animals would die from the tumour of interest as a measure of potency. Crouch and Wilson (1981) took the slope parameters in the one-hit model in (6.7) as a measure of potency. Noting that the probability of tumour induction $P(d) \simeq \lambda d$ for low levels of exposure d, Crouch and Wilson also proposed using the value of λ as a means of estimating the response rate P at a given dose d. This will be reasonable when the one-hit model, which is essentially linear even at moderate doses, provides an adequate description of the observed dose-response curve, but will be less satisfactory when the dose-response curve is highly nonlinear.

Sawyer *et al.* (1984) proposed a potency index based on the TD_{50}, or dose estimated to induce tumours in 50% of the exposed subjects. The method provides for the effects of both dose and time, and, like the method of Crouch and Wilson, is based on a simple exponential or one-hit model for the effects of dose. In cases where the time to tumour occurrence may not be directly observable, moreover, Sawyer *et al.* used the time to death with tumour present as a surrogate for the observable failure time (see Peto *et al.*, 1984, for further discussion of this point).

Fig. 6.8 Range of carcinogenic potency in male rats (from Gold *et al.*, 1984)



The index proposed by Sawyer *et al.* (1984) has recently been calculated using an extensive data base of known animal carcinogens compiled by Gold *et al.* (1984). Expressing the TD_{50} in terms of the daily intake of the compound relative to total body weight, this analysis revealed potencies varying over more than seven orders of magnitude (Fig. 6.8). Only a few nanograms of TCDD, for example, were required to induce a 50% tumour occurrence rate during the course of a rodent lifetime, whereas several grams of the food colours FD & C Red No. 1 and FD & C Green No. 1 were required to elicit the same rates of response.

6.3 Time-to-tumour models

Why use time-to-tumour models?

There are a number of situations in which the use of models based on time-to-tumour information has advantages. For example, when survival differs very greatly in different groups, one may be able to make fuller use of the data. In the extreme case, where near the end of an experiment there are survivors in only one group, most of the methods described so far cannot make use of any tumours detected in that group, as there is nothing with which to compare them. Where a parametric time-to-tumour model can be fitted, however, these data can be used to make more precise group comparisons.

Time-to-tumour models may allow one to present the results fully in a concise manner, which allows direct comparison with results from other comparable experiments. As we shall see, it is often possible to fit parametric models in which certain parameters, common to all groups, describe the general shape of the tumour incidence or survival curve, while a further single parameter, estimated separately for each group, describes the strength of the treatment effects. If common shape parameters are used, the strength parameters can be used to compare directly the results of different experiments, which would not be possible for methods based on testing a null hypothesis.

Graphical presentation of the observed and fitted time curves for tumour onset in treatment and control groups can be used to indicate whether there might be any interaction between the effect of treatment and time that should be investigated in more detail. The null hypothesis methods may, for example, not pick up a situation in which treatment increases tumour incidence early on in the study but decreases it later.

Time-to-tumour models are also of particular use in experiments specifically aimed at characterizing the mode of action of the carcinogen being tested. The shape of the time-response curve may assist in indicating whether the animal model used is apposite to the human situation where information on time-response may also be available. Furthermore, especially in more complex designs such as stopping experiments, it may give insight into whether the carcinogen had initiating or promoting action.

Finally, as noted in the previous section, time-to-tumour models may allow a more reliable method of low-dose extrapolation than those based on percentages of animals with tumours, especially where high doses markedly reduce mortality.

There are two main disadvantages of time-to-tumour models. One is the need to

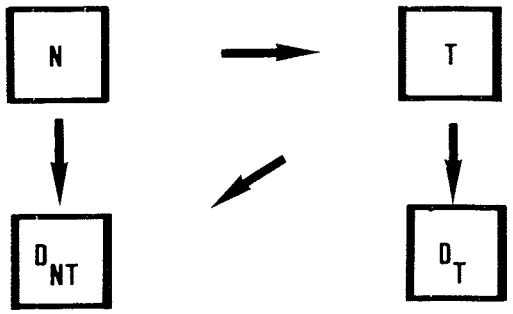
make the additional assumption, as compared with nonparametric methods, that tumour incidence follows a particular parametric relationship with time. A poor choice of relationship can affect conclusions as to the carcinogenicity of the treatment under test. The second is that, generally, far more extensive computing is required.

A general formulation

In the past, time-to-tumour models have mainly been applied to visible tumours. However, in recent years a number of attempts have been made to consider the more general situation where tumours may also be fatal or incidental (for example, Hartley & Sielken, 1977; Kodell *et al.*, 1982a; Peto *et al.*, 1984).

A simple way to model a long-term animal experiment is illustrated in Figure 6.9, where the possible states an animal can be in are given as boxes, and the connecting arrows indicate possible transitions (Kodell & Nelson, 1980). In this model, an animal starts in a normal disease-free state (N) and may at some time either develop a tumour (T) – or perhaps more precisely be in a state where a tumour can be detected – or die from a cause unrelated to the tumour of interest (D_{NT}). An animal with a tumour may also subsequently die from this unrelated cause, or may die because of the tumour (D_T).

Fig. 6.9 Illustration of illness and death states with possible transitions in rodent bioassay; N, normal; T, tumour; D_T, death from tumour; D_{NT}, death not from tumour (from Kodell & Nelson, 1980)



This model is aimed at the types of observation which can arise from a long-term experiment. The assumption that a transition is made from a normal state to a state where the tumour is detectable is simplistic, inasmuch as it does not take into account details of the underlying biological processes. However, it would not in practice be possible to identify all of these states, even by increasing the number of investigations. Even in the model proposed, the information on cause of death, required to distinguish between D_T and D_{NT} when a tumour is present, is often difficult or impossible to obtain.

Three random variables may be used to describe the above model:

(a) X: time to onset of tumour, or transition time from the normal state (N) to the tumour-bearing state (T)
(b) Y: time to death due to tumour, or transition time from the normal state (N) to the death-from-tumour state (D_T)

(c) Z: time to death from an unrelated cause, or transition time from the normal state (N) to the death-not-from-tumour state (D_{NT}).

The random variable Z is not of major concern in drawing statistical conclusions on the process of carcinogenesis and is not considered further, except with regard to assumptions about Z needed to form the likelihood functions upon which statistical inferences may be based. The two random variables X and Y, however, which have to satisfy the condition $X \leq Y$, are of major interest for the statistical inference.

Let $G_X(t)$ and $G_Y(t)$ be the distribution functions of X and Y, or $S_X(t) = 1 - G_X(t) = \text{pr}(X \geq t)$ and $S_Y(t) = 1 - G_Y(t) = \text{pr}(Y \geq t)$ the survivor functions of X and Y.

In addition, we consider for any index X or Y the density $f(t) = \text{d}G(t)/\text{d}t$, the hazard function $\lambda(t) = f(t)/S(t)$ and the cumulative hazard function, $\Lambda(t) = \int_0^t \lambda(u)\,\text{d}u$. Note that $S(t) = \exp[-\Lambda(t)]$.

The hazard function $\lambda(t)$ is an extremely useful tool for modelling distributions of random variables which represent the time to a well defined event. It can also be defined as the conditional probability that the event of interest occurs at time t, if it has not occurred before that time. Let X be the random variable of interest:

$$\lambda_X(t) = \lim_{\Delta t \to 0} \frac{\text{pr}(t \leq X < t + \Delta t \mid X \geq t)}{\Delta t} = \frac{f_X(t)}{S_X(t)}.$$

The hazard function is therefore also denoted the age-specific failure rate.

In a long-term animal experiment, considering visible or occult tumours, four different events can be observed at a certain time point:

A. appearance of a visible tumour
B. death caused by the tumour of interest (fatal context)
C. death from unrelated cause, tumour of interest present (incidental context)
D. death without tumour of interest.

It has to be noted that deaths under C and D, from a cause unrelated to the tumour of interest, can occur in scheduled sacrifices. Thus, animals killed because of the experimental design or other external reasons represent observations of type C and D, depending on whether the tumour of interest is seen or not seen in the particular animal.

The contribution to the likelihood function of the four types of observations, expressed in terms of the two random variables X and Y, are as follows:

A. density of X: $f_X(t)$
B. density of Y: $f_Y(t)$
C. $\text{pr}(X < t \leq Y)$: $S_Y(t) - S_X(t)$
D. survivor function of X: $S_X(t)$.

In a given experiment, let $t_1, t_2, \ldots, t_K$ be the distinct times at which events of the above type are observed; a_k, b_k, c_k and d_k ($k = 1, \ldots, K$) are the number of events of type A, B, C or D at time t_k. The latent-failure-time approach is taken in the formation of likelihoods below. Although this is the most common approach to the competing

risks problem, strong yet unverifiable assumptions are required to form the likelihoods for occult tumour data (Kalbfleisch *et al.*, 1983).

We consider four typical types of experiments:

(i) *Visible tumours*

As the time of onset is observable directly, there is no interest in the random variable Y; only events of the type A and D are observed, leading to a likelihood function

$$L_1 = \prod_{k=1}^{K} \{f_X(t_k)\}^{a_k} \{S_X(t_k)\}^{d_k}.$$

With $f_X(t) = \lambda_X(t) S_X(t)$ and $S_X(t) = \exp(-\Lambda_X(t))$, the log-likelihood function is often expressed as

$$LL_1 = \sum_{k=1}^{K} \{a_k \log[\lambda_X(t_k)] - (a_k + d_k)\Lambda_X(t_k)\},$$

using only the hazard function and its cumulative version, which are often the basic entities on which parametric models are formulated.

(ii) *Occult tumours – all tumours observed in a fatal context*

In this special situation, where no tumours are found in animals dying from unrelated causes, only events of types B and D are observed. The likelihood function is

$$L_2 = \prod_{k=1}^{K} \{f_Y(t_k)\}^{b_k} \{S_Y(t_k)\}^{d_k}$$

or, as above, the log-likelihood function

$$LL_2 = \sum_{k=1}^{K} \{b_k \log[\lambda_Y(t_k)] - (b_k + d_k)\Lambda_Y(t_k)\},$$

which are formally identical to L_1 and LL_1, respectively.

(iii) *Occult tumours – all tumours observed in an incidental context*

In this case $S_Y(t) = 1$, as no deaths due to the tumour of interest are occurring, only events of the types C and D being observed. In order to form a likelihood for this case it must be assumed that tumour-bearing and tumour-free animals of the same age have identical hazard functions for death unrelated to tumour (that is, intercurrent mortality). The likelihood function under this assumption is

$$L_3 = \prod_{k=1}^{K} \{1 - S_X(t_k)\}^{c_k} \{S_X(t_k)\}^{d_k}.$$

The log-likelihood function expressed in terms of $\Lambda_X(t)$ would be

$$LL_3 = \sum_{k=1}^{K} \{c_k \log[1 - \exp(-\Lambda_X(t_k)] - d_k \Lambda_X(t_k)\}.$$

(iv) *Occult tumours – observed in fatal or incidental context*

In this most general case, events of the types B, C and D are observed. In order to form a likelihood for this case, it must, of course, be assumed that each tumour can be reliably classified as either fatal or incidental. In addition, it must be assumed that tumour-bearing and tumour-free animals of the same age have identical hazard functions for death unrelated to tumour. The likelihood function under these assumptions is

$$L_4 = \prod_{k=1}^{K} \{f_Y(t_k)\}^{b_k} \{S_Y(t_k) - S_X(t_k)\}^{c_k} \{S_X(t_k)\}^{d_k}.$$

Kodell *et al.* (1982a) pointed out that

$$S_Y(t) - S_X(t) = [1 - Q(t)]S_Y(t),$$

with $Q(t) = S_X(t)/S_Y(t) = \text{pr}(X > t \mid Y > t)$ being the conditional probability of tumour onset after time t, given tumour-free survival through time t. Even under the strong assumptions noted above, the function $S_Y(t)$ does not have a simple interpretation in terms of time to death due to tumour alone (that is, independent of the influence of time to onset of tumour).

From this it follows that

$$L_4 = \prod_{k=1}^{K} \{f_Y(t_k)\}^{b_k} \{S_Y(t_k)\}^{c_k + d_k} \{1 - Q(t_k)\}^{c_k} \{Q(t_k)\}^{d_k},$$

which can be written as the product of two parts. One,

$$L_4^{(1)} = \prod_{k=1}^{K} \{f_Y(t_k)\}^{b_k} \{S_Y(t_k)\}^{c_k + d_k},$$

depends on $S_Y(t)$ only; the other,

$$L_4^{(2)} = \prod_{k=1}^{K} \{1 - Q(t_k)\}^{c_k} \{Q(t_k)\}^{d_k},$$

on $Q(t)$ only.

The two components of the log-likelihood function are

$$LL_4^{(1)} = \sum_{k=1}^{K} \{b_k \log[\lambda_Y(t_k)] - (b_k + c_k + d_k)\Lambda_Y(t_k)\},$$

which is similar to LL_2 above; and, taking $\Lambda_Q(t) = -\log Q(t)$, the second part

$$LL_4^{(2)} = \sum_{k=1}^{K} \{c_k \log[1 - \exp(-\Lambda_Q(t_k)] - d_k \Lambda_Q(t_k)\},$$

which is similar to LL_3.

In the case where all tumours are fatal or visible, nonparametric estimation of the survivor functions can be made by the Kaplan–Meier method, as discussed in Section 5.3. Where all tumours are incidental, nonparametric estimation can be done by the method of Hoel and Walburg, as discussed in Section 5.5. Kodell *et al.* (1982a), who consider the fourth, most general case where tumours may be either fatal or incidental, note that the Kaplan–Meier estimator can be applied to the first part, $L_4^{(1)}$ or $LL_4^{(1)}$, to estimate $S_Y(t)$. With the assumption that the ratio $Q(t)$ is monotonically decreasing, they then derive a nonparametric maximum likelihood estimation for $S_X(t)$ as well.

Dinse and Lagakos (1982) weaken the restriction of monotonicity of $Q(t)$ and propose to estimate $S_X(t)$ in the class of all survivor functions that are stochastically smaller than $S_Y(t)$. The Kaplan–Meier estimate of $S_Y(t)$ is used as a starting point in an iterative approach. Given this estimate, $LL_4^{(2)}$ is maximized for $Q(t)$ under the restriction that the resulting estimate must be a monotone decreasing survival function. This is enforced by applying techniques of isotonic regression. The estimate for $S_X(t)$ thus derived is then inserted into the original likelihood function, which is maximized again for $S_Y(t)$. This process is iterated until convergence.

Turnbull and Mitchell (1984) have proposed a simpler algorithm, by addressing the problem in terms of the joint distribution of time of onset of tumour (X) and time of death from tumour (Y), which have marginal distributions $1 - S_X(t)$ and $1 - S_Y(t)$. They point out that the joint distribution can have positive mass only on a finite set of disjoint intervals in the (x, y)-plane, which can be constructed from the observations. They then use the EM-algorithm to estimate the probability masses to be attributed to these intervals. The marginal distributions, and hence the survivor functions $S_X(t)$ and $S_Y(t)$, can then be derived from the estimate of the joint distribution.

Choice of parametric model

The ideal model to use would clearly be one derived from sound biological theories of the carcinogenic process. In practice, of course, understanding of the carcinogenic process is far from perfect, but one can still aim for a model which both fits in with available knowledge and fits observed data at least reasonably well. By far the most attention has been given to the multistage model and, in the case of carcinogenesis experiments in which the dose has been applied continuously or at regular intervals, to the use of the Weibull distribution derived from it. For this reason, and also because it has proved satisfactory for analysis of a considerable number of data sets, we will consider this in detail first and turn our attention to other models only at the end of this section.

Multistage models and derivation of the Weibull distribution

Armitage and Doll (1954) observed that, in humans, the age-specific incidence rates of many types of cancer are proportional to a power of age (or time from first exposure) and showed that this result would be expected under a multistage model. This model makes the following simplifying assumptions:

(i) that there is a large and constant number (N) of cells at risk,
(ii) that all the cells start in an identical state,

(iii) that at least one of them has to undergo a fixed number κ of stages or transformations before a tumour appears, and

(iv) that any cell in a given stage has the same very small, but constant probability per unit time of commencing the transformation into the next stage (kinetic rate-constants $\delta_1, \delta_2 \cdots \delta_\kappa$).

Under these assumptions, the number of cells that have undergone the first $\kappa - 1$ stages by time t is given by

$$N_{\kappa-1} = N \int_0^t \delta_{\kappa-1} \int_0^{u_{\kappa-1}} \delta_{\kappa-2} \cdots \int_0^{u_2} \delta_1 \, du_1 \, du_2 \cdots du_{\kappa-1} = \frac{N\delta_1\delta_2 \cdots \delta_{\kappa-1} t^{\kappa-1}}{(\kappa-1)!}.$$

If it is further assumed that each transformation takes a constant time $(w_1, w_2 \cdots w_\kappa)$, the formula for the incidence rate $I(t)$ becomes

$$I(t) = \beta\kappa(t-w)^{\kappa-1},$$

where $\beta = N\delta_1\delta_2 \cdots \delta_\kappa/(\kappa!)$ and $w = \sum_{j=1}^{\kappa} w_j$ is the sum of the constant times of the transformations, which can be seen as the minimal time to tumour. This is a particular way of parametrizing the Weibull distribution. The survivor function is then

$$S(t) = \exp(-\beta(t-w)^\kappa). \tag{6.31}$$

While this model is clearly an over-simplification, it may be expected that if cancer mechanisms in animals and humans are of this type, incidence rate data from laboratory animals, which are often inbred and presumably therefore more homogeneous, may follow a Weibull distribution even more clearly than for humans. The suggestion of using Weibull distributions to analyse continuous carcinogenesis experiments in animals where time-to-tumour is observable was first made by Pike (1966).

From a study of the way in which the formula was derived, it should be clear that κ and w are inherent properties of the process being studied and should not vary between groups within an experiment studying the same cancer type in the same species of animals. The parameter β, on the other hand, should be affected by treatment, assuming that the effect of continuous application of a carcinogen will be to alter at least one of the kinetic rate constants. In an experiment with several dose groups, $i = 1, \ldots, I$, say, one would consider different parameters $\beta_1, \ldots, \beta_I$ for the Weibull distribution of time to tumour for the animals from the respective groups, or, as will be outlined later, one would model the dependence of the β_i on the dose level or other covariates.

Fit of Weibull distributions to visible tumour data

Weibull distributions have been used mainly in the literature to analyse visible tumour data, often from mouse skin-painting experiments widely used in the 1960s and 1970s to evaluate the carcinogenicity of tobacco-smoke condensates and of other chemicals. In the next few sections we consider these applications in some detail,

before turning to more recent work using Weibull distributions in the analysis of tumours not visible in life. Some of the ideas used in the visible tumour analyses (for example, for significance testing of treatment effects) have natural analogues for nonvisible tumours, but are described in detail only for the former situation.

Maximum likelihood estimation of the parameters β, κ and w of the Weibull distribution is discussed in detail by Peto and Lee (1973). The contribution of the log-likelihood LL_1 of an animal dying without a tumour at time t' is given by

$$\log[S_X(t')] = -\beta(t' - w)^{\kappa},$$

while that of an animal getting a tumour at time t'' is given by

$$\log[\lambda_X(t'')S_X(t'')] = \log\beta + \log\kappa + (\kappa - 1)\log(t'' - w) - \beta(t'' - w)^{\kappa}.$$

To illustrate the fitting of Weibull models, consider the cigar-smoke condensate data (Section 4.2). The maximum likelihood estimates (and standard errors obtained from the inverse of the information matrix) of w, κ and β for the three cigar-smoke condensate groups are given in Table 6.3.

Table 6.3 Parameter estimates of Weibull models fitted to the three groups, from data on cigar-smoke condensate

	Low dose	Middle dose	High dose
$\hat{w} \pm SE$	19.5 ± 8.6	30.4 ± 10.1	34.5 ± 4.9
$\hat{\kappa} \pm SE$	1.7 ± 0.6	2.3 ± 0.8	2.0 ± 0.5
$(\hat{\beta} \pm SE) \times 10^4$	2.13 ± 5.95	0.62 ± 2.25	3.17 ± 6.61
Log-likelihood	-131.522	-164.084	-180.537

Because the estimates of κ and w have not been constrained to be equal in the three groups, the estimates in the table do not provide a particularly useful summary of the relationship between cigar-smoke condensate dose and tumour incidence. The standard error estimates are relatively large due to the fact that the likelihood functions are quite flat around their maxima. Accordingly, when making inferences about the Weibull parameters, likelihood ratio tests should be employed rather than tests based on the asymptotic normality of the maximum likelihood estimators.

Estimated percentiles often provide a useful summary of the Weibull time-to-tumour curves. Estimates of percentiles can be obtained directly from the estimated tumour incidence curves, and standard errors of estimated percentiles can be derived using the delta method (Miller, 1981b, pp. 25–27). Estimates of the 25th percentile (denoted T_{25}) and the 50th percentile (that is, the median, denoted T_{50}) for the three cigar-smoke condensate groups are as follows:

	Low dose	Middle dose	High dose
$T_{25} \pm SE$	84.5 ± 8.3	71.4 ± 4.2	62.5 ± 3.1
$T_{50} \pm SE$	127.6 ± 19.2	90.8 ± 4.8	77.5 ± 3.5

The percentiles provide a much more useful summary of the Weibull curves than was provided by the estimates of the three Weibull parameters. They indicate that the tumours are appearing earlier with increasing dose of cigar-smoke condensate. The standard errors of the percentile estimates, with the exception of the median for the low-dose group, are less than 10% of thei. corresponding estimates. The standard error for the estimated median of the low-dose group is large because the estimate, in this case, represents an extrapolation outside the observed time period.

To formulate the full log-likelihood function LL_1, in the situation of common values for w and κ but different parameters β_i $(i = 1, \ldots, I)$ for the different experimental groups, we have to extend our notation slightly. Let a_{ki} and d_{ki} be the number of events of type A and D observed at time t_k in group i $(k = 1, \ldots, K; i = 1, \ldots, I)$. Then we have

$$LL_1 = \sum_{k=1}^{K} \sum_{i=1}^{I} [a_{ki}\{\log \beta_i + \log \kappa + (\kappa - 1)\log(t_k - w)\} - (a_{ki} + d_{ki})\beta_i(t_k - w)^{\kappa}].$$

Maximization of this log-likelihood function, which depends on $\beta_1, \ldots, \beta_I$, κ and w, is not straightforward, but can be achieved satisfactorily by a modified Newton–Raphson iterative procedure.

For the cigar-smoke condensate data, the maximum likelihood estimates of $w(\pm SE)$ and $\kappa(\pm SE)$ are 17.5(±6.45) and 2.8(±0.5), respectively, and the maximum likelihood estimates of β_i are as follows:

	Low dose $(i = 1)$	Middle dose $(i = 2)$	High dose $(i = 3)$
$(\hat{\beta}_i \pm SE) \times 10^6$	2.28 ± 5.22	4.50 ± 10.22	7.69 ± 17.28

The parameter estimates now are indicative of an increasing tumour rate with increasing dose of cigar-smoke condensate. The log-likelihood for this model, with common w and κ but different β_i, is −481.173. Fitting a model with a common β in addition to common w and κ gives a log-likelihood of −491.353. Thus, the likelihood ratio test statistic for $H_0: \beta_1 = \beta_2 = \beta_3$ is computed as $2(491.353 - 481.173) = 20.36$. Comparing this computed value to a table of percentiles for the chi-square distribution with two degrees of freedom (in general, $I - 1$ degrees of freedom) we find that $p = 0.00004$, indicating a strong effect of cigar-smoke condensate on tumour incidence.

The sum of the log-likelihoods from the initial fits of separate Weibull models to each dose group is $-131.522 - 164.084 - 180.537 = -476.143$. Thus, a test of the null hypothesis that $w_1 = w_2 = w_3$ and $\kappa_1 = \kappa_2 = \kappa_3$ can be based on the likelihood ratio test statistic, which is computed as $2(481.173 - 476.143) = 10.06$. Comparing this computed value to a table of percentiles for the chi-square distribution with four degrees of freedom (in general, $2I - 2$ degrees of freedom), we find that $p = 0.039$, indicating some evidence of heterogeneity. It is interesting to note that in spite of this evidence that the assumption of common w and κ may be invalid, the likelihood ratio test statistic for $H_0: \beta_1 = \beta_2 = \beta_3$ differed only slightly from the log-rank test statistic calculated for the cigar-smoke condensate data in Section 5.6 (that is, $X^2_{LH} = 20.16$).

Peto and Lee (1973) discuss various possibilities for proceeding in the case of known or unknown parameters κ and β. Of particular importance is the situation in which the main interest is in between-treatment comparison, that is, in the relative magnitude of the β_i values. Then, rather than carry out full estimation of κ, w and β_i, it is reasonably satisfactory, and much simpler computationally, to fix κ and w values from previous experience and compute the β_i from the formula

$$\beta_i = s_i / v_i,$$

where $s_i = \sum_{k=1}^{K} a_{ki}$ is the number of animals bearing tumours in the ith group and $v_i = \sum_{k=1}^{K} (t_k - w)^{\kappa}$, the summation being over both times of tumour and times of death without tumour in group i. Pike (1966) has demonstrated that the ratio of β's between different groups is virtually independent of the actual values of κ and w chosen, provided the (κ, w) pair is not too far from the best fitted values. Where κ and w are known, the asymptotic variance of β_i is given by $\operatorname{var} \beta_i = \beta_i^2 / s_i$.

Goodness-of-fit to the Weibull distribution can be tested by dividing the experimental time period into J intervals $(T_0 = 0, T_1]$, $(T_1, T_2], \ldots, (T_{J-1}, T_J]$. Within each interval one compares the observed number of animals in the ith group first developing a tumour in the jth interval O_{ij} with the number expected E_{ij}. E_{ij} is calculated by $E_{ij} = \beta_i v_{ij}$, where j refers to the interval $(j = 1, \ldots, J)$, and v_{ij} is calculated by summing, for each animal of group i surviving and tumour-free at T_{j-1}, the term $(t^* - w)^{\kappa} - (T_{j-1} - w)^{\kappa}$, where $t^* = \min(T_j, t_k)$, that is, the time to death or tumour for animals that experience one of these events during the jth interval, or the upper limit of the interval for the remainder. If the numbers of tumours are too small per group, it will often be useful to combine these O_{ij} and E_{ij} values over groups for each time interval:

$$O_j = \sum_{i=1}^{I} O_{ij} \quad \text{and} \quad E_j = \sum_{i=1}^{I} E_{ij}.$$

The statistic $X^2 = \sum_{j=1}^{J} (O_j - E_j)^2 / E_j$ should then produce an approximate chi-square variable on $J - 1$ or $J - 3$ degrees of freedom, depending on whether κ and w were assumed to be known or were fitted from the data.

Treatment effects: estimation and significance testing

If κ and w are known or have been estimated from the data, then the log-likelihood for an I group experiment is given by

$$LL = \sum_{i=1}^{I} s_i \log \beta_i - \sum_{i=1}^{I} \beta_i v_i.$$

If the parameters β_i for each group depend on certain covariates as explanatory variables (dose, carcinogen, method of application, etc) $z_1 \cdots z_p$, where z_{iu} is the value of the uth variable in the ith group, then it is convenient to relate β_i to the z_{iu} by the expression

$$\log \beta_i = \sum_{u=1}^{p} \theta_u z_{iu}$$

or

$$\beta_i = \exp \sum_{u=1}^{p} \theta_u z_{iu},$$

where the θ_u are regression coefficients to be estimated. It can be said that a log-linear model for the β_i is used. The log-likelihood function then becomes

$$LL = \sum_{i=1}^{I} s_i \sum_{u=1}^{p} \theta_u z_{iu} - \sum_{i=1}^{I} v_i \exp \sum_{u=1}^{p} \theta_u z_{iu}.$$

Multiple regression methods based on maximum likelihood estimation are used for this problem. Likelihood ratio tests can be employed to investigate the significance of certain covariates. Let $LL^{(1)}$ be the log likelihood of a given model with a certain number of covariates. Inclusion of d further covariates alters the fitted log likelihood to $LL^{(2)}$. Under the null hypothesis that the regression coefficients θ_u for the newly included covariates are zero, $2(LL^{(1)} - LL^{(2)})$ should be approximately chi-square-distributed with d degrees of freedom. For detailed illustrations, see Peto and Lee (1973).

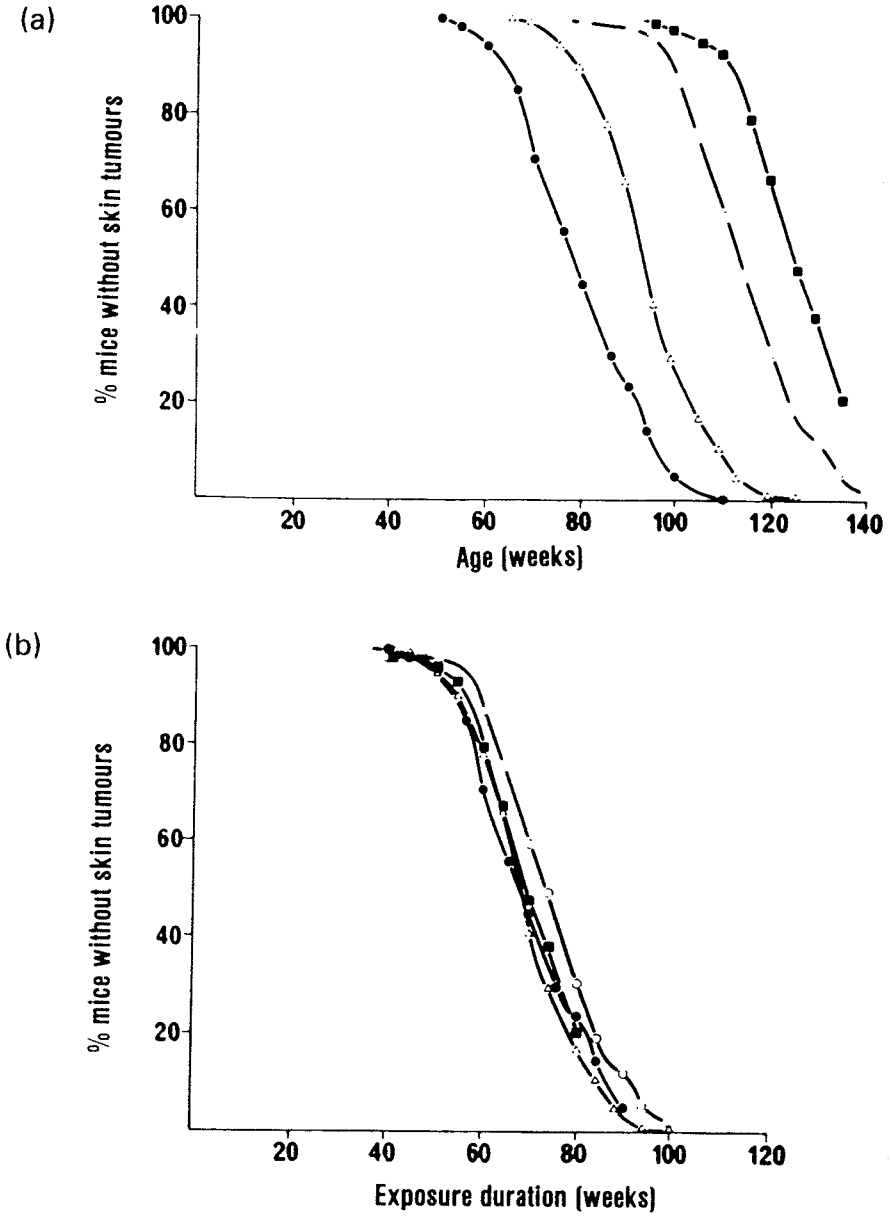
Aitkin and Clayton (1980) have published a computer program to fit such regression models to possibly censored failure-time data as they arise in this context. Their program is developed in the framework of the GLIM package (Baker & Nelder, 1978) for fitting generalized linear models.

Support for the model

The fourth assumption of the multistage hypothesis underlying the Weibull distribution implies that, if treatment is continuous, the kinetic rate-constants for each stage do not depend on the age of the animal. It follows that, provided the carcinogen affects the first stage of the process strongly enough for it to be a reasonable approximation to assume all observed tumours to have arisen because of this, the age-specific incidence rates will depend wholly on the duration of treatment and not at all on age *per se.*

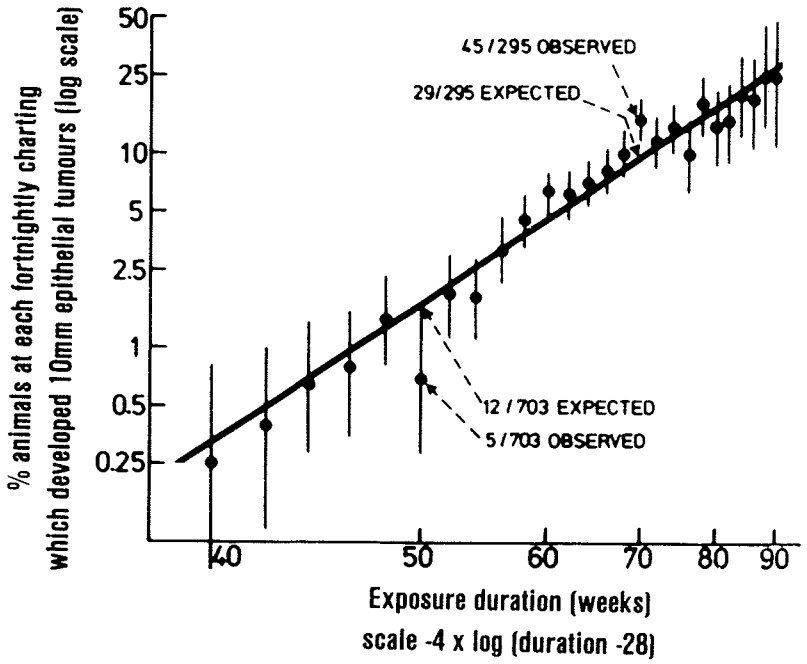
In a large experiment carried out by Peto *et al.* (1975), 3,4-benzo[*a*]pyrene (BP) was applied to the skin of mice in four groups of increasing size starting at 10, 25, 40 and 55 weeks old, respectively. As can be seen from Figure 6.10, the percentage of mice without a tumour, when plotted against age, differed markedly between the four groups. However, when plotted against treatment duration, the four groups were virtually identical, as expected from multistage assumptions. Having shown that the relationships between tumour incidence and duration in the four groups were not significantly different, Peto *et al.* (1975) combined the results of the four groups to illustrate the overall fit to the Weibull distribution. This is shown in Figure 6.11; it is obvious that, over the 100-fold range of incidences from 0.25% per fortnight up to the massive rate of 25% per fortnight observed after 90 weeks of regular BP administration, the points did approximately fit the theoretical straight line obtained by taking logarithms of the Weibull equation $I = \beta(t - w)^{\kappa}$. The multistage model also predicts that, if a carcinogen has an effect directly proportional to dose on each of c (of κ) kinetic-rate constants, and if the dose applied is sufficiently large for the background

Fig. 6.10 Percentage of tumourless mice against (a) age or (b) duration of exposure to benzo[*a*]pyrene (from Peto *et al.*, 1975)



rate-constants to be neglected for those c stages of the cancer process, the age-specific tumour incidence rate will then be proportional to dose to the power of c. Thus, if, say, stages two and three of a five-stage process are affected linearly by treatment, d is dose, δ_i are background rate-constants and ϕ_i is the increment in rate-constant per unit

Fig. 6.11 Incidence rates of 10-mm epithelial tumours at successive fortnightly chartings against duration of benzo[*a*]pyrene (BP) application, on a log-log scale from 28 weeks onwards. (The points are statistically independent, and 90% confidence intervals are indicated.) (from Peto *et al.*, 1975)



dose for affected stages, the Weibull parameter β will be proportional to

$$\delta_1(\delta_2 + \phi_2 d)(\delta_3 + \phi_3 d)\delta_4\delta_5.$$

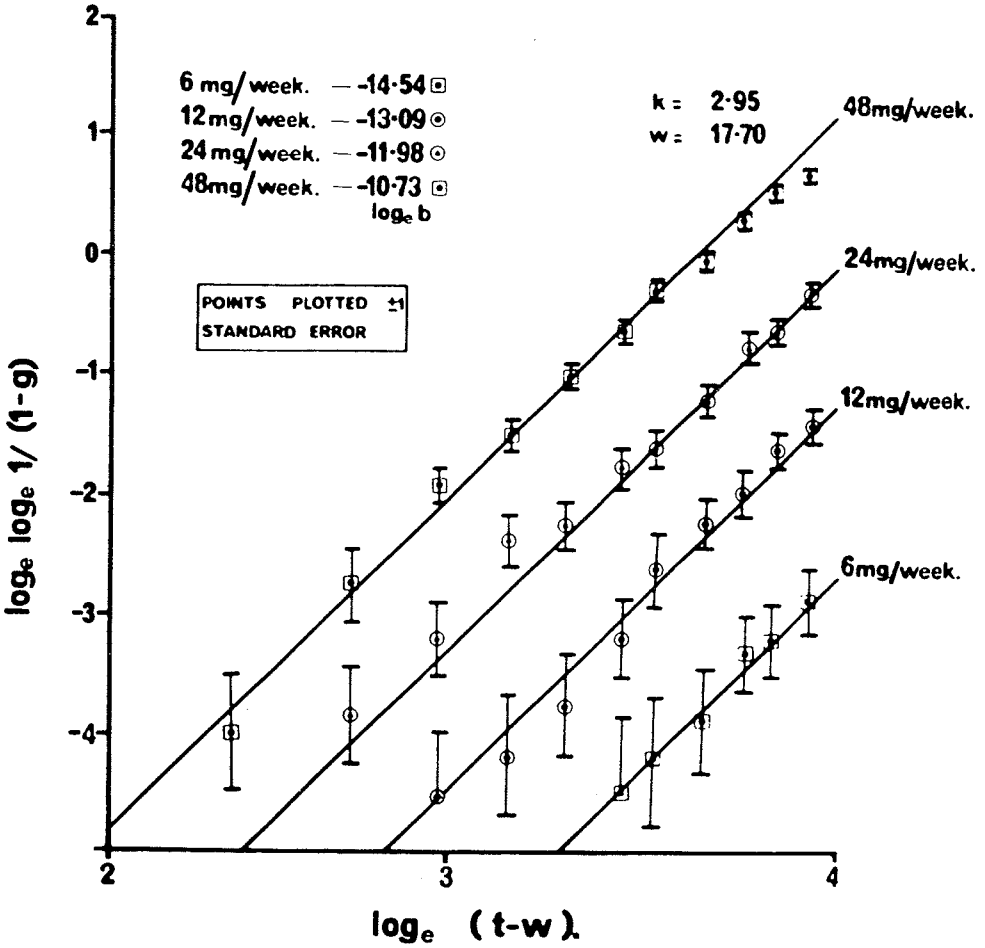
As d becomes large, this approximates to

$$\delta_1\delta_4\delta_5\phi_2\phi_3 d^2.$$

Another way of testing whether the observed failure times comply with the Weibull distribution is based on the fact that for the Weibull distribution $\log\log[1/S(t)] = \log\beta + \kappa\log(t-w)$. Note that the left-hand side can also be written as $\log[-\log S(t)]$. Using a nonparametric estimate of the survivor distribution $S(t)$ (Kaplan–Meier estimate discussed in Section 5.3) and plotting its above transform against the logarithm of time provides a simple check of the model.

Lee and O'Neill (1971) analysed an experiment in which BP was painted continuously at 6, 12, 24 and 48 μg per week on four groups of 300 mice. They found that not only could skin tumour incidence be well described by a Weibull distribution with κ and w common to all four groups, but that β was proportional to dose squared. As can be seen in Figure 6.12, the plots of $\log\log[1/\hat{S}(t)]$ against $\log(t-17.70)$ form approximately parallel equidistant straight lines. The slope of the lines, 2.95, estimates κ and is not significantly different from an integer value as suggested by the model. The average vertical difference between the lines is almost exactly twice the logarithm of the ratio of successive doses implying $c = 2$ and thus the results are consistent with a multistage hypothesis in which BP affects two out of three of the stages of the process.

Fig. 6.12 Fit of Weibull distribution to data from a skin-painting experiment with benzo[*a*]pyrene in mice (from Lee & O'Neill, 1971)



Lee *et al.* (1977) describe the analysis of a series of ten mouse skin painting experiments in which there were a total of 55 treatment groups consisting of either whole cigarette smoke condensate (SWS) or various fractions of it tested at varying dose levels. Common values of $\kappa = 3.05$ and $w = 11.29$ were fitted to the skin tumour data by maximum likelihood methods and the following linear model for the remaining Weibull parameter

$$\log \beta_{ij} = \mu + \alpha_i + q \ (\log \text{dose}_j)$$

was fitted to the responses for the ith treatment (fraction) and jth dose level. This approach leads to a simple description of the results in terms of the 'tumorigenic ratio' which measures the activity of the fraction relative to whole-smoke condensate on a weight-for-weight basis.

Druckrey (1967) reported quantitative dose-response relationships incorporating time to response information for a variety of chemical carcinogens and established the now well-known relationship

$$d \cdot t^n = \text{const.},$$

where d is the daily dose and t the median 'tumour induction time'. This empirical

relationship can be seen as a corollary of the Weibull model (6.31) (Carlborg, 1981). Consider $w = 0$ and the parameter β being proportional to some power of the daily dose d, that is, $\beta = \alpha \cdot d^m$. The absence of a constant, not dose-dependent term in this submodel for β implies a zero background response. Solving (6.31) for the median induction time gives

$$0.5 = \exp(-\alpha d^m t^\kappa)$$

or

$$dt^n = \{(\log 2)/\alpha\}^{1/m} = \text{const.}$$

with $n = \kappa/m$.

Noncontinuous exposure

Although the examples considered above concerned only the analysis of experiments in which skin tumours were produced in mice by regular skin painting, a Weibull distribution has also been successfully fitted to experiments with rats in which a single intrapleural inoculation of asbestos resulted in mesotheliomas of the pleura (Berry & Wagner, 1969; Wagner *et al.*, 1973). There are two theoretical reasons why a Weibull distribution might fit in this situation. One is that, although the injection of asbestos is given as a single dose, the asbestos is not easily destructible and remains in the animal for a considerable time after injection, thus simulating continuous exposure. The second is that, in the multistage model, if the effect of a single exposure is so large that a substantial proportion of cells at risk are transformed very rapidly through the first κ^* stages, with a subsequent tumour occurring only after background transformations cause the remaining $\kappa - \kappa^*$ transformations, the incidence rate will still obey a Weibull distribution but with a parameter $\kappa - \kappa^*$ and not κ.

In other experiments, such as those described by Day and Brown (1980), animals have been exposed for varying lengths of time and then treatment has been stopped. It is clear that in many of these experiments a simple Weibull distribution does not fit the observed response. This is not surprising, as the fourth assumption of the multistage hypothesis will not hold since kinetic rate constants of the stages affected by treatment will presumably change on stopping treatment. A number of workers have considered the mathematical implication of the application of multistage models to 'stopping experiments' (Lee, 1975; Whittemore & Keller, 1978; Day & Brown, 1980; Parish, 1981). For example, in a three-stage model in which treatment causing kinetic rate constants α_1, α_2 and α_3 was applied up to time S and then stopped, causing reversion to background kinetic rate-constants δ_1, δ_2 and δ_3, the incidence rate at time T ($>S + w$), in the simplified situation where the waiting time w all occurs after the final transformation, is given by

$$N\left[\frac{\alpha_1\alpha_2\delta_3 S^2}{2} + \alpha_1\delta_2\delta_3 S(T - S - w) + \delta_1\delta_2\delta_3\frac{(T - S - w)^2}{2}\right];$$

the above references give details of formulae in more general situations (κ stages rather than three; individual waiting times for each stage). It should be noted that the

shape of the incidence curve with time after stopping depends on which stages it is assumed that the carcinogen affects. Thus, if only the first stage is affected ($\alpha_1 > \delta_1$, $\alpha_2 = \delta_2$, $\alpha_3 = \delta_3$), the fall off in incidence compared with continuous exposure will be much less pronounced than if later stages are affected. In particular, incidence will tend to be approximately constant for some time after stopping if the penultimate stage only is affected, and will tend to drop sharply to background levels after stopping if the final stage is affected. Lee (1975) has used maximum likelihood methods in an attempt to distinguish formally between hypotheses in which a carcinogen does and does not affect a certain stage or stages. However, the computation involved is considerable.

Further discussion of details of analysis of these special experimental situations is outside the scope of this monograph, although it is worth pointing out that the methods for stopping experiments can also be applied to crossover experiments in which varying treatments are given in varying orders to the same animals.

Fitting Weibull distributions to data for internal tumours

In their analysis of data from the British Industrial Biological Research Association (BIBRA) nitrosamine study, some of which are described in Section 4.3, Peto *et al.* (1984) successfully used related Weibull distributions to describe the distribution of time X to onset of tumour and of time Y to death because of tumour. Referring back to the general formulation given above, they assumed $\Lambda_X(t) = \beta t^\kappa$ and $\Lambda_Y(t) = \beta f t^\kappa$. The additional parameter f they referred to as the 'fatality factor', ranging from 0 for a completely nonfatal tumour to 1 for an instantly fatal tumour.

Over the wide range of dose levels tested, f (and κ) appeared to be essentially invariant of dose. This allowed characterization of the dose-response relationship for liver and oesophageal tumours in terms of a single parameter β for each tumour type. While more experience is needed with this model, it appears to be a very useful approach.

Kodell and Nelson (1980) use the Weibull distribution within their simplified observational model for the carcinogenic process, which was introduced above (Fig. 6.9). They consider the transition time from N to T, which corresponds to the random variable X for time to onset of tumour, to follow a Weibull distribution. Their parameterization of the hazard function is $\beta_1 t^{\gamma_1}$. They further consider the transition time from T to D_T, which corresponds to the random variable $Y - X$, being of a Weibull type with hazard function $\beta_2 t^{\gamma_2}$ as well as transition from N or T to D_{NT} with hazard function $\beta_3 t^{\gamma_3}$. Within this framework, the likelihood function is developed considering both natural deaths as well as scheduled sacrifices. The likelihood function depends on the six parameters β_1, β_2, β_3, γ_1, γ_2 and γ_3 and can be maximized numerically.

Tolley *et al.* (1978) also chose a Weibull function to describe transition to the tumour state, but chose a Gompertz function for transition to death from other causes. More generally, Kalbfleisch *et al.* (1983) discuss likelihood estimation for an arbitrary parametric model without necessarily making the assumption of independent compet-

ing risks. In principle, however, the general formulation outlined earlier in this section, which does not require estimation of the distribution of time to death from causes other than tumour, seems simpler.

Limitations of multistage models and other modelling approaches

In an analysis of data from a mouse-skin painting experiment, it was assumed that the incidence rate followed a log-normal distribution with time (Day, 1967). Subsequent analysis by Peto *et al.* (1972) showed that Weibull distributions with a (κ, w)-pair common to all groups provided a significantly better fit to the data than did the log-normal.

Parish (1981) felt that it was unreasonable to expect animals to have an identical susceptibility to the effects of applied carcinogens, and suggested a model in which the parameter β had a gamma distribution. To gain an impression of the likely variation in susceptibility, she analysed data from the ageing experiment of Peto *et al.* (1975) referred to previously, looking at time to appearance of further tumours in animals according to how many tumours they already had. She concluded that the data were consistent with a 50-fold variation in susceptibility between the 5th and 95th percentile of the distribution. But even so, this variation was not large enough to make the distribution of time to tumour differ materially from a Weibull distribution, except where incidence rates were extremely high. The effect of susceptibility is to make the plot of log incidence against $\log(t - w)$ fall away from a straight line at high t values, and this may be why, in Figure 6.12, there is a discernible, slight drop-off with the 48 mg/week dose for the last four points plotted.

It has also been noted that in some circumstances the dose-response relationship is not of the form predicted by the multistage model. Davies *et al.* (1974), who tested response to seven dose levels of smoke condensate in a mouse-skin painting experiment, noted that there was a clear flattening off in response above doses of 180 mg/week. They suggested that high-dose levels were killing off a proportion of the cells at risk due to toxic effects, thus violating the first assumption of the multistage model that the number of cells at risk for each animal is the same for each group. In certain circumstances, it may be useful to modify the multistage model to allow for this possibility. Hulse *et al.* (1968) showed that the observed incidence of epidermal and dermal tumours in mice following superficial external β-irradiation may be accounted for by assuming that tumour incidence is proportional to the square of the dose and that potential tumour cells lose their reproductive integrity according to an exponentially decreasing relationship with dose. For example, the dose-dependent part of the hazard function may be of the form

$$(\beta_0 + \beta_1 d + \beta_2 d^2)\exp(-\alpha_1 d - \alpha_2 d^2).$$

Whittemore (1978) has reviewed a number of quantitative theories of carcinogenesis. She presented clear evidence of the inadequacy of theories not dependent on a multistage process, such as the single-stage theory of Iverson and Arley (1950) and the multicell theory of Fisher and Hollomon (1951), and considered a number of alternative versions of the multistage theory. She concluded that, although the

multistage theory has a number of limitations (failure to distinguish between benign and malignant tumours, to consider the possibility of cell repair or the action of the host's immune system, to consider the differences in susceptibility or to consider that the sensitivity of target cells to transformation may not be constant), it nevertheless provides a flexible, broad and biologically plausible framework in which to examine the gross behaviour of tumour data.

Moolgavkar and Knudson (1981) propose a two-stage model which incorporates the growth and differentiation of normal target cells and intermediate cells (that is, cells in which the first stage has occurred). They demonstrate that experimental animal data and human epidemiological data are consistent with their two-stage model, noting that previous inferences that there are more than two stages in the development of cancer can be explained by differences in the growth kinetics of intermediate cells. By incorporating differentiation into their model, the authors are able to explain the age-incidence curves for some cancers (for example, certain childhood cancers), which cannot be explained easily in terms of a simple multistage model.

In an attempt to use the full information from an animal experiment (including time to death) for the estimation of 'safe doses', Hartley and Sielken (1977) model the hazard function as a product of a dose-dependent and a time-dependent term

$$\lambda(t, d) = g(d) \cdot h(t),$$

where the time-dependent term is chosen to be

$$h(t) = \sum_{r=1}^{R} r b_r t^{r-1}.$$

Proportional hazards models

For the analysis of rapidly lethal or observable tumours, we showed in Section 5.6 that the appropriate methods correspond to those given in Section 5.3 for survival analysis. The only difference is that one uses death due to tumour, or the appearance of an observable tumour, as the experimental endpoint, rather than death from any cause. It is useful to adopt the terminology 'failure' to denote such well-defined events as appearance of an observable tumour or death, and 'failure time' to denote the time to occurrence of such an event. Methods have been developed for the analysis of censored failure times which require no distributional assumptions (such as Weibull distributed time-to-tumour). The most widely used method is the proportional hazards model (Cox, 1972).

Under the proportional hazards model, the hazard function, that is, the age-specific failure rate for an animal with covariates $\mathbf{z} = (z_1, \ldots, z_p)'$, is

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \cdot \exp(\boldsymbol{\theta}'\mathbf{z})$$

where $\lambda_0(t)$ is a completely unspecified hazard function, and $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_p)$ a vector of regression parameters, and

$$\boldsymbol{\theta}'\mathbf{z} = \sum_{u=1}^{p} \theta_u z_u.$$

For example, for a given animal, z_1 could be the administered dose level of a test compound, z_2 could be the initial body weight, z_3 could describe the row location and z_4 the column location of the animal's cage. Then the magnitude of the association between dose level of the compound and the failure, with adjustment for the remaining variables, can be measured by the estimate of the parameter θ_1 corresponding to z_1 in the proportional hazards model.

As in Section 5.3, suppose that failures are observed at K distinct times t_k, $k = 1, \ldots, K$. Let x_k denote the number of animals failing at t_k, and let $\mathbf{s}_k$ denote the sum of the covariate vectors $\mathbf{z}_{ik}$ corresponding to the animals failing at t_k, $i = 1, \ldots, x_k$. Then, if the number of ties (that is, $x_k > 1$) at each t_k is small, the parameter vector $\boldsymbol{\theta}$ can be estimated by maximizing the approximate likelihood (Breslow, 1974):

$$L = \prod_{k=1}^{K} \frac{\exp(\boldsymbol{\theta}'\mathbf{s}_k)}{\{\sum_{j \in R_k} \exp(\boldsymbol{\theta}'\mathbf{z}_j)\}^{x_k}},$$

where R_k denotes the set of indices corresponding to animals which survived to time t_k, and thus were at risk of failing at t_k. The approximate likelihood can be maximized using the Newton–Raphson method, and the covariance matrix for the resulting estimator $\hat{\boldsymbol{\theta}}$ can be estimated by the negative of the inverse of the matrix of second partial derivatives of $\log(L)$ (Kalbfleisch & Prentice, 1980, Chapter 4; Miller, 1981b, Chapter 6). To illustrate analyses based on the proportional hazards model, consider the cigar-smoke condensate data (Section 4.2). The data can be examined for evidence of a dose-related increase in tumour rates by fitting the model $\lambda(t; z) = \exp(\theta z)\lambda_0(t)$, where $z = 0$ for animals in the low-dose group, $z = 19$ for animals in the middle-dose group and $z = 44$ for animals in the high-dose group. As noted above, the hazard function for the low-dose group, $\lambda_0(t)$, need not be specified. Maximizing the likelihood L leads to the estimate $\hat{\theta} = 0.0261 \pm 0.0060$, indicating that cancer risk increases with increasing dose.

Pairwise comparison of the middle-dose group to the low-dose group and of the high-dose group to the low-dose group can be accomplished by fitting a model which includes two covariates $z_1 = 1$ if the animal is from the middle-dose group and $= 0$ otherwise, $z_2 = 1$ if the animal is from the high-dose group and $= 0$ otherwise. The model

$$\lambda(t; \mathbf{z}) = \exp(\theta_1 z_1 + \theta_2 z_2)\lambda_0(t)$$

provides estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ such that $\exp(\hat{\theta}_1)$ and $\exp(\hat{\theta}_2)$ are estimates of the relative risk of the middle-dose group compared to the low-dose group, and of the high-dose group compared to the low-dose group, respectively. The results are summarized in Table 6.4.

Table 6.4 Estimates of relative risks using the proportional hazards model, for data on cigar-smoke condensate

	Dose d_i		
	0	19	44
$\exp(\hat{\theta} z_i)$	1.0	1.88	3.20
$\exp(\hat{\theta}_1 z_1 + \hat{\theta}_2 z_2)$	1.0	1.64	3.15

As in Section 5.6, there is good agreement between the estimates $\exp(\hat{\theta}_1 z_1 + \hat{\theta}_2 z_2)$ and the estimates based on the regression parameter θ.

If there are no ties (that is, $x_k = 1$ for all k), L can be derived as the marginal likelihood based on the distribution of ranks, but in the presence of ties, the expression for the marginal likelihood is more complicated than L (Kalbfleisch & Prentice, 1980, Chapter 4). Estimates of survival curves based on the proportional hazards model are available (Kalbfleisch & Prentice, 1980, pp. 84–87; Miller, 1981b, pp. 133–136), and estimates of percentiles can be obtained from the estimated survival curves. Confidence intervals for percentiles can be calculated by analogy to the methods based on the Kaplan–Meier survival curves (Slud *et al.*, 1984), with substitution of the variance expression corresponding to the proportional hazards survival curves (Kalbfleisch & Prentice, 1980, pp. 116–117).

Regression models for tumour prevalence

The logistic regression model of Dinse and Lagakos (1983) for comparing treatment groups with respect to tumour prevalence was discussed in Section 5.5. Their regression model can be used in a more general setting when there are other covariates in addition to specific group membership or dose level. This more general regression context allows incorporation of several covariates and is computationally simple to analyse.

Formally, Dinse and Lagakos use the regression model

$$\psi(x, \mathbf{z}, t) = \frac{\text{pr}(Y = 1 \mid X = x, \mathbf{Z} = \mathbf{z}, T = t)}{\text{pr}(Y = 0 \mid X = x, \mathbf{Z} = \mathbf{z}, T = t)} = \exp\{\alpha x w(t) + \boldsymbol{\theta}'\mathbf{z} + \gamma(t)\}$$

to model the odds ratio of having ($Y = 1$) or not having ($Y = 0$) the tumour at death, with X a binary treatment indicator, $\mathbf{z}$ a p-vector of covariates and T survival time. The scalar α and the p-vector $\boldsymbol{\theta}$ are unknown parameters with $\gamma(t)$ and $w(t)$ prespecified functions of time. If treatment is assumed to have a constant effect on log-odds, $w(t)$ is set equal to 1 for all t. By setting $w(t) = 1 + f(t)\delta/\alpha$ where $f(t)$ is some function of time, such as t or $\log t$, one replaces $\alpha x w(t)$ by $\alpha x + \delta x f(t)$, and this allows a test of the hypothesis $\delta = 0$, that is, whether the proportional treatment odds relationship depends on time.

Integrated models for tumour prevalence and lethality

Another method, which differs from those described above, in that it takes into account the possibility of simultaneous study of more than one type of tumour, but is applicable only to experiments in which there are a number of scheduled sacrifices, has been described by Turnbull and Mitchell (1978) and by Mitchell and Turnbull (1979). For the purposes of their method, animals in one of R treatment groups ($r = 1, \ldots, R$) dying in M time intervals ($m = 1, \ldots, M$) are classified as being in one of K ($k = 1, \ldots, K$) 'illness states'. These illness states are defined in terms of whether an animal has or does not have particular tumour types. Thus, dealing with three tumours of interest, there are $2^3 = 8$ illness states.

Their statistical model is defined in terms of prevalence p_{kmr}, the probability that an animal from group r, alive at the beginning of the mth interval, is in illness state k at

that time, and lethality q_{kmr}, the conditional probability that an animal from group r dies during the mth interval, given that it was alive and in illness state k at the beginning of this interval. Data consist of w_{kmr}, the number of animals in group r sacrificed (withdrawn) in interval m and found in illness state k; d_{kmr}, the number of animals from group r dying in interval m and diagnosed with illness stage k, and s_{mr}, the number of animals from group r surviving through interval m. The distribution of such surviving animals among the illness states is not known, but can be characterized by s'_{kmr}, as these unobserved counts must be taken into account when fitting statistical models for the prevalence and for the lethality.

The numerator of the prevalence of an illness state is represented by $w_{kmr} + d_{kmr} + s'_{kmr}$, the term s'_{kmr} being estimated iteratively. Such values comprise data to which a log-linear model, formulated in terms of dependency of p_{kmr} on explanatory factors, such as treatment, time and the presence of tumours of each type can be fitted.

Similarly, a statistical model is fitted for lethality q_{kmr}. Since the lethality is a conditional probability, both numerator, d_{kmr}, and denominator, $d_{kmr} + s'_{kmr}$, have to be specified. The model used is a logistic one in which the dependency of q_{kmr} on a similar set of explanatory factors is studied.

The crucial problem of not knowing the s'_{kmr} is dealt with as follows. Firstly, the distribution of the s_{mr} survivors to the illness states is assumed to be as in the distribution observed in the animals which died or were killed, so that

$$s'_{kmr} = s_{mr}(d_{kmr} + w_{kmr}) \Big/ \sum_{k=1}^{K} (d_{kmr} + w_{kmr}).$$

Using these s'_{kmr}, prevalence and lethality models are then fitted, which give rise to estimates $\hat{p}_{kmr}$ and $\hat{q}_{kmr}$. These estimates are then used to re-assess the distribution of the s_{mr} survivors to the illness states by the formula

$$s''_{kmr} = s_{mr}\hat{p}_{kmr}(1 - \hat{q}_{kmr})/(1 - h_{mr}),$$

where $h_{mr} = \sum_{k=1}^{K} \hat{p}_{kmr}\hat{q}_{kmr}$ is the unconditional probability of dying in group r in interval m. s'_{kmr} can be replaced by s''_{kmr} and the same models fitted to the slightly modified data. This process can then be iterated until the distribution of the survivors into the illness state no longer changes.

It should be noted that the analysis makes the (technical) assumption that illness state changes are made only at the beginning of each of the M time intervals and that sacrifices occur only immediately after the illness state changes, in order for the prevalences and the lethalities to be defined. It is thus convenient to choose partitions of the time axis such that each interval covers one scheduled sacrifice.

A detailed application of this approach is given by Wahrendorf (1983). A very interesting feature of this method is that it allows simultaneous analysis of different tumour types. This opens interesting possibilities for assessing associations between different tumours. Berlin *et al.* (1979) also consider a general Markov model for multiple tumour types and discuss the question of identifiability.

6.4 Summary

The number of animals developing tumours during the course of a conventional two-to-three-year rodent carcinogen bioassay will depend on the dose level to which they are exposed. The overall shape of the dose-response curve can, however, vary widely, depending on the particular agent being evaluated. Although most dose-response relationships generally increase with dose, this increase may be either linear or distinctly nonlinear. In the latter case, dose-response curves which increase rapidly beyond a certain dose range may be noted, as with nasal tumours induced by exposure to formaldehyde. Conversely, a levelling off of the rate of response may be observed at high doses, as with liver tumours resulting from exposure to vinyl chloride. Combinations of these different shapes are also possible, as with the S-shaped dose-response seen for aflatoxin-induced liver tumours.

A variety of different mathematical dose-response models may be used to provide a parsimonious description of the observed dose-response relationship. Simple tolerance distribution models, although often sufficiently flexible to provide a good fit to dose-response data, are somewhat naive in terms of their underlying biological basis as a possible mechanism for carcinogenesis. Stochastic models based on the notion that carcinogenesis results from the random occurrence of one or more fundamental biological events are more appealing, but are necessarily based on strong but uncertain assumptions. Foremost among such mechanistic models is the Armitage–Doll multistage model, based on the assumption that a cell progresses through a number of distinct stages before becoming cancerous, the transition intensity function for each stage being a linear function of dose.

Since many compounds may require metabolic activation before being converted into their active form, consideration may be given to pharmacokinetic models for this process. This is particularly important when certain steps such as absorption, elimination, activation or detoxification are saturable. Even if the response rate is directly proportional to the dose level of the activated complex reaching the target tissue, such saturation effects may account for nonlinearity in the dose-response relationship when expressed as a function of the administered rather than of the delivered dose.

Given a suitable model for the dose-response relationship, estimates of certain quantiles of the curve may be of interest. Because human exposure to most environmental carcinogens is low, there has been considerable interest in obtaining estimates of risk in the low-dose region based on the downward extrapolation of results obtained at higher doses. Unfortunately, this is subject to considerable model uncertainty, and different models, all equally consonant with the observed data, give widely different projections upon extrapolation of low doses. Because of this, a linear extrapolation to low doses is often advocated as the most prudent approach. This will be particularly appropriate in cases where the background response rate can be considered to arise from an at least partially dose-wise additive model.

This low-dose model dependency may be circumvented by restricting attention to those quantiles lying well within the observable response range. Historically, quantiles such as the TD_{50} (the dose estimated to induce tumours in 50% of exposed animals)

have been used as the basis for various measures of carcinogenic potency. The TD_{50} itself has recently been used by Gold *et al.* (1984) to demonstrate variations in carcinogenic potency spanning eight orders of magnitude.

Time-to-tumour models attempt to describe the carcinogenic process in detail and make use of information on individual tumour occurrence and survival times. For observable tumours, parametric models addressing the time to first occurrence of this tumour have been developed on biological principles. The parameters of these models allow for inferences to be drawn on both the time and dose dependency of tumour incidence. For occult tumours the situation is more complicated: tumour incidence is not directly observable and its estimation depends on the experimental design, particularly the use of a series of interim sacrifices. However, within this framework, estimates for the relevant functions characterizing the carcinogenic response are derived, if identifiable. This provides the basis not only for unbiased statistical inference but also for clear biological conclusions. Time-to-tumour models make use of as much information as possible; they are in turn based on some assumptions, but they provide the most thorough description of the observed response in a long-term animal experiment.

LIST OF ESSENTIAL SYMBOLS – CHAPTER 6 (in order of appearance)

$P(d)$	probability of treatment-induced tumour at dose d
$G(t)$	tolerance distribution
$P^*(d)$	probability of spontaneous or treatment-induced tumour at dose d
π_0	background response rate (probability of spontaneous tumour)
$D_1(t)$	administered dose at time t
$D_2(t)$	activated dose at time t
D_2^*	effective dose under steady state conditions
$\boldsymbol{\theta}$	vector of parameters $\theta_1, \ldots, \theta_p$
$P^*(d, \boldsymbol{\theta})$	probability of response (tumour) at dose d dependent on parameter $\boldsymbol{\theta}$
d_i	dose levels ($d_0 = 0$, $d_1 < \cdots < d_I$)
x_i	number of animals with tumour at dose d_i
n_i	number of animals at dose d_i
$L(\boldsymbol{\theta})$	likelihood of observed outcome under any dose-response model $P^*(d; \boldsymbol{\theta})$
$\hat{\boldsymbol{\theta}}$	maximum likelihood estimate of $\boldsymbol{\theta}$
$\mathbf{0}$	vector of zeros
$\mathbf{V}$	variance–covariance matrix of $\sqrt{n}\,(\hat{\boldsymbol{\theta}}^1 - \boldsymbol{\theta})$
v^{rs}	(r, s)th element of the inverse of $\mathbf{V}$
$\Pi(d)$	added risk over background at dose d
d_q	dose corresponding to the added risk, q, over background
C_q	measure of carcinogenic potency based on d_q
N	disease-free state of animal
T	animal with tumour
D_{NT}	death from a cause unrelated to the tumour of interest
D_T	death due to tumour of interest

X time to onset of tumour (random variable)
Y time to death due to tumour (random variable)
Z time to death from an unrelated cause (random variable)
$f(t)$ density function
$\lambda(t)$ hazard function
$\Lambda(t)$ cumulative hazard function
$S(t)$ survival function
(N.B. A subscript X or Y to the quantities $f(t)$, $\lambda(t)$, $\Lambda(t)$ and $S(t)$ indicates the corresponding functions for the random variable X or Y)
t_k time of observation of an event such as death or the occurrence of tumour ($k = 1, \ldots, K$)
a_k number of animals with appearance of a visible tumour at time t_k
b_k number of animals with death caused by tumour of interest (fatal context) at time t_k
c_k number of animals with death from unrelated cause, tumour of interest present (incidental context), at time t_k
d_k number of animals dying without tumour of interest at time t_k
L_1 likelihood in the case of visible tumours
L_2 likelihood in the case of occult tumours all observed in a fatal context
L_3 likelihood in the case of occult tumours all observed in an incidental context
L_4 likelihood in the case of occult tumours observed in fatal or incidental context
(N.B. LL_i ($i = 1, 2, 3, 4$) denotes the logarithm of the likelihoods L_i ($i = 1, 2, 3, 4$))
κ number of stages or transformations required before tumour occurs
δ_j kinetic rate constants for the jth stage of multistage model ($j = 1, \ldots, \kappa$)
w_j constant time taken by the jth transformation of the multistage model ($j = 1, \ldots, \kappa$)
β_i Weibull shape parameter for group i ($i = 1, \ldots, I$)
LL log-likelihood for an experiment with I groups under the Weibull model with known w and κ
$\mathbf{z}$ vector of covariates observed for each animal
$\lambda(t; \mathbf{z})$ hazard function of animal with covariates $\mathbf{z}$
$\lambda_0(t)$ baseline hazard function