2. GENERAL CONSIDERATIONS ON THE EVALUATION OF ANIMAL CARCINOGENESIS STUDIES

CHAPTER 2

GENERAL CONSIDERATIONS ON THE EVALUATION OF ANIMAL CARCINOGENESIS EXPERIMENTS

2.1 Introduction

The primary purpose of a long-term carcinogenicity experiment is to determine if the administration of a test substance to animals of some species alters the normal pattern of tumour development in that species. In a typical long-term carcinogenicity experiment, a pool of animals is divided by randomization into several groups. One group serves as a concurrent control group, while the remaining groups are exposed to various dose levels of the test substance by some appropriate route of administration. The test animals are observed for a major portion of their lifespan, and all animals that die during the study are subjected to necropsy unless they are substantially cannibalized or autolysed. The experiment is terminated according to a predetermined stopping rule, for example, after a fixed period of time on study, or when mortality in the control or lowest-dose group exceeds a specified limit such as 50% (see IARC, 1980). At termination, all surviving animals are killed and subjected to necropsy. For each animal, tissues are taken from several organ sites and examined histopathologically. The basic data obtained from each animal are the times of appearance of any visible tumours, the time of death, the cause of death (in so far as the cause can be determined), a list of organs examined at necropsy, and the histopathological diagnoses for those organs examined. The goal of the statistical analysis of these survival and pathology data is to quantify the strength of evidence regarding the carcinogenic potential of the test agent. In this book the emphasis will be on evaluating the carcinogenicity of a test substance only in the species under study. For recent reports on the role of long-term animal tests in assessing the potential carcinogenic risk to man, see Weisburger and Williams (1981), Squire (1981), IARC (1982b), and Interdisciplinary Panel on Carcinogenicity (1984), as well as the remarks in Section 2.18.

2.2 Determination of relevant biological events

In long-term carcinogenicity studies, the experimental outcome of interest is the occurrence of a tumour at some target organ. Throughout this book, the word 'tumour' will be used quite generally to refer to a well-defined class of neoplastic lesions. The determination of which class of lesions at a given organ should be analysed for evidence of carcinogenicity can be extremely difficult and requires the judgement of experienced pathologists. Such determinations will vary from species to species and

among organs within a species, but generally it is advisable to restrict the grouping of lesions to tumours of the same histological type arising in the same type of tissue (IARC, 1980). In addition, analyses should be restricted to primary tumours rather than to a grouping of primary and metastatic tumours. The importance of and difficulty in defining the class of lesions which represents a carcinogenic response indicates the need for a close working relationship between pathologists and statisticians involved in the evaluation of animal carcinogenesis studies.

Although evidence of a carcinogenic effect can occur in any of the several organs examined, the effect of a carcinogenic agent is likely to be concentrated in one or a few target organs. A clear-cut carcinogenic response at one target organ may be obscured by an analysis based on the incidence of all tumours, regardless of their sites of occurrence. This is particularly true in species which have a high rate of naturally occurring tumours. Thus, unless there is information available prior to the evaluation of an experiment indicating that a test agent is likely to have carcinogenic potential at more than one organ or tissue type, the pooling of tumour incidence data from two or more sites should be avoided. In general, statistical analyses of tumour incidence data should be restricted to tumours which develop at a specific organ, or to tumours of a type known to have a multicentric origin (for example, leukaemias) (IARC, 1980). In exceptional cases, it may be reasonable to pool tumours at biologically related sites or tumours with common morphological characteristics, but this should be done only in close consultation with a pathologist.

The evaluation of a long-term carcinogenicity experiment is complicated by the fact that there is no single well-defined biological response that characterizes a carcinogenic effect. A carcinogenic agent may cause any of several types of alterations in the normal pattern of tumour development, and data from a carcinogenicity experiment must be examined carefully for each of the possible changes. The most important carcinogenic response is an increase in age-specific rates of tumour incidence in exposed animals over some portion of the lifespan of the test species, leading to an increased lifetime probability of developing a tumour. Another possible carcinogenic response is an acceleration of tumour development in exposed groups; that is, tumours similar to those occurring naturally may develop earlier in life in exposed animals than in control animals, while the lifetime probability of developing a tumour remains unchanged. The possibility of such an acceleration effect, which might arise if only a fixed proportion of the test animals are at risk of developing a tumour, has been proposed (Lagakos & Mosteller, 1981); however, there is little experimental evidence for such an effect in inbred strains. For some species, an increase in the number of tumours at some organ in exposed animals may indicate carcinogenicity (Tomatis *et al.*, 1973; Shimkin & Stoner, 1975; Ward & Weisburger, 1975; Peraino *et al.*, 1977; Ward & Vlahakis, 1978). Potential problems in the interpretation of data on tumour multiplicity are discussed by Peto *et al.* (1980). Qualitative morphological or biological differences between tumours found in exposed animals and tumours found in control animals may also provide evidence of a carcinogenic effect. For example, tumours that have unusual morphology may be observed in exposed animals but not in control animals, or tumours in exposed animals may be more prone to metastasize than tumours in control animals (Frith *et al.*, 1979; Reznik & Ward, 1979; Hoover *et al.*, 1980; Stinson *et al.*, 1981). In

many cases, the identification of unusual patterns of tumour development depends heavily on the knowledge and judgement of pathologists and toxicologists (Ward, 1984).

Although the statistical analysis of tumour and survival data is an essential component of the evaluation of carcinogenicity, care must be taken, in assessing carcinogenic potential, not to place total reliance on the finding of statistically significant results. For example, there may be minor changes in the normal tumour pattern, which might not be detected by standard statistical methods, but may nonetheless be evidence of carcinogenicity. Of particular importance is the possibility that a few exposed animals, but no control animal, may develop a tumour at an organ in which tumours very rarely occur naturally. Historical control data can sometimes be used to increase the sensitivity of statistical tests for rare tumours (see Chapter 7). The induction of tumours that are normally rare may be extremely important in assessing human risk (Squire, 1981). On the other hand, even a significant tumour increase in only one sex at an organ in which naturally occurring tumours are quite common may not provide convincing evidence of carcinogenicity in the test species (Fears *et al.*, 1977; Gart *et al.*, 1979). The evaluation of carcinogenicity in these and other cases may rest as much on biological, toxicological or pharmacological considerations as on the presence or absence of statistical significance.

2.3 Non-neoplastic precursors of a neoplasm

Although the emphasis in ensuing chapters is on the analysis of tumour data, it should be noted that evidence of carcinogenicity can also be obtained from data on non-neoplastic lesions, when such lesions are precursors of a neoplasm. When both neoplastic lesions and non-neoplastic precursors are discovered in animals in the same experiment, analyses of the non-neoplastic precursors separate from the neoplastic lesions must be interpreted with care. If a test agent is extremely efficient at converting non-neoplastic precursors to neoplasms, an analysis based solely on non-neoplastic lesions may show a dose-related decrease, while an analysis of neoplastic lesions shows a dose-related increase, indicating that the compound is carcinogenic. Two analyses may be informative in such a situation, one based only on animals with neoplastic lesions and the other based on animals with either a neoplasm or a non-neoplastic precursor.

2.4 Adjustment for intercurrent mortality

Identification of differences in overall mortality among exposure groups is an important step in evaluating a carcinogen bioassay. Even if a single biological response characterizes a carcinogenic effect, the analysis of changes in the normal pattern of tumour development may still be difficult. This is because the observed experimental outcome corresponding to a particular alteration in tumour development can vary from experiment to experiment, depending on whether the control and exposed groups differ in intercurrent mortality. Intercurrent mortality refers to interim deaths not related to the development of the particular type or class of tumours to be analysed for

evidence of carcinogenicity (Peto *et al.*, 1980). Adjusting for such intercurrent mortality is an important consideration in evaluating a carcinogenesis experiment. Consider an agent which causes increases in the underlying (and in general, unobservable) age-specific rates of tumour incidence throughout the lifespan of the test species. If intercurrent mortality rates are equal in the control and exposed groups, then the increased age-specific rate of tumour incidence in the exposed animals results in an increase in the observed proportion of exposed animals that develop tumours during the experiment. If, however, the intercurrent mortality rates are higher in exposed animals than control animals, then the observed proportion of exposed animals that develop tumours may not be increased, in spite of underlying age-specific rates of tumour incidence that are uniformly higher.

To illustrate this point, suppose that the test agent is quite toxic and that the agent induces tumours that do not shorten the lives of tumour-bearing animals. In a hypothetical experiment summarized in Table 2.1 with 100 animals exposed to the test

Table 2.1 Proportions of control and exposed animals dying with a tumour in early and late stages of a hypothetical carcinogenesis experiment (the denominator is the number of animals dying in each time period, and the numerator is the number of these dead animals in which tumours were found at necropsy)

	Control	Exposed
Died prior to 15 months	1/20 (5%)	18/90 (20%)
Died after 15 months	24/80 (30%)	7/10 (70%)
Total for experiment	25/100 (25%)	25/100 (25%)

agent and 100 unexposed control animals, suppose that, in the first 15 months, 90 exposed animals die and 18 of these exposed animals are discovered to have tumours, while only 20 control animals die and one of these control animals is discovered to have a tumour. Thus, during the first 15 months of the experiment, the observed percentage of animals with tumour is 20% in the exposed group and 5% in the control group. Suppose now that seven of the remaining ten exposed animals and 24 of the remaining 80 control animals develop tumours in the last months of the experiment, so that the observed percentage of animals with tumour among those surviving 16 months or longer is 70% in the exposed group and 30% in the control group. In spite of the fact that the proportion of exposed animals with tumour is higher in both the early and late stages of the experiment, the percentage of animals that are observed with tumour in the entire experiment is 25% (25/100) in both the exposed and control group. Since it is common in animal carcinogenesis experiments for exposed animals to have higher intercurrent mortality rates than control animals, and for prevalence rates in both the control and exposed groups to increase as an experiment progresses, this hypothetical example illustrates an outcome that can occur in practice. It is clear that, in the analyses of any changes in the pattern of tumour development, differences in longevity between control and exposed animals must be taken into consideration.

2.5 Combining analyses as opposed to pooling data

The hypothetical example given in Table 2.1 illustrates an important source of bias that can arise in the analysis of animal carcinogenicity experiments. This is the inappropriate pooling of data. Data pooling refers to the practice of calculating summary tumour rates by simply adding the number of animals with tumour in several distinct strata (for example, two time periods in the above example) and dividing this sum by the total number of animals (in all strata). The evaluation of the carcinogenicity of a test agent often involves the analysis of data from several strata. Some strata are naturally defined; for example, each sex of each species of test animal defines a stratum. Within each sex/species experiment, further stratification is possible, such as the subdivision of data by time period in Table 2.1. The spontaneous tumour rate will usually vary from stratum to stratum (this being the motivation for the stratification by time period in Table 2.1); however, under the null hypothesis that the test agent has no carcinogenic effect, the tumour rates within each stratum for all exposure groups should be equal. The example of Table 2.1 illustrates the danger of pooling data from different strata prior to statistical analysis. An effect which is present in all strata may be obscured when data from several strata are pooled. Analysis of data from each time period in Table 2.1 would show up the increased tumour rates in exposed animals in both strata. By using appropriate methods, analyses from several strata can be combined. Increases or decreases in tumour incidence in all strata provide enhanced statistical evidence regarding the carcinogenic potential of the test agent. This combining of analyses is preferable to analysing pooled data, and appropriate methods for combining analyses are discussed in Chapter 5.

2.6 Considerations related to sex and species

Typically, animal carcinogenesis experiments are carried out using both sexes of each test species. In the statistical analysis of long-term carcinogenicity tests, tumour incidence data from male and female animals should never be pooled. Because there are often large differences between sexes in the rates of naturally occurring tumours, such pooling of data from both sexes can lead to incorrect inferences (e.g., see Gart, 1962). In addition, hormonal differences may lead to carcinogenic risk in only one sex or to a substantially higher risk in one sex. Thus, the data for each sex should be analysed separately. The demonstration of a similar carcinogenic effect in both sexes of a species strengthens the scientific inference regarding carcinogenicity and can help rule out the possibility of a spurious (for example, false-positive) result. Where appropriate (that is, when a similar carcinogenic effect is observed in both sexes), the separate statistical analyses can be combined to obtain a summary quantification of risk in the test species.

An analogous situation arises if a compound has been tested in multiple experiments using different species or strains of test animals. Because of genetic differences, there can be considerable variability among species and among strains within a species with respect to their rates of naturally occurring tumours and their susceptibility to compound-induced tumours (Haseman & Hoel, 1979). Thus, as is the case with data

from both sexes within a species, data from different species and strains should never be pooled. The data from each species or strain should be analysed and evaluated separately. The finding of a carcinogenic effect in more than one species or strain strengthens the scientific inference, particularly if the carcinogenic effect is at the same organ or of the same histological type in different species. If similar patterns of tumour induction are observed in multiple species, then the separate statistical analyses can be combined across species to obtain a summary quantification of risk for the test compound.

2.7 Consideration of tumour dose-response

In many animal carcinogenesis experiments, groups of animals are exposed at multiple dose levels of a test substance. If the test substance is carcinogenic, one would, in most situations, expect tumour rates to increase with increasing dose level. If a test substance protects against tumour induction or development, one would, in most situations, expect tumour rates to decrease with increasing dose level. Certainly, a monotonic change in observed tumour rates with increasing dose should strengthen the inference that differences in tumour rates are due to exposure to the test substance, with steeper dose-response curves providing stronger evidence of an effect. Thus, in choosing statistical methods, priority should be given to methods that are more powerful (that is, are more likely to indicate statistical significance) when observed tumour rates increase monotonically with dose. Accordingly, in presenting methods for analysing multiple-dose experiments, emphasis will be on testing for monotonic trends in tumour rates rather than on more general heterogeneity tests. In interpreting the results of such trend tests, it should be noted that a significant trend does not necessarily imply increased cancer risk at very low doses. Also, in the presence of a significant trend test, the lack of an increase in observed tumour rates at lowest dose levels should not be taken as evidence of a threshold. With the small numbers of animals typically used in animal carcinogenesis experiments, reasonable inference can seldom be made regarding the shape of the tumour-response curve at very low doses (Portier & Hoel, 1983a).

2.8 Observable tumours

For certain tissues such as the skin, tumours are visible, and hence their development can be closely monitored. There are other organs in which the presence of internal tumours may be evident before the death of the tumour-bearing animals; for example, mammary tumours often can be identified by palpation in living rodents (Davis *et al.*, 1956). Once an observational endpoint has been defined clearly in such cases (for example, the occurrence of a skin tumour of some prespecified minimum diameter), then the age at which an animal obtains this endpoint can be observed relatively unambiguously. Such tumours have been termed mortality-independent tumours (Peto *et al.*, 1980), because their observation does not require the death of the tumour-bearing animals. Whether or not a given animal develops such a tumour, however, is dependent upon mortality. An animal that dies at an early age is less likely

to have developed such a tumour than is an animal that survives to old age. Thus, in subsequent discussions, these tumours will be referred to as 'observable' rather than 'mortality-independent'. More specifically, tumours of a certain type will be referred to as observable if, when they reach some standard point in their development, they can be identified in all living animals. The class of observable tumours includes those such as palpable tumours, which are not necessarily visible. For observable tumours, incidence data can be analysed using standard life-table methods (Peto *et al.*, 1980). These life-table methods test for earlier or more frequent observation of tumours in exposed groups while correcting for differences in intercurrent mortality rates. The interpretation of the life-table analyses for observable tumours is relatively straightforward; significant differences among control and exposed groups provide evidence of agent-related changes in tumour onset or development time or in the magnitude of age-specific rates of tumour incidence.

2.9 Problems with occult tumour data

Internal tumours that can be discovered only at necropsy are termed 'occult' tumours. Perhaps the most difficult aspect of the statistical analysis of data from a long-term carcinogenicity test is that inferences must be made concerning unobservable quantities, the ages at onset of occult tumours, when all that actually can be observed are the ages at death of the tumour-bearing animals. Whereas the analysis of observable tumours is relatively straightforward, analysis of occult tumours involves additional assumptions concerning the relationship between the observable outcome, age at death with a tumour, and the endpoint for which it is a surrogate, namely, age at onset of the tumour. Even with the additional assumptions, it is not, in general, possible to test directly hypotheses about the rates of tumour onset, that is, tumour incidence rates (McKnight & Crowley, 1984). It will be important to keep in mind throughout the following discussion the surrogate role played by age at death with tumour.

2.10 Contexts of observation for occult tumours

For occult tumours, the statistical analysis is complicated by the dependence of the observation of these tumours on the death of the animals. Differences in the ages at which particular tumours are observed can result from changes in age-specific tumour incidence or mortality rates or from changes in tumour growth rates, but they can also result from changes in age-specific mortality rates associated with causes unrelated to the development of that particular type of tumour. Thus, the early appearance of tumours in dying exposed animals may or may not provide evidence of carcinogenicity. If tumours at some organ are killing their host animals, the observation of tumours earlier in exposed groups than in the control group may be evidence of carcinogenicity of the test agent. If, on the other hand, tumours at a particular target organ are occurring with equal age-specific incidence rates in control and exposed groups, while animals are dying younger in exposed groups due to causes unrelated to these tumours (for example, toxicity or tumours at another site), then the observation of these

tumours earlier in exposed groups should not be considered as evidence of carcinogenicity at the target organ. The early appearance of tumours in such cases may reflect simply the higher intercurrent mortality rate in the exposed groups. Thus, it has been noted that the context of the observation of a tumour should be determined as accurately as possible (Hoel & Walburg, 1972; Peto *et al.*, 1980). Accordingly, in subsequent discussions, a tumour which either directly or indirectly kills its host will be said to have been observed in a fatal context. A tumour which is observed at necropsy of an animal which has died of some unrelated cause will be said to have been observed in an incidental context. There are many difficulties involved in determining the context of observation of a tumour (Gart, 1975; Gart *et al.*, 1979; Peto *et al.*, 1980); however, when such determinations cannot be made, the range of appropriate statistical methods and the interpretation of the statistical analysis may be limited. Thus, when possible, the effort should be made to determine the context of observation of each tumour (Peto *et al.*, 1980). It is important that such determinations be accurate, as errors can lead to bias in the statistical analysis (Lagakos, 1982).

2.11 Potential for bias in the analysis of fatal tumours

Even if the contexts of observation have been accurately determined, there still can be difficulties in the statistical analysis of occult tumours. The analysis of data on fatal tumours is based on the observed ages at which animals are killed by the tumours, and makes use of the same life-table methods used in the analysis of data on tumours that are observable. Significant differences between control and exposed groups are taken as evidence that the test agent is associated with changes in the age-specific rates of death caused by tumour (Peto *et al.*, 1980). As noted earlier, the observation, age at death caused by tumour, is used as a surrogate for a variable that cannot be observed, namely, either the age at onset of tumour or the age at which a tumour reaches some standard developmental stage. If a tumour is rapidly lethal, then, in fact, the age at death may closely approximate the age at onset. In classifying a tumour as occurring in a fatal context, however, no distinction is made between tumours which kill their hosts rapidly after onset and tumours which kill their hosts slowly, perhaps several months after onset. In interpreting the analysis of tumours observed in a fatal context, it should be considered that earlier deaths due to tumours do not necessarily imply earlier onset times or more rapid development of tumours, particularly in cases in which the tumours are killing their hosts slowly.

In the case of observable tumours which can be detected before they become life-threatening, the interpretation of the life-table analysis in terms of tumour development is relatively straightforward. However, the interpretation of fatal tumour analyses in terms of tumour development requires the assumption that control and exposed animals are equally likely to be killed by a tumour at any particular stage in the tumour's development. This assumption may not always be reasonable. Consider an experiment in which animals in an exposed group have age-specific incidence rates for some tumours that are identical to those for the control animals. Suppose that these tumours are all observed in a fatal context, but that the exposed animals die more rapidly than the control animals once they develop a tumour. This acceleration of

deaths due to tumour is not necessarily evidence of enhanced tumour development in the exposed animals. In many chemical carcinogenesis experiments, exposed animals are administered a test compound at the maximum tolerated dose – a high dose which may place the exposed animals under great physiological stress. It is possible that a tumour might kill such a stressed animal at an early stage in the tumour's development, a stage at which the tumour would be unlikely to kill a healthier control animal. Of course, if age-specific incidence rates are unaffected by the test agent, then the proportion of exposed animals observed with tumours during the entire experiment should not be higher than the proportion of control animals observed with tumours. Accordingly, analyses of fatal tumours which indicate a significant acceleration of death caused by tumour, without an accompanying increase in the proportion of animals observed with tumours, must be interpreted with care (Lagakos & Mosteller, 1981; Mantel, 1980).

2.12 Potential for bias in the analysis of incidental tumours

As in the case of tumours observed in a fatal context, the interpretation of the statistical analysis of occult tumours observed in an incidental context can be quite difficult. Incidental tumour data are analysed using methods which have been termed 'prevalence methods' (Hoel & Walburg, 1972; Peto, 1974; Peto *et al.*, 1980). With prevalence methods, the age range spanned by a carcinogenesis experiment is subdivided into discrete age intervals. For each experimental group of animals, a proportion is formed in each age interval as follows: the numerator is the number of animals that are found to have a tumour at necropsy after dying of unrelated causes during the interval, and the denominator is the total number of animals that die of causes unrelated to the tumour during the interval. The prevalence methods analysis tests for equality of the resulting proportions in control and exposed groups across all of the age intervals using standard contingency table methods. Although significant differences between control and exposed groups are taken as evidence that the test agent is associated with changes in the tumour onset rate (Peto *et al.*, 1980), the prevalence method can be shown to test for equality of the incidence rates of underlying tumours only under strict assumptions (McKnight & Crowley, 1984). Moreover, the proportions formed within each interval will estimate the true tumour prevalence rates only if animals dying of causes other than the tumour of interest are representative, with respect to tumour presence, of all animals surviving the interval (Lagakos & Ryan, 1985). Testing for equality of the proportions of animals dying with tumours is equivalent to testing for equal prevalence rates only if, within each age interval, the risk of dying for tumour-bearing animals relative to tumour-free animals is the same in control and exposed groups. If, for example, an exposed animal with a tumour is twice as likely as an exposed animal without a tumour to die during a particular age interval, then, for prevalence methods to be legitimately applied, a control animal with a tumour should also be twice as likely as a control animal without a tumour to die during that age interval.

If the relative risk of dying for tumour-bearing animals (relative to tumour-free animals) varies among control and exposed groups for some age intervals, then the

prevalence methods may not provide valid tests for equality of tumour prevalence rates. The assumption that these relative risks are equal in control and exposed groups for all age intervals may not be realistic for all carcinogenesis experiments. Suppose, for example, that the presence of a tumour accelerates death in exposed animals more than in control animals for some cause unrelated to tumour. Then the relative risk of dying for tumour-bearing animals relative to tumour-free animals will tend to be higher in exposed groups than in the control group, at least in early age intervals. The added burden of even a nonfatal tumour in an animal which is exposed to a maximum tolerated dose of some chemical may put the animal at a substantially higher risk of death due to some cause not related to a tumour, while the same tumour in a healthy control animal would result in little increased risk. As in the related situation for fatal tumours (previously discussed), such an occurrence could give the appearance of tumour acceleration; however, the apparent acceleration would reflect differences in mortality patterns, not an acceleration of tumour incidence or tumour development. As with fatal tumours, incidental tumour analyses that indicate significant acceleration of tumour onset without an accompanying increase in the proportion of animals observed with tumours must be interpreted with care.

2.13 Combining analyses of fatal and incidental tumour data

For any particular organ, it may be that tumours are observed both in a fatal and an incidental context. If, for example, a group of animals has a high intercurrent mortality rate, many potentially fatal tumours may be observed in an incidental context because tumour-bearing animals die from other causes before being killed by the tumours. A method of combining the analyses based on fatal and incidental tumours is presented in Chapter 5. When the contexts of observation of all tumours are to be used in evaluating an animal carcinogenesis experiment, the determination of carcinogenic potential at a particular organ must be based on this combination of fatal and incidental tumour analyses (Peto *et al.*, 1980). Evaluation of carcinogenicity based either on fatal tumours separately or incidental tumours separately can be misleading. For example, consider an experiment in which the age-specific incidence rates for some type of tumour are equal in the control and exposed groups, and the proportion of these tumours which would eventually be observed in a fatal context is the same in the control and exposed groups. Suppose, however, that the exposed animals with these tumours have a higher risk of dying from causes not related to tumours than do the control animals with the same tumours. Then, many of the exposed animals with these tumours that would eventually have been observed in a fatal context may have their tumours observed in an incidental context because the intercurrent mortality rates have been selectively increased. Thus, if only the fatal tumour analysis is considered, there might appear to be a deficit of fatal tumours in the exposed groups. Such a selective increased risk of intercurrent mortality, however, would lead to a higher relative number of exposed animals observed with tumours in an incidental context. Thus, in the combined analysis, the deficit of exposed animals with fatal tumours would, in some sense, be balanced by an excess of exposed animals with incidental tumours. The combination of the two analyses, one based on tumour death rates and the other based

ostensibly on tumour prevalence rates, may seem somewhat contrived and excessively complex. In fact, it is difficult to justify such an analysis rigorously (McKnight & Crowley, 1984). This analysis, however, is presently the best solution to the difficult problem of using information on the ages at death of tumour-bearing animals to test for differences in tumour incidence rates. This problem arises whenever the ages at observation of tumours are determined by the deaths of the tumour-bearing animals, because the ages at death may be influenced by factors unrelated to the tumours. Although animal sacrifice schemes can be formulated to alleviate the problem partially, it will be inherent in any experiment in which all animals are allowed to die a natural death (McKnight & Crowley, 1984).

In order to justify formally the use of the fatal and incidental tumour analyses to make inferences about tumour incidence rates, certain assumptions, some of which are difficult to verify, must be made. In particular, an assumption similar to that of noninformative censoring is required (Kodell *et al.*, 1982a; Lagakos, 1982). Such an assumption implies that, with regard to all life-shortening disorders (including toxicity) not caused by the presence of a tumour, a tumour-bearing and tumour-free animal are equally healthy. In other words, it is assumed that tumour-bearing and tumour-free animals are equally susceptible to, and have the same distribution for time to death due to, each life-shortening disorder. There is some experimental evidence to argue against such an assumption (Lagakos & Ryan, 1985), and there is additional indirect evidence which suggests it may not always be reasonable. The presence of a tumour has been shown to impair certain specific immune functions in rodents (Howell *et al.*, 1975; Nelson *et al.*, 1980; Perry & Greene, 1981); in addition, several carcinogens have been shown to be immunotoxic for a number of functions (Ball, 1970; Parmiani *et al.*, 1971; Vos & de Roij, 1972; Gainer & Pry, 1972; Koller, 1973). Thus, it is possible that tumour-bearing animals may be more susceptible than tumour-free animals to disease in some cases, and that this effect could be more pronounced in exposed groups than in control groups. Whether or not the assumption of noninformative censoring is reasonable can be determined only by experiments designed to compare the general health status of tumour-bearing and tumour-free animals. It should be noted, however, that the presence of immunotoxic effects such as those mentioned above may have little impact on mortality patterns in experiments in which test animals are housed under conditions free from pathogenic organisms.

2.14 Context of observation unavailable

In many studies conducted to date, the contexts of observation are unknown for all or most tumours. As definitive information in this regard will probably remain incomplete in many future studies, a method of adjusting for differential survival rates which does not require the accurate categorization of individual tumours as incidental or fatal would be useful. Although no such method is currently available, the lack of information regarding contexts of observation of tumours does not necessarily preclude a valid assessment of the carcinogenic potential of a test compound. In some cases, it may be possible to classify particular lesions as being almost always fatal or almost always incidental, although there is currently no consensus among pathologists on this

point. Another possible approach is to carry out two separate analyses, one taking all tumours to be incidental and the second taking all tumours to be fatal. In many cases, these two analyses will be in agreement, lessening concern regarding the lack of data on context of observation. It should be cautioned, however, that the actual level of significance attained by an analysis using accurate information on contexts of observation may not always be bracketed by the levels of significance that are attained in the separate analyses performed by assuming that all tumours are incidental or all tumours are fatal (Lagakos & Louis, 1985). Moreover, doubling the number of statistical tests will accentuate the problem of multiple comparisons. In some cases, no valid assessment of carcinogenic potential can be made without knowing the context of observation (Peto *et al.*, 1980). Therefore, attempts to acquire reliable information on context of observation are to be encouraged.

2.15 Analysis of crude tumour rates

In view of the difficulties involved in determining the context of observation of a tumour (that is, fatal or incidental) and in interpreting the subsequent analyses based on ages at death of tumour-bearing animals, it has been suggested that the analysis of crude tumour rates be performed, as a first step in evaluating tumour incidence data (Gart *et al.*, 1979). A crude tumour rate for an experimental group is defined as the number of animals in the group which develop a tumour (regardless of the context of observation) at some organ during the experiment, divided by the number of animals in the group in which that organ was examined for the presence of a tumour. To avoid any possible bias due to differential early mortality, it is often advisable to eliminate all the animals that died in the experiment prior to the occurrence of the first tumour when crude tumour rates are computed (Gart *et al.*, 1979). Although differences among groups with respect to longevity can seriously affect the analysis of crude tumour rates, one advantage of analysing the crude rates is that the impact of differences in mortality can readily be predicted. The assumption made in assessing the effect of mortality on crude rates is that the longer an animal survives, the greater is its probability of developing a tumour. Thus, if two groups of animals have identical (unobservable) age-specific tumour rates, but one group has higher intercurrent mortality rates, then the group with the higher intercurrent mortality rates will tend to have a lower crude tumour rate. The effect of differences in mortality upon the analysis of crude tumour rates is summarized in Table 2.2 (reproduced from Gart *et al.*, 1979). Under the column for tumour association with treatment, '+' indicates a significant increase in crude tumour rates associated with increasing exposure level, '0' indicates no significant change in the crude tumour rates with increasing exposure level, and '−' indicates a significant decrease in crude tumour rates associated with increasing exposure level. Under the column for mortality association with treatment, '+' indicates increasing mortality with increasing exposure level, '0' indicates no change in mortality with increasing exposure level, and '−' indicates decreasing mortality with increasing exposure level.

The table indicates certain situations in which the analysis of crude tumour rates may suffice to provide evidence of the carcinogenicity of a test agent. In particular, when

Table 2.2 Interpretation of the unadjusted analyses of tumour incidence in light of the survival analyses

Outcome type	Tumour: association with treatment	Mortality: association with treatment	Interpretation[a] of the unadjusted test of tumour incidence
A	+	+	Unadjusted test may underestimate tumorigenicity of the treatment
B	+	0	Unadjusted test gives a valid picture of the tumorigenicity of the treatment
C	+	−	Tumours found in treated groups may reflect the longer survivorship of the treated groups. A time-adjusted analysis is indicated
D	−	+	The apparent negative finding in tumours may be due to the shorter survivorship in the treated groups. A time-adjusted analysis and/or a re-test at lower doses is indicated
E	−	0	Unadjusted test gives a valid picture of the possible tumour-preventive capacity of the treatment
F	−	−	Unadjusted test may underestimate the possible tumour-preventive capacity of the treatment
G	0	+	High mortality in treated groups may lead to the unadjusted test missing a possible tumorigen. Adjusted analysis and/or a re-test at lower doses is indicated
H	0	0	Unadjusted test gives a valid picture of the lack of association with treatment
I	0	−	Longer survivorship in treated groups may mask a tumour-preventive capacity of the treatment

[a] Many of these interpretations assume that the maximum tolerated dose (MTD) was used and that a sufficient proportion of animals survived in sufficient numbers for an appropriate length of time

testing for increases in tumour incidence rates in exposed animals, a definitive statement regarding carcinogenicity can usually be based on the analysis of crude rates in outcomes A, B, E, F, H, and I given in Table 2.2. In situations with outcome A in which the differences among crude rates are of marginal significance, adjusted analyses may be desired to account for the differential mortality, thus improving the strength of the statistical evidence regarding carcinogenicity. Adjusted analyses refer to the previously discussed analyses of fatal and incidental tumours using the ages at death of tumour-bearing animals. Adjusted analyses usually are essential only in outcomes C, D, and G. There are clearly situations in which analyses of crude rates do not allow a definitive statement regarding carcinogenicity. However, in cases in which the crude analyses do suffice, the interpretation of the results is straightforward, and no unverifiable assumption is needed to justify the method of analysis. In some situations (for example, if it is known that a type of tumour is rapidly lethal or if acceleration is suspected), both adjusted and unadjusted analyses may be valid, but the adjusted analyses can have greater power.

2.16 Concomitant information

Although the analysis of data on survival and tumour pathology is of primary importance in evaluating a carcinogenesis experiment, other data which are sometimes useful in evaluating carcinogenic potential may be obtained from long-term animal experiments. For example, both survival and tumour rates can vary with varying patterns of weight gain and food consumption (Weindruch & Walford, 1982; Haseman, 1983a). Thus, observations of concomitant variables, such as animal weights, food consumption and measurements from haematology or other clinical chemistry examinations, are often recorded at various times during the course of an experiment (IARC, 1980). Whenever possible, such observations should be recorded for individual animals and not merely summarized by cage or group average. By use of stratification, analysis of tumour incidence data can be adjusted quite easily for differences in certain concomitant variables, such as initial body weight. Alternatively, logistic regression methods can sometimes be useful in adjusting for concomitant variables (Dinse & Lagakos, 1983). Formal incorporation into the statistical analysis of those variables for which observations are obtained at various times during the course of an experiment can be technically difficult. Furthermore, interpretation of such analyses is often not straightforward. If, for example, treatment causes an observable response which is on the causal pathway to carcinogenesis, then adjusting for this observable response in the analysis of tumour rates can lead to incorrect inferences regarding carcinogenicity. Nevertheless, observations on such variables often provide valuable information necessary to the valid interpretation of a carcinogenesis experiment (see, for example, the discussion section in Tarone *et al.*, 1981).

2.17 Need for interdisciplinary decision process

A variety of statistical methods for the analysis of tumour pathology data from a long-term animal carcinogenesis experiment will be presented in this book. The methods which are appropriate for a given experiment are determined primarily by the extent of the pathology reporting (that is, whether or not the contexts of observation of all tumours are reported) and by the presence or absence of differences between the control and exposed groups with respect to intercurrent mortality. Whatever statistical methods are used, it should always be kept in mind that the statistical analysis is only one component in the evaluation of an animal carcinogenesis experiment. The process of carcinogenesis is extremely complex, and the proper evaluation of tumour pathology data requires the careful appraisal of intricate patterns of lesions, some of which are malignant and some benign. Certain complicated patterns of response may not be quantified easily using statistical methods. Thus, the evaluation of long-term carcinogenicity experiments must be an interdisciplinary process, incorporating the input of pathologists, toxicologists and other scientists.

2.18 Overall evaluation of carcinogenicity

For the overall evaluation of the carcinogenicity of a given exposure, say a chemical, many considerations are required. As far as data from long-term animal experiments

are concerned, there may be several studies available, and, in addition, there will also be information on short-term tests and epidemiological studies. Thus, within the IARC programme on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, the assessments of evidence for carcinogenicity from studies in experimental animals are classified into four categories (IARC, 1982b):

'(i) *Sufficient evidence* of carcinogenicity, which indicates that there is an increased incidence of malignant tumours: (a) in multiple species or strains; or (b) in multiple experiments (preferably with different routes of administration or using different dose levels); or (c) to an unusual degree with regard to incidence, site or type of tumour, or age at onset. Additional evidence may be provided by data on dose-response effects, as well as information from short-term tests or on chemical structure.

'(ii) *Limited evidence* of carcinogenicity, which means that the data suggest a carcinogenic effect but are limited because: (a) the studies involve a single species, strain, or experiment; or (b) the experiments are restricted by inadequate dosage levels, inadequate duration of exposure to the agent, inadequate period of follow-up, poor survival, too few animals, or inadequate reporting; or (c) the neoplasms produced often occur spontaneously and, in the past, have been difficult to classify as malignant by historical criteria alone (e.g., lung and liver tumours in mice).

'(iii) *Inadequate evidence,* which indicates that because of major qualitative or quantitative limitations, the studies cannot be interpreted as showing either the presence or absence of a carcinogenic effect; or that, within the limits of the tests used, the chemical is not carcinogenic. The number of negative studies is small since, in general, studies that show no effect are less likely to be published than those suggesting carcinogenicity.

'(iv) *No data* indicates that data were not available to the Working Group.

'The categories *sufficient evidence* and *limited evidence* refer only to the strength of the experimental evidence that these chemicals are carcinogenic and not to the extent of their carcinogenic activity nor to the mechanism involved. The classification of any chemical may change as new information becomes available.'

Similar criteria applicable to the epidemiological context are used in the assessment of evidence for carcinogenicity from studies in humans. The final evaluation of carcinogenic risk to humans relies strongly on the epidemiological information but also incorporates the evidence from short-term tests and from studies in experimental animals.