Chapter 4 in "Testing for Toxicity" 559
Edited by J.W. Gorrod
Taylor & Francis 1981 ISBN 0-85066-218-4

4. Testing *in vivo* for general chronic toxicity and carcinogenicity

pp 29-43

F. J. C. Roe

Introduction

The concept of testing a chemical agent for chronic toxicity and carcinogenicity depends on the assumption that it is possible to distinguish changes that are outside the normal range for unexposed control animals. In practice, many of the changes that can be produced by toxic agents are also seen from time to time in untreated control animals. This may be because the background environment of the control animals is contaminated by toxic agents, or because some forms of toxicity represent no more than the enhancement or advancement in time of age-related changes. In any event, properly constructed and completely valid contemporary control groups are a *sine qua non* of all general toxicity and carcinogenicity studies *in vivo*. It is virtually never safe to rely on previous experience with animals of the same strain, even in the same laboratory and even if the animals concerned are from the same inbred stock.

In theory, there is no limit to the number of different kinds of toxic effect that may occur in response to toxic agents. In practice, a limit is set by the methods used for detecting them (e.g. ultrastructural changes are only recognized if e.m. studies are undertaken).

Observations in chronic toxicity tests fall into two classes: those made during life and those made at or after death. During life, animals need to be examined daily for general condition. This is not so much to provide information on toxicity, but for humane reasons and to ensure minimization of risk of loss through cannibalism and autolysis. More relevant to the assessment of toxicity are measurements of body weight, food and water consumption, urinalysis, blood sampling for haematological changes and changes in serum chemical and biochemical parameters, and periodic opthalmoscopic examinations, etc. Apart from the daily checks of general condition, less frequent (e.g. once weekly) but more detailed clinical examinations need to be made. In the course of these the condition of the coat, the state of the skin and natural orifices and the presence of visible or palpable lesions should be recorded. Protocols for chronic toxicity tests usually refer to observations on 'behaviour'. However, there seems to be no agreed system for classifying behavioural disturbances. Few observations are recorded under this heading as a rule, and many of those that are refer to disturbances recognizable by other means (e.g. the observation that an animal

spends a lot of time at its water bottle is recognizable and more accurately assessed by measurements of water consumption and urine output). Three problems relate to the observation of behavioural changes. Firstly, the scientists and technicians responsible for the conduct of toxicological tests often lack formal training in this area. Secondly, nocturnal animals are usually observed during the day when they are asleep. Thirdly, many untreated caged animals show behavioural disturbances anyway because of the fear, boredom, confinement and lack of sexual fulfilment that are features of life in laboratory cages.

Apart from the normal schedule of routine observations, there is no limit to the number of special investigations that can be included in the protocols for a chronic toxicity study (e.g. function tests referable to various body systems). However, one thing should be stressed: it is not possible to carry out all possible measurements on the same set of animals without grossly interfering with the conditions of the test and without incurring the risk of introducing into the results artefacts which might be confused with effects attributable to exposure to the test agent. It is essential, therefore, for those designing chronic toxicity tests to be very selective in the list of measurements *in vivo* that they propose should be made in any one study.

Because of the possibility that submission of animals to test procedures will alter tumour risk, some authorities are opposed to the concept of combined chronic toxicity–carcinogenicity studies in laboratory rodents. A common compromise is to design a study in which the majority of animals in each group are not subjected to any clinical tests, but which includes supplementary satellite sub-groups (possibly only top dose and control) for clinical measurements. These provide the *in vivo* data dependent on blood sampling or other interference with the well-being of animals. This leaves the option open at the end of the study to include or exclude tumour-incidence data derived from the satellite groups when evaluating the data as a whole for evidence of carcinogenic activity.

Terminal observations include body weight, final blood tests, marrow smears, full records of macroscopic findings at necropsy, selected organ weights and microscopic examination of tissues. The macroscopic observations made at necropsy should, in my opinion, be regarded as the principal and most important 'harvest' of data derivable from any experiment. If organs and tissues are macroscopically normal, equivocal measurements and effects of doubtful significance observed during life need not be taken too seriously. Comparison of groups for tumour incidence should be based on the numbers of macroscopically visible swellings that are subsequently, on microscopy, proved to be neoplasms, for it is difficult to assess the significance of the finding by chance of a neoplasm of microscopic dimensions in an arbitrarily selected 5 μm-thick slice of a large organ such as the liver.

Route of administration

It is now generally accepted that realistic routes of administration of test substances should be used in tests for chronic toxicity or carcinogenicity. The subcutaneous or intramuscular injection of a substance to which man is,

or will be, exposed by the oral route introduces two kinds of problem in interpretation. Firstly, the parental route bypasses the liver, with the result that the administered substance may reach sites in the body which it would never reach after oral administration. Secondly, there is now plenty of evidence that sarcomas may arise *non-specifically* at the site of subcutaneous or intramuscular injection of substances which exhibit no carcinogenic activity when given by mouth.

In practice, it is not always possible to use the same route of administration as for man. Toxicological studies with tobacco smoke provide an interesting example. Despite the association between smoking and risk of lung cancer in humans, it has not proved possible to induce lung cancers in animals by exposing them to tobacco smoke by the inhalation route. Man voluntarily avoids serious toxicity from nicotine and carbon monoxide by adjusting the volume of smoke he takes into his lungs from each puff of a cigarette and the frequency at which he takes puffs, etc. Animals exposed to tobacco smoke in a specially designed smoke-exposure chamber have no such control over the doses of nicotine and carbon monoxide which they inhale. If smoke concentrations are too high, they die from nicotine or CO overdose, but if smoke concentrations are reduced to avoid such overdose, the highest level of exposure of the animals is equivalent to no more than that for human very light smokers. Apart from this, in many exposure systems involving intermittent exposure to smoke (e.g. 15 seconds exposure, during each minute) the 'smarter' animals soon learn to hold their breath while the smoke is present in the chamber. They then make up for their oxygen deficit during the smoke-free intervals in the exposure regime. Because of these difficulties, it is now generally agreed that the most practical way to compare the smoke from different tobacco products for potential carcinogenicity is to apply solution–suspensions of smoke condensates repeatedly to the dorsal skin of mice, and measure response in terms of the development of skin tumours.

Oral administration may be achieved either by admixture of the test agent with the diet or by the introduction of measured doses of it (in a vehicle if necessary) directly into the stomach (i.e. by gavage). Admixture with the diet requires less skill, but makes it much more difficult to ensure that animals receive prescribed doses. Also, 'dosing' is spread throughout the 24 hours, which may not represent human exposure to a drug that is given, say, once or twice a day by mouth after meals. Finally, there is the possibility of direct interaction between the test material and food constituents. There is no simple solution to this quandary. From the armchair of the academic pharmacologist, it is easy to say 'choose the method of exposure which best matches human exposure in terms of absorption from the gut, and blood and tissue concentrations throughout the 24 hours'. In practice, neither method may do this well, and in any case the advice is irrelevant in the case of a carcinogenicity study involving exposure to the maximum tolerated dose and/or to levels tens, hundreds or thousands of times those to which man is to be exposed.

Vehicle

It is all too easy to forget how small a mouse is when choosing the vehicle for

administering a test agent in a toxicological study. Take, for example, the decision to administer an agent in once-daily doses by gavage to mice weighing 25 g, using a volume of 0·1 ml ethanol as the vehicle. This dose is equivalent to a human inbibing, during less than one minute each day, 280 ml absolute alcohol! In other words, the unfortunate choice of vehicle results in the toxicity test being carried out in *chronically alcoholic mice*. Similarly, arachis oil may profoundly alter the percentage of lipid ingested by animals each day, and this in turn may determine, *inter alia*, both risk of liver damage and risk of liver tumour development.

Doses and dose ranges

Toxicologists who have basic training in pharmacology will naturally be attracted by experimental designs involving multiple logarithmically spaced dose levels covering a range that includes the human exposure level. Furthermore, they will seek to match dose levels in different species in the light of comparative pharmacokinetic data. Theoretical oncologists, on the other hand, are apt to demand that tests for carcinogenicity be conducted at maximum tolerated dose (MTD) levels, irrespective of human exposure levels. I do not at this point wish to get embroiled in arguments concerning the scientific justification or wisdom of this demand. Suffice it to say, it is usually possible to take account of both concepts in a single experimental plan. In any case, it is advisable to undertake preliminary tests, not only to determine the MTD under the conditions of the proposed main study, but to ensure that the proposed dose ranges are sensible. I know of many examples of a disastrously high death rate at the start of a large study designed on the basis of wrong assumptions concerning general toxicity.

Choice of species

Armchair toxicologists glibly recommend, for chronic toxicity testing purposes, the choice of a species that mimics man in the way that it metabolizes the substance to be tested. In practice, this is rarely more than an unattainable ideal. Frequently, comparative metabolic data are not available at the time chronic toxicity tests need to be started. Alternatively, when such data are available, there is found to be no animal species which mimics man, or the only species that do behave in the same way as man are too large and long-lived for chronic toxicity or carcinogenicity testing purposes. In fact, there are really only three species—rat, mouse and hamster—that can be used for carcinogenicity tests. Tests lasting two to seven years in dogs and monkeys may seem to the scientists and toxicologists concerned to be long, but unless animals are studied well into the last quarter of their available lifespan, chronic toxic effects and effects on cancer risk may be missed.

How long should tests last?

There is no straightforward answer to this question. The incidence of most kinds of cancer and of many degenerative disorders increases exponentially

with age, and treatment-related effects on the incidence of such may not become evident until late in life. For manifestations of chronic toxicity or cancer risk which are detectable *in vivo*, there is no conflict—the longer an experiment continues, the more data accumulate. However, for toxic changes, including slow growing internal tumours, which are only detectable at necropsy (i.e. incidental findings at necropsy) it is possible to miss an effect of treatment on age-standardized risk by continuing an experiment for too long. One rather costly way out of this dilemma is to include, in the design of a study, subgroups to be killed at intervals throughout its course. Otherwise, it is necessary to try to reach a sensible compromise in the light of the particular objectives of the study. In any event, it is generally accepted nowadays that a carcinogenicity study in the rat should continue for not less than two years, and one in the mouse or the hamster for not less than 18 months. In some laboratories the time of termination of studies is determined by actual survival experience (e.g. the time at which there are only 25% of animals surviving at the second dose level down, the two sexes being regarded as separate experiments). A disadvantage with this approach is that the workloads of laboratory staff cannot be scheduled in advance.

Dosimetry

Needless to say, every effort should be made in any chronic toxicity study to establish the actual, as distinct from the theoretical, dose to which animals are exposed. For this purpose, diets containing test substances need to be analysed for adequacy of admixture with the diet and stability of the test substance, etc. Information on absorption and excretion should be collected throughout studies, and it should not be assumed that mature animals chronically exposed to a test substance handle it in the same way as young animals or as naive animals exposed to it only once. Also, it should not be assumed that high doses are handled either quantitatively or qualitatively in the same way as low doses.

When is an effect a toxic effect?

It should be obvious that not all observed effects of test substances in chronic toxicity studies are toxic effects. Some are adaptive and, in the case of drugs, some are manifestations of their known pharmacological activity. I have been told of a case where a governmental regulatory body nearly banned a proposed drug on the grounds that it caused 'leucopenia' in rats. The drug, a very useful and effective antibiotic, had benefited rats in the test by reducing the incidence and severity of various infections which they carried, and in doing so reduced the pathologically high white-cell counts seen in untreated controls to more 'normal' levels! During pregnancy, liver weights increase, hence hormones which bring about a hormonal status similar to that existing during pregnancy must be expected to increase liver weight relative to body weight. Toxicology is fundamentally more a biological science than a mathematical or chemical one. Therefore, the final stage of the assessment of a toxicity study should not be left solely either to the statistician or to the

chemist. There is still an important job to do after the statistician has identified differences between treated and control groups which achieve statistical significance. A toxicologist with adequate training and experience in physiology, pharmacology and pathology is required to consider what the findings are likely to mean biologically, both for the animals in the test and for humans. When doing this he will need to consider the precise conditions of the test, the nature of the test substance and all other available information about it, and to draw on all his resources of fundamental biological knowledge and experience.

Diet and effects of overnutrition

During the past few years I have become increasingly convinced that the conditions under which most chronic toxicity and carcinogenicity studies are conducted are absurdly unnatural, with the result that numerous artefactual changes are seen in both control and treated animals. The fact that animals are confined without exercise in small cages and deprived for all or most of their lives of normal sexual fulfilment is no doubt of considerable importance. However, possibly even more important is the fact that they are consistently overfed, sometimes with unnecessarily nutritious food overladen with protein, fat and sometimes minerals. The formulae of diets commonly used today were worked out many years ago, before pathogen-free animals became generally available. As diseases have been eradicated, animals' nutritional requirements have no doubt changed. Also, diets that were deemed optimal for young, growing, conventionally maintained animals are probably entirely unsuited for geriatric SPF animals. To make matters worse, feed compounders have tended to increase the lipid content of diets to enable them to pasteurize food pellets without the risk of their breaking up. Gellatly (1975) saw far higher incidences of liver tumours in mice fed on a high-fat diet than in mice fed on a similar diet with lower fat content.

Pelvic nephrocalcinosis is nowadays a common finding in laboratory rats of both sexes, while cortico-medullary nephrocalcinosis has come to be regarded by many pathologists as a 'normal' finding in female laboratory rats. I am quite sure that neither kind of nephrocalcinosis is 'normal'. In reality, they are consequences of paying inadequate attention to the mineral requirements of the animals in question. Many laboratory diets contain too much calcium and phosphorus, but too little magnesium. The results of chronic toxicity tests on substances which themselves alter mineral balance become difficult to interpret against a background of manifest mineral imbalance in control animals.

Animals in the wild rarely go for long periods during which they do not have to fight or take risks to obtain their food. To survive they need to remain slim, agile and alert. By an accident of history, it is the universal custom to provide laboratory rodents under test with food *ad libitum*. Such provision, along with the facts that the diet provided tends to be overnutritious and that the animals have no exercise and are disease-free, results in widespread obesity. In laboratory rodents, *ad libitum* feeding and the consequent obesity are associated with greatly increased cancer risk and decreased

survival. I strongly urge those not familiar with these facts to read recently published papers by Tucker (1979), Conybeare (1980) and Roe (1979). Tables 4.1 and 4.2 summarize data from the first two of these papers.

Table 4.1. Effects of dietary restriction on tumour incidence in specified-pathogen-free Wistar rats (Tucker 1979).

	Males		Females	
	Ad lib.	20% restricted	*Ad lib.*	20% restricted
Survival to two years (%)	72	90	68	88
Tumour bearing at or before two years (%)	66	24***	88	56*
Number of tumours per rat (mean)	0·94	0·27***	1·18	0·76**
Rats with pituitary tumours (%)	32	0***	66	39**
Rats with mammary tumours (%)	0	0	34	6***

* $P < 0{\cdot}05$, ** $P < 0{\cdot}01$, *** $P < 0{\cdot}001$, statistical significance (ignoring better survival of diet-restricted groups).

Table 4.2. Effect of simple dietary restriction on survival and tumour incidence in Swiss mice fed on commercially supplied cubed diets (Conybeare 1980).

	Males		Females	
	Ad lib.	20% restricted	*Ad lib.*	20% restricted
Number of mice	160	160	160	160
Survival to 83 weeks (%)	58	66	62	77*
Any tumour at any site (%)	44	22·57**	31	11***
Any malignant tumour at any site (%)	11	4**	14	4***
With lung tumour (%)	19	12*	15	5***
With liver tumour (%)	29	7·5***	4	0·6***
With lymphoma (%)	2·5	0·6**	7	2·5***
With other tumours (%)	5·0	2·5*	7·5	2·5*

* $P < 0{\cdot}05$, ** $P < 0{\cdot}01$, *** $P < 0{\cdot}001$, statistical significance (ignoring better survival of diet-restricted groups).

Interestingly, two of the kinds of tumour associated with overfeeding in rats (i.e. pituitary tumours and mammary gland tumours) are both associated with high circulating levels of prolactin. In non-pregnant humans, circulating prolactin levels rarely exceed about 40 ng/ml. In *ad libitum*-fed female

rats aged three months or over, serum prolactin concentrations begin to rise to levels as high as 500–600 ng/ml in old age. In *ad libitum*-fed males, concentrations also rise, but only by about half as much. As far as I know, no-one has studied the effects of dietary restriction on prolactin levels, but I would wager a silk shirt that lower concentrations would be found in diet-restricted animals. Furthermore, as an article of faith, I believe that the astronomically high prolactin levels seen in *ad libitum*-fed animals are totally abnormal and render such animals unsuitable as models for many kinds of toxicity study.

In my opinion, a very serious reconsideration is long overdue of the conditions under which we conduct chronic toxicity and carcinogenicity studies; we must seek for a way to maintain animals without their becoming obese, overladen with minerals, and grossly hormonally abnormal in a variety of ways. I suspect that to achieve this it will be necessary to restrict dietary intake *as a rule*. In addition, it may be necessary to provide animals with some outlet for their sexual urges. I predict that the time will come when we look back at the conditions under which most chronic toxicity tests are conducted today and regard them as ludicrously inadequate!

Statistical analysis

Richard Peto's (1974) guide-line paper on methods of analysis of carcinogenicity data should be required reading for those involved in undertaking or interpreting carcinogenicity studies. Tumours (e.g. skin or subcutaneous) that are evident during life, rapidly growing tumours that kill, and slowly growing tumours of internal organs that are only discovered when animals die for other reasons, present different problems when it comes to comparing treated and control groups for tumour risk. Although Peto's words of wisdom relate to the handling of tumour data, they apply with equal force to other kinds of manifestation of toxicity.

Quantification of histopathological data

'A pathologist who regards statistics as unnecessary is a menace.' I make no apology for inventing this quotation, and give my permission for it to be used freely in the future (preferably with attribution) wherever appropriate!

Many veterinarians and medical men come to experimental pathology without relevant experience in the art of experimentation and without really understanding that the purpose of much of the work in which they find themselves engaged has little to do with diagnostic pathology. Moreover, it is not uncommon for pathologists of this ilk to find themselves in the position of mere slide readers with little or no responsibility for the drawing-up of protocols, for the supervision of the running of the experiments, or even for the conduct of the necropsies which provide the sections they have to read.

Thus, in the past it often happened that even the most quantitatively minded slide-reading pathologist could not produce statistically analysable data, because of fundamental faults in experimental design and execution over which he had no control prior to the point when the slides were produced

for him to read. The position is generally not much better today, although the Good Laboratory Practice (GLP) Guidelines should do much to improve the situation. Some of the commoner faults are as follows:

(*a*) failure to randomize animals between experimental and control groups by an acceptable method;
(*b*) failure to handle and/or observe control animals as often as treated animals;
(*c*) keeping treated and control animals in different rooms, in different places in the same room, or on different shelves or racks;
(*d*) having different people responsible for treating and/or observing control and treated animals;
(*e*) having different people responsible for deciding when animals in control and treated groups should be killed because they are moribund;
(*f*) having different people responsible for carrying out necropsies on Saturdays and Sundays than on other days in the week, so that more animals from groups affected by toxicity have weekend necropsies than from control groups.

In some cases a combination of these faults could render any attempt at a quantitative histopathological evaluation quite pointless.

The purposes of animal experiments

Animal experiments may be conducted for a variety of reasons. It is obviously important, therefore, for the pathologist reading slides to know precisely what the purpose or purposes are. Very often the aim is simply to detect effects—physiological or pathological—which are manifested as differences between control and treated groups, between high-dose and low-dose groups or between groups exposed either to substance A or substance B, etc. In all such cases, it is important to be able to assess whether a difference is attributable to treatment or chance, and the soundest basis for making such an assessment is a statistical one. Of course, not all differences are susceptible to useful or meaningful statistical analysis and there is always a need for both common sense and experience. But there is only a fine line between experience and blind prejudice, unless the former is quantitative rather than anecdotal.

The material provided

To assess whether a difference between two groups of animals in the incidence of a lesion is likely to be real or due to chance, it is necessary to be able to compare like with like. The pathologist whose involvement in an experiment only begins when he receives a pile of sections is often unable to fulfil this requirement, not only because of faults in the design and conduct of experiments, but also because of biases introduced at the time of post-mortem or subsequently during tissue processing. In a well run laboratory every attempt is made to take standard samples of a standard list of tissues which are then processed in a standard way to provide a standard set of slides.

Furthermore, tissues will be orientated in a standard way in blocks so that approximately the same area of, say, lung or liver, is available for scrutiny by the pathologist. In other laboratories, however, uniformity is lacking, with the result that the amount, orientation and quality of particular tissues available for comparison from different animals is highly variable. In such circumstances the pathologist cannot compare 'like' with 'like'.

Personally, I believe that the macroscopic findings are extremely important—often more important than the microscopic ones—and that in any case meaningful histopathological evaluation must take into full account macroscopic data. It is thus vital to check that sections are available from all macroscopically undiagnosable lesions and that these are additional to, and not substitutes for, standard sections of organs, unless the standard section includes the lesion.

Some of the ways in which the slide reader may be forced to depart from the principle of comparing 'like' with 'like' are as follows.

(*a*) There may be non-standard samples of tissue (e.g. left lobe of liver from one animal and right lobe from another).

(*b*) There may be differences in the area of tissue on the slide.

(*c*) There may be differences in the orientation of tissue on the slide (e.g. transverse section of trachea from one animal and longitudinal section from another).

(*d*) There may be several samples of a tissue from one animal but only one sample from another. (*Example*: at necropsy the liver is seen to be pale so that five samples are taken instead of the usual two. In one of the extra bits a minute tumour which was not seen macroscopically is discovered. Such a tumour should be excluded from any comparison of tumour incidence between groups.)

(*e*) A ribbon of several sequentional sections of the same organ is mounted for one animal but only one section of the organ for another animal. (*Example*: the last section in the ribbon shows a lesion not present in the first. But how can one distinguish the first from the last? It is impossible to decide whether the lesion should be taken into account or not when comparing the risk of developing the lesion in the two different animals.)

(*f*) There may be differences in thickness, quality of cutting, and quality of staining.

(*g*) Tissues may be present in some animals but not in others. (*Example*: the protocols state that sections of the thyroid glands of rats will be prepared for each animal. In some cases, sections of one or both parathyroid glands happen to be included, in other cases neither parathyroid is present. A minute parathyroid adenoma, not noted macroscopically, is discovered under the microscope. Does this mean that the animal concerned was at higher risk of developing this kind of tumour or only that one is more likely to find parathyroid adenomas if sections of thyroid happen to include the parathyroid glands?)

(*h*) Further sections of a particular tissue are requested for some animals but not for others. (*Example*: three Leydig cell tumours are found among 20 rats in a treated group but none are found in the 20 controls. Serial sections

of the testes are requested for the 20 treated rats but not for the controls, and two further tumours are found. The discovery of these two additional tumours is uninterpretable unless serial sections are also cut from the testes of the 20 control rats.)

How the non-quantitative pathologist can add to the problems of interpretation

The most usual ways in which the pathologist does this can be listed as follows.

(*a*) He follows no standard procedure of reporting.

(*b*) He describes the same lesion in a variety of different ways.

(*c*) He records only the most severe or interesting lesion present. (*Example*: in one rat with a liver tumour, moderate bile-duct proliferation and widespread fatty degeneration, he mentions only the liver tumour, whereas in another rat which has no liver tumour he mentions the other two lesions. There is no logic in this, since there is no special relation between the three kinds of lesion in the sense that fatty degeneration or bile-duct hyperplasia are preneoplastic. So, by omitting to report the lesser lesions in the rat with the liver tumour, he is under-reporting their occurrence.)

(*d*) He adopts different criteria when looking at slides from treated and control animals.

(*e*) He uses imprecise criteria for the diagnosis of lesions and does not guard against 'diagnostic drift' as he works through the material from a large study.

(*f*) He fails to indicate clearly and in a standard way the size, severity, multiplicity and other characteristics of lesions, so that a real difference in severity between groups is overlooked.

Which parameters and how many of them?

Given adequate time and the possibility of examining the specimen at various levels of magnification, it would be easy to evolve a thousand or more different parameters for use in comparing samples of the same tissue from different animals. The art of pathology, as of most other pursuits, lies in the exercise of selectivity based on knowledge, experience and common-sense. In practice, a preliminary assessment of sections from high-dose and control animals usually serves to identify the parameters most likely to be useful.

Personally, I dislike having to record more than about six things at any one time when looking through a batch of sections of a particular organ or tissue. If for some reason it is necessary to record more parameters than this, I usually choose to go through sections more than once.

In choosing which parameters to record, the pathologist should take into account:

(*a*) the object of the study;

(*b*) the protocols for the study;

(*c*) knowledge of the strain of animal used with reference to background disease incidence;
(*d*) clinical, biochemical and haematological data; and
(*e*) necropsy data and observations made when tissues are trimmed after fixation and before further processing.

For each tissue, I list obligatory parameters but leave room for any additional comments (i.e. arbitrary parameters) which may be warranted. Obligatory parameters include:

(*a*) whether the tissue is present or not;
(*b*) how many sections of it are available for assessment; and
(*c*) whether the section (sections) are of adequate quality for making an assessment.

This information provides an essential basis for calculating the incidence of any lesion.

I have found it convenient to deal with obligatory parameters in tabular form, using one or more columns and a series of readily understandable abbreviations and hieroglyphics to indicate presence, absence, multiplicity, size, severity, etc. of lesions. With a little ingenuity, it is possible to reduce a lengthy histopathological report to a handful of numbers and symbols in a series of suitably headed columns.

It is my practice to have a trial run with each tissue, with a view to selecting the most important and appropriate parameters. The trial run consists of looking quickly through sections from a few high-dose and control animals of each sex in the full knowledge of the treatment they have received. Thereafter, I plough blindly through all the sections of the tissue in question—that is to say, without knowing to which group they belong. Of course, it often happens that new parameters are recognized as being important and relevant some time after the ploughing process has begun. Where this happens and a new obligatory parameter is created, I have to examine again all the sections looked at prior to its creation to assess them in respect of the new parameter. The other thing that happens is that one comes to recognize that one of the parameters originally thought to be important is not so. This presents no problem to anyone bold enough to cross out a whole column on a working table! I always leave myself a generous 'other comments' column for one-off lesions and poetic descriptions of oddities which defeat my diagnostic prowess.

Computerization of data

Most large contract laboratories and companies which undertake large-scale animal toxicology are already equipped with, or presently in the process of becoming equipped with, on-line computer facilities for dealing with haematological, clinical chemistry, urinalysis, body-weight data and other clinical observations. Also, there already exist several computer programmes for handling histopathological data. However, most of these are either unsatisfactory or cumbersome. In theory, a system into which the pathologist can

feed data directly without requiring the services of an intermediary to prepare data transfer sheets might seem the most economical. But this is only so if it is not expensive in terms of the pathologist's time. If the pathologist has to look up every lesion he sees in a code book before he can write down something the computer understands, then the time he requires to do the job is liable to be multiplied. At the other extreme, a system which is too simple is apt to produce reports which are unacceptable from a pathological viewpoint.

A chronic toxicity–carcinogenicity study involving say, 400 rats from which 40 tissues are subjected to microscopic examination, is likely to produce of the order of 100–200 000 (i.e. $400 \times 40 \times 10$) items of histopathological data, concerning presence of section(s), number of sections, quality of sections, presence or absence of lesion A, lesion B, lesion C, etc., multiplicity of A, B, C lesions, severity of lesions A, B, C, etc. It is possible to analyse even this large amount of data without the use of a computer. Indeed, it is better to do so than to use an inefficient or otherwise unsatisfactory programme.

My colleague, Mr Peter Lee, and I have prepared a system with the following features.

(*a*) The pathologist records his findings on a series of tables, each line of which represents one animal.

(*b*) Organs and tissues are examined in logical order from a series of sections presented to the pathologist in that same order.

(*c*) For each tissue the first few boxes deal with the presence, number and quality of available sections. The pathologist is then free to use as many other columns as he needs for other parameters. Each column he uses becomes an obligatory parameter for all animals of the same sex in the study. Unusual findings are recorded separately.

(*d*) The tables are designed in such a way that the data can be fed directly into a computer programmed to prepare summarizing tables, analyse data statistically, and print out full histopathological reports for all organs of all animals.

(*e*) The unusual findings referred to in (*c*) can either be fed directly into the computer as they are, so that they will appear with printed-out histopathological reports of individual animals, or they can simply be bound into the report of the study without going through the computer at all.

(*f*) The system is extremely flexible and leaves the pathologist largely free to choose which parameters to use.

(*g*) A two letter code is used to describe tissues (e.g. LU = lung, KI = kidney). Only rarely is there possible confusion (e.g. TH = thymus, so that TY has to be used for thyroid).

(*h*) Sizes of lesions are recorded, where appropriate, as mean diameters (e.g. in mm). Numbers of lesions are recorded as such up to an arbitrary maximum, and appropriate severity is recorded using a 6-point scale (i.e. 0, 1, 2, 3, 4 or 5), the highest being the most severe.

(*i*) Meaningful abbreviations are used as headings for parameter columns and for describing unusual lesions in the separate list.

(*j*) For each experiment, the pathologist provides a glossary of all abbreviations used and a list of criteria for the diagnosis of all grades of all lesions.

The above system is not designed to do anything but handle as efficiently and effectively as possible the data derived from one particular experiment. It is not primarily designed for making comparisons of data derived from different experiments.

Summary of main points

(*1*) Valid contemporary control groups are an essential requirement for chronic toxicity studies.

(*2*) It is not possible or sensible to try to carry out measurements for every conceivable form of toxicity in a single experiment. A sensible selection of parameters has to be made. This selection should be based on the nature of the test agent and the precise object of the study, rather than on which tests are easiest to perform.

(*3*) Toxic effects need to be distinguished from adaptive responses and from pharmacological effects.

(*4*) Observations made macroscopically at necropsy constitute the most important 'data harvest' from any chronic toxicity test. A high standard of microscopical observation cannot remedy deficiencies resulting from careless or inexperienced post-mortem examinations.

(*5*) A realistic route of administration should usually be chosen. If this is not possible, artefacts associated with an unrealistic route need to be recognized. The very real possibility of vehicle-associated effects must not be overlooked.

(*6*) The idea that the species of choice for chronic toxicity/carcinogenicity studies is one that mimics man in the way that it handles the agent metabolically is usually unhelpful. The rat, the mouse and the hamster are virtually the only species suitable for lifespan carcinogenicity studies and none of these may mimic man.

(*7*) Tests should usually last well into the last quarter of the available lifespan. However, this may pose problems in the interpretation of slowly developing internal changes which increase logarithmically in incidence in old age and which are only discoverable at necropsy.

(*8*) The need to use different statistical methods for the analysis of incidental and non-incidental tumour and chronic toxicity data is stressed.

(*9*) Overnutrition associated with *ad libitum* feeding shortens life and magnifies spontaneous tumour incidences. Overfed animals exhibit very high serum prolactin levels. Therefore, in the author's opinion, *it would probably be wise to conduct all chronic toxicity/carcinogenicity studies under conditions of slight dietary restriction.*

(*10*) Research is needed into the reasons for, and ways to avoid, the very high incidence of hormonal disturbances prevalent in untreated control animals of both sexes.

(*11*) Animal diets are often unbalanced in respect of minerals. Attention is drawn to the high incidence of nephrocalcinosis resulting from this.

(*12*) An approach to the quantification and computerization of histopathological data is outlined.

References

Conybeare, G. (1980), *Fd. Cosmet. Toxicol.*, **18**, 65.
Gellatly, J. B. M. (1975). In *Mouse Hepatic Neoplasia*, Editors: Butler, W. H. & Newberne, P. N., p. 77, Amsterdam: Elsevier.
Peto, R. (1974), *Br. J. Cancer*, **29**, 101.
Roe, F. J. C. (1979), *J. Human Nutr.*, **33**, 405.
Tucker, M. J. (1979), *Int. J. Cancer*, **23**, 803.