Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

04-Sep-2002

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

Cancels & replaces the same document of 28 August 2002

OECD Environment, Health and Safety Publications.
Series on Testing and Assessment No. 35 and Series on Pesticides No. 14

GUIDANCE NOTES FOR ANALYSIS AND EVALUATION OF CHRONIC TOXICITY AND CARCINOGENICITY STUDIES

JT00130828

ENV/JM/MONO(2002)19
Unclassified

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment No. 35

And

Series on Pesticides No. 14

Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies

Environment Directorate

Organisation for Economic Co-operation and Development

Paris 2002

Also Published in the Series Testing and Assessment:

No. 1, Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995)

No. 2, Detailed Review Paper on Biodegradability Testing (1995)

No. 3, Guidance Document for Aquatic Effects Assessment (1995)

No. 4, Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)

No. 5, Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)

No. 6, Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)

No. 7, Guidance Document on Direct Phototransformation of Chemicals in Water (1997)

No. 8, Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)

No. 9, Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides During Agricultural Application (1997)

No. 10, Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)

No. 11, Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)

No. 12, Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)

No. 13, Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries (1998)

No. 14, Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)

No. 15, Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)

No. 16, Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)

No. 17, Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)

No. 18, Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)

No. 19, Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (November 1999)

No. 20, Revised Draft Guidance Document for Neurotoxicity Testing (in preparation)

No. 21, Detailed Review Paper : Appraisal Of Test Methods for Sex Hormone Disrupting Chemicals (2000)

No. 22, Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)

No. 23, Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)

No. 24, Draft Guidance Document on Acute Oral Toxicity Testing (2000)

No. 25, Detailed Review Document Hazard Classification Systems for Specific Target Organ Toxicity Following Single or Repeated Exposure in OECD Member Countries (2000)

No. 26, Revised Analysis of Responses Received from Member Countries to the Questionnaire on Data Requirements for Acute Oral Toxicity Testing (2000)

No. 27. Guidance Document On The Use Of The Harmonised System For The Classification Of Chemicals Which Are Hazardous For The Aquatic Environment (2001)

No. 28, Guidance Document for the Conduct of Skin Absorption Studies (2000)

No. 29, Draft Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)

No. 30, Detailed Review Document on Hazard Classification Systems for Mixtures (2000)

No. 31, Detailed Review Paper on Non-genotoxic carcinogens detection: The Performance of in vitro cell transformation assays (2001)

No. 32, Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)

No. 33, Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)

No. 34, Guidance document on the Development, Validation and Regulatory Acceptance of New and Up-dated Internationally Acceptable Test Methods in Hazard Assessment *(Draft)*

No. 35, Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies.

No. 36, Report of the OECD/UNEP Workshop on the use of Multimedia models for estimating overall Environmental Persistence and Long-range Transport in the context of PBTS/POPS Assessment.

No. 37, Detailed Review Document on Hazard Classification Systems for Substances which pose an Aspiration Hazard.

Also published in the Series on Pesticides:

No. 1, Data Requirements for Pesticide Registration in OECD Member Countries: Survey Results (1993)

No. 2, Final Report on the OECD Pilot Project to Compare Pesticide Data Reviews (1995)

No. 3, Data Requirements for Biological Pesticides (1996)

No. 4, Activities to Reduce Pesticide Risks in OECD and Selected FAO Countries. Part I: Summary Report (1996)

No. 5, Activities to Reduce Pesticide Risks in OECD and Selected FAO Countries. Part II: Survey Responses (1996)

No. 6, OECD Governments' Approaches to the Protection of Proprietary Rights and Confidential Business Information in Pesticide Registration (1998)

No. 7, OECD Survey on the Collection and Use of Agricultural Pesticide Sales Data: Survey Results (1999)

No. 8, Report of the OECD/FAO Workshop on Integrated Pest Management and Pesticide Risk Reduction (1999)

No. 9, Report of the Survey of OECD Member Countries' Approaches to the Regulation of Biocides (1999)

No. 10, Guidance Notes for Analysis and Evaluation of Repeat-dose Toxicity studies

No. 11, OECD Series On Pesticides Number 11 Survey Of Best Practices In The Regulation Of Pesticides In Twelve Oecd Countries

No. 12, OECD Series On Pesticides Number 12 Guidance For Registration Requirements For Pheromones And Other Semiochemicals Used For Arthropod Pest Control

No. 13, Report of the OECD Workshop on Sharing the Work of Agricultural Pesticide Reviews.

Published Separately:

OECD Guidance for Country Data Review Reports on Plant Protection Products and their Active Substances - Monograph Guidance (1998, revised 1999)

OECD Guidance for Industry Data Submissions on Plant Protection Products and their Active Substances - Dossier Guidance (1998, revised 1999)

Report of the Pesticide Aquatic Risk Indicators Expert Group (2000)

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation composed of 30 industrialised countries in North America, Europe and the Pacific. The OECD works to co-ordinate and harmonise government policies, address issues of mutual concern, and respond to international problems.

The work of the OECD related to chemical safety is carried out in the **Environment, Health and Safety Programme**. As part of its work on chemical testing, the OECD has issued several Council Decisions and Recommendations (the former legally binding on Member countries), as well as numerous Guidance Documents and technical reports. The best known of these publications, the **OECD Test Guidelines**, is a collection of methods used to assess the hazards of chemicals and of chemical preparations. These methods cover tests for physical and chemical properties, effects on human health and wildlife, and accumulation and degradation in the environment. The OECD Test Guidelines are recognised worldwide as the standard reference tool for chemical testing.

The Pesticide Programme was created in 1992 within the OECD's Environment, Health and Safety Division to help OECD countries: 1) harmonise their pesticide review procedures, 2) share the work of evaluating pesticides, and 3) reduce risks associated with pesticide use.

The Pesticide Programme is directed by a body called the Working Group on Pesticides, composed primarily of delegates from OECD Member countries, but also including representatives from the European Commission and other international organisations (e.g. United Nations Food and Agriculture Organisation, United Nations Environment Programme, World Health Organisation, Council of Europe), and observers from the pesticide industry and public interest organisations (NGO's).

In addition to the **Series on Testing and Assessment** and **the Series on Pesticides**, the Environment, Health and Safety (EHS) Division publishes documents in six other series: **Good Laboratory Practice and Compliance Monitoring**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Chemical Accidents**; **Pollutant Release and Transfer Registers**; **and Emission Scenario Documents**. More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (see next page).

This publication was produced within the framework of the Inter-Organisation Programme for the Sound Management of Chemicals (IOMC). It was approved for derestriction by the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, the governing body of the Environment, Health and Safety Division.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 by UNEP, ILO, FAO, WHO, UNIDO and the OECD (the Participating Organisations), following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. UNITAR joined the IOMC in 1997 to become the seventh Participating Organisation. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

**OECD Environment Directorate,
Environment, Health and Safety Division**

**2, rue André-Pascal
75775 Paris Cedex 16
France**

Fax: (33-1) 45 24 16 75

E-mail: ehscont@oecd.org

PREFACE

The aim of this OECD project is to develop harmonised guidance on conducting independent evaluations of, and writing reviews of, chronic oral toxicity and carcinogenicity tests. The following guidance notes contain much material that is identical to or derived from *Guidance Notes for the Analysis and Evaluation of Repeat-dose Toxicity Studies*, because many concepts apply to both types of study. Furthermore, while the notes are chiefly to assist in the interpretation and transparent reporting of toxicological data on pesticides, they may also be of use in the evaluation of studies in other programmes, allowing studies of different groups of chemicals, *e.g.* pesticides, biocides and industrial chemicals, to be assessed in the same way.

The aim of the guidance notes is to outline core concepts in order to obviate the need to consult large numbers of text books, while still pointing the reader to sources of more detailed or specific information. They are intended to complement OECD Test Guidelines and other OECD publications, including *Guidance for Industry Data Submissions* and *Guidance for Country Data Review Reports on Plant Protection Products and their Active Substances – Monograph Guidance* (OECD, 1998a, b). However, whereas the latter provides guidance on the format and presentation of entire evaluation reports (or monographs), including acceptability criteria for industry data submissions (or dossiers), terminology and structure, these notes emphasise data interpretation, scientific judgement and report writing in the context of regulatory toxicology evaluations.

Although OECD Test Guidelines 451–453 (OECD, 1981a, b, c) already provide information on experimental design these guidance notes also address the subject. This is because toxicologists may have to assess studies that pre-date the development of OECD test guidelines, were designed according to other test guidelines, or do not conform to any test guidelines.

The document is based in part on the US EPA SEP document Toxicity Potential (Guidance for Analysis and Evaluation of Subchronic and Chronic Exposure Studies) (EPA-540/9-85-020).

With respect to harmonisation of test guidelines, the guidelines used by the US Office of Pesticide Programs (OPP) and the Office of Pollution Prevention and Toxics (OPPT) have been harmonised with those of the OECD (see *Federal Register*, 20 June 1996).

Acknowledgement

The First draft of this document was prepared by the Chemicals Unit, Department of Health and Ageing, Canberra, Australia, as part of an OECD project to develop harmonised guidance on the preparation of data reviews for toxicity studies. Subsequent drafts incorporated useful amendments and additions suggested by agencies and organisations in a large number of OECD countries. The authors wish to thank all who provided comment and advice, which has made a significant contribution to the development of the final manuscript.

TABLE OF CONTENTS

1. INTRODUCTION

Among the chemicals to which humans are exposed, pesticides are unique because of their deliberate introduction into the environment to kill or otherwise control life forms considered detrimental to human welfare. Experimental animals have served as useful models for detection of potential human responses to these hazardous or potentially hazardous substances. Regulations governing the conducting and reporting of toxicology studies in animals, and guidelines on acceptable and useful experimental designs (or protocols) for evaluating possible adverse health effects (or hazards) of pesticidal agents, have been published by a number of national agencies and international bodies.

Chronic exposure studies aim to determine toxic effects and potential health hazards following prolonged, repeated exposure (US EPA,1982). This type of study is usually required if humans are likely to be exposed to a substance over a significant portion of their life span, as is potentially the case with pesticide residues in the diet.

This document has two specific purposes. First, it is intended as a general guide to the analysis and evaluation of data from chronic exposures of toxicity test species to pesticides and other chemicals. Second, it outlines the kind of information an independent assessment of toxicity studies should encompass so as to maximise the scientific integrity and transparency of the assessment report. A report that includes this information will not only contain enough detail to establish regulatory end-points (such as the human Acceptable Daily Intake), but will also help clarify the mode of action of the test chemical, especially when considered alongside assessments of other studies in the data set.

More generally, these guidelines aim to foster a common approach among evaluators working in different national regulatory agencies, and so increase confidence and mutual understanding and contribute to the harmonisation activities undertaken by the OECD and other agencies, such as the WHO.[1]

This document should be consulted alongside other national guidance and requirements documents. It provides broad guidance on approaches to hazard assessment and on some of the problems and pitfalls that may arise during an assessment of possible compound-related changes in parameters measured in toxicity studies. The text reflects scientific understanding and standards as at the date of issue. In time, the scientific community will gain a better understanding of the mechanisms of toxicity, and this may lead to changes in both methodology and interpretation of results; analysis and evaluation of toxicity studies should reflect scientific consensus at the time data are reviewed.

This document provides guidance on *hazard* assessment. It is not a substitute for the many excellent texts on toxicology, clinical chemistry and pathology, nor does it consider the

[1] An example is the IPCS project *Harmonisation of Approaches to the Assessment of Risk from Exposure to Chemicals*, which has developed a *Conceptual Framework for Cancer Risk Assessment*. The framework (presented in Appendix VIII) is an analytical tool for judging whether the available data support a postulated mode of carcinogenic action.

multiplicity of effects likely to be encountered in chronic exposure studies. It provides only minimal comment on *risk* assessment, which may vary according to national preferences and priorities. Furthermore, while the guidance notes are for use in the assessment of pesticides, they may have some applicability in evaluating studies in other programmes, *e.g.* industrial chemicals.

1.1 Definitions

- **Toxicity** means the intrinsic capacity of a chemical substance or a mixture of substances to induce injury.

- **Hazard** means the observed toxic manifestation(s) induced by a known quantity of a substance under known exposure conditions.[2] The term is frequently used interchangeably with "intrinsic toxicity".

- **Risk** means the probability that an identified hazard or hazards will or will not be encountered under anticipated exposure conditions.[3] The basic approach to risk assessment can be expressed by the simple formula:

$$\text{Risk} = \text{Hazard} \times \text{Exposure}$$

Thus, if either hazard or exposure can be reduced or minimised, the risk, or likelihood, of harm can be reduced or minimised. The identification of hazard and the assessment of risk in respect of a given substance are informed judgements. Such judgments are usually based on data relating to toxicity, proposed uses and anticipated exposure conditions. Use of a pesticide product and expected exposure conditions define the type and probable extent (*i.e.* duration and degree) of exposure, as well as the size and composition of the exposed population. A particular product may pose one or more risks depending on use(s) and exposure conditions.

Paracelsus (1493-1541) stated that "All substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy." (Amdur *et al.*, 1991). This is a fundamental principle of toxicology and hazard assessment. The risk a pesticide poses to humans and the environment depends on exposure conditions and does **not** equate directly to the pesticide's intrinsic toxicity. This is illustrated by the following: imagine the perfect containment system, which prevents absolutely any exposure of humans and the environment to a toxic substance. Since exposure is zero, the risk to humans and the environment is also zero, although the toxicity of the substance remains unchanged.

- **Dose** refers to a stated quantity or concentration of a substance to which an organism is exposed. A dose is most commonly expressed as the amount of test substance per unit weight of test animal (*e.g.* mg/kg bw).

- **Dosage** is a general term that refers to dose, its frequency and the duration of dosing. It is properly applied to any rate or ratio involving a dose. Dosage often involves the

[2] A more general definition of "hazard" is any threat to people and what they value.

[3] Defined more broadly, risk is a measure of the likelihood of harm or loss as a result of hazards, *i.e.* the word "risk" implies both the existence of a threat and its potential for happening.

dimension of time (*e.g.* mg/kg bw/d), but the meaning is not restricted to this relationship (Hayes, 1991).

- **Dose–Response Relationship** means the correlative association between a dose and the response (or effect), or range of responses, it produces. The concept expressed by this term is indispensable to the identification, evaluation and interpretation of most pharmacological and toxicological responses to chemicals. The assumptions that underlie it are: (a) the response is a function of the concentration of the test chemical at a site; (b) the concentration is a function of the dose; and (c) response and dose are causally related (Klaassen & Eaton, 1991). The existence of a dose-response relationship for a particular biological or toxicological response provides a defensible conclusion that the response is a result of exposure to a known substance.

- **Chronic Toxicity Studies** (also known as **Long-term Toxicity Studies**) are repeat-dose studies performed to identify target organs for toxicity, to determine the interrelationship between ageing and toxicity, to investigate the effects of treatment using a range of doses, and to establish a No-Observed-Effect Level (NOEL) that may be used for predicting an Acceptable Daily Intake (ADI) for humans [or, in the USA, a Reference Dose (RfD)]. They are usually conducted in mice, rats and dogs, although hamsters and non-human primates are also used. Other than in dermal toxicity testing and developmental studies, rabbits are not generally accepted as non-rodents for toxicology testing unless satisfactory evidence of their particular suitability is provided. Although some authors (*e.g.* Stevens & Mylecraine, 1994) define "long-term" as more than 3 months in duration, or longer than 10% of the life span of rodents, a minimum duration of 12 months is specified by OECD Test Guideline 452 (OECD, 1981b) and US EPA Health Effects Test Guidelines 870.4100 (US EPA, 1998a). For hazard assessment of pesticides, most chronic toxicity studies last 52 to 104 weeks.

- **Carcinogenicity Studies** (also known as **Oncogenicity Studies**) are performed to determine the carcinogenic potential and dose–response relationships of the test chemical. They yield data on the production of tumours as well as pre-neoplastic lesions and other indications of chronic toxicity that may provide evidence of treatment-related effects and insights into the mechanism of carcinogenesis (US EPA, 1996). Given that development of tumours is age-related and that large groups are required to detect increases in treated animals, carcinogenicity studies are normally conducted in small rodents (usually mice and rats) over most of their life span. Syrian golden hamsters have been used in studies of carcinogenesis in the respiratory and urinary tract (OECD 1981a), but are seldom employed in the hazard assessment of pesticides. Mice are generally exposed to the test chemical for 18–24 months, rats for 24–30 months, exposure being longer for strains of greater longevity or with a lower spontaneous tumour rate. However, under OECD Test Guideline 451 (OECD 1981a) a study may be terminated if the survival rate in the lower-dose or control groups declines to 25%, while the US EPA Health Effects Test Guildelines 870.4200 (US EPA, 1998b) specify that survival in any group should not fall below 50% at 15 months in the case of mice and 18 months in the case of rats, or below 25% at 18 and 24 months respectively. In addition, the WHO (1990) recognises a further type of carcinogenicity study that continues until mortality in the most susceptible group reaches a fixed level, usually 80%.

- **Combined Chronic Toxicity/Carcinogenicity Studies** encompass both neoplastic effects and general toxicity, including neurological, physiological, biochemical, haematological and pathological effects. Typically, rats are used for combined chronic

toxicity/carcinogenicity assessment except in respect of the dermal route, for which mice are preferred (US EPA, 1998c). The study design incorporates satellite groups of treated and control animals scheduled for interim sacrifice after 12 months on study for investigation of pathological abnormalities that are uncomplicated by age-related changes. OECD Test Guideline 453 (OECD, 1981c) and US EPA Health Effects Test Guidelines 870.4300 (US EPA 1998c) specify the same duration of exposure as in carcinogenicity studies.

- **Short- and medium-term bioassays for carcinogenesis** screen for carcinogenic or pre-neoplastic effects in animal models without exposing them to the test chemical for most of their life span. Advances in understanding chemical carcinogenesis have led to the development of short- and medium-term assays that use neoplasia or lesions that are precursors to neoplasia as end-points. These tests do not require chronic exposure to the test chemical, yet can provide information on its mechanism or mode of action and thus help in the assessment of the risk to humans of developing cancer. IARC (1999) and Dearfield & Benz (1999) have reviewed these tests. Both reviews also address the role of genetic toxicology data in the prediction of carcinogenic hazard, and warn that the many available tests do not have equal significance for this purpose.

Some of the medium-term tests for carcinogenicity involve the development of proliferative lesions in a single tissue, *e.g.* foci of alteration in the liver (Williams *et al.*, 1982; Goldsworthy *et al.*, 1986; Ito *et al.*, 1989). Others use tumour end-points, such as induction of lung adenomas in the A-strain mouse (Maronpot *et al.*, 1986) or induction of tumours in initiation–promotion studies using various organs, including the skin, bladder, intestine, liver, lung, mammary gland and thyroid (see reviews by Enzmann *et al.*, 1998a & 1998b; IARC, 1992 & 1999). A further category of study is the "start/stop" protocol. Here, an agent is administered for a limited period to induce particular effects or lesions; the progression or reversibility of these is then observed in the absence of further treatment (Todd, 1986; Marsman & Popp, 1984).

Assays in genetically engineered rodents have also been developed following the identification of genes, such as proto-oncogenes and tumour-suppressor genes, that are highly conserved across species and associated with a wide variety of human and animal cancers. They involve activated oncogenes that are introduced (transgenic) or tumour suppressor genes that are deleted (knocked out). If appropriate genes are selected, these assay systems may provide information on mechanisms of tumour formation or serve as selective tests for carcinogens. The modified transgene is expected to accelerate carcinogen-induced cancer development without interfering with other relevant genetic and/or epigenetic steps. High spontaneous tumour incidence in control animals is a major confounding factor of the conventional bioassay; the presence of the transgene itself does not induce high spontaneous tumour incidence in the short time span of the assay. These assays have been extensively reviewed in publications, including a single-theme issue of Toxicological Pathology (26 (4), 1998) and others (Tennant *et al.*, 1995; Contrera & DeGeorge, 1998; Eastin, 1998; Bucher, 1998; Eastin &Tennant, 1998).

If regulators are to gain confidence in these newly developed tests, the tests must be validated and detailed guidance provided for performance and interpretation. Acceptance of any test is based on the attainment of scientific consensus regarding the relationship of the new end-points to animal and human health and to outcomes in established tests, as well as the operating characteristics of the test in terms of reproducibility, accuracy, sensitivity and robustness. Regulators dealing with pharmaceuticals have concluded that there is sufficient experience with some *in vivo* short- and medium-term rodent carcinogenicity models to support their application as complementary second-species studies in conjunction with a single two-year rodent carcinogenicity study. The guidelines of the International

Conference on Harmonisation (ICH) Expert Working Group (1997) nominate models of initiation–promotion in rodents or models of carcinogenesis using neonatal or transgenic rodents as potentially suitable for this purpose. The choice of species for the conventional carcinogenicity study is based on a number of considerations, including pharmacology, repeat-dose toxicology, metabolism, toxicokinetics and route of administration. The rat is recommended in the absence of clear evidence favouring one particular species.

Regulators of pesticides may need to consider the following points if medium-term tests are submitted in support of, or as a substitute for, a long-term bioassay. The selection of an alternative carcinogenicity study is expected to be scientifically justified; and should be based on how the study can contribute additional mechanistic information that may be useful for interspecies extrapolation and in the weight-of-evidence assessment of carcinogenic potential. It is essential to address issues related to the route of administration; operational characteristics of the model; level of characterisation of, and degree of experience with, the model; and any relevant toxicologic issues associated with the particular pesticide, such as genotoxicity. The test method must be validated; general criteria for validation of toxicology assays include assessment of repeatability within a laboratory and reproducibility of the method at multiple laboratory sites. There may be less scope for the use of short- and medium-term tests in the regulation of pesticides or industrial chemicals than in the regulation of pharmaceuticals, because the available supporting information, such as the results of pharmacokinetic and metabolism studies, is likely to be more limited (Blain *et al.*, 1998).

1.2 Concepts

1.2.1 Dosing Regimen

The purpose of a long-term exposure study is the detection of biological evidence of any toxic or oncogenic potential of the substance being investigated. Protocols should therefore maximise the sensitivity of the test without significantly altering the accuracy and interpretability of the biological data obtained. The dose regimen has an extremely important bearing on these two critical elements.

Since one of the objectives is determination of the dose–response relationship in respect of any observed effects, OECD Test Guidelines 451–453 require at least three dose levels, as well as controls. Dose selection should be based on the findings of subchronic or other range-finding studies.

The highest dose to be used in a chronic toxicity or carcinogenicity study needs to be carefully considered and **the reasons for the final choice clearly defined**. Ideally, the doses selected will maximise the detection of dose–response relationships and facilitate the extrapolation of these to potential hazards for other species, including humans. The largest administered dose should **not** compromise the biological interpretability of the observed responses. For example, it is generally considered that it should not:

(a) in a chronic study, exceed the **maximum tolerated dose** (or **MTD**) defined as the highest dose to produce toxic effects without causing death **and** to decrease body weight by no more than 10% relative to controls (Derelanko, 2000);

(b) in a dietary study, exceed 5% of the total diet because of the possibility that higher levels will cause nutritional imbalances, or;

(c) produce severe toxic, pharmacological, behavioural or physiological effects that might shorten the duration of the study or otherwise compromise the study results;

(d) in a carcinogenicity study, significantly affect the survival rate except through tumour production, or cause a body weight decrement greater than 10–12% of concurrent control values, because larger decreases can mask, reduce, delay or prevent the development of tumours (DeGeorge, 1999).

Data on pharmacokinetics or metabolism may be helpful in determining dose levels, particularly if there is evidence of bioaccumulation of the test compound or metabolites, or evidence of dose-dependent changes in absorption or detoxification.

In inhalation studies involving respirable particles, evidence of impaired ability to clear particles from the lung should be considered along with other signs of toxicity to the respiratory airways to determine whether the appropriate highest exposure concentration has been selected. In dermal studies, evidence of skin irritation may indicate the highest dose is adequate (US EPA, 1996).

The International Life Sciences Institute (ILSI) Risk Sciences Working Group on Dose Selection has published the principles to be considered in the selection of doses in chronic rodent bioassays (Foran, J.A., and the ILSI Risk Sciences Working Group on Dose Selection, 1997). The article puts forward five principles intended to discourage reliance on an MTD as the sole criterion. These are summarised below:

Principle 1: Sound toxicological principles must underlie the choice of doses. Increasing a dose can increase the ability to detect an effect; therefore, doses for chronic rodent bioassays should be selected from a reasonable range so as to maximise the sensitivity of the study. However, it is possible that mechanisms of toxicity or chemical modes of action active at higher doses are not relevant to humans exposed to lower doses, leading to results inappropriate for human risk assessment. Dose selection should therefore be made in the light of all relevant information from "pre-chronic" studies and other sources, be based on an understanding of the test chemical's mechanisms or modes of action, and draw on good scientific principles to enhance the soundness of judgements concerning risk to humans.

Principle 2: One aim in the selection of the highest dose is to minimise the likelihood of a false negative result. However, the qualitative nature of the hazard may itself be dose-dependent. Scientists should encourage innovative approaches to dose selection by considering appropriate study designs, mechanistic data and other information in the design and interpretation of studies. Use of additional end-points and other information must be based on sound scientific rationale, and such designs should be evaluated on their individual merits.

Principle 3: Selection of the middle and lower doses should take account of human exposure, the test chemical's mechanism or mode of action, toxicokinetics and the other factors listed under Principles 4 and 5. It should not be made simply by specifying a fraction of the highest dose. Where human exposure influences dose selection, account should be taken of the human exposure route, mode, frequency, duration and dose relative to the chronic study. Subpopulations that may be more highly exposed than the general population, or that are genetically more susceptible, should also be considered. If the test substance is expected to exhibit a toxicity threshold, or if carcinogenicity and chronic toxicity are being evaluated in combination, the study should include one dose that does not elicit adverse effects, *i.e.* a NOAEL) (No-Observed-Adverse-Effect Level). However, caution must be exercised to ensure the NOAEL is not an artefact of a small sample size or poor study design.

Principle 4: Additional information from "pre-chronic" studies should be considered, including histopathology, toxicokinetics, cellular growth, death and regeneration, disturbances of physiology or homeostasis, clinical chemistry, haematology, urinalysis, body weight and organ weights.

Principle 5: Choice of the highest dose will also be influenced by the test compound's physico-chemical characteristics (*e.g.* its solubility and vapour pressure), bio-availability, palatability and potential to cause adverse effects (*e.g.* irritation, erosion, ulceration) at the site of administration. Doses should be selected so as to minimise or avoid adverse nutritional, physical, organoleptic and irritant effects.

Although it may be argued that responses observed at doses far in excess of levels experienced under real or potential exposure conditions fall within the bounds of the classic dose–response concept, there are valid scientific concerns that such doses introduce biases of considerable significance into the already difficult task of evaluating dose-response relationships in animals and the assessment of their relevance in the identification of human hazard and risk (Paynter, 1984). Doses that overwhelm normal mechanisms of metabolism, detoxification or excretion, or produce severe tissue damage (*e.g.* necrosis or demyelination), can make interpretation difficult or lead to erroneous conclusions.

It is commonly accepted that the lowest dose should not produce any evidence of toxicity, *i.e.* it should allow the establishment of a NOEL or NOAEL.

1.2.2 Dosing Route

Chronic studies most commonly use the oral route, administration being by dietary admixture, by gavage or in capsules (for non-rodents). However, depending on the possible route of exposure of occupationally exposed workers or the public, the test substance may need to be investigated in toxicity studies using the dermal or inhalational route.

When the objective of a study is to establish an ADI for humans, diet is the preferred route of administration, provided the test chemical is sufficiently stable and not unpalatable at the concentrations used. If the concentration of the test chemical is held constant throughout the study, the achieved dose will decline by approximately 2.5-fold over the dosing period; this may diminish the severity of toxic effects as the study progresses (WHO, 1978).

In both humans and laboratory animals, diet has a direct bearing on health, and many neoplastic and non-neoplastic diseases are caused (or prevented) by dietary factors, including variations in the composition and amount of feed consumed. The association in rats of calorific consumption, the spontaneous formation of tumours and life span is well established. Although the zero-dose group may be expected to control for the influence of diet, dietary constituents may still profoundly affect the outcome of an experiment. For example, the insecticide propoxur was tumorigenic to the rat urinary bladder when administered in Altromin 1321 diet, but not when administered in a casein-based semi-synthetic diet. This discrepancy is believed to arise from urinary pH, which is slightly alkaline in rats consuming Altromin 1321 but acidic in rats fed a casein-based diet. Cohen *et al.* (1994) have demonstrated that a reduction of 1 pH unit or more is enough to inhibit urothelial proliferation and tumour formation in rats treated with propoxur.

Although administration by gavage or in capsules may overcome problems arising from instability or unpalatability, and may permit delivery of a more precise and consistent dose, bolus administration also has the potential to introduce artefacts, such as irritation in the oesophagus or stomach. Furthermore, the test chemical's pharmaco- and toxicokinetic behaviour following bolus

administration and incremental ingestion of the same daily dose in feed or water may differ, leading to dissimilar toxicological effects. Chloroform induces hepatocellular cytotoxicity, regenerative proliferation and liver cancer in mice when administered by gavage in corn oil at doses that do not cause development of these lesions when administered in drinking water (see review by Butterworth *et al.*, 1998).

For dermal exposure the material, in a suitable vehicle, is applied evenly to a clipped or shaved area of skin of approximately 10% of the total body surface area. The site may be occluded with polyethylene sheeting and gauze patches, or semi-occluded, in order to prevent dislodgement of material and oral ingestion, which could affect the validity or usefulness of the study. With volatile or semi-volatile materials, application and covering procedures should minimise the possibility of evaporation. There are useful chapters or sections on dermal toxicity testing in the standard toxicology textbooks, *e.g.* Derelanko & Hollinger (1995) and Hayes (1994).

The surface area of the respiratory membrane is large, estimated at approximately 50-100 square metres in the normal adult compared with the estimated area of the small intestine at 250 square metres (Guyton, 1991) and much more air (about 5000 times, by volume) is inhaled each day than food or water is ingested (McClellan & Henderson, 1989). Exposure to airborne material through inhalation is therefore potentially greater than exposure to material that enters the body via the dermal or oral route. Airborne material can be gases or vapours, liquid droplets or solutions, aerosols (with both solid and vapour components), or dry fibres or powders. Consequently, in inhalational toxicity studies, the mechanisms needed to deliver chemicals to a test chamber in a form that can be inhaled are quite complex, particularly given the need to include measuring devices that can establish particle size and concentration and the form of the material in the exposure chamber. Furthermore, many factors can influence the inhalation of materials and their deposition and retention in the respiratory tract. Therefore, conducting inhalational studies is a considerably more complex matter than undertaking equivalent studies using the dietary or dermal routes. Of critical importance, in both the conduct and assessment of such studies is the need to establish what portion of the material delivered to the exposure chamber is in a respirable form. OECD Test Guidelines 451–453 and their US EPA equivalents provide some guidance on the conduct of chronic inhalation studies, including discussion on the relative merits of continuous and intermittent exposure. In addition to standard toxicology texts, useful references on inhalation toxicology include McClellan & Henderson (1989), Mohr *et al.* (1988) and Salem (1987).

1.2.3 Treatment-related Responses

Responses produced by chemicals in humans and experimental animals may differ according to the quantity of the substance received and the duration and frequency of exposure, *e.g.* responses to acute exposures (a single exposure or multiple exposures occurring within twenty-four hours or less) may be different from those produced by subchronic and chronic exposures. Not all observed responses within a study, irrespective of exposure duration or frequency, will represent toxicity *per se*. They may encompass a range of effects from physiological through to toxic manifestations. Although it sometimes may be difficult to make a clear distinction between these responses, an attempt to do so should be made. If an evaluator is uncertain of the type or the biological significance of a response, he/she should not hesitate to obtain competent advice for resolving the uncertainty. It is essential that all relevant toxicity end-points (statistically and/or biologically significant) be identified for consideration when evaluating data for the presence or absence of non-toxic levels.

Classification of treatment-related responses

Contrasting approaches are available to the classification of the types of responses that a living organism can manifest during or after exposure to a xenobiotic. **The way in which treatment-related responses are described or classified may differ between agencies, depending on national policy considerations.**

Adverse *vs.* Non-adverse Responses.[4] Some agencies prefer to classify responses as either **adverse** or **non-adverse**, which has the perceived advantage of simplicity and direct relevance to hazard and risk assessment of pesticides. In this paradigm, an adverse response is defined as "any treatment-related response that results in change in the morphology, physiology, growth, development or life span of an organism, which results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other environmental influences" (OECD/IPCS, 1998). The decision on how to classify effects as adverse or non-adverse is made on a case-by-case basis by reference to the overall weight of evidence. The definition of adverse would cover any toxic response, but would also encompass an event such as increased activity of the hepatic cytochrome-P450-containing mono-oxygenase system (enzyme induction), *if* it altered hormonal homeostasis and caused tumour production, for example, or increased the organism's susceptibility to injury by other chemicals.

However, an alternative conceptual scheme may be preferred by agencies which assess industrial or consumer chemicals, biocides and/or veterinary pharmaceuticals in addition to pesticides, and which use common criteria to regulate dietary exposure to residues of pesticides and veterinary medicines. This second scheme is based on a more general biological science viewpoint, which recognises **physiological responses**, **toxic responses**, and under some circumstances, **pharmacological responses**. Thus, this concept is especially helpful for the toxicological assessment of chemicals intended for pharmaceutical use.

- **Physiological** responses are non-adverse. They vary within limits which are in accord with the normal functioning of a living organism; examples of such response are the usual respiratory and pulse rate increases associated with increased physical activity, systemic changes associated with normal pregnancy, and those associated with homeostatic mechanisms. These variable factors are not important toxicity end-points in subchronic and chronic exposure studies unless their fluctuations are abnormally altered by a dose regimen. If such alterations occur at a particular dose or are part of a dose–response relationship, they should be correlated with other toxicity end-points that may be present. Altered physiological functions arising from interaction of a xenobiotic with a cellular receptor site, are often referred to as **pharmacological** responses if they are reversible and of limited duration. Whilst some of these responses may be undesirable under certain circumstances, they are distinguished from toxic (adverse) responses by generally not causing injury. Provided there were no adverse consequences, enzyme induction could be considered to be an example of a pharmacological response. But if enzyme induction caused tumorigenesis, it would have to be regarded as an adverse response and be accounted for in the hazard and risk assessment process.

- **Toxic** responses are, by definition, adverse. They may be reversible or irreversible but are distinguished from other types of responses by being injurious and therefore adverse

[4] An ECETOC technical report entitled *Recognition of, and Differentiation between, Adverse and Non-adverse Effects in Toxicology Studies* is due for publication in 2002.

and harmful to living organisms or tissues. A chemical that causes a physiological or pharmacological effect may produce a toxic response if the exposure is prolonged and/or if the dose is increased beyond a certain level. The reversibility or otherwise of such responses may also depend on these two factors. The reversibility or irreversibility of a histopathological change will depend on the ability of the injured organ or tissue to regenerate. For example, liver has a relatively great ability to regenerate and many types of injury to this organ are reversible. By contrast, differentiated cells of the central nervous system are not replaced and many injuries to the CNS are irreversible.

1.2.4 Carcinogens and Carcinogenesis

Adult tissues, even those composed of rapidly replicating cells, maintain a constant size and number of cells. They do this through regulation of the rate of cell replication, the differentiation of cells to assume specialised functions, and programmed cell death (or apoptosis) (US EPA, 1996). **Cancers** are diseases in which somatic mutation of genes critical to the maintenance of control over cell division leads to loss of control over cell replication, differentiation and death. Such uncontrolled cell replication can cause the growth of **tumours** (or **neoplasms**) — masses of abnormal, disorganised cells that arise from pre-existing tissue and are characterised by excessive and uncoordinated proliferation and abnormal differentiation.

Tumours are classified as either **benign** or **malignant**. Malignant tumours invade or infiltrate surrounding tissues, often damaging or destroying them. They may also spread by dissemination via the circulatory and vascular systems to distant sites, a process known as **metastasis.** Growth may be rapid. The morphology of malignant tumours is variable. Some well-differentiated examples bear a resemblance to their parent tissues, but recognisable features are progressively lost in moderately and poorly differentiated malignancies. Undifferentiated, or anaplastic, tumours are composed of pleomorphic cells that do not resemble normal tissue. Benign tumours, by comparison, show a close morphological resemblance to their tissue of origin, grow by slow expansion and form circumscribed and (usually) encapsulated masses. They may stop growing or even regress, and do not metastasise or invade surrounding structures, although they may compress them. However, benign tumours may become malignant (UK DoH, 1991).

Definitions of **carcinogenesis** vary. The *Merriam–Webster Medical Dictionary* (1997) and the *On-line Medical Dictionary* (1998) define it as the production or generation of malignant tumours, while *Lewis' Dictionary of Toxicology* (1998) states that it is "any process that produces malignant neoplasms". By contrast, other sources [e.g. UK DoH (1991) and Derelanko (1995)] regard carcinogenesis as the production of malignant **or** benign tumours. The US National Cancer Institute[5] defines it as "the process by which normal cells are transformed into cancer cells", without drawing any distinction between benign or malignant tumours. **These guidance notes assume carcinogenesis may give rise to both benign and malignant tumours**, given that benign tumours may be an intermediate stage in the development of malignancies, and that both are frequently observed together in experimental animals. Furthermore, no distinction is drawn between benign and malignant tumours in the IARC definition of **chemical carcinogenesis**, which reads: "[a] the induction by chemicals of neoplasms that are not usually observed, [b] the earlier induction by chemicals of neoplasms that are commonly observed, and/or [c] the induction by chemicals of more neoplasms than are usually found".

[5] CancerNet at www.nlm.nih.gov

Carcinogenesis is also referred to as **tumorigenesis** or **oncogenesis.** Some authors use the term "tumorigenesis" when referring specifically to the induction of benign tumours, but, for the reasons stated above, this document makes no such distinction. A chemical or other agent that causes cancer is known as a **carcinogen**, **oncogen** or **tumorigen.** It should be noted, however, that some cancers do not form solid tumours.

Multi-stage model of carcinogenesis. The following definition is from Derelanko (1995). In the current multi-stage model of carcinogenesis, development of a malignant tumour occurs in three stages: initiation, promotion and progression. Initiation involves an irreversible change in a normal cell (usually an alteration of the genome), allowing unrestricted growth. The initiated cell may remain latent for months or years. During this period it is phenotypically indistinguishable from surrounding cells of the same type. Development into a neoplastic cell requires a period of promotion. Under the influence of a promoter, tumour formation is accelerated through clonal expansion of the initiated cell. Promoters, which do not interact directly with DNA, are a diverse group of agents believed to act through a variety of mechanisms that most often result in increased cell proliferation. Promotion is considered reversible and requires prolonged and repeated exposure to promoter agents. Progression is the final step in which pre-neoplastic foci develop into malignant cells. In this stage, tumour development is characterised by karyotypic changes, increased growth rate and invasiveness. Progression may be spontaneous, influenced by environmental factors, or mediated by progressors. Resulting tumours may be either benign or malignant.

Spontaneous carcinogenesis. In both humans and experimental rodents, carcinogenesis may occur in the absence of exposure to exogenous carcinogens. **Background** tumour formation is a normal observation in control animals in rodent carcinogenicity studies. The incidence of spontaneous tumours varies between tissues, and the susceptibility of a given tissue or organ varies between species and strain, and can be influenced by other factors, including diet. For example, Thurman, Moeller & Turturro (1995) found in lifetime studies that the incidence of testicular interstitial cell adenoma is 49% in F344 rats compared with 9% among FBNF1 rats. A 40% food restriction lowered incidences in these strains to 19% and 4% respectively. In the context of an oncogenicity study using laboratory animals, a chemical is considered carcinogenic if it causes an incidence of tumours higher than the background level (usually ascertained by comparison with a concurrent control group), or causes tumours that seldom develop spontaneously in the study species or strain. Further possible indications of carcinogenicity include an increase in the ratio of malignant to benign tumours and a reduction in the age at which tumours form.

Appendix V contains a list of references that provide data on the background incidence of neoplasia in laboratory species.

1.2.5 Mechanisms of Carcinogenesis [adapted from Derelanko (1995)]

- Genotoxic (through interaction with genetic material)

1. *Direct-acting* or "*primary*" carcinogens interact directly with DNA or chromosomal material, causing alterations in DNA structure or in chromosomal structure or number. Generation of reactive electrophiles and covalent binding (adduction) with DNA may be involved. This may lead to depurination, depyrimidation or breakage of DNA strands; such lesions can develop into mutations with a round of DNA synthesis and cell division. Other agents may have the same effect through intercalation onto the DNA helix, or bring about changes in gene expression by methylating DNA. Examples of direct-acting carcinogens are bischloromethyl ether, beta-propiolactone and ethylene imine.

2. *Procarcinogens*, or *"secondary"* carcinogens, require biotransformation to form a direct-acting carcinogen. Examples are nitrosamines, ethylene dibromide and vinyl chloride.

3. *Inorganic* carcinogens, such as nickel and cadmium, exert direct effects on DNA through interference with DNA replication.

- Epigenetic (no evidence of interaction with genetic material)

4. *Cytotoxic* carcinogens mediate their effect through cytolethality and consequent regenerative cell proliferation. Mutations may occur secondarily via several mechanisms, including release of nucleases, generation of reactive oxygen radicals, DNA replication before adduct repair, preferential growth of pre-neoplastic cells through selective lethality towards normal cells, and the expression of oncogenes. Examples are nitrilotriacetic acid and chloroform.

5. *Mitogens* stimulate cell proliferation directly or via a cellular receptor. Mutations may occur secondarily as a result of increased cell proliferation. Changes in the rate of cell birth or death may cause preferential growth of pre-neoplastic cells. Phenobarbital and alpha-hexachloro-cyclohexane are examples.

6 *Peroxisome proliferators* perturb lipid metabolism and increase generation of oxygen radicals within the cell (especially in rodent liver). Diethylhexyl phthalate and clofibrate are examples. Further information about this carcinogenic mechanism is presented in Table 6.

7. *Immunosuppressors* enhance the development of virally enhanced, transplanted and metastatic neoplasms, possibly through impairment or loss of natural and acquired resistance to tumours. Azathioprene, cyclosporin A and 6-mercaptopurine are examples.

8. *Hormones* and *hormonal-altering agents* cause chronic stimulation of cell growth through activation of regulatory genes, promotional effects resulting from alteration of hormonal homeostasis, inhibition of apoptotic cell death, and generation of reactive radicals. Estrogens, diethylstilbestrol and synthetic androgens are examples.

9. *Solid-state* carcinogens, such as some polymers, metal foils and asbestos, have an uncertain mode of action. The physical size and shape of the agent is critical. Generally, only mesenchymal cells or tissues are affected.

10. *Co-carcinogens* enhance the activity of genotoxic carcinogens when both are administered simultaneously. Possible mechanisms include enhanced biotransformation of a pro-carcinogen, inhibition of detoxification of a primary carcinogen, and enhanced absorption or decreased elimination of a genotoxic carcinogen. Phorbol esters, catechol and ethanol are co-carcinogens.

11. *Promoters*, when administered after a genotoxin, promote the formation of tumours through enhancement of the clonal expansion of preneoplastic cells. Numerous mechanisms have been proposed. Phorbol esters, saccharin and croton oil are promoters.

12. *Progressors* influence the development of initiated and promoted cells. Their activity may be associated with alterations in biochemical and morphological characteristics, increased growth rate, invasiveness and metastases, or direct or indirect induction of karyotypic changes. Examples are arsenic salts, benzene and hydroxyurea.

1.2.6 Effect and No-effect Levels

- **No-Observed-Effect Level (NOEL).** Also known as the **No-Observable-Effect Level**, is the highest dose of a substance administered to a group of experimental animals at which there is an absence of observable effects on morphology, functional capacity, growth, development or life span, which are observed or measured at higher dose levels used in the study. Dosing at the NOEL should therefore produce no biologically significant differences between the group of chemically exposed animals and an unexposed control group maintained under identical conditions. The NOEL is expressed in milligrams of chemical per kilogram of body weight per day (mg/kg bw/d) or, in a feeding study, in ppm in food (converted to mg/kg bw of compound intake by measured or estimated food intake over the period of the study).

The NOEL has been simply defined as the highest dose of a substance which causes no changes distinguishable from those observed in normal (*i.e.* control) animals (WHO, 1990).

- **No-Observed-Adverse-Effect Level (NOAEL).** The No-Observed-Adverse-Effect Level is the highest dose of a substance at which no toxic (*i.e.* adverse) effects are observed (WHO, 1990). It may also be worded in more detail thus: The NOAEL is defined as the highest exposure at which there is no statistically- or biologically-significant increase in the frequency of an adverse effect when compared to a control group (National Academy of Sciences/National Research Council, 1994). The definition of "NOEL" is equivalent, but with "adverse" removed. In considering which of the two terms to use, the issue is often to decide whether a compound-related effect is necessarily adverse. For example, some toxicologists consider enlargement of the liver associated with cytochrome P450 induction an adaptive pharmacological response. Others consider it an adverse effect because of its potential to enhance the toxicity of other xenobiotics, to disrupt hormonal homeostasis or to cause hyperplastic or neoplastic responses through enhanced cellular turnover. National policy considerations may also influence judgement. For example, some agencies consider plasma cholinesterase inhibition an adverse effect, whereas others do not. Therefore, while these guidance notes use both "NOEL" and "NOAEL", the two terms may not be equivalent.[6]

- **LOEL.** The Lowest-Observed-Effect Level is the lowest dose of a substance to cause changes distinguishable from those observed in normal (*i.e.* control) animals (WHO, 1990).

- Generally, a defensible presumption that the observed effect is induced by the substance is based primarily on the detection of a trend away from the normal for the species or strain of animal used (using concurrent and/or historical control data) and a demonstration of a dose–response relationship for the effect.

[6] Consider the following hypothetical case. In a 90-day study a test chemical is administered in doses of three different sizes. The lowest dose has no observed effects, the mid dose and above causes enzyme induction, enlargement of the liver and plasma cholinesterase inhibition, and the top dose causes erythrocyte cholinesterase inhibition. The bottom dose is designated the NOEL. If the effects of the mid dose are regarded as non-adverse, this dose is termed the NOAEL or LOEL. If the effects of the mid dose are regarded as adverse, this dose is termed the LOAEL.

- **LOAEL.** The Lowest-Observed-Adverse-Effect Level is the lowest dose of a substance to cause adverse changes distinguishable from those observed in normal (*i.e.* control) animals. The term "adverse" may often be a point of contention.

It should always be borne in mind that in any study the NOEL/NOAEL and LOEL/LOAEL will be determined by the doses selected, and should be chosen on the basis of scientific judgement and never with reference to statistical considerations alone. A frequent problem is that a clear dose–response trend is evident but the effect at the lowest dose is too subtle to be statistically significant.

- **Threshold Dose.** The acceptability and usefulness of the concept of the experimental NOEL or NOAEL depend on the scientific rationale supporting the existence and demonstrability of a **threshold** for responses produced by biologically active agents. As used here, the term "threshold" means the level of a stimulus that comes just within the limits of perception, and below which **no** recognisable response is elicited. In experimental toxicology studies, the threshold dose is a function of the toxin's mechanism of action, the sensitivity of the methods used to detect signs of toxic injury, the susceptibility of the tissue or species being observed, and factors such as the choice of doses, the method of administration, the duration of the study, the sample size and the choice of statistical analysis.

Some non-genotoxic carcinogens induce tumours as a secondary event following a toxicological effect that has a threshold. These substances do not present a carcinogenic hazard at doses that do not produce the primary toxicological event (UK DoH, 1991). As is usually done with toxins that have non-carcinogenic effects, it is feasible to set an ADI for humans for a non-genotoxic carcinogen by applying a safety factor to the NOEL or NOAEL (see below), provided the carcinogen's mode of action is sufficiently well characterised.

It is generally assumed that genotoxic carcinogens have the potential to damage DNA at any level of exposure, and that such damage may lead to tumour development. It is therefore prudent to treat genotoxic carcinogens as having no discernible threshold dose (UK DoH, 1991).

1.2.7 Acceptable Daily Intake For Humans (ADI)

It is accepted that the absolute safety of chemicals for humans cannot be established because, while it is possible to prove a chemical can produce a toxicological effect, it is not possible to determine the absolute absence of a toxicological effect.

The Acceptable Daily Intake of a chemical is defined as the daily intake that during an entire lifetime, on the basis of the information available at the time, appears to be without appreciable risk. It is expressed in milligrams of the chemical per kilogram of body weight (mg/kg bw). "Without appreciable risk" means that adverse effects will not result even after a lifetime of exposure. Furthermore, for a pesticide residue the ADI is a guide to the maximum amount that can be taken daily in food without appreciable risk to the consumer. Accordingly, the figure is derived as far as possible from feeding studies in animals.

The determination of an ADI entails the establishment of an overall NOEL or NOAEL, which is generally the lowest NOEL or NOAEL in the most sensitive species. (As discussed above, the definition of "adverse" may be agency specific, and the NOEL and NOAEL are not necessarily the same. Note also that in occupational risk assessment consideration of the relevant route of exposure may influence the NOEL or NOAEL.) This approach is justified unless there is evidence (1) from

toxicokinetic/metabolic studies that the toxicokinetic behaviour of the most sensitive species differs from that of humans and the species is therefore less relevant as a predictor of human toxicity than another toxicity test species; or (2) that the toxic effect which has the lowest NOEL or NOAEL is not relevant to humans; or (3) that the lowest NOEL or NOAEL is derived from an inadequate or invalid study. The full database must therefore be used, and all relevant findings correlated, in determining the most appropriate health end-point.

An ADI is then derived from the NOEL or NOAEL; the qualitative approach taken follows the principles outlined in IPCS Environmental Health Criteria Monographs Nos. 104 and 210 (WHO, 1990 & 1999). The uncertainty inherent in extrapolation between and within species has generally been dealt with by using a safety (*i.e.* uncertainty) factor. This is usually 100 but may range from 10 to 5 000, depending on the source and quality of data, the biological relevance of the endpoint, and the hazard assessment (carried out on a case-by-case basis).

Safety factors are not necessarily rigidly applied. When based on studies in animals, the safety factor is usually 100, derived by multiplying a factor of 10 for species extrapolation with a factor of 10 for individual variation in human populations. In general terms only, a safety factor of 10 applies when appropriate human data are available. Further safety factors may have to be incorporated to provide additional protection for special risk groups (*e.g.* infants), or where the toxicological database is of poor quality. Further safety factors may also be used when the toxicology database is incomplete (*e.g.* in the field trial of a new chemical where it is proposed that produce from treated plants or animals be consumed), or the nature of the potential hazards indicates the need for additional caution. These supplementary safety factors may range up to 10, 20 or even 50, giving an overall safety factor of 1 000–5 000.[7] The ADI is calculated by dividing the NOEL or NOAEL by the safety factor. This approach assumes that exposure at less than the ADI is without appreciable risk, but there is no attempt to quantify the level of risk.

1.2.8 Reference Dose (RfD)

Although regarded by many as being synonymous with "ADI", the RfD is in fact distinctly defined. It was developed by a US EPA work group for assessment of risks associated with systemic toxicity, but not carcinogenicity. The RfD is, in general, an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects during a lifetime. Usually, doses less than the RfD are not likely to be associated with adverse health risks, and are therefore less likely to be of regulatory concern. As the frequency and/or magnitude of the exposures exceeding the RfD increase, the probability of adverse effects in a human population increases. However, all doses below the RfD are not assumed to be "acceptable" (or risk-free), and nor are all doses that exceed the RfD necessarily "unacceptable" (*i.e.*, result in adverse effects).

The RfD is derived by dividing the NOAEL or LOAEL by an Uncertainty Factor (UF) that reflects the data set upon which the RfD is based. In practice, the standard UFs used in determining RfDs for

[7] The European Commission Co-ordination (ECCO) peer review team advocates a stepped approach. It recommends the steps 100 (the usual safety factor), 250, 500 and 1000, but clear, written rules for the justification of safety factors higher than 100 must be defined (*e.g.* a safety factor of 1000 might be applied to the LOEL for carcinogenic and teratogenic effects) (1174/ECCO/BBA/97, 5 January 1999: "Documentation, clarification and updating information with respect to Section 4 of Part B of the ECCO-Manual").

pesticides are 10 to account for interspecies extrapolation and 10 for intra-species variation. Additional UFs may be applied for the use of a LOAEL instead of a NOAEL, and when extrapolating from shorter than chronic animal studies. In addition, a Modifying Factor (MF) of greater than zero but less than or equal to 10 is sometimes also applied, based on a professional judgement of the *entire* database of the chemical. The equation is RfD = NOAEL or LOAEL / (UF x MF), expressed in mg/kg bw/d.

For a detailed explanation of the RfD, how it varies from the ADI, and its role in risk management, the reader is referred to the relevant background document (US EPA, 1993; see http://www.epa.gov/ngispgm3/iris/rfd.htm).

Where datasets allow appropriate analysis, alternative concepts such as the Benchmark Dose (BMD) or Effective Dose (EDx) are under consideration by regulatory agencies calculating health end-points. For a discussion of the BMD and the methods used, the reader is referred to Barnes *et al.* (1995). Commonly, a threshold dose for an adverse effect is established and a NOEL derived. The BMD method has been developed to take into account the shape of the dose–response curve and the size of the sample; it aims to quantify the risk associated with doses at or above the NOEL. It is not a method of risk extrapolation, *i.e.* it does not make numerical extrapolations to extremely low doses or levels of risk.

1.2.9 Regulatory Decision Making

Regulatory actions may be based on some or all of a number of distinct processes, summarised as follows:

- **Hazard Assessment.** Assessment of the inherent properties of chemicals and their capacity to harm humans and the environment.

- **Risk Assessment.** Estimation of the probability of any harm occurring and its likely extent. Risk assessments contain some or all of the following steps:

– Hazard Identification. The determination of whether a particular chemical is or is not causally linked to particular health effects.

– Exposure Assessment. The determination of the extent of human exposure before or after application of regulatory controls. For agricultural and veterinary chemicals, this is often calculated from national food consumption data and market-basket surveys of pesticide residue in foodstuffs. Daily intake calculations are based on the procedures outlined in *Guidelines for Predicting Dietary Intake of Pesticide Residues* (revised 1997) prepared by the Global Environment Monitoring – Food Contamination Monitoring and Assessment Programme in collaboration with the Codex Committee on Pesticide Residues and published by the WHO (1997). See also Chapter 5 of IPCS Environmental Health Criteria Monograph No. 210. Formal programmes to assess human exposure to pesticides in the home garden/home veterinary setting vary from country to country; in general, pesticides for home garden/home veterinary use are of low toxicity and their availability, packaging and labelling are subject to appropriate controls. If additional assessment of exposure from home garden/home veterinary use is required, the evaluator should take account of the proposed use pattern and application rate of all the formulations proposed for marketing. Such information is critical in establishing potential human exposure and the routes of exposure associated with different uses.

- Dose–response Assessment. The determination of the relation between the magnitude of exposure and the probability of occurrence of the health effects in question. Dose–response relationships play an important role in hazard identification, because a positive correlation between the dose of a chemical and the incidence or extent of a toxic effect is the best evidence of a causal link; *i.e.* evidence that the chemical is hazardous.

- **Risk Characterisation.** The description of the nature, and often the magnitude, of human risk, including attendant uncertainty, in which the exposure and dose–response assessments are combined to produce a risk estimate, and in which the strengths and weaknesses, major assumptions, judgements and estimates of uncertainties are discussed.

- **Benefit Assessment.** Analysis of the possible advantages of a certain use of a chemical product.

- **Assessment or Analysis of Consequences.** Prediction of the consequences of a certain choice.

- **Risk–Benefit Assessment.** Assessment based on an acceptable level of risk from the standpoint of society.

- **Risk Management.** The process by which the risk from an identified hazard is controlled and exposure is maintained at a level that presents an acceptably low risk. Risk management activities are concerned generally with identifying subpopulations at risk, establishing exposure standards for air, water and food, and taking remedial action following excessive exposure.

The various steps in the process, leading to the management of the identified risks, are made independently of each other and in a manner that distinguishes facts and scientific observations from general viewpoints. Thus, the process of risk assessment uses the factual base (*i.e.* the database of the hazard and exposure assessments) to define the likely effects of hazardous materials and situations on the health of individuals or populations under given exposure conditions. Risk management is the process of weighing up policy alternatives and selecting the most appropriate regulatory action, integrating the result of risk assessment with social, economic and political concerns to reach a decision (NAS/NRC, 1983; see also Chapter 6 of IPCS Environmental Health Criteria Monograph No. 210).

Separating risk assessment and risk management as much as possible allows evaluators to concentrate on analysis, evaluation and interpretation of toxicological data according to sound scientific principles and without regard for subsequent regulatory decisions and risk management actions. At least some of the controversy surrounding regulatory actions has resulted from a blurring of the distinction between the risk assessment process and risk management policy (NAS/NRC, 1983).

2. DOCUMENTATION AND DATA ACCEPTANCE

Quality, integrity and thoroughness in the reporting of experimental data from toxicity studies are essential to the subsequent independent analysis and evaluation of submitted studies. In essence, quality evaluations expected of regulatory agencies have their foundations in the submitted evidential documentation. A qualitative assessment of the acceptability of study reports is therefore an important part of the process of independent evaluation.

While directed at laboratories conducting toxicology studies, IPCS Environmental Health Criteria Monograph No. 141, *Quality Management for Chemical Safety Testing* (WHO, 1992), is a useful reference in regard to issues of quality control. Data screening pro formas are included in the OECD Pesticide Forum Working Group's *Forms for Screening Test and Study Reports and Summaries for Completeness* (OECD, 1998f). Further guidance on completeness checks is provided in *Guidance for Industry Data Submissions* and *Guidance for Country Data Review Reports on Plant Protection Products* (OECD, 1998a, b).

To be acceptable to a regulatory agency, studies must be of an adequate standard. The EC's technical guidance document on risk assessment for new and existing substances (EC, 1996) defines the adequacy of an experiment in terms of its *reliability* and *relevance*.

2.1 Reliability

Reliability covers the inherent quality of the study, relating to test methods and the way that the conduct and results of the study are described. Parameters included here are: the observational and experimental methods; frequency and duration of exposure; the species, strain, sex and age of the animals used; the numbers of animals used per dosage group; dose, route and frequency of dosing; and the conditions under which the substance was tested.

Many guidelines for the generation of scientifically valid data concern good experimental design, laboratory practice and reporting, *e.g.* OECD and US EPA guidelines and accepted codes of Good Laboratory Practice, or GLP (OECD, 1982 & 1993; US EPA, 1983). They can be helpful in determining the acceptability of reports and data. However, an evaluator needs to judge how well a study *in toto* facilitates the identification of potential adverse effects, or lack thereof, of the substance being evaluated, rather than how precisely it fits a prescribed test guideline or "recipe". The experience of senior evaluators can be helpful in allaying concerns about the acceptability or otherwise of the conduct or reporting of a study.

An evaluator should read a report, including supporting data presentations, and make a judgement as to whether the study in question was well conducted and reported or whether significant deficiencies exist. If there are obvious deficiencies that would lead the reviewing toxicologist to consider the study invalid, further evaluation may be a waste of resources. The procedures to be adopted with regard to deficient studies are likely to differ from one country to another.

An evaluator should also consider any effects of modifying factors that may result in major inequalities between control and treated animals. This qualitative consideration has more to do with the evaluation and interpretation of data than with acceptability of documentation. It is mentioned here because determination of the factors that may have a major influence on toxicological data needs to be made prior to the analysis of the data. Many factors influence the responses of experimental animals to chemical substances; some of these are discussed by Doull (l980). Some influences may be quite subtle, as exemplified by studies performed by Thompson *et al.* (1982), in which it was noted that the onset of acute pulmonary oedema in rats being used in immune hypersensitivity studies was sudden and seasonal. Subsequent studies revealed the reasons for this. Circadian rhythms and seasonal physiological variations can subtly influence experimental results. Short days enhance the immune function of all species, lymphatic organs becoming largest in late autumn or early winter and diminishing prior to breeding. Such influences on animal responses can be troublesome when their effects are confused with, or misinterpreted as, toxic responses to treatment.

Occasionally, detailed analysis of the data subsequent to the initial reading of the report will indicate deficiencies that were not obvious previously. These should be noted and the analysis completed as far as possible.

If an evaluator has any doubts about the thoroughness or competency of the execution and reporting of a study, he/she should discuss them with his/her supervisor before taking appropriate action in line with national or regional procedures.

2.2 Relevance

The *relevance* of a study is the extent to which it is appropriate to a particular hazard or risk assessment. To assess the relevance of data, it is necessary to judge if an appropriate species has been studied, if the route of exposure is relevant to the population and exposure scenario under consideration, and if the test chemical is representative of the chemical to which the population is or will be exposed. The test chemical should therefore be properly identified and any significant impurities described (EC, 1996).

If the test chemical is a pesticide, most studies will have been performed on animals, and there will often be no data on its metabolism, toxicokinetics or toxicity in humans. Under these circumstances, adverse effects observed in animals will normally be assumed to occur in humans, even if the threshold level of exposure is unknown. Clear and well-documented evidence of a species-specific effect is therefore required before an animal study is deemed irrelevant to humans. [The EC's technical guidance document on risk assessment for new and existing substances (EC 1996) cites the example of light hydrocarbon-induced renal nephropathy in male rats.]

In cases where human data are available on the test chemical or a close structural analogue (*e.g.* in the case of parasiticides, antibiotics and some organophosphates), it may be possible to judge the relevance of animal data on the basis of comparative metabolism and toxicokinetics or clinical experience. An example is abamectin, an agricultural acaricide and veterinary parasiticide that has enhanced developmental toxicity in P-glycoprotein-deficient CF-1-strain mice. The relevance of studies in the CF-1 mouse has been dismissed by the Joint Meeting on Pesticide Residues (JMPR) following comparative studies with abamectin in P-glycoprotein-normal CD-1 mice, and veterinary and human clinical experience obtained with its close structural relative, ivermectin.

These qualitative considerations establish the acceptability or otherwise not only of specific reports but also of the eventual evaluation, interpretation, judgements and risk assessments made by toxicologists.

Whether reports and other technical information submitted to regulatory agencies are acceptable is primarily a scientific judgement. The submitters of the information deserve to know the rationale behind any rejection of data, hence this should be succinctly stated in the evaluation document.

3. ANALYSIS AND EVALUATION OF ADVERSE EFFECTS IN CHRONIC TOXICITY AND CARCINOGENICITY STUDIES

It is important that all toxicity data, and the methods by which they are obtained, be subjected to critical and independent scientific assessment by the regulatory evaluator. As the primary emphasis is on independent assessment, it is the evaluator's responsibility to ensure evaluation reports, including any company summaries and company-sponsored "expert reports" which may be used, comprehensively document study results, interpretations and conclusions in an accurate, clear and transparent manner.

Valuable guidance documents for evaluating data and conducting assessments are IPCS Environmental Health Criteria Monograph No. 104, *Principles for the Toxicological Assessment of Pesticide Residues in Food* (WHO, 1990), and related monographs, *e.g.* IPCS EHC 6, 70 & 141 (WHO, 1978, 1987, 1992). Evaluators may also refer to *OECD Guidelines for the Testing of Chemicals* (OECD, 1993) to check the adequacy of certain studies. Another useful document is *Guidelines for the Preparation of Toxicological Working Papers for the Joint FAO/WHO Expert Committee on Food Additives*, Geneva, July 1987 (ICS/89.41) (FAO/WHO, 1987).

3.1 Analysis and Evaluation of Major Study Parameters

It needs to be borne in mind that not all observed effects of test substances are necessarily toxic effects (see definition of NOAEL, Subsection 1.2.6). Rather, they may be physiological (*e.g.* liver enzyme induction leading to hepatic enlargement, although see Subsection 1.2.3, "Treatment-related Responses"), or may be a manifestation of a pharmacological effect (*e.g.* in an animal colony suffering from low-grade infections, an antibiotic will lower leucocyte counts in treated animals relative to controls; obviously it is not appropriate to describe this as a leucopenic effect of the chemical).

Concurrent control groups should always be used. Notwithstanding the value of historical control mean/range data, it is generally not appropriate to rely on statistical comparisons with historical controls because biological parameters, including the incidence of spontaneous lesions, can vary significantly over time (and even between concurrent randomised control groups). Controls and treatment group animals must be of comparable age because some forms of toxicity represent no more than acceleration and/or enhancement of age-related changes. Examples of pathological changes in aged rodents that may be affected by compound administration include chronic progressive glomerulonephropathy, peripheral nerve degeneration, amyloidosis and various neoplasms.

The use of non-treated and vehicle-treated control groups aids the assessment of effects due to vehicles or excipients. When a vehicle is used, the need for vehicle-treated controls is paramount. Since some parameters can be affected by animal handling (*e.g.* serum ALT rose in mice that were grasped round the body compared with unhandled or tail-handled mice; Swaim *et al.*, 1985), control animals should be treated in the same way as test animals.

For reasons already stated, control animals must receive as much attention during the analysis and evaluation process as treated ones. Any untreated animal or group of animals may exhibit signs of abnormality or drift from the norm for the species or strain in question. Because of the possibility that statistically significant differences between treated and control groups are the result of abnormal values among the controls, such differences should usually be dose-related and should delineate a trend away from the norm for the particular stock of animals if they are to be indicative of a true compound-related effect.

Historical control data may be useful in evaluating the acceptability of the "normal" data obtained from control groups (Haseman *et al.*, 1984 & 1997; Paynter, 1984; Sumi *et al.*, 1976; Tarone, 1982). Any departure from the norm in the control groups should be discussed in the evaluation document and taken into consideration, especially in any statistical analysis. Consistent departures could necessitate investigation of the source of the animals.

Ideally, all historical control data submitted for consideration are obtained from the laboratory at which the study being assessed was carried out, and relate to animals of the same strain, age and sex, and obtained from the same supplier, as those used in the study. They should come only from studies conducted within five years, or two to three years either side, of the study under review. Any study methodology that could have affected the results should be identified. Relevant parameters include pre-sampling conditions such as fasting or non-fasting, haematology and clinical chemistry assay methods, histopathological criteria for lesion identification, and time of terminal sacrifice. European requirements for submission of historical control data are fully described in Section 5.5, Annex II, of Directive 91/414/EEC: *Plant Protection Products*. Where historical data are used in an assessment, they should be clearly identified (see Section 5.2).

Basic parameters, *e.g.* body-weight gain, food consumption and conversion efficiency, are important and delineate the LOEL in a high proportion of studies. It is not uncommon for such parameters to be affected earlier in a study, and at a lower dose, than many other markers. Weil & McCollister (l963) analysed toxicity end-points, other than oncogenicity, obtained from short- and long-term tests and concluded that only a relatively small number of end points were effective in delineating the LOEL. Body weight, liver weight, kidney weight and liver pathology delineated the LOEL for 92% of test chemicals in subchronic studies and for 100% in chronic studies. To reach 100% efficiency in subchronic studies, renal and testicular histopathology had to be included. Heywood (l981) surveyed the toxicological profiles of 50 compounds in rodent and non-rodent species and confirmed these observations. These criteria should therefore receive careful attention in the analysis and evaluation process. However, there is no implication that they delineate *all* the stress markers or toxicity end-points likely to appear, particularly since toxicology testing has undergone significant development since the study reviews referred to were reported, especially with respect to clinical chemistry and neurotoxicology. Evaluators should therefore be aware that effects on any end-points may be important.

As noted above (Subsections 1.1.1 and 1.1.2), the existence of a dose–response relationship in respect of a particular biological or toxicological effect provides strong evidence that the response is a result of exposure to the agent being tested. In hazard assessment and the reporting of studies, correlations between external dose and the incidence and/or intensity of toxicological end-points need to be considered and reported.

3.1.1 Mortality/Survival

Death is a highly definitive end-point but is no longer regarded as ethically acceptable unless it is unavoidable. In addition to possible treatment-related effects, a considerable number of animals develop spontaneous disease during chronic experiments. In full life-span experiments, even in the absence of lethal treatment-related effects, all animals eventually die of spontaneous disease. Timely humane killing is the usual means of terminating pain and distress when chemical analgesia cannot be used and the prospect of recovery is poor, or if the condition is likely to interfere with the experiment. Should a severe health disorder develop in a group of animals, termination of the group, or even the entire experiment, may be necessary. The OECD's *Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation* (OECD, 2000) should be consulted for further information about the euthanasia of experimental animals used in safety evaluations.

Reasonable efforts should be made to determine the cause, or likely cause, of individual deaths. The evaluation of pathological lesions or morphological changes in belatedly observed deaths is frequently complicated by post-mortem autolysis. OECD Test Guidelines 451 and 453 specify that in order for any negative carcinogenicity study to be acceptable, no more than 10% of any group can be lost through autolysis, cannibalism or management problems. They further stipulate that survival of all groups should be no less than 50% at 18 months for mice and hamsters and at 24 months for rats. It is important to separate deaths caused by factors unrelated to exposure to the test agent (*e.g.* acute or chronic infections, age or disease-related degenerative processes, anatomical abnormalities, negligent handling or accident) from toxicity-induced deaths. All data relating to the study life of moribund or dead animals, as well as the results of post-mortem examinations, should be scrutinised in an attempt to make this distinction.

Analysis of mortality requires more than a statistical treatment of incidence at the termination of a study. Survival/mortality data can be influenced by factors other than the test substance. Changes in the protocol during the course of a study can complicate the analysis, *e.g.* alterations in dosage levels can produce a confusing mortality pattern.

The maintenance of statistical power in inter-group comparisons is dependent on the survival of sufficient animals throughout the study. Excessive reductions in group size increase the "critical difference" between groups; in other words, they reduce a study's sensitivity to inter-group differences in the parameters under analysis, including the incidence of cancer and other diseases. Furthermore, the likelihood of making a false positive or false negative error rises as group sizes fall (WHO, 1987).

If high mortality has affected a study, evaluators must clarify in their assessment report whether it was associated with toxicity (*i.e.* whether the MTD was exceeded) or caused by infectious disease or some other factor, such as the genetic background of the test animals. The reader also needs to know if it was confined to one group or occurred in several groups, and must be given a comparison of findings among controls and animals administered low, intermediate and high doses.

The data submitter should explain any unusual mortality pattern on biological or toxicological grounds. If overall mortality is significantly greater than expected for the particular colony and strain, or for a particular group within a study, a credible explanation should be provided. (If this is not the case, national agencies may consider conducting a laboratory and database audit.) It is the prerogative of individual agencies to accept or reject studies affected by high mortality, depending on their policy and the circumstances of each case. Even if an agency does not have fixed criteria for the acceptance or rejection of studies, the evaluator should note if the validity of a study's conclusions has been compromised through premature loss of test animals.

It is important to evaluate mortality patterns within each group. Mortality may be clustered early or late in a study, be intermittent and scattered throughout, or be higher in one sex than in the other. Analysis of the cause of individual deaths helps determine the toxicological significance of these various patterns. Treated animals that die early may simply be among the more susceptible in the test population. Alternatively, in experiments in which the quantity of test substance in the diet is kept constant, such deaths may indicate changes in compound intake per unit body weight. Relative to body weight, young rats eat more than older rats, and hence ingest relatively more of the test substance. Early deaths may therefore be the result of the higher exposure, on a mg/kg bw/d basis, of young animals than of older animals.

Deaths clustered in a specific period may indicate a spontaneous disease epidemic of limited duration. If high mortality in treated groups is associated with infectious agents, but there is no such evidence in the concurrent control group, this could indicate an immuno-suppressive action of the chemical being tested.

Although the discussion so far has focused on the deleterious impact of high mortality, a study's validity can be reduced by high *survival*. This might happen in a carcinogenicity or combined chronic toxicity/carcinogenicity study that employs insufficiently high doses of the test compound or is terminated prematurely, before sufficient progression of the disease. The implications of high survival vary, depending on the underlying causes and which group or groups show unusually low mortality. In this context, the effect of dietary intake on mortality needs to be considered. A compound administered in the diet may make the laboratory chow more or less palatable, may have a pharmacological stimulant or depressant effect on appetite, or may affect the partitioning of nutrients in the food. Likewise, decreased water consumption (*e.g.* if an unpalatable compound is administered in the water) leads to reduced food consumption. These effects may have a significant influence on longevity, since it has been clearly shown in animals that long-term dietary restriction increases life span very significantly (*e.g.* Tucker, 1979). Conversely, excessive *ad libitum* intake of highly nutritious diets can reduce life span (see Table 3 for comparison of the life span of laboratory and wild rats). Regulatory authorities have yet to formulate recommendations on the feeding of restricted or *ad libitum* diets in toxicity studies; however, there are some useful references on this topic, including Keenan (1998; see also related articles by the same author), Christian *et al.* (1998), Klinger *et al.* (1996), Masoro (1992) and Thurman *et al.* (1995).

3.1.2 Clinical Observations

Adverse clinical signs (*i.e.* gross observations) noted during the exposure period may correlate with toxicity end-points or disease processes. These can be used as supportive evidence of a dose–response relationship and may play a role in the determination of the NOEL or NOAEL. However, not all adverse clinical signs correlate with pathological or morphological changes in organs or tissues. Some are caused by biochemical or physiological effects. For example, incoordination, muscle twitching, tremor or diarrhoea may indicate acetylcholinesterase inhibition without any morphological changes being evident in nervous tissue. Non-specific signs such as reduced activity, hunched posture and piloerection are also commonly observed. These are probably behavioural responses to malaise induced by the test chemical, and may not correlate with any specific alteration in biochemistry or organ histology. Useful information on gross behavioural observations in laboratory animals and abnormal behaviour patterns can be found in Bayne (1996).

Most clinical signs observed during physical examination of individual animals are determined without the aid of instruments. It is important, therefore, that all deviations from the "normal"

observed in control and treatment groups are adequately described and recorded during the study and presented in the study report.

Many qualitative signs can be counted, scored for intensity and tabulated as incidences. However, statistical analysis is of limited value. The evaluator must rely on the number of individuals per group exhibiting signs of a particular type, and both the frequency and intensity of the responses, to gain an impression of a dose–response relationship.

Clinical observations that might be associated with neoplasia (*e.g.* haematuria, abdominal distension or impaired respiration), as well as palpable tumours, may be useful in defining the time a tumour was first suspected of being present. They may also be of help in the evaluation of decreased tumour latency in long-term rodent studies and in determining the cause of death. **An evaluation should draw attention to any correlation, or lack thereof, between clinical signs and specific toxicity end-points.**

An ophthalmoscopic examination should be made of at least the control and high-dose animals both before a test substance is administered and at the conclusion of a study. The limited usefulness of gross examination of the outer parts of the eye by ophthalmoscope should be borne in mind, particularly in studies of compounds that are potentially toxic to the visual system, *e.g.* organophosphorus compounds. Examination of deeper parts of the eye, including the fundus, is indispensable, as it provides good information on toxic effects on the eye, although not necessarily on the optic nerve and visual system in the CNS. Histopathological data on the eye and optic nerve, and, if available, any electroretinographic data, should therefore be considered in conjunction with clinical ophthalmoscopy.

3.1.3 Body Weight Changes and Food and Water Consumption

During the course of a study it is important to compare changes (both gains and losses) in the body weight of individual animals and groups of animals with changes in the concurrent controls (Heywood, 1981; Roubicek *et al.*, 1964; Weil & McCollister, 1963). Such changes are usually related to food intake, and analysis of one but not the other is of limited value. Weight loss or decreased body weight gain may not always be related to toxicity *per se* (Seefeld & Petersen, 1984). Occasionally, incorporation of the test substance into the diet reduces the palatability of the diet to many individuals in all treatment groups or to the majority of individuals in the higher dietary level groups. Food spillage needs to be considered in the evaluation of palatability and substance intake. The same considerations apply if the test substance is administered in drinking water.

Reduced palatability is often apparent during the first two or three weeks of a study. Sometimes animals in the affected group(s) grow accustomed to the diet and a gradual increase in group weight gain occurs (Nolen, 1972). But even if the gains of individual animals per gram of food consumed are favourable (*i.e.* the animals' food conversion efficiency is high), group weight or weight gain may continue to lag, producing a statistically significant difference between the affected group(s) and the concurrent controls that is unrelated to toxicity of the test substance (McLean & McLean, 1969). Sometimes the test substance interacts with one or more essential nutritional elements in the diet, causing weight gain decrements or altering toxic responses (Casterline & Williams, 1969; Conner & Newbern, 1984; Rogers *et al.*, 1974). This phenomenon may be encountered in subchronic studies, and can usually be overcome by acceptable means before a chronic study is initiated. Occasionally, control values for weight gain (at one or more time points) can be low, causing the other value to appear unusually high.

Weight loss is a sensitive, objective sign of health problems. It may point to wasting diseases (cancer, chronic renal disease, etc.), pain and distress, or inability to eat (due to oral ulcers, for instance), and is a leading criterion in the making of decisions about animal euthanasia [see the OECD's guidance document on humane end-points (OECD, 2000)].

Diet, food and water consumption and body weight gains *per se* can also have an important influence on many aspects of animal responses, including shifts in metabolic, hormonal and homeostatic mechanisms (Kennedy, 1969), disease processes (Berg & Simms, 1960; Paynter, 1984; Ross & Bras, 1965; Tannenbaum, 1940; Cohen *et al.*, 1994) and maturation (Innami *et al.*, 1973). (See also Subsection 1.2.2.) These should be considered when unusual effects are observed in the absence of any indication of injury to organs or other vital systems. Immunotoxicological indices may also be affected. For example, depressed total white blood cell (WBC) count and thymic and splenic T-cell count, reduced reactivity to T-cell mitogens, decreased relative spleen weight, lymphoid atrophy and increased NK cell and macrophage phagocytic activity have been observed in animals as responses to weight loss caused by decreased food consumption. Although specialised immunological parameters are not usually measured during chronic toxicity or carcinogenicity studies, evidence of consequentially impaired resistance to infectious disease may be seen among affected animals, and may influence the reviewer's interpretation of the results.

Of significant help in the evaluation of body weight changes and attendant effects is the graphical presentation of group mean body weights and food consumption *vs.* compound consumption (on a mg/kg bw basis). This allows quick identification of any unusual or sudden changes in gain or loss by any group. Evaluators should be aware that body weight, body weight gain and food conversion efficiency data are usually of greatest value during the early to middle part of rodent chronic and carcinogenicity studies, and must be interpreted with caution as a study extends past week 78 and approaches termination. By this time, rodents have ceased to grow and are entering the geriatric phase of their lives. Rats in particular may lose weight because of ageing lesions, such as renal disease or pituitary tumours; heavier rats tend to die earlier than lighter rats, thereby selecting for lighter individuals and biasing the group mean body weights downwards.

3.1.4 Haematological, Clinical Chemistry and Urinary Measurements

Regulatory guidelines generally suggest that haematological, clinical chemistry and urinary parameters be routinely measured in chronic toxicity studies. However, for **carcinogenicity** studies, OECD Test Guideline 451 (OECD, 1981a) does not specify clinical chemistry or urinalysis, and the only haematological parameter called for is a differential blood count.

Because of normal biological variation in inter-animal values, and the alteration of values in response to a variety of inputs, evaluators have to contend with much "noise" in this area; they are frequently presented with statistically significant but scattered effects, in the absence of any evidence of clinically significant relationships with specific toxicity end-points. For example, Pearl *et al.* (1966) restrained rats for six hours and monitored aspartate aminotransferase (AST) and alanine aminotransferase (ALT) activity in the serum. The activity of both enzymes rose considerably, indicating they are susceptible to stress factors. AST appeared particularly susceptible, not returning to its basal level even within six days. To deal with noise it is necessary to examine whether an effect is within the normal range of variation, using concurrent and historical controls. Note that some parameters can vary significantly without clinical manifestation, while others, *e.g.* the level of serum potassium, have a very narrow normal clinical range, meaning small changes can be important.

Frequently these data show apparently random changes in individual groups, or, less commonly, trends in changes across several groups that are unrelated to dose. If, as an aid to evaluation, historical control data are used for comparison, it must be kept in mind that "normal values" in haematological and clinical chemical measurements depend on the specific methods used to generate the data. Thus only values obtained using identical methods at the same laboratory are valid in such comparisons. Values from literature that do not specify the methods employed to obtain them should be used with caution. These comments underline the importance of concurrent control data for clinical chemistry, haematology and urinalysis parameters.

The following example illustrates how differences in experimental methods can influence the results of clinical laboratory tests, even within the same animal(s). Dameron *et al.* (1992) collected blood samples from the orbital venous plexus (OVP) and the posterior vena cava (PVC) of adult male Sprague–Dawley rats. There were biologically significant differences in coagulation times (as measured by prothrombin time and activated partial thromboplastin time) and serum Mg and P levels between the samples from the OVP and the samples from the PVC.

Haematology assays provide information on bone marrow activity and on the status of other organs governing the synthesis, function and destruction of components of the circulatory system. The battery of haematology tests usually carried out in toxicity studies evaluates erythrocytes, leukocytes, platelets and coagulation. The following exposition of some of the more common effects on haematological parameters is adapted from Hall (2001).

Disturbances in erythrocyte parameters often reflect an imbalance between the production and loss of red blood cells (RBCs). *Non-regenerative anaemias* arise from reduced erythrocyte production. Direct injury to haematopoietic stem cells results in aplastic anaemia, which is characterised by decreased erythrocyte, leukocyte and platelet counts accompanied by hypocellular bone marrow. Chronic inflammatory disease and dysfunction of the kidney, liver and endocrine system have a negative effect on erythropoiesis and erythrocyte survival, and can be associated with anaemia characterised by a low RBC count but normal mean corpuscular haemoglobin concentration (MCHC) and mean corpuscular volume (MCV). Megaloblastic anaemia is a non-regenerative anaemia characterised by abnormally large erythrocytes and hypersegmented "giant" neutrophils. This macrocytosis occurs because developing erythrocytes undergo fewer divisions than normal before maturation. Vitamin B_{12}, or folate, deficiency and impaired DNA synthesis are the most common causes of megaloblastic anaemia. Finally, non-regenerative anaemia is often a feature of leukaemia because of competition between proliferating neoplastic and normal haematopoietic cells for nutrients and space in the bone marrow.

Regenerative anaemias result from two general causes: blood loss and erythrocyte destruction. In both cases there is likely to be an increased proportion of circulating reticulocytes relative to mature erythrocytes, reflecting higher RBC production in response to homeostatic regulation. Higher MCV may also be observed, sometimes together with decreased MCHC. The most common mechanisms of chemical-induced haemolysis are direct damage to RBC membranes, oxidation of haemoglobin (causing the formation of Heinz bodies and possibly methaemoglobinaemia) and immune-mediated RBC destruction.

Differential WBC counts enumerate neutrophils, lymphocytes, monocytes, eosinophils and basophils. When interpreting and reporting differential WBC counts, it is essential to use the absolute, not relative, cell counts, as the latter are of little value in assessing an animal's condition. A marked increase in circulating WBCs is of major importance in the diagnosis of leukaemia, which, in addition to being caused by some chemical carcinogens, is a common spontaneous lesion in ageing rodents (especially F344 rats). Evaluators should be aware that WBC counts can also rise in animals frightened by handling and the taking of blood samples. Neutrophil counts increase in response to

infection and inflammation, and can be decreased by cytotoxic test chemicals. An elevated lymphocyte count is seldom associated with test chemicals, although it may be observed in conjunction with chronic inflammation or following administration of chemicals that elicit an immune response. Lymphocyte counts can be depressed by corticosteroids and prolonged stress, and are frequently low in moribund animals. Eosinophils usually increase in number as a secondary phenomenon in response to parasitosis or hypersensitivity, while monocyte counts can rise because of inflammation, tumour-associated necrosis or haemolytic anaemia. Test chemicals rarely have an effect on basophil counts.

Effects on blood clotting mechanisms are assessed both functionally (by measurement of thromboplastin, clotting and prothrombin times) and by enumeration of thrombocytes (*i.e.* platelets). Unless the test chemical is a haematopoietic growth factor, an increased thrombocyte count is seldom a primary effect; it is more likely to be a result of stimulation of the bone marrow caused by haemolytic anaemia, blood loss or infection. Decreased production or increased consumption of platelets may cause thrombocyte counts to fall. Test chemicals that inhibit RBC and WBC formation also frequently inhibit platelet production. Platelet consumption can increase following extensive haemorrhage, especially from multiple sites. If lesions affecting blood vessels are severe and widespread, disseminated intravascular coagulation may develop, and platelet counts will decrease markedly. Immune-mediated thrombocytopenia is also mediated by some xenobiotics, sometimes in conjunction with immune-mediated haemolytic anaemia. For haematology values, see Derelanko & Hollinger (1995), Chapter 14, Tables 14–20, and Chapter 22, Tables 17–19.

Urinalysis is primarily concerned with kidney function and is valuable in the identification of chemically induced renal toxicity, but can also provide information about the functioning of organ systems outside the urinary tract. An increase in the volume of urine and a decrease in its specific gravity often indicate impairment of the kidney's ability to concentrate urine. Urinary pH can be used to ascertain disturbances in the bodily acid/base balance but is unreliable if the urine specimen is not fresh. Urine specimens that have been collected overnight at room temperature may lose carbon dioxide or become contaminated with ammonia-producing bacteria. Injury to renal tubules can lead to sloughing of tubular epithelial cells, which subsequently appear in the urine. Proteinuria may arise from injury to glomerular membranes, renal tubules or both. Glomerular injury can impair the retention of proteins of high molecular weight, such as albumin, while damaged or dead renal cells can release enzymes and other proteins into the urine. Proteinuria is also common among ageing rats, being a consequence of progressive nephropathy, and may complicate interpretation of results during the second year of a chronic study. High levels of protein in the urine may also be associated with the presence of casts; hyaline casts contain protein alone, while cellular casts consist of aggregated erythrocytes, leukocytes or epithelial cells.

Elevated urinary glucose or nitrate concentrations may follow injury to the renal proximal tubule, which is the site of reabsorption of glucose and amino acids. Glucosuria may also result from diabetes mellitus, although this is not a common finding in toxicology studies. If the test chemical or its metabolites are present in the urine at high concentrations, it/they may precipitate from solution and form solid aggregates visible on examination. Irritation, mechanical abrasion, or inflammation of the kidney, ureters or bladder are likely to cause the presence of erythrocytes, leukocytes and/or haemoglobin in the urine. Urinalysis can also provide information on toxicity outside the kidney and bladder. Elevation of bilirubin in the urine is suggestive of hepatocellular dysfunction and jaundice. Increased urinary excretion of ketones is an indication of diabetes, a diet low in carbohydrate or impaired fatty acid oxidation (Tyson & Sawhney, 1985; Bennett, 1997; Tarloff & Kinter, 1997; Hall, 2001). For urinalysis values and test parameters for renal function in rats and other experimental animals, see Derelanko & Hollinger (1995), Chapter 8, Tables 3–13, and Chapter 22, Table 16.

Clinical chemistry assays are extremely useful for diagnosing organ system dysfunction in living animals, and because repeated measurements are possible from the same individual, they can provide data on the time course of toxicity. To gain maximum information from enzyme determinations it is important to consider the most appropriate enzymes, together with their distribution among and within organs and their subcellular location. Evaluators should also be aware that the sensitivity and specificity of changes in serum enzyme activity are greatly influenced by the species selected for testing (see e.g. Clampitt, 1978; Tyson & Sawhney, 1985), and that species differences are of great importance when specific clinical chemistries are selected for inclusion in toxicity studies. For clinical chemistry values in laboratory species, see Derelanko & Hollinger (1995), Chapter 14, Tables 7–13.

The following paragraphs are adapted from Hall (2001).

Liver function may be assessed by monitoring a variety of parameters. Several of the enzymes normally present within hepatocytes can be released into the circulation if the integrity of cell membranes is compromised. The enzymes most frequently used to assess hepatocellular injury are ALT, AST, sorbitol dehydrogenase, glutamate dehydrogenase and lactate dehydrogenase (LDH). In general, ALT is the most useful in detecting hepatocellular injury in most laboratory animal species except guinea pigs, in which its hepatocyte concentration is relatively low. However, increased serum ALT activity does not always indicate primary hepatocellular injury. Biliary toxicity and bile duct obstruction may cause increased ALT activity through the effects of retained bile salts on the cell membranes of neighbouring hepatocytes. Severe muscle damage can increase serum ALT activity in the absence of liver injury, especially in primates. Some corticosteroids and anticonvulsants appear to induce ALT production, leading to a proportional rise in serum ALT activity.

Serum AST and LDH tend to parallel serum ALT with respect to liver injury, but occur in high concentrations in other tissues (especially muscle) and are hence less specific than ALT. Increases in serum AST activity caused by hepatotoxicity are usually less pronounced than concurrent increases in serum ALT activity. Since a proportion of intracellular AST is located within mitochondria, a more severe injury may be necessary for the release of a like quantity of AST. Corticosteroids and anticonvulsants induce AST production, as they do ALT production. Decreased serum ALT and AST activity is occasionally observed in toxicology studies. The most widely recognised cause involves an effect by the test chemical on vitamin B_6, a coenzyme required for full catalytic activity of aminotransferases. However, decreased serum aminotransferase activity has not been shown to correlate with significant hepatotoxicity.

Serum sorbitol dehydrogenase and glutamate dehydrogenase are also good indicators of hepatotoxicity in laboratory species. Sorbitol dehydrogenase is a cytosolic enzyme, whereas glutamate dehydrogenase is located within mitochondria. Alkaline phosphatase (AP) is present within biliary and canalicular membranes, kidney, intestine and bone (Tyson & Sawhney, 1985; Evans & Lake, 1998). Isozyme analysis is often used to differentiate between bone injury and damage to organ tissue, as these sources produce slightly different forms of AP. In adult dogs AP is a sensitive test for biliary function, but it is of little diagnostic value in rats, in which serum AP levels are relatively high and vary with diet, as serum AP is principally derived from the intestines (Tyson & Sawhney, 1985; Evans & Lake, 1998). AP activity in rat serum also tends to decline with age due to reduced bone AP synthesis with increasing age. As with the aminotransferases, hepatic AP production can be increased by some anticonvulsants and exogenous and endogenous corticosteroids.

Serum gamma-glutamyl transferase (GGT) activity is an effective indicator of biliary toxicity in rats and is more specific than AP. Although the highest concentrations of this membrane-localised enzyme are in the kidney and pancreas, serum elevations have been reported only with hepatobiliary toxicity and following induction by xenobiotics that stimulate mixed-function oxidase activity. Whereas serum

AP activity can rise following toxic effects on bone formation, serum GGT activity is unaffected. Furthermore, serum GGT activity is less likely to increase because of primary hepatocellular toxicity or intrahepatic cholestasis caused by hepatocellular swelling than AP. In rodents, basal serum GGT activity is low, and even small increases may be significant.

Hall (2001) cautions that liver enzyme activity may not change even in the presence of hepatotoxicity or hepatic dysfunction. The liver can be dysfunctional without significant cholestasis or hepatocellular degeneration and necrosis. Animals with end-stage liver cirrhosis can exhibit normal serum enzyme activity. For this reason, serum total bilirubin concentration is the liver-function test of choice. Conjugated hyperbilirubinaemia is a result of impaired secretion of bilirubin or cholestasis, or both. When increased bilirubin concentrationis caused by cholestasis, serum AP activity may also rise. However, relatively severe haemolysis can overwhelm the secretory process and also cause an increase in serum unconjugated bilirubin. Evaluators should therefore refer to the findings of haematology and tissue pathology to confirm the cause of any observed hyperbilirubinaemia. By contrast, serum bilirubin concentration may fall following administration of mixed-function oxidase inducers, which may enhance the metabolism and excretion of bilirubin.

Since the liver is responsible for the synthesis of many substances, severe hepatocellular dysfunction can cause decreased serum concentrations of urea nitrogen, glucose, albumin and total protein, together with prolonged coagulation times. Conversely, hyperglobulinaemia may occur. Depending on the type of hepatotoxicity, blood cholesterol levels may fall or rise. Examination of the entire biochemical profile, together with other clinical and anatomical pathology findings, is necessary to properly evaluate liver toxicity.

The principal clinical chemistry parameters used to monitor renal function are blood urea nitrogen (BUN) and creatinine concentrations. BUN is affected by the rate of urea production, the glomerular filtration rate and the rate of urine flow through the renal tubule. It is most likely to rise in response to dehydration (which decreases the glomerular filtration rate), renal diseases, toxicity to the renal parenchyma and blockage of the urinary outflow tract by calculi or other obstructions. Creatinine is formed by non-enzymatic breakdown of creatine, and changes in its serum concentration following alterations in renal blood flow, renal function or urine outflow tend to parallel those of BUN. Elevated blood creatinine is a reliable indicator of impaired glomerular filtration. Significant impairment of renal function may also raise serum phosphate concentration and lower sodium and chloride concentrations.

The standard battery of clinical chemistry assays includes measurement of serum proteins, carbohydrates and lipids. In toxicology studies, the most frequent reason for increased serum protein concentration in test animals is reduced hydration. In this case, albumin and globulin concentrations should remain constant relative to one another. Conversely, serum protein concentration may decrease following a prolonged deficit in food consumption, reduced protein synthesis (*e.g* due to maldigestion, malabsorption or hepatic dysfunction) or elevated protein loss (*e.g.* through renal dysfunction, haemorrhage or dermal toxicity). Inflammatory conditions can stimulate globulin production but may also decrease serum albumin concentration. Especially in rodents, a small decrease in serum albumin concentration can occur if the dose of the test chemical is poorly tolerated, but this effect is usually an indication of general poor health rather than of a specific toxic mechanism.

Interpretation of serum glucose levels is complicated by blood sampling artefacts. The concentration of serum glucose rises in response to fear, and is higher in non-fasted than fasted animals. On the other hand, failure to process blood samples promptly after collection results in a loss of glucose through consumption by erythrocytes. Diabetes mellitus, pancreatitis and hyperadrenocorticism raise serum glucose concentration, but modest treatment-related depression in this parameter is sometimes seen in

animals that fail to thrive and gain body weight. Conditions that cause hypoglycaemia include malabsorption, hepatic disease, endotoxaemia, and some tumours, in particular insulinomas and hepatomas.

Effects on serum cholesterol and triglyceride concentrations are often encountered in toxicology studies. However, they are easier to detect in young rats than older animals, which exhibit much variability in these parameters because of the influence of food consumption and assimilation, body weight and composition, liver function and thyroid and pituitary hormone balance. Cholesterol and triglyceride levels rise in hypothyroidism and diabetes mellitus. Cholestasis and lesions of the nephron can give rise to hypercholesterolaemia.

In contrast with many other clinical chemistry parameters, circulating mineral and electrolyte concentrations are regulated within narrow limits, and even small perturbations can be statistically and biologically significant. Increased serum calcium concentration is relatively uncommon in toxicology studies unless the test chemical specifically targets calcium metabolism, behaves similarly to vitamin D or causes hyperparathyroidism or renal disease. Because roughly half of circulating calcium is bound to serum albumin, hypercalcaemia can also arise from dehydration, while depression of calcium and albumin tend to occur concomitantly. Hypoparathyroidism, pancreatitis and renal disease can lower serum calcium. Serum inorganic phosphorus concentration is very sensitive to the glomerular filtration rate, and simultaneous elevations of phosphorus and BUN are common. Dietary phosphate intake also has a major influence on the level of circulating phosphorus, which may decrease following reduced food consumption.

Changes in serum sodium and chloride concentrations tend to occur in parallel when they are associated with relative water content, and both electrolytes become depleted by fluid loss. Serum chloride is disproportionately affected by disturbance of the acid/base balance, and increases following metabolic acidosis resulting from diarrhoea. Serum potassium concentration is highly sensitive to dietary intake. Increases may be observed following metabolic acidosis, severe tissue necrosis and renal failure, while decreases result from fluid loss and conditions that cause alkalosis.

Creatinine phosphatase (CPK) activity in the serum is measured primarily as a marker for toxicity to the skeletal muscle, while in studies on organophosphate and carbamate pesticides, assays of AchE activity in RBCs, blood and brain are undertaken to measure the extent of inhibition caused by the test chemical.

In rodent studies, interpretation of clinical chemistry data often becomes more problematic after the animals have reached the age of 18 months, because of the onset and progression of age-associated disease. As a study population approaches the end of its life span, rodents tend to display a general pattern of declining organ function, rising tumour burden and impaired production and/or regulation of hormones (especially those associated with thyroid function and sexual reproduction). This frequently results in a time-related loss of uniformity in the results obtained from different individuals. Increased within-group variation inevitably leads to higher standard deviations, resulting in a loss of statistical sensitivity and making treatment-related effects more difficult to detect. It is therefore not uncommon to encounter clinical chemistry parameters that show a statistically significant inter-group difference at the sixth and 12th months of a chronic study, but lose statistical significance at 18 or 24 months despite a continuation of the underlying toxicity. In such cases, evaluators are advised to compare the patterns of data shown by the controls with those of the various treated groups, and to examine the histology findings for correlations with clinical chemistry results.

In an overview of the usefulness of clinical chemistry data in subchronic toxicity studies, the relative sensitivities of eight clinical chemistry parameters commonly used to detect potential toxic liver and

kidney effects were evaluated for a series of 61 subchronic rat studies conducted by the US National Toxicology Program (Travlos *et al.*, 1996). Liver and kidney lesions were reported in 31% and 41% of the studies respectively. There was an association between treatment-related increases in ALT and sorbitol dehydrogenase activity and histopathological changes in the liver; changes in sorbitol dehydrogenase had greater positive and negative predictive value than similar changes in ALT. There was an association between treatment-related increases in BUN and creatinine concentration and morphological changes in the kidney; changes in creatinine had greater positive predictive value than similar changes in BUN. It is important to understand that clinical chemistry, haematology and urinalysis may fail to detect minor injury or may reflect only transient or reversible changes. Evaluation and interpretation of test results must therefore be performed carefully and correlated with histopathological findings.

When analysis and evaluation of clinical data indicate a dose–response relationship or a biologically important drift from concurrent control values, the evaluator should attempt to correlate the effect(s) with other manifestations of toxicity, and should indicate whether or not a correlation could be made.

Standard veterinary (*e.g.* Bush, 1991; Duncan *et al.*, 1994; Evans, 1996; Fox *et al.*, 1984; Jain, 1993) and human (*e.g.* Fischbach, 1996; Henry, 1984; Tyson & Sawhney, 1985; Walach, 1996) clinical manuals should be consulted for information about laboratory diagnostic tests and for assistance in the evaluation of potential correlations between clinical chemistry, haematological and urinary data and adverse effects.

Regulatory guidelines often specify which laboratory parameters should be measured; those most commonly measured are included in the table in Appendix II. Although immunotoxicological and neurotoxicological parameters are not usually measured in chronic toxicity and carcinogenicity studies, there may be circumstances in which they are investigated. If so, the reviewer should discuss effects on immuno- and neurotoxicity in separate additional sections in the evaluation report, referring to the relevant test guidelines [OECD Test Guideline 424 (1998e) and US EPA OPPTS 870.6200 (1998d) & 870.7800 (1998e)].

3.1.5 Absolute and Relative Organ Weights

It has been stated that the most efficient criteria for evaluation of the LOEL are changes in liver, kidney and body weights (Weil & McCollister, l963; Heywood, l981). Organ weights are usually reported as both absolute and relative weights (relative to body weight and/or brain weight).

Factors such as circadian rhythms, food intake, dehydration, diet, age of animals, organ workload, stress, and time and method of killing influence organ and body weights and the variability of such data. The review of this subject by Weil (1970) should be consulted. The most important factors appear to be the method of killing and the timing of necropsy. The killing method not only affects the appearance of the tissue, important in describing gross necropsy observations, but may also, in conjunction with the timing of necropsy, cause postmortem changes in organ weights (Boyd & Knight, 1963; Pfeiffer & Muller, 1967). A uniform exsanguination technique was described and evaluated by Kanerva *et al.* (1982) that significantly reduced ($P<0.05$) the absolute and relative liver and kidney weights compared with the same weights in animals that were not exsanguinated. The standard deviations of the mean absolute and relative liver weights were also significantly reduced ($P<0.05$). Exsanguination did not appear to affect the absolute or relative weights, nor the standard deviations of the mean weights, of the heart, brain or spleen. In addition, the use of fasted animal body weights can

reduce the variability of organ weight/body weight ratios. Adkins *et al.* (1982) discuss the standardisation of the technique for determination of testes weights to reduce variability.

A not uncommon problem in the interpretation of study findings is the misinterpretation of changes in relative organ weight. For example, there is no sense in reporting an increase in relative brain weight if the test chemical is causing significant body weight loss or reduced body weight gain. Because the brain is spared under conditions that lead to reduced body weight, the relative brain weight obviously increases! Similarly, the relative weight of other organs may change as a function of changes in body weight rather than as a result of a specific compound effect. Useful tables showing the relationship of relative organ weights to various levels of reduced body weight (produced by dietary restriction) in rats can be found in Sharer (1977) and related references; some data are also given in Table 2. It must be borne in mind, therefore, that when growth is markedly affected in a toxicity experiment, alterations in organ weight/body weight ratios have to be expected as a physiological response to decreased nutrient intake; these changes must be differentiated from organ weight changes caused by primary toxic effects of the compound being tested.

Furthermore, the interpretation of organ weight changes must not be made solely on the determination of a statistically significant difference between the concurrent control value and a treatment group value. A proper evaluation will also include consideration of any correlation between organ weights (absolute and relative) and histopathological and metabolic/pharmacodynamic data. It is generally considered that histopathology is more sensitive for establishing the LOEL than organ or body weight changes. Changes in organ weights are therefore a more reliable indication of toxicity if there are accompanying histopathological abnormalities. Any such correlations should be discussed in the evaluation documentation. Appendix III contains a list of organs that should be weighed and/or examined histopathologically.

3.1.6 Post-mortem Observations

Pathologists have an important role in toxicology since they provide information on the differences in tissue and organ morphology that establish the presence or absence of lesions and whether there are dose–effect relationships in respect of such lesions. This information is critical in establishing the toxic and other effects of a substance. Zbinden (1976) discussed the role of the pathologist in some detail, as well as the use of semi-quantitative methods and more accurate morphometric methods for rating the severity of lesions. He cautioned that, even using such methods, care should be taken in evaluating tissue and organ pathology/lesions because of the lack, at that time, of generally and internationally accepted nomenclature in toxicological pathology. Haseman *et al.* (l984) also discussed the problems created by different nomenclature. Although much progress has been made in the standardisation of nomenclature, to minimise any difficulties in this area an experienced pathologist will describe each significant lesion type, at least once, in such detail that another competent pathologist can perceive a mental picture of the lesion and form a judgement as to its relevance to the histopathology induced by the chemical being tested.

To assist in the uniform description of pathologies, several initiatives have been undertaken:

- A series of *Guides for Toxicologic Pathology* has been published by the US Society of Toxicologic Pathologists (STP), in cooperation with the Armed Forces Institute of Pathology (AFIP) and the American Registry of Pathology (ARP), introducing the *Standardized System of Nomenclature and Diagnostic Criteria* used by toxicologic pathologists around the world. These monographs are used to diagnose proliferative and non-proliferative lesions in laboratory animals. Arranged by organ system, the guides

contain morphologic descriptions and colour photomicrographs of spontaneous and induced lesions seen in evaluations of safety and efficacy in laboratory animals. Available monographs include guides to non-proliferative lesions in rats in bone, cartilage, tooth and synovium; the alimentary canal; soft tissue and skeletal muscle; the central nervous system; and kidney and the lower urinary tract. Further details, including information about additional guides as they become available, may be found at: http://www3.afip.org/cgi-bin/bookstore.cgi.

- The Registry Nomenclature Information system (RENI), developed at the Fraunhofer Institute of Toxicology and Aerosol Research. This comprehensive electronic system presents the IARC's *International Classification of Rodent Tumours* (also available as a CD-ROM and in hard copy format as IARC Scientific Publication 122), together with additional information (Mohr & Moraweitz, 1995). The use of RENI is specifically required by the EC (see Annex IIA, VI. 6.7 of the TNG of Directive 98/8/EC). RENI can be accessed on-line at: http://www.ita.fhg.de/reni.

- The Registry of Industrial Toxicology Animal-data (RITA), also maintained by the Fraunhofer Institute, containing validated data on tumours and other proliferative lesions from over 11 000 control mice and rats and using standardised nomenclature and diagnostic criteria (Moraweitz *et al.*, 1992; Mohr & Moraweitz, 1995; Bahnemann *et al*., 1995; Mohr, 1999). Further information about RITA can be obtained from the RENI Web site above.

- The North American Control Animal Database (NACAD), which has the same purpose, structure and diagnostic criteria as RITA, but uses a slightly modified nomenclature aimed at amalgamating the STP and RITA systems. See:

- http://www.ita.fhg.de/reni/nacad_d.html.

- International Harmonization of Rat Nomenclature, which is a project intended to reconcile the various nomenclatures of proliferative lesions in rats. This can be accessed at the Fraunhofer ITA Web server: http://www.ita.fhg.de/reni/rat_nomenclature/index.htm, or the STP Web site: http://www.toxpath.org/nomen/index.htm.

- Tumour pathology, diagnostic criteria and classification in laboratory animals is covered extensively in two series of publications by the IARC: *International Classification of Rodent Tumours, Part 1. The Rat* (IARC, 1992–1997); and *Pathology of Tumours in Laboratory Animals*, second edition (1990, 1994 & 1996).

While it is highly important that the study pathologist uses standardised diagnostic criteria, the reliability of diagnosis is greatly enhanced by comprehensive quality assurance and peer review. Some laboratories, such as the US National Toxicology Program, institute these procedures to ensure diagnostic criteria are applied consistently. This subject is discussed further by Ward *et al.* (1995), Boorman *et al.* (1985 & 1986) and Hardisty & Boorman (1986).

Pathology data can facilitate the interpretation of other data, such as organ weight changes or haematology findings (*e.g.* Krinke *et al.*, 1991), and evaluators should always make it clear if there are any associations between pathological abnormalities and other findings of physiological significance. Nevertheless, not all changes in tissue morphology are accompanied by abnormalities in the haematology, serum/urine biochemistry or other measured parameters. And although the test chemical may cause significant perturbation in organ biochemistry (*e.g.* in liver or kidney function), there will not necessarily be accompanying changes in the histological appearance of the affected organ(s).

Age-related changes, especially those associated with the geriatric stage, can have an extremely important effect on histopathology, as well as on clinical chemistry, metabolic and pharmacokinetic parameters (Grice & Burek, 1983; Mohr *et al.*, 1992, 1994 & 1996; Capen *et al.*, 2001). They may therefore have overt, and frequently subtle, influences on observed physiological and toxic responses during the latter part of any long-term study. As indicated earlier, spontaneous degenerative lesions, especially when misinterpreted as toxic effects, can cause major difficulty in hazard evaluation. It is essential to differentiate between spontaneous and/or age-associated lesions and lesions induced by treatment. References such as Grice & Burek (1983) and Benirschke *et al.* (1978), which provide detailed descriptions of histopathological changes that can be caused by toxic substances, or by spontaneous, degenerative or other diseases, and their incidences in experimental animals, are very helpful in this respect, as is advice from a competent and experienced pathologist.

Sometimes the test chemical may simply elicit an increased incidence of age-associated lesions, which can be detected by inter-group comparison even if they are not morphologically distinguishable from spontaneous lesions. Alternatively, the test chemical may increase the severity of ageing lesions, even in the absence of any marked effect on their incidence. Furthermore, the test chemical may accelerate the development of ageing lesions, which could become apparent at interim sacrifice or among premature decedents.

Paradoxically, under some circumstances the test chemical may have the opposite effect and *decrease* the incidence of age-associated lesions, *e.g.* where life span is shortened; fewer treated than control animals survive long enough to develop lesions of old age. Given the nexus between body weight gain, reproductive senility and age-related disease, a test chemical that reduces body weight gain could also protect against some of the conditions commonly found in obese, geriatric rodents.

An overview of factors, including the physiological and environmental, that can complicate the interpretation of findings in a toxicity study may be found in the *Handbook of Toxicologic Pathology* (Bucci, 1991).

3.2 Toxicokinetic and Metabolism Data

Toxicokinetic (absorption, distribution and elimination) and metabolism data on the handling of a substance in a test species can be very useful in the evaluation and interpretation of data from a chronic exposure study, as discussed by Paynter (1984) and references cited therein. References in this paper also discuss dose-dependent effects in the absorption process and in biotransformation interactions (Levy, 1968), the potential difficulties presented by impurities, the overloading of detoxification mechanisms (Munro, 1977) and other important experimental considerations (Dayton & Sanders, 1983).

The following serves as an example of a correlation between toxicokinetics and toxicology findings. A pesticide produced a particular target organ pathology in a repeat-dose study which did not show any

significant dose–response association. Subsequent analysis of the toxicokinetic data showed that the test substance, which was highly lipophilic and metabolically very stable, reached saturation in the target organ after only a few doses, even at the lowest dose.

With respect to plasma levels of the test chemical in a toxicity study, it is important to note that in rats sex hormones have a marked influence on liver biotransformation (see *e.g.* Chhabra & Fouts, 1974). In general, males metabolise xenobiotics (as well as endogenous substrates) faster than females — something not generally observed in other species. Thus rat studies may reveal sex-related differences in plasma kinetics and in the clinical and toxicological effects of a test chemical that are not relevant to human exposure.

A number of toxicology textbooks include chapters on pharmacokinetics and toxicology assessment, *e.g.* Sharma & Coulombe (1996). *Science and Judgement in Risk Assessment (*National Academy of Sciences/National Research Council, 1994), has useful sections on the impact of pharmacokinetic information in risk assessment.

3.3 Statistical Tests

It must be borne in mind that the objective of a toxicology study is to demonstrate responses of biological importance. Where statistical analyses are used to reach a judgement, an awareness of the validity of the tests employed and the degree of certainty (*i.e.* confidence) pertaining within the context of the study should be demonstrated.

The use of statistics in toxicology has limitations (Gad & Weil, 1986): (1) statistics cannot make poor data better; (2) statistical significance may not imply biological significance; (3) an effect that may have biological significance may not be statistically significant; (4) the lack of statistical significance does not prove safety. The importance and relevance of any effect observed in a study must be assessed within the limitations imposed by the study design and the species being studied. For example, dog studies use relatively few animals, limiting the power and reliability of statistical analysis. Group and individual data from dog studies therefore require careful examination to ensure the assessment report's conclusions are consistent with any biological effects of the test chemical.

Appendix IV lists some common statistical tests. If statistical tests either have not been used or appear inappropriate, unusual or unfamiliar, this should be noted in the assessment report (or other action taken, *e.g.* re-analysis of data by the sponsor or the assessor).

A number of textbooks and papers on the application of statistics in experimental toxicology and the life sciences are available; these include Dickens & Robinson (1996), Gad & Weil (1986), Gad & Weil (1989), Gart *et al.* (IARC, 1986), Lee (1993), Salsburg (1986), Tallarida & Murray (1987) and Waner (1992).

3.4 Completion of Analysis

By this stage an evaluator should have formulated judgements of a study and supporting rationale concerning: (a) the adequacy of the study design, including whether the dose selection was appropriate for measuring the test chemical's carcinogenicity potential; (b) the adequacy of the conduct of the study and the reliability of the data generated; (c) the existence of biologically important toxic effects; (d) the relevance of any factors noted during the evaluation which might have had some bearing on the

outcome of the study and modified the findings in some way; and (e) the likelihood that any of the observed effects were induced by the administered substance.

The evaluator should succinctly summarise the critical toxicokinetic and toxicological data, together with any modifying factors for the study under review. He/she should state the lowest, or most appropriate, NOEL or NOAEL, or the absence thereof (see Subsection 1.2.7, 'Acceptable Daily Intake for Humans'), clearly indicating the effect(s) on which it is based (*i.e.* the LOEL should be apparent). It is important to correlate findings from different studies; while this is done in the final summary of all toxicity studies, it is often appropriate to mention cross-study correlations (or the unexpected or unexplained absence of them) in individual study summaries. Possible or proven mechanisms of toxicity should also be discussed and included in the final summary.

4. EVALUATION OF THE WEIGHT OF EVIDENCE

The essential purpose of a chronic toxicity study is the detection of biological evidence of the toxic and/or oncogenic potential of the substance being investigated. The evaluation of the weight of evidence[8] produced by toxicity studies is the process of considering the cumulative data so as to arrive at a level of concern about the potential adverse effects of a substance. It is composed of a series of judgements concerning the adequacy, validity and appropriateness of the methods used to produce the database (i.e. the relevance and reliability of studies, as discussed in Part 2, 'Documentation and Data Acceptance'), and judgements that bring all the data considered into causal, complementary, parallel or reciprocal relationships. Because (a) knowledge of toxicity mechanisms is still developing, (b) good epidemiological evidence is seldom available, and (c) animal studies are not always conclusive, the information available at a given time may provide only "persuasive" rather than "hard" evidence, one way or the other, about the potential health effects of a substance under given exposure conditions. It is therefore necessary to succinctly discuss the rationale for judgements and conclusions contained in a risk assessment, together with any associated uncertainties. This becomes important when new data or new scientific knowledge require(s) re-evaluation of the database or a change in a previous risk assessment or regulatory action.

At present, there is no acceptable substitute for informed judgement, based on sound scientific principles, in the analysis, evaluation, interpretation and weighting of biological and toxicological data derived from animal toxicity studies conducted according to currently available protocols.

While the prediction of human carcinogenic hazard is a separate subject beyond the scope of this document, it is worth observing that a carcinogenic response in experimental animals is more significant for human health if it occurs in more than one species and/or in both sexes. The formation of tumours at several sites is viewed with greater concern than tumour formation at a single site. A carcinogenic response confined to one species assumes greater human significance if it is seen in two or more studies conducted at different times, in different laboratories or under different protocols. Animal carcinogens that are genotoxic, or structurally similar to known human carcinogens, also assume greater significance. A carcinogen that increases the incidence of a neoplastic disease that is rare in the test species or strain is of greater concern than a carcinogen that increases the incidence of a neoplasm having a high spontaneous incidence. At first, it may appear logical that a carcinogen that increases only benign tumours in experimental animals is of lesser significance to human health than a test chemical that causes malignancies. However, it should never be assumed that an agent that causes benign tumours in animals will not cause malignancy in humans. In any case, benign tumours are potentially serious, even lethal, diseases, depending on their size, growth rate and site of origin.

There are also factors that reduce concern about the human relevance of a carcinogenic response in animals; for instance, if the metabolism and toxicokinetic behaviour of a chemical in humans is fundamentally different from its behaviour in the species in which it is carcinogenic, or if the animal

[8] As opposed to the "strength of evidence", commonly taken to mean the degree of conviction regarding the outcome of a particular experiment, *e.g.* as expressed by the US National Toxicology Program's "clear evidence", "some evidence", "equivocal evidence" and "no evidence" of carcinogenicity. "Weight of evidence" involves integration of **all** available data, not just those from one study.

study employs an inappropriate route of administration, or demonstrates carcinogenic activity only at doses that cause excessive toxicity.

In addition to identifying toxic effects, and the doses at which they do or do not occur, chronic exposure studies may yield insight on the mode of action by which toxicity is mediated. (The *mode* of action is a general description of the key events and processes, starting from interaction between the test chemical and the cell and ending with the full manifestation of a toxic response. The *mechanism* of action, by contrast, is a detailed description of these key events and processes at the molecular level.) An evaluator should comment on this aspect of a study to the extent possible, correlating pathological findings with any effects on the various urological, biochemical, haematological or other parameters measured. Given that chronic exposure studies seldom exist in isolation, the evaluator may also be able to refer to other data (especially metabolism and toxicokinetic data; acute, short-term, repeat-dose data; and data from subchronic and genotoxicity studies), and should draw on his/her wider knowledge of disease processes and/or experience already gained with similar chemicals.

However, the assessment report may have been prepared before the evaluator has assessed other studies that contain significant information; or, as is the practise in some agencies, the workload may be divided among several individuals, none of whom sees the entire database. For these reasons, analysis of the mode and mechanism of action is best left to a later stage in the evaluation process, usually the discussion of the entire body of knowledge of the test compound.

It is at this point that the IPCS's Conceptual Framework for Cancer Risk Assessment (see footnote 1 and Appendix VIII) should be applied. The framework is an analytical tool that provides a logical, structured approach to the assessment of the overall weight of evidence for a postulated mode of carcinogenic action. Its use should increase the transparency of the analysis by ensuring the facts and reasoning, including any inconsistencies and uncertainties in the data, are clearly documented. After describing the **key events** on the path to cancer that have been measured, the next stages in the analysis consider the **dose-response relationships** observed for the key events and tumours, and the **temporal association** between the key events and formation of tumours. Subsequently, the **strength, consistency and specificity of association** of tumour response with the key events is discussed, leading to examination of the **biological plausibility and coherence** of the postulated mode of action in light of what is known about carcinogenesis generally and for the case, specifically. **Alternative modes of action** are also discussed before arriving at a final **assessment of the postulated mode of action**. The framework also includes a section on **uncertainties, inconsistencies and data gaps** identified during the analysis. Following development of the framework at a workshop at the IARC, Lyon, France, in February 1999 (the draft report of which is available at http://www.ipcsharmonize.org/cancer/cancer2-0299.html), details have been published in the scientific literature (Sonich-Mullin *et al.*, 2001). For further assistance, evaluators should refer to the IPCS scheme for predicting and quantifying human carcinogenic and mutagenic hazard, and for indicating the probable mechanism of action of carcinogens (Ashby *et al.*, 1996).

The US EPA's *Proposed Guidelines for Carcinogen Risk Assessment* (US EPA, 1996) incorporates a similar "Framework for Evaluating (a) Postulated Carcinogenic Mode(s) of Action", modelled on the frameworks developed by Bradford Hill in the examination of cigarette smoking and cancer.

It is important to appreciate that, while these two framework analyses are performed with the entire available database (and are hence "above" the level of an assessment of any individual study), their usefulness is highly dependent on the amount of detail included in the assessment reports on which they are based.

Although the IPCS's conceptual framework has been developed to assist in the assessment of carcinogenic end-points, the principles on which it is based are broad; this should mean it can be used to analyse the modes of action of *non*-neoplastic diseases. Irrespective of the nature of the disease process, characterising the mode of action makes it easier to reach judgements about the human relevance of toxicological findings, the possible need for further data, risk quantification and appropriate regulatory standards for the chemical in question. For example, hazard communication through labelling and the publishing of data on the safety of materials is facilitated by the OECD's *Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances* (OECD, 1998g), which presents a "Harmonized System for the Classification of Chemicals which Cause Cancer", based on the strength and weight of evidence criteria discussed above.

5. THE EVALUATION REPORT: STRUCTURE AND FORMAT

This part should be read in conjunction with OECD Test Guidelines 451–453 (OECD, 1981a, b, c) and the OECD guidance documents for country data review reports and industry data submissions (OECD, 1998a, b).

Where there are a number of studies in different species within a study classification, reports should be grouped by species, preferably in order of increasing species size, in the summary section as well as in the main body of the report.

5.1 Study Identification

An evaluation of a toxicity study should include the following information so the study can be clearly identified if it is referred to or resubmitted by the sponsor company at a later date, or submitted by another company. The information can be incorporated into the heading and the first paragraph of the evaluation.

a.	Title of study (should identify study species, dose route and study duration)
b.	Report/study number
c.	Laboratory report/project number
d.	Study sponsor (usually the registrant)
e.	Testing laboratory and brief address
f.	Authors' names (if available/appropriate)
g.	Date of report
h.	Period over which the study was conducted
i.	Test guidelines/protocol followed
j.	GLP status (or QA statement) and relevant authority
k.	Indication of whether the report is published or unpublished

Items a–g and k should also be included in the bibliography to the evaluation report/monograph.

5.2 Level of Study Reporting

Both the methodology and the study findings should be presented in sufficient detail for a reader to form an independent conclusion. Ideally, a report obviates the need, during a subsequent review of the test chemical, to refer back to the original study data. Application of the IPCS's Conceptual Framework for Cancer Risk Assessment is aided greatly by tabular summaries of comparative data on the incidence, or severity, of intermediate end-points (so-called "key events") and tumours. Successful use of the framework is also heavily reliant on adequate documentation of a) dose–response relationships in respect of tumours and associated end-points, and b) the temporal association between the key events and the eventual formation of tumours.

Note: Until OECD Member countries have had the opportunity to exchange a number of reviews prepared according to these guidelines, evaluators, if in doubt as to how much detail or explanation of findings to report should include more, rather than less, information.

The importance of assessing whether observed changes or differences are treatment related must be stressed: if they are clearly unrelated to the test compound, coincidental findings should not be mentioned. However, if there is any concern that an effect could possibly be related to dosing, it should be mentioned, with a comment about the lack of a dose–response relationship or other

unequivocal evidence. Tabulation of the data in question is useful in this situation: it enables a peer reviewer to determine a level of concern without returning to the raw data in the original study report.

Within the body of the evaluation report the following (minimum) information should be recorded:

1. A brief statement of the objective of the test or study, including any special or unusual reason for conducting it.

2. The identity (including batch number) and purity of the test material, including its common (*i.e.* generic) name. If the report is being prepared at an early stage in the development of a chemical, a common, or generic, name may not be available. If so, the manufacturer's code or chemical name should be given. The composition of the test chemical should be stated in a confidential data annex.

Note: The chemical names of the compound (IUPAC, CAS and common names), as well as any synonyms, the CAS number, company code names and numbers, any trade names, the empirical formula, the structural formula, the molecular weight and all available physicochemical data should all be given at some point in the evaluation report. Information about identified impurities, isomer ratios and the stability of the pure compound should also be included.

3. Details of the composition of solvents or dosing excipients, or, in the case of compounds included in the diet, the trade name and manufacturer of the diet together with a brief description of the diet preparation (including information on any vehicle used and the frequency of preparation). Analyses of the stability, homogeneity and concentration of the compound in the dietary admix should also be reported. While mixing, determination of the stability of the test material and mixtures, storage conditions and administration to the animals are covered by Standard Operating Procedures (SOPs) and quality control procedures (see *e.g.* EHC 141), information on these matters is important in determining compound intake over the course of a study and should be briefly reported. If it appears there may have been problems with any of these procedures, they should be reported in more detail, with a discussion of how the veracity or conclusions of the study might be affected.

4. The species, strain, source, and initial age and body weight of the test animal used. (Given the significant influence of genetic background on growth, baseline physiological or biochemical parameters and disease development, information about strain and source might be needed to compare study groups with historical controls.)

5. The number of animals per sex and per group, as well as numbers of animals in any additional subgroups or recovery groups.

6 In the case of a non-GLP study, or if the evaluator has identified possible problems or deviations in protocol, information on the housing, environmental conditions and acclimation period of the study animals should be included. In a GLP study these issues should be covered by SOPs and quality control procedures (see *e.g.* EHC 141), so minimal detail is appropriate (see sample assessments in Appendices VI and VII).

7. The dosing route and doses used (including the vehicle employed for negative controls), together with the rationale for the dose selection (which could summarise the results of range-finding or subchronic studies, or refer to these if they appear elsewhere in the evaluation report). Dietary levels, in ppm, and measured or estimated daily intakes, in mg/kg bw/d, should also be reported. Where no calculation based on food intake has been made, the conversion factors in Table 1

should be used. As outlined in IPCS Environmental Health Criteria Monograph No. 104 (WHO, 1990), if dietary intake is measured, JMPR evaluations indicate that X ppm in the food is **equal** to Y mg/kg bw, but if there are inadequate food intake data and the tabulated conversion factors are used, it is reported that X ppm in the food is **equivalent** to Y mg/kg bw.

Details of dosing methods, especially in dermal and inhalational studies, should be recorded. In the case of dermal studies, the application procedure, including the site of treatment, the manner in which the skin was treated (shaved, abraded or non-abraded), whether the site was occluded (and, if it was, the method of occlusion), and the size of the area treated, should be reported. In the case of inhalational studies (see OECD Test Guidelines 451–453), the following information should be recorded: (1) the methods used for generating the test atmosphere and a description of the test chamber, including whether whole-body or nose-only exposure was used; (2) the time to equilibration of the test atmosphere; (3) the test atmosphere concentration; (4) the determination, distribution and consistency of particle size over the course of the study. The last of these is particularly important, since if the majority of particles were larger than the respirable limit, exposure via the inhalational route will have been inadequate. The interim criterion of the US EPA is that at least 25% of particles should measure less than 1 μm (Whalan & Redden, 1994). Whalan & Redden provide much other useful information on the inhalation of particulate material in toxicology studies. Yet further information can be found in Chapter 5 of the *CRC Handbook of Toxicology* (Derelanko & Hollinger, 1995).

8. The duration and timing of dosing.

9. Sacrifice times.

10. The observations made, the parameters measured, and the frequency with which these operations were carried out.

11. The treatment-related effects on:

– mortalities	(with examination for cause of death)
– gross observations of behaviour and appearance	("clinical signs")
– food and water consumption	(Water consumption is not specifically requested under OECD Test Guidelines 451–453, but is essential for estimating intake of the test chemical if it is administered in drinking water.)
– body weights/body weight changes	
– functional investigations	(*e.g.* ECG, motor activity, neurological tests, ophthalmology, blood coagulation)
– blood biochemistry (*)	
– haematology (*)	
– urinalyses (*)	
– serum chemical concentrations	(if measured*)
– macroscopic pathology	
– organ weights	(absolute and relative)
– microscopic histopathology	
– any other special investigations	(*e.g.* hepatic cytochrome P-450 levels)

(* Depending on the regulatory guideline, at intervals during the study and at term)

12. The percentage (and absolute) change relative to controls, the dose- relatedness of changes, the biological and statistical significance of findings, and the suspected mode of action. Where parameters have been measured repeatedly throughout the study, it is advisable to indicate whether there are consistent time-related trends in any deviation from control values, or whether such variation is transient.

13. It may be necessary to compare results from the treated and control groups with historical control means or ranges. This can be particularly important in judging the biological significance of rare or unusual tumours and non-neoplastic abnormalities, and in cases where the concurrent control group displayed results that were atypical for untreated animals of the same strain, age, etc. It should be borne in mind that among the factors that influence the reported incidence of spontaneous lesions are diet, genetic background, housing conditions, age, and the techniques used to prepare and examine biological tissue. Even though adoption of harmonised GLP and test guidelines should be reducing the differences between laboratories, it remains highly desirable to use historical control data generated with the same strain of the test species as used in the study under review, at the same study laboratory, and no more than two to three years before or after the study under comparison. (The US FDA permits a five-year span within which the study under evaluation can fall anywhere.) Information should be provided on the source of the historical data and how closely or otherwise it matches the study being evaluated. The EC has formal requirements regarding the submission of historical control data, described in Annex II, V. 5.5 of Directive 91/414/EEC. The relative weight given to concurrent and historical controls will depend on the circumstances, and should be made clear to the reader. Appendix V lists a selection of sources of historical control data for laboratory animals.[9]

14. Tabulation of data is useful, and incidences of findings should be given in sufficient detail to allow independent assessment purely on the basis of the report. All parameters that may have been affected by treatment should be included. As an aid to interpretation, the severity of histological abnormalities should be described, while it is helpful if incidence data for benign and malignant neoplasms of the same histiogenic origin found at the same site are presented together with incidence data for related hyperplastic or other pre-neoplastic lesions. Any narrative accompanying such tabular data should address the toxicological significance of the results, not repeat what is presented in the table(s).

15. The statistical method used to evaluate each parameter. Appendix IV lists many of the statistical tests used in toxicology studies. If there is a paucity of statistical testing, or if tests appear inappropriate, the assessor may consider it necessary to comment in the report, conduct a re-analysis or request the study sponsor, or data owner, to provide further comment and/or analysis. Evaluators should be aware that low or differential survival may have an important bearing on the interpretation of carcinogenicity data. Low survival will compromise the statistical power of a study. An unequal reduction of the number of "at risk" animals can lead to underestimation of the test compound's oncogenic potential (Peto *et al.*, 1980; Portier & Bailer, 1989).

16. If possible, the identity and a discussion of the target organs and mechanism(s) of action. The NOEL or NOAEL for each study (as appropriate) should be recorded, with a statement on the toxicological end-points upon which it has been set (so that the LOEL is clear). An evaluator should determine the NOEL or NOAEL on the basis of his/her own assessment of the data, not that of the company scientist(s); however, if the NOEL or NOAEL of the company scientist(s) differs, it should be given, with an explanation of the assessment on which it is based.

17. Comment on the adequacy of the study, including the suitability of the dose selection for establishing carcinogenicity and other toxicological end-points. Any deficiencies should be discussed in detail and comment made on the regulatory relevance of the study. If there is no GLP certification, the evaluator should at least note whether the study was inspected by a Quality Assurance Unit (there should be a signed QA statement) and make some comment on the apparent

[9] One means of obtaining a historical control database is to compile relevant material supplied by sponsoring companies as part of study reports.

quality of the protocol and the adequacy of the methods used. Reasons for rejecting a study as part of a regulatory package should be clearly stated; for regulatory purposes, some countries also append a one-word descriptor indicating whether a study is acceptable or not.

Example headings and first paragraphs follow. Superscript letters refer to the points listed in Section 5.1, superscript numbers to the points listed directly above. Much of the information in points 1–8 can be incorporated into the first paragraph, either as text (upper example) or in point form (lower example).

Burke ED & Wills HO[f] (1985): Chronic (52-week) Dietary Toxicity Study in Dogs[a]. Project No. 2174-112[c] Report no. 5638[b] Lab: Burton Labs Inc, Rocky River, NJ USA[e] Sponsor: Pesticide Corp, Research Triangle Park, NC USA[d] Expt Date: 15 Sept 1983 to 20 Sept 1984[h] Unpublished[k] Report Date: 26 Jan 1985[g] (QA: Yes GLP: US EPA and Japan MAFF)[j]

Purebred beagle dogs obtained from AnimaLabs Inc., NY USA[4] (5/sex/group, initially aged 3 months and weighing 3.9 to 4.8 kg)[5] were dosed with chlortoxane[2] (Pesticide Corp, Rayleigh NC USA; Batch no. #34-CD; 98.9% purity) in the feed[6] at levels of 0, 200, 2500 and 50000 ppm for 52 wk[7]. Mean compound consumption[6] ranged between 6.9 to 9.6, 80 to 115 and 1698 to 2494 mg/kg bw/d (both sexes). The study was conducted in accord with OECD Test Guideline No. 452[i 1].

Burke ED & Wills HO[f] (1985): Chronic (52-week) Dietary Toxicity Study in Dogs[a]. Project No. 2174-112[c] Report no. 5638[b] Unpublished[k] Report Date: 26 Jan 1985[g]

Test Chemical:	Chlortoxane[2] (Pesticide Corp, Rayleigh NC USA)
Batch no:	#34-CD
Purity: 98.9%	
Test Species:	Purebred beagle dogs (AnimalLabs, NY USA[4], 5/sex/group, initially aged 3 months and weighing 3.9 to 4.8 kg)[5]
Dose:	In feed[6] at levels of 0, 200, 2500 and 50000 ppm for 52 wk [equal to between 6.9 to 9.6, 80 to 115 and 1698 to 2494 mg/kg bw/d (both sexes)][7]
Lab:	Burton Labs Inc, Rocky River, NJ USA[e]
Sponsor:	Pesticide Corp, Research Triangle Park, NC USA[d]
Study Duration:	15 Sept 1983 to 20 Sept 1984[h]
QA:	Yes
GLP:	Yes (US EPA and Japan MAFF)[j]
Guidelines:	***OECD Test Guideline No. 452[i 1]***

5.3 Layout and Formatting

It is not the intention here to specify a standard format and layout for reports, since the needs of national agencies differ as regards presentation of data to advisory or other committees and, to a lesser extent, the styles and preferences of individual evaluators. **However, a report should be structured to allow ready access to all significant and relevant points arising from the assessment.** The

OECD's *Guidance for Country Data Review Reports on Plant Protection Products and their Active Substances* (OECD, 1998b) should be consulted for guidance on general report organisation.

An evaluation of a chronic study should include comment on the effects of treatment on mortality and morbidity, clinical signs, food and water consumption, body weights, ophthalmoscopy, clinical chemistry, haematology, urinalysis, organ weights, gross and microscopic histopathology (non-neoplastic, pre-neoplastic and neoplastic) and any other measured parameters.

It is hard to locate specific data in a large block of solid text. Study findings are therefore best sorted into paragraphs, e.g. Mortality/Clinical Signs; Body Weights; Food Consumption; Ophthalmoscopic Findings; Haematology; Clinical Chemistry; Urinalysis; Organ Weights; Gross Pathology; Histopathology (non-neoplastic, pre-neoplatic and neoplastic); Conclusions; plus paragraphs on any extra investigations (e.g. Neurobehavioural Assessment) as necessary. Some findings can be presented in the same paragraph (e.g. Body Weights and Food Consumption; Haematology, Clinical Chemistry and Urinalysis; Gross and Microscopic Pathology), particularly if a lack of effects means individual paragraphs may only be a sentence or two long.

Alternatively, if toxicity is mainly confined to a specific organ or system and causes a range of related gross histological or biochemical, haematological and urological effects, the evaluator may find it advantageous to group these findings rather than separate them into individual end-points.

Sub-headings before the above paragraphs will improve clarity, particularly in the evaluation of studies in which there are a large number of positive findings that need to be reported at length.

Parameters in some of the above areas of investigation often do not change; such negative results should be reported only briefly, to indicate measurements were performed. An alternative approach is to list, at the beginning of the study evaluation, all the types of observation that were conducted, and then report on only the compound-related (or possibly compound-related) findings. Sample evaluations are provided in Appendix VI and VII.

In an evaluation report on a package of toxicology studies, chronic toxicity studies should be grouped by route of administration and presented by species (in the order mouse, rat, rabbit, dog, monkey). Within each species, studies should be arranged in order of increasing duration.

5.4 Terminology Used in Evaluation Reports

To avoid losing important aspects of an independent assessment in an unnecessarily long document, evaluation reports should be as concise and precise as possible, consistent with adequate reporting as outlined above. Abbreviations are acceptable provided any not in widespread use are clearly defined. Nevertheless, excessive use of abbreviations can make a document tedious, and sometimes difficult, to read, especially if reader and writer speak different native languages and translation makes abbreviations less clear.

Appendix I lists commonly accepted abbreviations of terms used in toxicology studies. (Reference is also made to a more detailed set of abbreviations in the relevant appendix of the OECD/EU guidance documentation on the preparation of dossiers and monographs.)

With respect to abbreviated terminology, the term "clinical chemistry/haematology/urinalysis parameters" is taken to refer to those measurements listed in Appendix II. If there were no changes in any of the parameters in this particular test battery, it is sufficient to state that "there were no changes

in any clinical chemistry/haematology/urinalysis parameters", without specifying all the individual parameters measured. If one or several of the parameters listed were not included in the test battery, such omissions should be noted with, if necessary, a statement on the adequacy of the parameters assessed.

While reporting and interpreting all relevant compound-related findings in sufficient detail, evaluation reports need to be as succinct as possible. Useful shorthand expressions include:

- *n*/sex/group = number of animals per sex in each dosage group
- dose-related = effects of the compound were dose-related
- compound-related = effects were compound-related but *not* necessarily dose-related
- po, iv, ip, sc, im = oral, intravenous, intraperitoneal, subcutaneous, intramuscular
- *x*/*y* = *x* animals affected out of *y* animals examined, *e.g.* incidences of the finding in respective groups were 1/10, 2/8, 3/7 and 5/10.

Standard abbreviations for haematology and clinical chemistry parameters may also be used.

5.5 Bibliographic Citations

The following format should be used for the citation of company data and data for which the standard information is unavailable or unclear. The objective is to provide a unique identifier for each study. Content and order are the key points to note; minor changes in the format may be made, depending on individual country preferences.

Company Data

Author:	Surname, Initial
Date:	Year report written (not year submitted)
Title:	Full title as it appears in the report
Testing Laboratory:	(where different from Company Name)
Report Number:	(and full date of the report, which is useful as a study identifier)
Company Name, City and/or Country:	(data submitter/data owner)
Unpublished:	(if an unpublished report)
(Country Code):	(individual country/agency identifier, if applicable)

Examples

Hartley M & Murray W (1994): S-1234 (Technical-grade) twenty-one day dermal study in rabbits. Contract Labs, London, United Kingdom. Report No. 007. Dated 13 December 1994. Pesticide Company, Bilthoven, Netherlands. Unpublished. (Country/Agency identifier)

Ebert M & Leist A (1985): 21-Day dermal study in Wistar rats. Report No. 84.0223, Dated 2 February 1985. Hoechst AG, Germany. Unpublished. (Country/Agency identifier)

Default entries

If standard citation information is unavailable or unclear, the following default entries should be used. If authors are not identified, use the name of the submitting company.

Pesticide Company (1994): S-1234 (Technical-grade) twenty-one day dermal study in rabbits. Contract Labs, London, United Kingdom. Report No. 007, Dated 13 December 1994. Pesticide

Company, Bilthoven, Netherlands. Unpublished. (Country/Agency identifier)

Another possibility if authors are unidentified is to name the study director.

If the report number is not specified, say so. If another identifier is used (e.g. the study number), state that.

Hartley M & Murray W (1994): S-1234 (Technical-grade) twenty-one day dermal study in rabbits. Contract Labs, London, United Kingdom. Report No. not specified, Study No. 2468, Dated 13 December 1994. Pesticide Company, Bilthoven, Netherlands. Unpublished. (Country/Agency identifier)

Note: In the main body of an evaluation report, the headings for individual studies may contain more detail than the above citation for an unpublished report, *i.e.* headings should include all the above information plus the start and end dates of the experimental phase of the study, study numbers and/or any other report identifiers.

Literature References

It is suggested that references be detailed as follows: author, year, title, journal, volume, and pagination.

White D, Ruehl KJ, Borman SA & Little J (1988) Effects of methylmercury on the microtubule system of mouse lymphocytes. Toxicol Appl Pharmacol **94**(1): 66-75.

This citation style is used by the IPCS in its EHC monographs and is consistent with recommendations to reduce keystrokes by eliminating unnecessary full stops, commas, etc. Names of journals should be abbreviated according to the International Serials Data System (ISDS) list of Serial Title Word Abbreviations, or given in full.

5.6 General Comments

Part 3 of this document contains detailed comments on the analysis and evaluation of toxicology studies. A number of further, more general, comments follow.

If possible, compound-related changes in biochemical, haematological and urinalysis parameters should be linked to changes in organ weight, gross pathology and/or histopathology. A number of reference books are useful in this regard, *e.g. Organ Function Tests in Toxicology Evaluation* (Tyson & Sawhney, 1985).

The following points should also be noted in the evaluation of chronic toxicity data:

- Findings should be considered on the basis of both statistical significance and likely biological significance. It is important to bear in mind the variability of biological data when assessing a statistically significant result: statistical significance does not necessarily equate to biological significance. Conversely, a finding that is not statistically significant may have biological significance when considered in the light of the likely toxicological or pharmacological action of the test compound, or when considered alongside results from other studies. Evaluators should therefore report trends or transient changes in parameters if there is an indication that these may be related to dosing with the compound in some way (see Part 4 for more detailed

discussion). This information may be useful in cross-study comparisons of results and in the consideration of the overall significance or relevance of an observed effect, *e.g.* in one study an effect may be only a trend while in another it may be clearly treatment-related.

- A particularly vexing problem for evaluators arises when studies that produce either clearly positive or clearly negative results have to be considered flawed. There may be questionable components in any long-term study, and toxicologists must learn to recognise what is useful and discard what is not (see Part 2 for more detailed discussion). The use of a seriously flawed negative study may provide only a false sense of security. On the other hand, a flawed positive study may be entitled to some weight; how much is a matter of judgement (Task Force of Past Presidents, 1982). Data obtained from studies carried out many years ago should not be dismissed out of hand simply because they do not meet today's standards; they may still provide useful information. Again, this is a matter for scientific interpretation and judgement on a case-by-case basis.

- In many cases issues arise that cannot be elucidated using the "standard" battery of toxicity tests. Evaluators should consider the need for any special studies, *e.g.* to investigate specific toxic effects in detail, such as the ocular toxicity of a new organophosphorus compound (this may include a comparative study of other chemicals of the same class already on the market), or to investigate effects on the immune system (immunosuppressant effects may help explain increased incidences of infections, mortalities and/or tumours).

- If an evaluator believes any such studies are necessary but they have not been provided, he/she should highlight this in the final summary and assessment; in some cases the sponsoring company may be able to provide the relevant information before completion of the assessment report. If an evaluator refers to data (*e.g.* historical control data) not included in the study report, he/she should list the source in the bibliography.

5.7 Interpretation of Carcinogenicity Data

Interpretation of carcinogenicity study results is profoundly affected by experimental design, strain and source of experimental animals, the purity and physical state of the test chemical, and exposure conditions (especially dose selection, vehicle and method of administration) (UK DoH, 1991). Inappropriate dose selection is a major shortcoming: some studies are not positive for carcinogenicity because of failure to administer a sufficient dose; in others, dose levels have a significant impact on survival. On the other hand, overt toxicity or inappropriate toxicokinetics as a result of excessive dosing may result in tumorigenesis that is secondary to toxicity rather than directly attributable to the agent (US EPA, 1996). Differences in food intake associated with unpalatability can influence tumour incidence: it may be difficult to separate the effect of the test chemical from the confounding influence of reduced consumption. Test substances that are insoluble in water may be dissolved in a vehicle to facilitate their incorporation into the diet or administration by gavage. Increased incidence of tumours may be associated with the vehicle, as in the case of pancreatic acinar cell adenomas in male F344 rats given corn oil. The vehicle and route of administration may also exert major effects on the toxicokinetics and metabolism of the test chemical (UK DoH, 1991).

The comprehensiveness of pathological examination and reporting also influences interpretation. If tumours were not detected by the investigators, or were not reported accurately or in sufficient detail, an evaluator can be misled as to the study outcome. Although current GLP and test guidelines do not guarantee high scientific standards, studies that comply with these are usually the most reliable and detailed sources of data. However, evaluators often have to assess studies that predate GLP and

modern test guidelines, or papers published in the open literature. It is essential that evidence from these sources is scrutinised carefully (UK DoH, 1991) and any limitations taken into account.

The overall pattern of the data commonly yields insight into whether tumours arose spontaneously or were induced by the test compound, and into the possible mode of action. Spontaneous tumours are more often benign and singular (per animal) than induced tumours, which are more commonly malignant and multiple (per animal). Where tumours are a result of cytotoxicity, there is often a continuum of effects in the target tissue, ranging from injury, regenerative activity, hyperplasia and pre-neoplastic lesions to benign and malignant tumours. There may be a similar continuum of hyperplastic, pre-neoplastic and neoplastic effects following exposure to carcinogens that act via the stimulus of cell division (mitogenesis). This is the case in the hormonal stimulation of the thyroid gland and exocrine pancreas, and the stimulation of smooth muscle by bronchodilating agents.

Often, but not invariably, there will be a range of hyperplastic, pre-neoplastic and neoplastic, findings within individual animals. Within the study population, a dose–response relationship may be evident with respect to the incidence and/or severity of cytotoxic, hyperplastic, pre-neoplastic or neoplastic lesions. If this is the case, pre-neoplastic and neoplastic lesions may be confined to the highest dose or doses, while at lower doses, cytotoxic and/or hyperplastic lesions may appear in the absence of tumour development. Where interim sacrifices have been performed, there may be evidence of a temporal association between a sequence of events and the eventual development of cancer. These events could include tissue injury, repair activity and development of hyperplastic or pre-neoplastic lesions. Withdrawal of treatment may reverse cytotoxic or hyperplastic lesions. Conversely, lesion development may be hastened by increasing doses.

Evaluators should therefore be particularly careful to search for underlying patterns in the experimental findings. If they detect any, they should emphasise the interrelationship of the various effects in their report.

The clearest indication of a positive carcinogenic response is obtained when the incidence of tumours rises above concurrent and historical control levels in both sexes and is higher at higher doses. Further significant observations in treated animals include an increase in rare types of tumour, metastases, reduced latency and the presence of tumours at multiple sites.

However, rodent carcinogenicity data may not be so clear-cut. The choice of doses may prevent a dose–response relationship or a NOEL for tumorigenicity being established, while some types of tumour are sex limited, particularly those arising from perturbation of endocrine hormone levels. In rodents, gender-related differences in metabolism of the test compound may also lead to sex-related differences in sensitivity to tumours, as in the case of thyroid tumours in rats and liver tumours in rats and mice. The incidence of some types of tumour may also increase with age, or with body weight differences between the treated groups and their concurrent controls (see the example in Subsection 1.2.4 under “Spontaneous carcinogenesis”). Examples of common rodent tumours that have well-characterised modes of formation are listed in Table 6.

In modern chronic toxicity and carcinogenicity studies, statistical analysis is usually performed by the study authors, and plays an important part in determining whether exposure to the test chemical is associated with increased tumour incidence. The US EPA’s *Proposed Guidelines for Carcinogen Risk Assessment* (1996) advises that:

"Statistical analysis should be performed for each tumour type separately. The incidence of benign and malignant lesions of the same cell type, usually within a single tissue or organ, are considered separately and are then combined when scientifically defensible (McConnell *et al.*, 1986). Trend tests

and pairwise comparison tests are the recommended tests for determining whether chance, rather than a treatment-related effect, is a plausible explanation for an apparent increase in tumour incidence. A trend test such as the Cochrane-Armitage test (Snedcor & Cochrane, 1967) asks whether the results in all dose groups together increase as the dose increases. A pairwise comparison test such as the Fisher exact test (Fisher, 1932) asks whether an incidence in one dose group is increased over the control group. By convention, for both tests a statistically significant comparison is one for which $p<0.05$ that the increased incidence is due to chance. Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result. A statistically significant response may or may not be biologically significant or vice versa. The selection of a significance level is a policy choice based on a trade-off between the risks of false positives and false negatives. A significance level of greater or less than 5% is examined to see if it confirms other scientific information. When the assessment departs from a simple 5% level, this should be highlighted in the risk characterisation. A two-tailed test or a one-tailed test may be used".

Two-tailed tests examine for the significance of differences from control in either direction, whereas a one-tailed test examines for departure in one specific direction only, and has greater power to detect a difference in that direction than does a two-tailed test (Rosner, 1995). If we wish to know whether the tumour incidence in a group of treated animals significantly *exceeds* that among controls, it is considered preferable to employ a one-tailed test. A two-tailed test is appropriate for analysing whether tumour incidence in the treated group is significantly *different* (*i.e.* greater *or* less) than among controls. Evaluators should comment upon the appropriateness of statistical methods used by the study authors, bearing in mind the potential effects of low or differential survival on statistical power and the likelihood of underestimating the test compound's oncogenic potential (Peto *et al.*, 1980; Portier & Bailer, 1989). For this reason, survival-adjusted methods are strongly preferred for comparison of tumour incidences (see Appendix IV).

Although statistical comparisons are of treated animals and concurrent controls, additional insights into the significance of tumours can be obtained from examination of historical control data (Tarone, 1982). Such data can add to an analysis particularly by enabling identification of uncommon types of tumour or the high spontaneous incidence of a tumour in a given animal strain. Caution should be exercised in simply comparing the test group response with the historical range, because the range data ignore differences in the survival of animals among studies and is related to the number of studies in the database (US EPA, 1996). The more studies in the database, the wider the range.

In analysing the results for uncommon tumours in a treated group that are not statistically significant in comparison to concurrent controls, the evaluator can use the experience of historical controls to conclude that the result was in fact unlikely to be due to chance. In the analysis of results for common tumours, a different set of considerations comes into play. Generally speaking, statistically significant increases in tumours should not be discounted simply because incidence rates in treated groups are within the range of historical controls, or because incidence rates in the concurrent controls are somewhat lower than average. Random assignment of animals to groups and proper statistical analysis should have ensured that statistically significant results are unlikely to have arisen by chance alone. However, caution should be exercised in interpreting results that are barely statistically significant, or in which incidence rates in concurrent controls are unusually low relative to historical controls (US EPA, 1996). As stated previously (Section 5.2, point 13), the most relevant historical data come from the same laboratory as, and from studies that used animals from the same supplier as, and were conducted within two to three years of, the study under review.

5.8 Interpretation of Carcinogenicity Studies other than Long-term Rodent Bioassays

Caution should be exercised in the interpretation of results from the various short-term *in vivo* protocols, including initiation–promotion studies, start/stop studies and transgenic mutagenesis assays (see Section 1.1). These are of comparatively short duration and may lack a complete histological examination, while tumours may not develop fully before study termination. Experimental manipulation of the carcinogenic process may limit the relevance of such studies to "real life" exposure scenarios. However, their results may assist the interpretation of data from carcinogenicity bioassays or other studies, particularly in regard to potential modes of action, or make available quantitative estimates of exposure, internal dose and mutation (Morrison & Ashby, 1994; Sisk *et al.*, 1994; Hayward *et al.*, 1995), (US EPA, 1996).

Analysis and interpretation of the various *in vitro* and *in vivo* genotoxicity studies, either alone or in combination, is beyond the scope of these guidance notes. Their results, however, often yield insight into the mode of action of a chemical carcinogen, particularly with regard to the its potential genotoxicity *in vivo*.

If an *in vivo* carcinogen is mutagenic, clastogenic or positive for other genotoxic end-points at a cellular level, its mode of action probably involves genotoxic activity. This conclusion would be reinforced if data were available showing at a molecular level that the chemical, its metabolite(s) or close structural analogues were capable of direct interaction with DNA or other chromosomal material.

If an *in vivo* carcinogen has yielded negative results in genotoxicity studies, this suggests that its mode of action in animals does not involve genotoxic activity. However, even *in vivo* genotoxicity studies may fail to duplicate the dose, duration or route of exposure used in carcinogenicity bioassays, and *in vitro* studies cannot reproduce the physiological environment found within a live animal. Hence, differences in metabolism, toxicokinetics, cellular susceptibility, excretion or other factors could cause a chemical to be carcinogenic *in situ* yet not demonstrate activity in genotoxicity studies. The degree of confidence an evaluator can have in such "negative evidence" will therefore vary from case to case, depending on the range of genotoxicity studies performed with the test chemical. [The definition of what constitutes "sufficient" weight of evidence in genotoxicity studies varies between regulatory agencies and schemes. At the absolute minimum, the genotoxicity database has to comprise data on gene mutation in bacteria and/or mammalian cells, and on clastogenicity in mammalian cells. For more detailed discussion, see Ashby *et al.* (1996).]

Sometimes, an *in vivo* carcinogen may not have been tested in a sufficient range of genotoxicity studies, one or more of the genotoxicity studies may have been technically inadequate, or the carcinogen may have yielded conflicting or equivocal results when tested for genotoxic activity. Under these circumstances, regulators should never exclude genotoxicity as a possible mode of action.

TABLE 1:

APPROXIMATE RELATION OF PARTS PER MILLION IN DIET TO MG/KG BODY WEIGHT/DAY*

Animal	Weight (kg)	Grams food consumed per day (liquids omitted)	Type of diet	1 ppm in food equivalent to, in mg/kg bw/d	1 mg/kg bw/d equivalent to, in ppm of diet
Mouse	0.02	3	Dry	0.150	7
Chick	0.40	50	laboratory	0.125	8
Rat, young	0.10	10	chow diets	0.100	10
Rat, older	0.40	20		0.050	20
Guinea pig	0.75	30		0.040	25
Rabbit	2.00	60		0.030	33
Dog	10.00	250		0.025	40
Cat	2	100	Moist	0.050	20
Monkey	5	250	semi-solid	0.050	20
Dog	10	750	diets	0.075	13
Man	60	1500		0.025	40
Pig or Sheep	60	2400	Relatively	0.040	25
Cattle, maintenance	60	7500	dry grain-	0.015	65
Cattle, fattening	500	15000	forage	0.030	33
Horse	500	10000	mixtures	0.020	50

* From Lehman (1954), as reproduced in IPCS Environmental Health Criteria Monograph No. 70 (WHO, 1987).

As outlined in IPCS Environmental Health Criteria Monograph No. 104 (WHO, 1990), if dietary intake is measured, JMPR evaluations indicate that X ppm in the food is **equal** to Y mg/kg bw, but if there is inadequate food intake data and the tabulated conversion factors are used, it is reported that X ppm in the food is **equivalent** to Y mg/kg bw.

TABLE 2:

EFFECT OF RESTRICTED FOOD CONSUMPTION IN RATS ON BODY WEIGHT, RELATIVE ORGAN WEIGHTS AND SOME LABORATORY PARAMETERS

Strain	**Ibm (RORO)**	**CD**		
Period of deprivation	4–13 weeks	14 days		
Wt at study commencement	ll0 g	l40–160 g		
Sex	male	female/male		
Food (% control)	65–38	83/62	53/42	30/25
Body wt (% control)	73–66	88/74	67/58	45/38
Relative Organ Weight (as % of control r.o.w.)*				
brain	127–140	112/130	143/148	202/230
stomach	–	118/162	140/162	157/188
lung	–	–	–	134/
kidney	nc	–/nc	108/nc	122/nc
submaxillary gland	–	–/123	116/126	157/188
thymus	–	–	126/82	19/25
thyroid	110–118	–	–	133/129
spleen	84–103	–	–/80	51/54
liver	64–77	90/75	78/68	64/60
uterus	–	–	76/–	67/–
large intestine	–	nc/nc	nc/nc	nc/nc
small intestine	–	nc/nc	nc/nc	nc/nc
heart	nc	nc/nc	nc/nc	nc/nc
adrenals	135–154	nc/nc	nc/nc	nc/nc
ovaries	–	nc/nc	nc/nc	nc/nc
pituitary	nc	–	–	–
testes	ll8–138	–/136	–/159	–/191
seminal vesicles/prostate	l04–71	–	–/72	–/32
Laboratory parameters (data, for males only, given as % of "control" values)				
AST	–	–	–	533
ALT	–	62	–	393
BUN	–	54	71	215
SAP	–	70	64	52
Hb/Hct	–	"increasing"	"increasing"	"increasing"
WBC	–	63	55	53

r.o.w. = "relative organ weight", *i.e.* organ weight expressed relative to body weight
– = "not determined"
nc = "no change"
Data for Ibm (RORO) strain taken from Sharer (l977) and for CD rats from Dairman (l978).

*Values are expressed as percentages of control r.o.w., *e.g.*the 127–140% value for brain reflects the sparing of brain in animals that have lost significant body weight (down to 66–73% of control animals). Similar trends are evident in stomach, submaxillary gland and testes.

TABLE 3:

SOME INFORMATION ON *RATTUS NORVEGICUS*

Parameter	Value	Parameter	Value
Adult wt (g) – male	200–500	Type of ovulation	spontaneous
– female	250–350	Fertilisation after ovulation (h)	7–10
Average life span (yr) – laboratory	2–3	No. of eggs shed	10+
– wild	4–5	Egg viability (h)	10–12
Approx. dietary consumption/d (g)	15–20	Gestation period (d)	21–22
Water consumption/d (mL)	24–35	Usual litter size	9–11 (range 6–15)
Approx. urine volume/d (mL)	11–15	Litter frequency/yr	7–9
Approx. faecal mass/d (g)	9–15	Weight at birth (g)	4–5
Body temperature (˚deg C)	37.3	Optimum weaning age (d) 21	(range 18–23)
Heart rate (/min) 300–375	(range 260-600)	Optimum weaning wt (g)	35–40
Ventilation rate (/min) 100	(range 66–210)	Menopause (mo)	15–18
Tidal volume (mL) 0.86	(range 0.60–1.25)	Chromosome no. (diploid)	42
O_2 consumption (mL/g/h)	2.0	Age at first oestrus (d)	36
Basal metab. rate (kcal/m^2/d) (300 g rat)	802	Age at first ovulatory oestrus cycle (d)	77 (range 45–147)
Arterial BP (mm Hg) –systolic	116		
–diastolic 90		Parturition length (h)	1–4
RBC count (million/mL)	7.0–9.7	Puberty (d) 50–60	
Hct (%)	46	Wt at maturity (g) – male	170–210
Cardiac output (mL/min)	50	– female	150–170
Blood vol./100g body wt	5.6–7.1	Recommended min. breeding age (wk)	12
Mean blood pH 7.40		(range 9–14)	
Plasma osmolarity (mOsm/kg)	288–336	Recommended max. breeding age (mo)	12–15
Lymphocytes – % of WBCs	86	Time of ovulation after oestrus (h)	8–11
Platelets (million/mm^3) 0.5–1(range 9–20)		Oestrus cycle (d)	4–5
Leucocytes (/μL) 9000 (range 6000–18000)		Oestrus (heat) duration (h)	12

Data from several authors collected in the *Whole Rat Catalogue*, l983, Harvard Bioscience, USA.

Note: Values will vary for different rat strains.

The following information may also be useful:

– Liver plasma flow in a rat is approximately 7 mL/min per 200 g body wt (Altman & Dittmer, l974).
– The first oestrous cycle in dams after delivery is 20–24 h postpartum.
– The tail of a rat may reach 85% of body length; it is longer in females than males.
– Hair growth is cyclic, with a resting and a growing phase lasting about l7 d each.
– Ossification is not complete until after the first year of life.

TABLE 4:

INFORMATION ON DOGS: SOME PHYSICAL AND PHYSIOLOGICAL PARAMETERS

Parameter	Value
Lifespan	12–14 yr
Adult weight	6–25 kg
Birth weight	300–500 g
Adult food consumption	250–1 200 g/d
Adult water consumption	100–400 mL/d
Breeding age (males)	9–12 mo
Breeding age (females)	10–12 mo
Oestrous cycle	biannual, monoestrus
Gestation period	56–58 d
Weaning age	6–8 wk
Litter size	4–8
Blood volume (adult)	8–9%, 75–110 mL/kg
Maximum safe bleed	8–10 mL/kg
Red cell count	5.5–8.5 x 10^6/mm^3
White cell count	6–14 x 10^3/mm^3
Haemoglobin	13–18 g/dL
Haematocrit	38–52%
Platelet count	200–600 x 10^3/mm^3
Heart rate	80–140 beats/min
Respiration rate	10–30 breaths/min
Rectal temperature	38.5°C
Urine pH	7.0–7.8
Urine volume	25–45 mL/kg
Chromosome number	2n=78

This table was adapted from Derelanko & Hollinger (1995); data sources are cited in that text. While there is a wide variety of breeds and strains of dogs, the pure-bred beagle is the most commonly used in toxicology studies, owing to its uniform and relatively small size, docile temperament, physiological similarity to humans and ability to adapt to life in large cages or pens.

Other useful metabolic, physiological and biochemical information on toxicology test species may be found in, among others, Siglin & Rutledge (1995) and Hollinger (1995).

TABLE 5:

INFORMATION ON MICE: SOME PHYSICAL AND PHYSIOLOGICAL PARAMETERS

Parameter	Value
Lifespan	1–2 yr
Male adult weight	20–35 g
Female adult weight	20–35 g
Birth weight	1.0–1.5 g
Adult food consumption	3–6 g/d
Adult water consumption	3–7 mL/d
Male breeding age/weight	6–8 wk/20–35 g
Female breeding age/weight	6–8 wk/20–30 g
Placentation	Discoidal endotheliochorial
Oestrous cycle	4–5 d, polyoestrus
Gestation period	19–21 d
Weaning age/weight	21 d/8–12 g
Litter size	10–12
Mating system(s)	1:1 or 1 male:multiple females
Adult blood volume	6–7% of bw
Maximum safe bleed	7–8 mL/kg
Red cell count	7–12 x 10^6/mm^3
White cell count	3–12 x 10^3/mm^3
Haemoglobin	13–17 g/dL
Haematocrit	40–54%
Mean corpuscular volume	43–54
Mean corpuscular haemoglobin	13–18
Mean corpuscular haemoglobin concentration	31–34
Platelet count	1000–1600 x 10^3/mm^3
Heart rate	300–600 beats/min
Respiration rate	90–180 breaths/min
Rectal temperature	37.5°C
Urine pH	6.0–7.5
Urine volume	1–3 mL/d
Chromosome number	2n=40

This table was adapted from Derelanko & Hollinger (1995); data sources are cited in that text.

TABLE 6:
SOME EXAMPLES OF COMMON RODENT TUMOURS AND THEIR MODES OF FORMATION

Tumour site/ species	**Mechanism/mode of formation**	**Relevance to humans***	**References**
Forestomach/rat	Local irritation, inflammation, ulceration leading to regenerative hyperplasia (e.g. chlorothalonil).	Probably low; anatomical differences between rats and humans.	Wilkinson & Killeen (1996)
Kidney/rat	• Direct genotoxicity (e.g. N-nitroso compounds). • Indirect DNA damage mediated by oxidative stress (e.g. potassium bromate, ferric nitrilotriacetate). • Sustained stimulation of cell proliferation, in response to: – Renal cell tubule injury/death (e.g. chloroform), or – In male rats, promotion of initiated cells by regenerative hyperplasia consequent to cytotoxicity due to accumulation of hyaline droplets containing ligand-bound α2μ-globulin (e.g. gasoline, d-limonene). Note, however, this is a hypothetical mechanism only and applies to a limited range of carcinogens — see discussion by Melnick & Kohn (IARC 1999, pp. 119–37).	Relevance of male rat-specific nephropathy is low; humans do not synthesise α2μ-globulin, the association of gasoline with human renal cancer is equivocal, and citrus (which contains d-limonene) is not carcinogenic to humans. The association of smoking and occupational exposure to PAH with renal cancer suggests that direct genotoxicity is a highly relevant mechanism for humans. Obesity, phenacetin intake, a diet high in protein and fat, reproductive factors and asbestos are also associated with human renal cancer. Mechanisms are not fully characterised but may include peripheral oestrogen formation, oxidative stress and chemical injury to renal tubule.	US EPA (1991) IARC (1999)
Urinary bladder/ mouse, rat	Regenerative hyperplasia arising from mechanical injury to urothelium, caused by formation of solid aggregates within the urinary tract (e.g. sodium saccharin, sulfosulfuron). Mitogenic effects on rat bladder urothelium have also been reported in the absence of urinary precipitates (e.g. propoxur).	Probably low–moderate. Anatomical differences between rat and human bladder decrease the likelihood of prolonged residence of uroliths in human bladder, but there is still an epidemiological association between urinary tract stones and cancer. Humans may also develop bladder cancer through hyperplastic response to infection. Strongest risk factors for human bladder cancer are smoking and occupational exposure to aromatic amines, suggesting that genotoxicity is a highly relevant mode of action for humans.	Rodent bladder carcinogenesis working group (1995) IARC (1999) Cohen *et al.* (1994)
Thyroid/rat	Genotoxicity [e.g. nitrosamines, acetochlor(?)]. Disturbance of the thyroid-pituitary axis occurring in response to: • Toxicity to thyroid follicular cells (e.g. PCBs). • Enhanced hepatic metabolism and biliary excretion of thyroid hormone (e.g. acetochlor, clofentazine, fenbuconazole, fipronil, pendimethalin, pentachloronitrobenzene, prodiamine, pyrimethanil, thiazopyr). • Inhibition of thyroid peroxidase (e.g. amitrole, mancozeb, ethylene thiourea), 5'-monodiodinase (e.g. erythrosine), iodine uptake (amitrole, ethiozin, mancozeb, ethylene thiourea, pentachloronitrobenzene) or thyroid hormone release (e.g. lithium, excess iodide). NB: Some agents have multiple mechanisms of action.	High in the case of genotoxins. Low otherwise: • Humans are less sensitive than rodents to disturbance of thyroid-pituitary axis. • There are species differences in thyroid biochemistry and physiology — rodents lack thyroxine-binding globulin, and thyroxine turnover is much more rapid in rodents than humans. • No chemical is known to be carcinogenic to human thyroid.	US EPA (1998f) Hurley *et al.* (1998) IARC (1999)

Liver/mouse, rat	Peroxisome proliferation (e.g. DEHP, lactofen, trichloroethylene), mediated by activation of Peroxisome Proliferator Activated Receptor (PPAR). Receptor activation may be a direct effect of the peroxisome proliferator or occur through peturbation of lipid metabolism. Proposed mechanisms of carcinogenesis include activation of genes coding for peroxisomal enzymes, increased lipid metabolism, H_2O_2 production and oxidative stress, inhibition of apoptosis, hepatocellular proliferation and oncogene activation.	Probably low; human hepatocytes contain PPAR but are less responsive than rodent hepatocytes to the peroxisome proliferator. Therapeutic doses of hypolipidaemic drugs produce little if any peroxisome proliferation in human liver.	Melnick *et al*. (1996), Green (1995), Cattley & Preston (1995) IARC (1995)
Liver/mouse, rat	Cytotoxicity, necrosis followed by regenerative hyperplasia (e.g. chloroform). Mitogenicity. Enzyme induction (e.g. phenobarbitone).	Low if non-genotoxic. Rodent liver cancers are often high-dose phenomena, dependent on hepatomegaly, enhanced cell turnover and/or prolonged cytotoxicity. These factors are unlikely to occur under normal conditions of human exposure. Experience with phenobarbitone shows humans are resistant to liver cancer in response to enzyme induction. Alcohol, aflatoxin and viral hepatitis are the main risk factors for hepatic cancer in humans. There is no known example of a non-genotoxic mouse liver carcinogen that is carcinogenic in humans.	Melnick *et al*. (1996) Butterworth *et al.* (1998) Carmichael *et al.* (1997)
Testis (Leydig cells)/mouse, rat	• Agonism of oestrogen (e.g. diethylstilbestrol, methoxychlor), gonadotropin-releasing hormone (GnRH) (e.g. buserelin, leuprolide) and dopamine receptors (e.g. mesulergine, norprolac). • Antagonism of androgen receptors (e.g. linuron, vinclozolin). • Inhibition of 5α-reductase (e.g. finasteride), testosterone biosynthesis (e.g. metronidazole, vinclozolin) and aromatase [e.g. formestane, letrozole (in dog)]. Most of these mechanisms ultimately involve an increase in serum luteinizing hormone (LH) and/or the responsiveness of Leydig cells to LH as proximate mediators.	Regulation of hypothalamo-pituitary-testis axis is similar in rodents and humans. However, humans have a comparatively lower incidence of Leydig cell tumours, and the human Leydig cell appears to be less responsive to proliferative stimuli and less susceptible to spontaneous and xenobiotic-induced tumours. Several chemicals causing Leydig cell tumorigenesis in rats appear not to induce testicular neoplasms in humans. GnRH and prolactin receptors are expressed less in human than in rat testis, so human Leydig cells are comparatively insensitive to GnRH and dopamine agonists. Available data suggest that non-genotoxic Leydig cell tumorigens in rodents probably have low relevance to humans.	Cook *et al.* (1999)

***Note:** Evaluators should **not** assume that all tumours at the sites listed arise via the mechanisms given. The possible significance of these tumours to humans cannot be dismissed, and should be judged case by case on the basis of all available data, preferably using the IPCS's Conceptual Framework for Cancer Risk Assessment (see Appendix VIII).

REFERENCES

Adkins AG, Alden CL & Kanerva RL (1982) Optimization and standardization of male gonad weight determinations in rats. Toxicol Pathol **10**: 33–37.

Altman PL & Dittmer DS (Eds) (1974) Biology Data Book No. VIII. Federation of American Societies of Experimental Biology, Bethesda, Md.

Amdur MO, Doull JD & Klaassen CD (1991) In: Casarett and Doull's Toxicology: The Basic Science of Poisons, 4th Ed. p. 5. Pergamon Press, New York, N.Y.

Ashby J, Waters MD, Preston J, Adler ID, Douglas GR, Fielder R, Shelby MD, Anderson D, Sofuni T, Gopalan HN, Becking G & Sonich-Mullin C (1996) IPCS harmonization of methods for the prediction and quantification of human carcinogenic/mutagenic hazard, and for indicating the probable mechanism of action of carcinogens. Mutat Res **352**: 153–157.

Bahnemann R, Jacobs M, Karbe E, Kaufmann W, Morawietz G, Nolte T, Rittinghausen S (1995) RITA – Registry of Industrial Toxicology Animal-data – Guides for organ sampling and trimming procedures in rats. Exp Toxic Pathol **47**: 247–266.

Barnes DG, Daston GP, Evans JS, Jarabek AM, Kavlock RJ, Kimmel CA, Park C & Spitzer HL (1995) Benchmark dose workshop: Criteria for use of a benchmark dose to estimate a reference dose. Reg Tox Pharmacol **8**: 471–486.

Bayne K (1996) Revised guide for the care and use of laboratory animals. American Physiological Society. Physiologist **39**: 199, 208–211.

Benirschke K, Garner FM & Jones TC (Eds) (1978) Pathology of Laboratory Animals, Vols. I & II. Springer-Verlag, New York, N.Y.

Bennett WM (1997) Overview: Clinical aspects of nephrotoxicity. In: Sipes IG, McQueen CA & Gandolfi AJ (Eds) Comprehensive toxicology, Vol. 7, pp. 1–9. Elsevier Science Ltd, Oxford.

Berg BN & Simms HS (1960) Nutrition and longevity in the rat. II Longevity and onset of disease with different levels of food intake. J Nutr **71**: 225–263.

Blain PG, Battershill JM, Venitt S, Cooper CC, & Fielder RJ (1998) Consideration of short-term carcinogenicity tests using transgenic mouse models. Mutat Res **403**: 259–263.

Boorman GA, Montgomery CA, Eustis SL, Wolfe MJ, McConnell EE & Hardisty JF (1985) Quality assurance in pathology for rodent carcinogenicity studies. In: Milman HA & Weisberger EK (Eds) Handbook of Carcinogen Testing, pp. 345–357. Noyes Publications, Park Ridge, N.J.

Boorman GA & Eustis SL (1986) The pathology working group as a means for assuring pathology quality in toxicological studies. In: Hoover BK, Baldwin JK, Uelner AF, Whitmire CE, Davies CL & Bristol DW (Eds) Managing conduct and data quality of toxicology studies, pp. 271–275. Princeton Scientific Publishing Co., Princeton, N.J.

Boyd EM & Knight LM (1963) Postmortem shifts in the weight and water levels of body organs. Toxicol Appl. Pharmacol **5**: 119–128.

Bucci TJ (1991) Evaluation of altered morphology. In: Hascheck WM & Rousseaux CG (Eds) Handbook of Toxicologic Pathology, ch. 3, pp. 23–25. Academic Press, San Diego, Cal.

Bucher JR (1998) Update on national toxicology program (NTP) assays with genetically altered or "transgenic" mice. Environ Health Perspect **106**: 619–621.

Bush BM (1991) Interpretation of Laboratory Results for Small Animal Clinicians. Blackwell Scientific Publications, London.

Butterworth BE, Kedderis GL & Conolly RB (1998) The chloroform cancer risk assessment: a mirror of scientific understanding. CIIT Activities **18**: 1–10.

Capen CC, Benjamin S, Hahn F, Dungworth DL & Carlton WW (2001) Pathobiology of the aging dog, Vols. 1 & 2. Iowa State University Press.

Carmichael NG, Enzmann H, Pate I & Waechter F (1997) The significance of mouse liver tumor formation for carcinogenic risk assessment: Results and conclusions from a survey of ten years of testing by the agrochemical industry. Environ Health Perspect **105** (11): 1196–1203.

Casterline JL, Jr, & Williams CH (1969) Effect of pesticide administration upon esterase activities in serum and tissue of rats fed variable casein diets. Tox Appl Pharmacol **14**: 266–275.

Cattley RC & Preston RJ (1995) Does DNA damage play a role in rodent liver cancer induced by peroxisome proliferators? CIIT Activities **15** (11): 1–6.

Chhabra RS & Fouts JR (1974) Sex differences in the metabolism of xenobiotics by extrahepatic tissues in rats. Drug Metab Dispos **6**: 375–379.

Christian MS, Hoberman AM, Johnson MD, Brown WR & Bucci TJ (1998) Effect of dietary optimization on growth, survival, tumor incidences and clinical pathology parameters in CD Sprague–Dawley and Fischer 344 rats: a 104-week study. Drug Chem Toxicol **21**: 97–117.

Clampitt RB (1978) An investigation into the value of some clinical biochemical tests in the detection of minimal changes in liver morphology and function in the rat. Toxicol. (Sup. l): 1-13.

Cohen SM, Cano M, Johnson LS, St John MK, Asamoto M, Garland EM, Thyssen JH, Sangha GH & Van Goethem DL (1994) Mitogenic effects of propoxur on male rat bladder urothelium. Carcinogenesis **15** (11): 2593–2597.

Conner MW & Newbern PM (1984) Drug–nutrient interactions and their implications for safety evaluations. Fund Appl Toxicol **4**: 5341–S356.

Contrera JF & DeGeorge JJ (1998) *In vivo* transgenic bioassays and assessment of the carcinogenic potential of pharmaceuticals. Environ Health Perspect **106** (Sup. 1): 71–80.

Cook JC, Klinefelter GR, Hardisty JF, Sharpe RM & Foster PMD (1999) Rodent Leydig cell tumorigenesis: A review of the physiology, pathology, mechanisms, and relevance to humans. Crit Rev Toxicol **29** (2): 169–261.

Dairman W & Balazs T (1970) Comparison of liver microsome enzyme systems and barbiturate sleep times in rats caged individually or communally. Biochem Pharmacol **9**: 951–5.

Dameron GW, Weingand KW, Duderstadt JM, Odioso LW, Dierckman TA, Schwecke W & Baran K (1992) Effect of bleeding site on clinical laboratory testing of rats: Orbital venous plexus versus posterior vena cava. Lab Animal Sci **42**: 299–301.

Dayton PG & Sanders JE (1983) Dose-dependent pharmacokinetics: Emphasis on phase I metabolism. Drug Metab Rev **14**: 347–405.

Dearfield KL & Benz RD (1999) Can the new genetic toxicology tests be used for regulatory safety decisions? Environ Mol Mutagen **33**(2): 91–93.

DeGeorge J (1999) Personal communication.

Derelanko MJ (2000) Toxicologist's Pocket Handbook. CRC Press, Boca Raton, Fla.

Derelanko MJ & Hollinger MA (Eds) (1995) CRC Handbook of Toxicology. CRC Press, Boca Raton, Fla.

Dickens A& Robinson J (1996) Statistical approaches. In: Evans GO (Ed) Animal Clinical Chemistry: A Primer for Toxicologists, pp. 45–48. Taylor & Francis, London.

Doull J (1980) Factors influencing toxicology. In: Casarett and Doull's Toxicology, 2nd Ed., pp. 70–83. Macmillan Pub. Co., New York, N.Y.

[**Note:** This reference is to an old edition of this well-known textbook but the latest edition (on page 22) still refers to this chapter in the 2nd edition for detailed discussion of factors influencing animal responses to toxic substances.]

Duncan JR, Prasse KW & Mahaffey EA (1994) Veterinary Laboratory Medicine: Clinical Pathology, 3rd Ed. Iowa State Press, Ames, Ia.

Eastin WC & Tennant RW (1998) In: Kitchin KT (Ed) Carcinogenicity: Testing Predicting & Interpreting Chemical Effects, pp. 395–410. Marcel Dekker Inc., New York, N.Y.

Eastin WC (1998) The U.S. National Toxicology Program evaluation of transgenic mice as predictive models for identifying carcinogens. Environ Health Perspect **106**(Sup. 1): 81–84.

EC (1996) Technical Guidance Document in Support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances and Commission Regulation (EC) No 1488/94 on Risk Assessment for Existing Substances, Part 1. Office for Official Publications of the European Communities, Luxembourg.

Enzmann H, Bomhard E, Iatropoulos M, Ahr HJ, Schlueter G & Williams GM (1998a) Short- and intermediate-term carcinogenicity testing – a review. Part 1: The prototypes mouse skin tumour assay and rat liver focus assay. Food Chem Toxicol **36**: 979–995.

Enzmann H, Iatropoulos M, Brunnemann KD, Bomhard E, Ahr HJ, Schlueter G & Williams GM (1998b) Short- and intermediate-term carcinogenicity testing – a review. Part 2: Available experimental models. Food Chem Toxicol **36**: 997–1013.

Evans GO (1996) Animal clinical chemistry: A primer for toxicologists. Taylor & Francis, London.

Evans JG & Lake BG (1998) The digestive system II: The hepatobiliary system. In: Turton J & Hooson J (Eds) Target Organ Pathology: A Basic Text, ch. 3, pp. 61–97. Taylor & Francis, London.

FAO/WHO (July 1987) Guidelines for the preparation of toxicological working papers for the joint FAO/WHO expert committee on food additives, Geneva (Document ICS/89.41).

Fischbach FT (1996) A Manual of Laboratory & Diagnostic Tests, 5th Ed. Lippincott, Philadelphia, Pa., and New York, N.Y.

Foran JA & The International Life Sciences Institute (ILSI) Risk Sciences Working Group on Dose Selection (1997) Principles for the selection of doses in chronic rodent bioassays. Env Hlth Persp **105**: 18–20.

Fox JG, Cohen BJ & Loew FM (1984) Laboratory Animal Medicine. Academic Press Inc, Orlando, Fla.

Gad S & Weil CS (1986) Statistics and experimental design for toxicologists. Telford Press, N.J.

Gad SC & Weil CS (1989) Statistics for toxicologists. In: Hayes AW. (Ed) Principles and Methods of Toxicology, 3rd Ed., ch. 7, pp. 221–274. Raven Press, New York, N.Y.

Goldsworthy TL, Hanigan MH & Pitot HC (1986) Models of hepatocarcinogenesis in the rat-contrasts and comparisons. Crit Rev Toxicol **17**: 61–89.

Green S (1995) PPAR: a mediator of peroxisome proliferator action. Mutat Res **333**: 101–109.

Grice HC & Burek JD (1983) Age associated (geriatric) pathology: Its impact on long-term toxicity studies. In: Current Issues in Toxicology, Springer-Verlag, New York, N.Y.

Guyton AC (1991) Textbook of Medical Physiology, 8th Ed. WB Saunders Co., Philadelphia, Pa.

Hall RL (2001) Principles of clinical pathology for toxicology studies. In: Hayes AW (Ed) Principles and Methods of Toxicology, 4th Ed., ch. 21, pp. 1001–1038. Taylor & Francis, Philadelphia, Pa.

Hardisty JF & Boorman GA (1986) National Toxicology Program pathology quality assurance procedures. In: Hoover BK, Baldwin JK, Uelner AF, Whitmire CE, Davies CL & Bristol DW (Eds) Managing Conduct and Data Quality of Toxicology Studies, pp. 263–269. Princeton Scientific Publishing Co., Princeton, N.J.

Haseman JK, Boorman GA & Huff J (1997) Value of historical control data and other issues related to the evaluation of long-term rodent carcinogenicity studies. Toxicol Pathol **25**(5): 524–527 (editorial).

Haseman JK, Huff J & Boorman GA (1984) Use of historical control data in carcinogenicity studies in rodents. Toxicol Pathol **12**: 126–135.

Hayes AW (Ed) (1994) Principles and Methods of Toxicology, 3rd Ed. Raven Press, New York, N.Y.

Hayes WJ (1991) Dosage and other factors influencing toxicology. In: Hayes WJ & Laws ER (Eds) Handbook ofToxicology, Vol. 1: General Principles, pp. 40–41. Academic Press, San Diego.

Hayward JJ, Shane BS, Tindall KR & Cunningham ML (1995) Differential *in vivo* mutagenicity of the carcinogen–noncarcinogen pair 2,4- and 2,6-diaminotoluene. Carcinogenesis **16**: 2429–2433.

Henry JB (Ed) (1984) Todd, Sanford, Davidsohn. Clinical Diagnosis and Management by Laboratory Methods, 17th Ed. Saunders Co., Philadelphia, Pa.

Heywood R (1981) Target organ toxicity. Toxicol Lett **8**: 349–358.

Hollinger MA (1995) In: Derelanko MJ & Hollinger MA (Eds) (1995) CRC Handbook of Toxicology, ch. 22, pp. 771–920. CRC Press, Boca Raton, Fla.

Hurley PM, Hill RN & Whiting RJ (1998) Mode of carcinogenic action of pesticides inducing thyroid follicular cell tumors in rodents. Environ Health Perspect **106**(8): 437–445.

IARC (1986) The Design and Analysis of Long-term Animal Experiments. Gart JJ *et al.* IARC Scientific Publications **79**, Lyon.

IARC (1990) Pathology of Tumours in Laboratory Animals, 2nd Ed., Vol. 1. Tumours of the Rat. Mohr U & Turusov VS (Eds). IARC Scientific Publications **99**, Lyon.

IARC (1992) Mechanisms of Carcinogenesis in Risk Identification. Vainio H, Magee P, McGregor D & McMichael AJ (Eds). IARC Scientific Publications **116**, Lyon.

IARC (1994) Pathology of Tumours in Laboratory Animals, 2nd Ed., Vol. 2. Tumours of the Mouse. Mohr U & Turusov VS (Eds). IARC Scientific Publications **111**, Lyon.

IARC (1995) Peroxisome Proliferation and its Role in Carcinogenesis. IARC Technical Report **24**, Lyon.

IARC (1996) Pathology of Tumours in Laboratory Animals, 2nd Ed., Vol. 3. Tumours of the Hamster. Mohr U & Turusov VS (Eds). IARC Scientific Publications **126**, Lyon

IARC (1992–1997) International Classification of Rodent Tumours, Part 1. The Rat (set of 10 fascicles). Mohr U (Ed). IARC Scientific Publications **122**, Lyon.

IARC (1999). The Use of Short- and Medium-term Tests for Carcinogens & Data on Genetic Effects in Carcinogenic Hazard Evaluation. McGregor DB, Rice JM & Vennitt S (Eds). IARC Scientific Publications **146,** Lyon.

IARC (1999). Species Differences in Thyroid, Kidney and Urinary Bladder Carcinogenesis. Capen CC, Dybing E, Rice JM & Wilbourn JD (Eds). IARC Scientific Publications **147,** Lyon.

ICH (1997) Guidance for Industry S1B Testing for carcinogenicity of pharmaceuticals. US FDA, CDER/CBER, Rockville, Md. http://www.fda.gov/cder/guidance/index.htm

Innami S, Yang MG, Mickelsen O & Hafs HD (1973) The influence of high-fat diets on estrous cycles, sperm production and fertility of rats. Proc Soc Exp Biol Med **143**: 63–68.

Ito N, Imaida K, Hasegawa R & Tsuda H (1989) Rapid bioassay methods for carcinogens and modifiers of hepatocarcinogenesis. Crit Rev Toxicol **19**: 385–415.

Jain NC (1993) Essentials of Veterinary Hematology. Lea & Fibiger, Philadelphia, Pa.

Kanerva RL, Alden CL & Wyder WE (1982) The effect of uniform exsanguination on absolute and relative organ weights, and organ weight variation. Toxicol Pathol **10**: 43–44.

Keenan KP, Laroque P & Dixit R (1998) Need for dietary control by caloric restriction in rodent toxicology and carcinogenicity studies. J Toxicol Environ Health B Crit Rev **1**(2): 135–148.

Kennedy GC (1969) Interactions between feeding behavior and hormones during growth. Ann NY Acad Sci **157**: 1049–1061.

Klaassen D & Eaton DL (1991) Principles of toxicology. In: Casarett & Doull's Toxicology: The Basic Science of Poisons, 4th Ed. p. 12–49. Pergamon Press, New York, N.Y.

Klinger MM, MacCarter GD & Boozer CN (1996) Body weight and composition in the Sprague–Dawley rat: Comparison of three outbred sources. Lab Animal Sci **46**: 67–69.

Krinke GJ, Perrin LPA & Hess R (1991) Assessment of toxicopathological effects in ageing laboratory rodents. Arch Toxicol (Sup. 14) 43–49.

Lee PN (1993) Statistics. In: Anderson D & Conning DM (Eds) Experimental Toxicology: The Basic Issues. Royal Society of Chemistry, Cambridge.

Lehman AJ (1954) Assoc Food & Drug Officials Quarterly Bulletin **18**: 66.

Levy G (1968) Dose dependent effects in pharmacokinetics. In: Important Fundamental Principles in Drug Evaluation, pp. 141–172.

Lewis RA (1998) Lewis' Dictionary of Toxicology. CRC Press, Boca Raton, Fla.

Maronpot RR, Shimkin MB, Witschi HP, Smith LH & Cline JM (1986) Strain A mouse pulmonary tumor test results for chemicals previously tested in the National Cancer Institute carcinogenicity tests. J Natl Cancer Inst **76**: 1101–1112.

Marsman DS & Popp JA (1984) Biological potential of basophilic hepatocellular foci and hepatic adenoma induced by the peroxisome proliferator, Wy-14,643. Carcinogenesis **15**: 111–117.

Masoro EJ (1992) Aging and proliferative homeostasis: Modulation by food restriction in rodents. Lab Animal Sci **42**: 132–137.

McClellan RO & Henderson RF (Eds) (1989) Concepts in Inhalation Toxicology. Hemisphere Publishing Corp, New York, N.Y.

McLean AEM & McLean EK (1969) Diet and toxicity. Br Med Bull **25**: 278–281.

Melnick RL, Kohn MC & Portier CJ (1996) Implications for risk assessment of suggested nongenotoxic mechanisms of chemical carcinogenesis. Environ Health Perspect **104**(Sup. 1): 123–134.

Merriam–Webster Medical Dictionary (1997) On-line edition at http://www.intelihealth.com/IH/ihtIH/WSIHW000/9276/9276.html

Mohr U, Dungworth DL & Capen CC (Eds) (1992) Pathobiology of the Aging Rat, Vol. 1. ILSI Press, Washington, D.C.

Mohr U, Dungworth DL & Capen CC (Eds) (1994) Pathobiology of the Aging Rat, Vol. 2. ILSI Press, Washington, D.C.

Mohr U, Dungworth DL, Ward J, Capen CC, Carlton W & Sundberg J (Eds) (1996) Pathobiology of the Aging Mouse, Vols. 1 & 2. ILSI Press, Washington, D.C.

Mohr U, Dungworth D, Kimmerle G, Lewkowski J, McClellan R & Stöber W (Eds) (1988) Inhalation Toxicology: The Design and Interpretation of Inhalation Studies and their Use in Risk Assessment. ILSI Monograph. Springer-Verlag, New York, N.Y.

Mohr U & Morawietz G (1995) Recent progress in the field of toxicologic pathology in Europe and the need for standardisation. J Toxicol Pathol **8**: 219–226.

Mohr U, (1999) RITA – Registry of Industrial Toxicology Animal-data: Optimization of carcinogenicity bioassays. Exp Toxic Pathol **51**: 461–475 (editorial).

Morawietz G, Rittinghausen S, Mohr U (1992) RITA – Registry of Industrial Toxicology Animal-data – Progress of the working group. Exp Toxic Pathol **44**: 301–309.

Morrison V & Ashby J (1994) A preliminary evaluation of the performance of the muta™ mouse (lacZ) and Big Blue™ (lacI) transgenic mouse mutation assays. Mutagenesis **9**: 367–375.

Munro IC (1977) Considerations in chronic toxicity testing: The chemical, the dose, the design. J Environ Path Toxicol **1**: 183–197.

National Academy of Sciences/National Research Council (1983) Risk Assessment in the Federal Government: Managing the Process. National Academy Press, Washington, D.C.

National Academy of Sciences/National Research Council (1994) Science and Judgement in Risk Assessment. National Academy Press, Washington, D.C.

Nolen G (1972) Effects of various restricted dietary regimens on the growth, health and longevity of albino rats. J Nutr **102**: 1477–1494.

OECD (1981a) Guideline 451. Carcinogenicity studies (17 pages; adopted 12 May 1981). In: OECD Guidelines for the Testing of Chemicals (1993) Section 4: Health effects. Vol. 2. Organisation for Economic Cooperation & Development, Paris.

OECD (1981b) Guideline 452. Chronic toxicity studies (14 pages; adopted 12 May 1981). In: OECD Guidelines for the Testing of Chemicals (1993) Section 4: Health effects. Vol. 2. Organisation for Economic Cooperation & Development, Paris.

OECD (1981c) Guideline 453. Combined chronic toxicity/carcinogenicity studies (15 pages; adopted 12 May 1981). In: OECD Guidelines for the Testing of

Chemicals (1993) Section 4: Health effects. Vol. 2. Organisation for Economic Cooperation & Development, Paris.

OECD (1982) Good Laboratory Practice in the Testing of Chemicals, 62 pp. Organisation for Economic Cooperation and Development, Paris.

OECD (1993) OECD Guidelines for the Testing of Chemicals. Section 4: Health effects. Vol. 2. Organisation for Economic Cooperation & Development, Paris.

OECD (1998a) Guidance for Industry Data Submissions on Plant Protection Products and their Active Substances – Dossier Guidance (Guidelines and criteria for industry for the preparation and presentation of complete dossiers and of summary dossiers for plant protection products and their active substances in support of regulatory decisions in OECD countries) Organisation for Economic Cooperation and Development, Paris http://www.oecd.org/ehs/PestsGD01.htm

OECD (1998b) Guidance for Country Data Review Reports on Plant Protection Products and their Active Substances – Monograph Guidance (Guidelines and criteria for the evaluation of dossiers and preparation of reports by regulatory authorities in OECD countries relating to the evaluation of active substances, the registration of plant protection products and the establishment of maximum residue limits (MRLs) and import tolerances) Organisation for Economic Cooperation and Development, Paris http://www.oecd.org/ehs/PestsGD01.htm

OECD (1998c) Guideline 408. Subchronic oral toxicity – rodent: 90-day study (10 pages; adopted 21 September 1998). In: OECD Guidelines for the Testing of Chemicals Section 4: Health effects. Vol. 2, 10th Addendum, October 1998. Organisation for Economic Cooperation & Development, Paris.

OECD (1998d) Guideline 409. Subchronic oral toxicity – non-rodent: 90-day study (9 pages; adopted 21 September 1998). In: OECD Guidelines for the Testing of Chemicals Section 4: Health Effects. Vol. 2, 10th Addendum. Organisation for Economic Cooperation & Development, Paris.

OECD (1998e) Guideline 424. Neurotoxicity study in rodents (15 pages; adopted 21 July 1997). In: OECD Guidelines for the Testing of Chemicals Section 4: Health Effects. Vol. 2, 10th Addendum. Organisation for Economic Cooperation & Development, Paris.

OECD (1998f) Pesticide Data Submissions to OECD Countries: Forms for Screening Test and Study Reports and Summaries for Completeness (Revised 21 September 1998), PFWG Document 98(10)

OECD (1998g) Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances. Organisation for Economic Cooperation & Development, Paris. http://www.oecd.org//ehs/Class/hclfinaw.doc

OECD (2000) Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Organisation for Economic Cooperation & Development, Paris. http://www.oecd.org/ehs/test/monoweb19.doc

OECD/IPCS (1998) Joint OECD/IPCS Project on Harmonization of Hazard/Risk Assessment Terminology. Critical analysis of survey results.

On-line Medical Dictionary (1995-98) at *omd@www.graylab.ac.uk*

Paynter OE (1984). Oncogenic Potential: Guidance for Analysis and Evaluation of Long-term Rodent Studies. Evaluation procedure #1000.1 Office of Pesticide and Toxic Substances, EPA, Washington, D.C.

Pearl W, Balazs T & Buyske DA (1966) Effects of stress on serum transaminase activity in the rat. Life Sci **5**: 67–74.

Pesticide Programs: Good Laboratory Practice Standards. Final Rule F.R. 46, No. 230, pp. 53946-53969, Tuesday 11/29/83.

Peto R, Pike MC, Day NE, Gray RG, Lee PN, Parish S, Peto J, Richards S & Wahrendorf J (1980) Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In: Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal. IARC Monographs Supplement **2**, pp. 311–426. IARC, Lyon.

Pfeiffer CJ & Muller PJ (1967) Physiologic correlates dependent upon mode of death. Toxicol Appl Pharm **10**: 253–260.

Portier CJ & Bailer AJ (1989) Testing for increased carcinogenicity using survival adjusted quantal response test. Fund Appl Toxicol **12** 731–737.

Rodent Bladder Carcinogenesis Working Group (1995) Report of the Rodent Bladder Carcinogenesis Working Group. Fd Chem Toxic **33**(9): 701–802.

Rogers AE, Sanchez O, Feinsod FM & Newberne PM (1974) Dietary enhancement of nitrosamine carcinogenesis. Cancer Res **34**: 96–99.

Rosner BA (1995) Fundamentals of Biostatistics. Wadsworth Publishing Co., Belmont, Ca.

Ross MH & Bras G (1965) Tumour incidence patterns and nutrition in the rat. J Nutr **87**: 245–260.

Roubicek CB, Pahnish OF & Taylor RL (1964) Growth of rats at two temperatures. Growth **28**: 157–164.

Salem H (Ed) (1987) Inhalation Toxicology: Research Methods, Applications and Evaluation. Marcel Dekker Inc., New York, N.Y., and Basel.

Salsburg DS (1986) Statistics for Toxicologists. Marcel Dekker Inc., New York, N.Y., and Basel.

Seefeld MD & Peterson RE (1984) Digestible energy and efficiency of feed utilisation in rats treated with 2,3,7,8,-tetrachlorodibenzo-p-dioxin. Toxicol Appl Pharmacol **74**: 214–222.

Sharer K (1977) The effect of chronic underfeeding on organ weights in rats. How to interpret organ weight changes in cases of marked growth retardation in toxicity tests? Toxicol **7**: 45–56.

Sharma RP & Coulombe RA, Jr (1996) Pharmacokinetics and risk assessment. In: Fan AM & Chang LW (Eds) Toxicology and Risk Assessment, ch. 7, pp. 81–99. Marcel Dekker Inc., New York, N.Y.

Siglin JC & Rutledge GM (1995) In: Derelanko MJ & Hollinger MA (Eds) (1995) CRC Handbook of Toxicology, ch. 1, pp. 28–49. CRC Press, Boca Raton, Fla.

Sisk SC, Pluta LJ, Bond JA & Recio L (1994) Molecular analysis of lacI mutants from bone marrow of B6C3F1 transgenic mice following inhalation exposure to 1,3-butadiene. Carcinogenesis **15**: 471–477.

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, Grant D, Hartley M, Knaap A, Kroese D, Mangelsdorf I, Meek E, Rice JM & Younes M (2001) IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Reg Toxicol Pharmacol **34**: 146–152.

Stevens KR & Myelcraine L (1994) Issues in Chronic Toxicology. In: Hayes AW (Ed) (1994) Principles and Methods of Toxicology, 3rd Ed. Raven Press, New York N.Y.

Sumi N, Stavrou D, Frohberg H & Jochmann G (1976) The incidence of spontaneous tumours of the central nervous system of Wistar rats. Arch Toxicol **35**: 1–13.

Swaim LD, Taylor HW & Jersey GC (1985) The effect of handling techniques on serum ALT activity in mice. J Appl Toxicol **5**: 160–162.

Tallarida RJ & Murray RB (1987) Manual of Pharmacologic Calculations with Computer Programs, 2nd Ed. Springer-Verlag, New York, N.Y.

Tannenbaum A (1940) Relationship of body weight to cancer incidence. Arch Pathol **30**: 509–517.

Tarloff JB & Kinter LB (1997) *In vivo* methodologies used to assess renal function. In: Sipes IG, McQueen CA & Gandolfi AJ (Eds) Comprehensive Toxicology, Vol. 7, pp. 99–119. Elsevier Science Ltd, Oxford.

Tarone RE (1982) The use of historical control information in testing for a trend in proportions. Biometrics **38**: 215–220.

Task Force of Past Presidents (1982) Animal data in hazard evaluation: Paths and pitfalls. Fund Appl Toxicol **2**: 101–107.

Tennant RW, French JE & Spalding JW (1995) Identifying chemical carcinogens and assessing potential risk in short-term bioassays using transgenic mouse models. Environ Health Perspect **103**: 942–950.

Thompson GE, Scheel LD & Lowe D (1982) Seasonal alteration in response to stress or physiological change. Drug & Chem. Toxicol **5**: 189–199.

Thurman JD, Moeller RB & Turturro A (1995) Proliferative lesions of the testis in *ad-libitum*-fed and food-restricted Fischer F-344 and FBNF1 rats. Lab Animal Sci **45**: 635–640.

Todd GC (1986) Induction and reversibility of thyroid proliferative changes in rats given an antithyroid compound. Vet Pathol **23**: 110–117.

Travlos GS, Morris RW, Elwell MR, Duke A, Rosenblum S & Thompson MB (1996) Frequency and relationships of clinical chemistry and liver and kidney histopathology findings in 13-week toxicity studies in rats. Toxicol **107**: 17–29.

Tucker MJ (1979) The effect of long-term food restriction on tumours in rodents. Int J Cancer **23**: 803–807.

Tyson CA & Sawhney DS (1985) Organ function tests in toxicology evaluation. Noyes Publications, Park Ridge, N.J.

UK DoH (1991) Guidelines for the evaluation of chemicals for carcinogenicity. Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment. UK Department of Health. Report on Health and Social Subjects 42. HMSO London. ISBN 0 11 321453 7.

US EPA (1982) Pesticide Assessment Guidelines Subdivision F. Hazard Evaluation: Human and Domestic Animals. Office of Pesticide and Toxic Substances. p. 107. Washington, D.C.

US EPA (1983) Good laboratory practice standards: Toxicology testing. Fed Reg. 48: 53922.

US EPA (1991) Alpha $_{2u}$-globulin: Association with chemically induced renal toxicity and neoplasia in the male rat. US Environmental Protection Agency, EPA/625/3-91/019F, September 1991.

US EPA (1993) Reference Dose (RfD): Description and Use in Health Risk Assessments. US Environmental Protection Agency, Background Document 1A, March 15, 1993. http://www.epa.gov/ngispgm3/iris/rfd.htm

US EPA (1996) Proposed guidelines for carcinogen risk assessment. US Environmental Protection Agency, EPA/600/P-92/003C, April 1996.

US EPA (1998a) Health Effects Test Guidelines OPPTS 870.4100 Chronic Toxicity. US Environmental Protection Agency, EPA 712-C-98-210, August 1998. http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998b) Health Effects Test Guidelines OPPTS 870.4200 Carcinogenicity. US Environmental Protection Agency, EPA 712-C-98-211, August 1998. http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998c) Health Effects Test Guidelines OPPTS 870.4300 Combined Chronic Toxicity/Carcinogenicity. US Environmental Protection Agency, EPA 712-C-98-212, August 1998. http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998d) Health Effects Test Guidelines OPPTS 870.6200 Neurotoxicity Screening Battery. US Environmental Protection Agency, EPA 712-C-98-238, August 1998. http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998e) Health Effects Test Guidelines OPPTS 870.7800 Immunotoxicity. US Environmental Protection Agency, EPA 712-C-98-351, August 1998. http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998f) Assessment of thyroid follicular cell tumors. United States Environmental Protection Agency, EPA/630/R-97/002, March 1998.
http://www.epa.gov/ncea/thyroid.htm

US EPA (1998g) Health Effects Test Guidelines OPPTS 870.3100 90-day Oral Toxicity in Rodents. US Environmental Protection Agency, EPA 712-C-98-199, August 1998.
http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998h) Health Effects Test Guidelines OPPTS 870.3150 90-day Oral Toxicity in Non-rodents. US Environmental Protection Agency, EPA 712-C-98-200, August 1998.
http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998i) Health Effects Test Guidelines OPPTS 870.3250 90-day Dermal Toxicity. US Environmental Protection Agency, EPA 712-C-98-202, August 1998.
http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

US EPA (1998j) Health Effects Test Guidelines. OPPTS 870.3465 Chronic Toxicity. US Environmental Protection Agency, EPA 712-C-98-204, August 1998.
http://www.epa.gov/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

Walach JB (1996) Interpretation of Diagnostic Tests, 6th Ed. Little Brown, Boston, Mass.

Waner T (1992) Current statistical approaches to clinical pathology data from toxicological studies. Toxicol Pathol **20**(3), Part 2.

Ward JM, Hardisty JF, Hailey JR & Street CS (1995) Peer review in toxicologic pathology. Toxicol Pathol **23**: 226–234.

Weil CS (1970) Significance of organ-weight changes in food safety evaluation. In: Metabolic Aspects of Food Safety, pp. 419–454. Academic Press, New York, N.Y.

Weil SC & McCollister DD (1963) Relationship between short- and long-term feeding studies in designing an effective toxicity test. Agr Food Chem **11**: 486–491.

Weingand K, Brown G, Hall R, Davies D, Gossett K, Neptun D, Waner T, MatsuzawaT, Salemink P, Froelke W, Provost JP, Dal-Negro G, Batchelor J, Nomura M, Groetsch H, Boink A, Kimball J, Woodman D, York M, Fabianson-Johnson E, Lupart M & Melloni E (1996) Harmonization of animal clinical pathology testing in toxicity and safety studies. The Joint Scientific Committee for International Harmonization of Clinical Pathology Testing.
Fund App Toxicol **29**(2): 198–201.

Whalan JE & Redden JC (1994) Interim Policy for Particle Size and Limit Concentration Issues in Inhalation Toxicology Studies. Health Effects Division, Office of Pesticide Programs, US EPA, February 1.

WHO (1978) IPCS Environmental Health Criteria 6: Principles and methods for evaluating the toxicity of chemicals. World Health Organisation, Geneva.

WHO (1987) IPCS Environmental Health Criteria 70: Principles for the safety assessment of food additives and contaminants in food. World Health Organisation, Geneva.

WHO (1990) IPCS Environmental Health Criteria 104: Principles for the toxicological assessment of pesticide residues in food. World Health Organisation, Geneva.

WHO (1992) IPCS Environmental Health Criteria 141: Quality management for chemical safety testing. World Health Organisation, Geneva.

WHO (1997) Guidelines for Predicting Dietary Intake of Pesticide Residues (Revised). Prepared by the Global Environment Monitoring – Food Contamination Monitoring and Assessment Programme, in collaboration with the Codex Committee on Pesticide Residues.

WHO (1999) IPCS Environmental Health Criteria 210: Principles for the Assessment of Risks to Human Health from Exposure to Chemicals. World Health Organization, Geneva.

Whole Rat Catalogue (1983) USA, Harvard Bioscience.

Wilkinson CF & Killeen JC (1996) A mechanistic interpretation of the oncogenicity of chlorothalonil in rodents and an assessment of human relevance. Regul Toxicol Pharmacol **24**(1): 69–84.

Williams GM, Laspia MF & Dunkel VC (1982) Reliability of the hepatocyte primary culture/DNA repair test in testing of coded carcinogens and noncarcinogens. Mutat Res **97**: 359–370.

Zbinden G (1976) The role of pathology in toxicity testing. In: Progress in Toxicology 2, pp. 8–18, Springer-Verlag, New York, N.Y.

APPENDIX I

Commonly used abbreviations and acronyms*

TIME		**WEIGHT**	
d	Day	**bw**	Body weight
h	Hour	**g**	Gram
Min	Minute	**kg**	Kilogram
Mo	Month	**Mg**	Microgram
Wk	Week	**mg**	Milligram
S	Second	**Ng**	Nanogram
Yr	Year	**Wt**	Weight
Length		**Dosing**	
Cm	Centimetre	**Id**	Intradermal
M	Metre	**Im**	Intramuscular
Mm	Micrometre	**Inh**	Inhalation
Mm	Millimetre	**Ip**	Intraperitoneal
Nm	Nanometre	**Iv**	Intravenous
		po	Oral
		sc	Subcutaneous
		MG/KG BW/D	Milligram per kilogram body weight per day
Volume		**Concentration**	
L	Litre	**M**	Molar
Ml	Millilitre	**Ppb**	Parts Per Billion
Ml	Microlitre	**Ppm**	Parts Per Million

Clinical chemistry & haematology

A/G Albumin/globulin ratio
ALT Alanine aminotransferase (SGPT)
AP Alkaline phosphatase
AST Aspartate aminotransferase (SGOT)
BUN Blood urea nitrogen
ChE Cholinesterase
CPK Creatine phosphatase (phosphokinase)
GGT Gamma-glutamyl transferase
Hb Haemoglobin
Hct Haematocrit
LDH Lactate dehydrogenase
MCH Mean corpuscular haemoglobin
MCHC Mean corpuscular haemoglobin concentration
MCV Mean corpuscular volume
NTE Neurotoxic target esterase
PCV Packed cell volume (Haematocrit)
PT Prothrombin time

RBC Red blood cell/erythrocyte
T_3 Triiodothyroxine
T_4 Thyroxine
TSH Thyroid stimulating hormone (thyrotropin)
WBC White blood cell/leucocyte
WBC–DC White blood cells–Differential count

ANATOMY
CNS Central nervous system
GIT Gastro-intestinal tract

CHEMISTRY
DMSO Dimethyl sulfoxide
GC Gas chromatography
GLC Gas liquid chromatography
HPLC High performance liquid chromatography
MS Mass spectrometry
RIA Radio-immuno-assay
TLC Thin layer chromatography

TERMINOLOGY
ADI Acceptable Daily Intake
ARfD Acute Reference Dose
GLP Good Laboratory Practice
LOEL Lowest-Observed-Effect Level
MRL Maximum Residue Limit or Level
NOEL No-Observed-Effect Level
NOAEL No-Observed-Adverse-Effect Level
op Organophosphorous pesticide
RfD Reference Dose

ORGANISATIONS & PUBLICATIONS
CAC Codex Alimentarius Commission
ECETOC European Chemical Industry Ecology and Toxicology Centre
FAO Food and Agriculture Organisation of the UN
IARC International Agency for Research on Cancer
IPCS International Programme on Chemical Safety
JECFA FAO/WHO Joint Expert Committee on Food Additives
JMPR Joint Meeting on Pesticide Residues
NCI National Cancer Institute
NTP National Toxicology Program
OECD Organisation for Economic Cooperation and Development
US EPA United States Environmental Protection Agency
WHO World Health Organisation

For a comprehensive list of standard terms and abbreviations, see Appendix 1 of the OECD document *Guidance for Country Data Review Reports on Plant Protection Products and their Active Substances – Monograph Guidance* (OECD, 1998b).

APPENDIX II

Haematology, clinical chemistry and urinalysis parameters measured in chronic toxicity and carcinogenicity studies

(Lists adapted from Canadian pro forma)

HAEMATOLOGY PARAMETERS

Haematocrit (Hct) Haemoglobin (Hb) Leucocyte count (WBC) Erythrocyte count (RBC) Platelet count Blood clotting measurements (Thromboplastin time), (Clotting time) (Prothrombin time)	Leucocyte differential count* Mean corpuscular Hb (MCH) Mean corpuscular Hb concentration (MCHC) Mean corpuscular volume (MCV) Reticulocyte count

* Minimum required for carcinogenicity studies (only on control and high dose unless effects are observed) based on US EPA OPPTS or OECD test guidelines.

CLINICAL CHEMISTRY PARAMETERS*

ELECTROLYTES	OTHER
Calcium Chloride Magnesium Phosphorus Potassium Sodium	Albumin Blood creatinine Blood urea nitrogen (BUN) Total cholesterol Globulins Glucose Total bilirubin Total serum protein Triglycerides Serum protein electrophoresis
ENZYMES	
Alkaline phosphatase (AP) Cholinesterase (ChE) Creatine phosphatase (CPK) Lactate dehydrogenase (LDH) Alanine aminotransferase (ALT, also SGPT) Aspartate aminotransferase (AST, also SGOT) Gamma-glutamyl transferase (GGT) Glutamate dehydrogenase Sorbitol dehydrogenase	

* Not required for carcinogenicity studies based on US EPA OPPTS or OECD test guidelines.

Urinalysis parameters*

Appearance Volume Specific gravity pH Sediment (microscopic) Protein	Glucose Ketones Bilirubin Blood Nitrate Urobilinogen

* Not required for carcinogenicity studies based on US EPA OPPTS or OECD test guidelines.

The joint international committee established to provide advice on clinical pathology testing of laboratory animals in regulated toxicity and safety studies has published its conclusions, including recommendations on which parameters should be measured (Weingand *et al.*, 1996). While these have yet to be formally incorporated into national or international guidelines, they are noted here as follows:

In repeat-dose studies in rodents, clinical pathology testing is necessary at study termination. Interim study testing may not be necessary in long-duration studies provided it has been done in short-duration studies using dose levels not substantially lower than those used in the long-duration studies. For repeat-dose studies in non-rodents, clinical pathology testing is recommended at study termination and at least once at an earlier stage. In studies of two to six weeks' duration in non-rodents, testing is also recommended within seven days of the start of dosing, unless it compromises the health of the animals. If a study contains recovery groups, clinical pathology testing at study termination is recommended.

CLINICAL CHEMISTRY

The core clinical chemistry tests recommended are glucose, urea nitrogen, creatinine, total protein, albumin, calculated globulin, calcium, sodium, potassium, total cholesterol and hepatocellular and hepatobiliary tests. For hepatocellular evaluation, measurement of at least two of the following blood tests is recommended: alanine aminotransferase, aspartate aminotransferase, sorbitol dehydrogenase, glutamate dehydrogenase, total bile acids. For hepatobiliary evaluation, measurement of at least two of the following blood tests is recommended: alkaline phosphatase, gamma-glutamyltransferase, 5'-nucleotidase, total bilirubin, total bile acids.

Note: Cholinesterase determinations in blood and RBCs are required if there is evidence that the test chemical is likely to have an effect on this enzyme (as, for example, in the case of organophosphate and carbamate pesticides). Apart from Japanese Ministry of Health and Welfare (JMHW) guidelines for pharmaceuticals and Japanese Ministry of International Trade and Industry (MITI) guidelines for industrial chemicals, subchronic or chronic study guidelines do not routinely require tests for blood or serum triglycerides.

HAEMATOLOGY

The core haematology tests recommended are total leucocyte (white blood cell) count, absolute differential leucocyte count, erythrocyte (red blood cell) count, evaluation of red blood cell morphology, platelet (thrombocyte) count, haemoglobin concentration, haematocrit (or packed cell volume), mean corpuscular volume, mean corpuscular haemoglobin and mean corpuscular haemoglobin concentration. If facilities for the automated counting of reticulocytes are unavailable, blood smears from each animal should be prepared for reticulocyte counts. Bone marrow cytology slides from each animal should be prepared at termination. The minimum recommended haemostasis tests are prothrombin time, activated partial thromboplastin time (or appropriate alternatives) and platelet count.

NOTE: Blood smears are especially useful when animals show evidence of anaemia from other haematology evaluations. Although most of the 1998 US EPA OPPTS toxicology guidelines do not require smears, blood cells are usually examined during WBC differential counts and abnormal findings reported.

URINALYSIS

Urinalysis should be conducted at least once during a study. For routine urinalysis, an overnight collection (approximately 16 h) is recommended. The core recommended tests include assessment of urine appearance (colour and turbidity), volume, specific gravity or osmolality, and pH, and either the quantitative or semi-quantitative determination of total protein and glucose.

APPENDIX III

Organs for organ weight determination and for histopathological examination

Organs weighed	Tissues Examined		
adrenals*# brain*# epididymides* gonads# heart* kidneys*# liver*# ovaries* spleen* testes* thymus* uterus*	adrenals*# accessory genital organs# aorta* bone marrow*# brain (3 levels)*# caecum*# colon*# duodenum*# epididymides* eyes# eyes (retina, optic nerve)* femur (including joint)# gall bladder (when present)* gonads# heart*#	ileum*# jejunum*# kidneys*# larynx* liver*# lungs*# lymph nodes*# female mammary gland*# muscle (skeletal)* musculature# nerve (peripheral)*# nose* oesophagus*# ovaries* pancreas*# pharynx* pituitary*#	prostate* rectum*# salivary gland*# seminal vesicle* skin*# spinal cord (cervical, thoracic, lumbar)*# spleen*# sternum# stomach*# testes* thymus# thyroid (with parathyroid)*# trachea*# urinary bladder*# uterus*# gross lesions*#

Minimum OECD requirements; thyroid (with parathyroid) weights are also required in studies in non-rodent species. In inhalation studies, the entire respiratory tract should be examined, including nose, pharynx and larynx. Full histopathology should be carried out on the preserved organs and tissues of all animals in the control and high dose groups and all animals that die or are killed during a study. If treatment-related changes are observed in the high dose group, these examinations should be extended to animals in all other dosage groups.

* Minimum US EPA requirements; US EPA chronic toxicity and carcinogenicity guidelines also require the lungs to be weighed if the test substance is administered by inhalation. In inhalation studies, the entire respiratory tract, including nose, pharynx, larynx and paranasal sinuses, should be examined and preserved. Thyroid (with parathyroid) weights are required in chronic studies in non-rodent species. Full histopathology should be carried out on the preserved organs and tissues of all animals in the control and high dose groups and all animals that die or are killed during a study. Gross lesions and target organs in all animals should also be examined.

APPENDIX IV

Some applicable statistical tests*

Tests for non-normality
Chi-square test
Kolmogorov–Smirnov test
Shapiro–Wilk test

Tests for homogeneity of variance
Bartlett's test
Levene's test

Assumed normally distributed data

1. Overall tests

Analysis of variance (ANOVA***,*** *fixed effects model; Model 1 ANOVA most common, but other models may be encountered*)
Analysis of covariance (ANCOVA)
Pearson's correlation coefficient
Linear regression *(Tests the relationship of the two parameters in dependent data. Used to examine trends in dose effects; tests the significance of the regression slopes,* e.g. *whether responses vary according to sex.)*

2. Pairwise comparisons

Duncan's multiple range test
Dunnett's t-test (*Compares control to each other group mean)*
Scheffe's test *(Multiple comparison; less power than Newman–Keuls multiple range test)*
Williams' t-test
Student's t-test
Fisher's least significant difference (LSD) test

Non-parametric procedures (percentage values, ranks, etc.)

Kendall's coefficient of rank correlation
Mann–Whitney U-test *(Analogous to t-test)*
Wilcoxon signed-rank test *(Paired data, matched paired data)*
Kruskal–Wallis ANOVA
Distribution-free multiple comparisons tests (*e.g. Dunn's test, Shirley's test*)
Jonckheere's test

*The authors wish to acknowledge Dr Joseph Haseman of NIEHS, Research Triangle Park, NC USA, for his valuable assistance in the preparation of this appendix.

Quantal data (mortalities, pathology findings, etc.)

Fisher's exact test
R x C chi-square test
Litchfield & Wilcoxon test (*Confidence limits of ED50, etc.*)

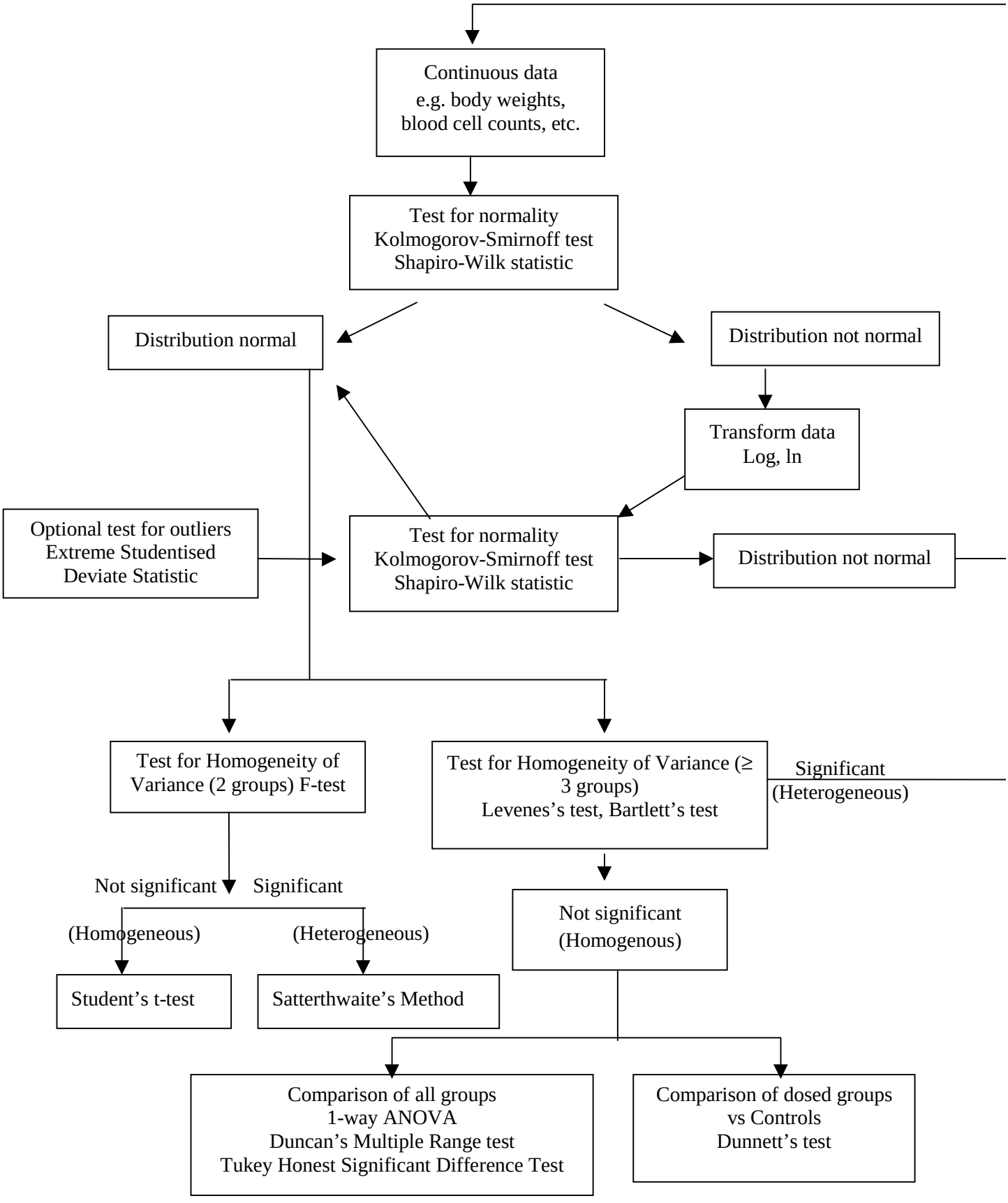
Multivariate methods

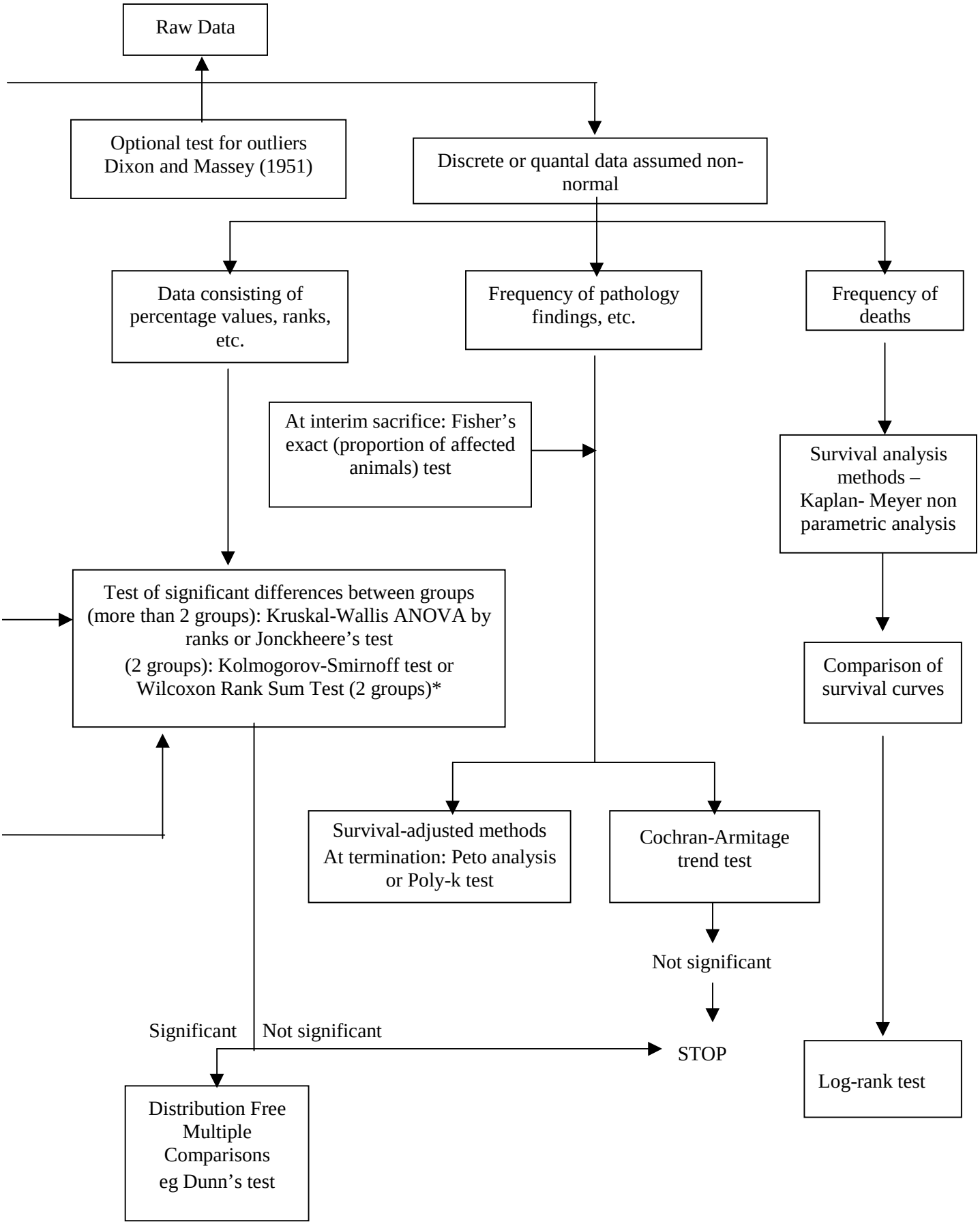
Hotellings T^2
MANOVA

Survival-adjusted procedures for analysis of carcinogenicity data

Peto's test (*See Peto* et al., *1980*)
Life table test (*For fatal cancers or cancers with observable onset times*)
Hoel–Walberg procedure
Logistic regression
Poly-k test (*Portier & Bailer,1989*)

A statistical decision tree is laid out on the next two pages, summarising the most common statistical procedures used for analysis of data in toxicology studies, together with their most likely outcomes. The pages should be read side by side.


Continuous data
e.g. body weights,
blood cell counts, etc.
Test for normality
Kolmogorov-Smirnoff test
Shapiro-Wilk statistic
Distribution normal
Distribution not normal
Transform data
Log, ln
Optional test for outliers
Extreme Studentised
Deviate Statistic
Test for normality
Kolmogorov-Smirnoff test
Shapiro-Wilk statistic
Distribution not normal
Test for Homogeneity of
Variance (2 groups) F-test
Test for Homogeneity of Variance (≥
3 groups)
Levenes's test, Bartlett's test
Significant
(Heterogeneous)
Not significant
Significant
(Homogeneous)
(Heterogeneous)
Not significant
(Homogenous)
Student's t-test
Satterthwaite's Method
Comparison of all groups
1-way ANOVA
Duncan's Multiple Range test
Tukey Honest Significant Difference Test
Comparison of dosed groups
vs Controls
Dunnett's test



* Also known as the Mann–Whitney U-test

APPENDIX V

Sources of historical control data on tumours in laboratory animals

Given the sensitivity of tumour formation to genetic, temporal and environmental factors, evaluators are advised to use historical control data from age-matched animals of the same strain, and from the same laboratory, as used in the study under review, and generated no more than five years from the study date. However, such data are not always available or may be too limited in scope. Evaluators may also be interested in examining changes in the rate of spontaneous tumour formation over time. The following references may be of help in this regard.

Baum A, Pohlmeyer G, Rapp KG & Deerberg F (1995) Lewis rats of the inbred strain LEW/Han: life expectancy, spectrum and incidence of spontaneous neoplasms. Exp Toxic Pathol **47**(1): 11–18.

Bode G, Hartig F, Hebold G & Czerwek H (1985) Incidence of spontaneous tumors in laboratory rats. Exp Pathol **28**(4): 235–243.

Bomhard E & Mohr U (1989) Spontaneous tumors in NMRI mice from carcinogenicity studies. Exp Pathol **36**(3): 129–145.

Chandra M & Frith CH (1992) Spontaneous neoplasms in aged CD-1 mice. Toxicol Lett **61**(1): 67–74.

Chandra M & Frith CH (1992) Spontaneous neoplasms in B6C3F1 mice. Toxicol Lett **60**(1): 91–98.

Chandra M & Frith CH (1992) Spontaneous neoplasms in aged control Fischer 344 rats. Cancer Lett **62**(1): 49–56.

Chandra M, Riley MG & Johnson DE (1992) Spontaneous neoplasms in aged Sprague–Dawley rats. Arch Toxicol **66**(7): 496–502.

Goodman DG, Ward JM, Squire RH, Chu KC & Linhart MS (1979) Neoplastic and nonneoplastic lesions in aging F344 rats. Toxicol Appl Pharmacol **48**(2): 237–248.

Gopinath C (1994) Spontaneous tumour rates: their use to support rodent bioassays. Toxicol Pathol **22**(2): 160–164.

Haseman JK, Arnold J & Eustis SL (1990) Tumor incidences in Fischer 344 rats. In: Boorman GA, Eustis SL, Montgomery C & Elwell M (Eds) Pathology of the Fischer 344 rat, pp. 555–564. Academic Press, San Diego, Cal.

Haseman JK, Elwell MR & Hailey JR (1999) Neoplasm incidences in B6C3F1 mice: NTP historical data. In: Maronpot RR (Ed) Pathology of the B6C3F1 Mouse, pp. 679–689. Cache River Press, Vienna, Ill.

Haseman JK, Hailey JR & Morris RW (1998) Spontaneous neoplasm incidences in Fischer 344 rats and B6C3F1 mice in two year carcinogenicity studies: A National Toxicology Program update. Toxicol Pathol **26**(3): 428–441.

Haseman JK, Huff JE, Rao GN, Arnold JE, Boorman GA & McConnell EE (1985) Neoplasms observed in untreated and corn oil gavage control groups of F344/N rats and (C57Bl/6N x C3H/HeN)F1 (B6C3F1) mice. J Natl Cancer Inst **75**(5): 975–984.

Kaspareit-Rittinghausen J, Deerburg F & Rapp K (1987) Mortality and incidence of spontaneous neoplasms in BDII/Han rats. Zeitschrift für Versuchstierkunde **30**(5–6): 209–216.

Lohrke H, Hesse B & Goerttler K (1984) Spontaneous tumors and lifespan of female NMRI mice of the outbred stock Sut:NMRT during a lifetime study. J Cancer Res Clin Oncol **108**(2): 192–196.

Maekawa A, Kurokawa Y, Takahashi M, Kokubo T, Ogiu T, Onodera H, Tanigawa H, Ohno Y, Furukawa F & Hayashi Y (1983) Spontaneous tumors in F-344/DuCrj rats. Gann. **74**(3): 365–372.

Maita K, Hirano M, Harada T, Mitsumori K, Yoshida A, Takahashi K, Nakashima N, Kitazawa T, Enomoto A, Inui K *et al.* (1988) Mortality, major cause of moribundity, and spontaneous tumors in CD-1 mice. Toxicol Pathol **16**(3): 340–349.

Maita K, Hirano M, Harada T, Mitsumori K, Yoshida A, Takahashi K, Nakashima N, Kitazawa T, Enomoto A, Inui K *et al.* (1987) Spontaneous tumors in F344/DuCrj rats from 12 control groups of chronic and oncogenicity studies. J Toxicol Sci **12**(2): 111–126.

NTP Historical Control Information database. http://ehis.niehs.nih.gov/ntp/docs/ntp_hcrs.html

Poteracki J & Walsh KM (1998) Spontaneous neoplasms in control Wistar rats: a comparison of reviews. Toxicol Sci **45**(1): 1–8.

Pour P, Althoff J, Salmasi SZ & Stepan K (1979) Spontaneous tumors and common diseases in three types of hamsters. J Natl Cancer Inst **63**(3): 797–811.

The Registry of Industrial Toxicology Animal-data (RITA), maintained by the Fraunhofer Institute, containing validated data on tumours and other proliferative lesions from over 11 000 control mice and rats. Further information about RITA can be obtained from the RENI Web site (see Subsection 3.1.6).

Rittinghausen S, Kaspareit J & Mohr U (1997) Incidence and spectrum of spontaneous neoplasms in Han:NMRI mice of both sexes. Exp Toxicol Pathol **49**(5): 347–349.

Sass B, Rabstein LS, Madison R, Nims RM, Peters RL & Kelloff GJ (1975) Incidence of spontaneous neoplasms in F344 rats throughout the natural life-span. J Natl Cancer Inst **54**(6): 1449–1456.

Sher SP, Jensen RD & Bokelman DL (1982) Spontaneous tumors in control F344 and Charles River–CD rats and B6C3HF1 mice. Toxicol Lett **11**(1-2): 103–110.

Solleveld HA, Haseman JK & McConnell EE (1984) Natural history of body weight gain, survival, and neoplasia in the F344 rat. J Natl Cancer Inst **72**(4): 929–940.

Takaki Y, Kitamura S, Uekusa T, Honma S, Aze Y, Wakabayashi K, Kuwubara N & Fukuda Y (1989) Spontaneous tumors in F-344/Jcl rats. J Toxicol Sci **14**(3): 181–195.

Tamano S, Hagiwara A, Shibata MA, Kurata Y, Fukushima S & Ito N (1988) Spontaneous tumors in aging (C57BL/6N x C3H/HeN)F1 (B6C3F1) mice. Toxicol Pathol 16(3): 321–326.

Ward JM (1983) Background data and variations in neoplasm rates of control mice and rats. Prog Exp Neopl Res 26: 241–258.

Ward JM, Goodman DG, Squire RA, Chu KC & Linhart MS (1979) Neoplastic and nonneoplastic lesions in aging (C57Bl/6N x C3H/HeN)F1 (B6C3F1) mice. J Natl Cancer Inst 63(1): 849–854.

APPENDIX VI

Sample evaluation 1

The following is an actual assessment report, although the names of the chemical, the sponsor, the conducting laboratory and the study, as well as the laboratory codes, have been altered to ensure confidentiality.

Davies LP & Dempsey J (1999): Oncogenicity study of Chemical S Administered in the diet to CD-1 mice for 18 months. Study no. MJ99117 Report no. 93099 Lab: Wagner Environmental Health Laboratory, Canberra, Australia Sponsor: Jenner Chemical Pty Ltd, Adelaide, Australia Expt date: 20 Apr 1996 – 15 Nov 1997 Unpublished Report date: 13 March 1999 (QA: Yes GLP: OECD, USEPA, EEC and Japan MAFF; Test guidelines US EPA Subdivision F 83-2, OECD 451, EEC 1988 & Japan MAFF 59 NohSan 4200)

[The evaluation does not make use of paragraph headings, unlike the example in Appendix VII.]

Chemical S was administered in the diet for approximately 18 months to CD-1 strain mice (Charles River Laboratory, Portage, MI USA, initially aged approximately 6 wk and weighing 30 [M] or 24 [F] g). An interim sacrifice was performed with 10/60 mice/sex/group at month 12. Mice were housed individually in suspended steel cages and were allowed *ad libitum* access to water and feed. Animal housing and husbandry were stated to be in accordance with the provisions of "Guide to the care and use of laboratory animals" (USPHS-NIH Publication No. 86-23). Weekly, the test material was mixed into the diet to form a premix, and this mixture was used to prepare the test diets. A negative control group received plain diet. Prepared diets were stored under refrigeration or kept in the animal room until use. The stability of chemical S and the stability and homogeneity of chemical S in the dietary mixtures were analysed by HPLC at study commencement, at months 6 and 11, and at termination.

Animals were observed twice daily and examined weekly for clinical signs of toxicity. Feed consumption and body weight were also measured weekly until wk 13 and at least every 4 wk thereafter. At interim sacrifice and study termination, mice were asphyxiated with CO_2 and a fasted blood sample was obtained from the posterior vena cava. Haematology and clinical chemistry parameters were measured in 10 mice/sex/group and are listed at the end of the evaluation. Gross pathological examination was performed on all animals and the adrenals, brain, kidneys, liver, spleen and testes were weighed. Tissues examined histologically under light microscopy are listed at the end of the evaluation.

The following parameters were analysed statistically: body weight, body weight change, food consumption (2-tailed Dunnett's Multiple Comparison Test); incidence of histopathological lesions (1-tailed Fisher's Exact Test with Bonferroni Inequality Procedure); haematology, clinical chemistry, terminal body weights, absolute and relative (to body and brain) organ weights [Normality Test or if n<20 Bartlett's Test, followed by either parametric (Dunnett's Test and linear regression) or non-parametric (Kruskall–Wallis, Jonckheere's &/or Mann–Whitney Tests) procedures]; mortality (2-tailed SAS LIFETEST) and organ weight outliers (Grubb's Test).

[Note that statistical analysis of clinical signs and gross pathology was not performed by the study authors, and the incidence of histopathology lesions was not subjected to a trend test, despite the desirability of this procedure. Should the study statistician have used the Bonferroni inequality procedure when testing the significance of histopathological lesions?]

Results

Chemical S was found to be stable throughout the study in its pure state, and was also stable in the dietary preparations for 14 d at room temperature and for 37 d when stored frozen. The mean dietary concentrations throughout the study were 30, 716, 3020 and 6790 ppm. Homogeneity analyses revealed that the within-batch coefficient of variation in test article concentration was approximately 6% at 30 and 7000 ppm. No data were presented for the intermediate concentrations.

The mean achieved doses were [M/F] 4.0/6.5, 93/153, 394/635 and 944/1388 mg/kg bw/d at the 30, 700, 3000 and 7000 ppm feeding levels, respectively. One male control died accidentally. Including this animal, the numbers of premature deaths were [M/F, control to high dose] 11/10, 8/13, 6/10, 9/11 and 12/10. Thus, survival was not compromised by the test compound. Abdominal swelling or distension, urine stained fur and abnormal penile erections were observed more frequently among 7000 ppm males (see Table). These findings were attributed by the study authors to urinary calculi. There were no treatment-related clinical signs in females.

[The evaluator establishes a correlation between clinical observations and pathological effects.]

No adverse effects were noted on mean body weight in either sex throughout the study. Cumulative body weight gain was up to 20% greater in some treated groups than controls, but statistical significance was only attained by 700 and 3000 ppm males and 700 and 7000 ppm females, on one or two occasions during the study. There were no statistically significant differences at study termination (see Table). A parallel trend was observed in feed consumption of the treated groups. The increases above control values were approximately 5–20%, and attained significance on 6 occasions in 700 ppm females and one or two occasions in the remaining treated groups (except 30 ppm males and 7000 ppm females), but were not statistically significant at study termination. These findings are therefore considered not to be of toxicological significance.

[The reasoning behind the evaluator's judgement is made clear; in considering whether there were treatment-related effects on body weight and food consumption, the evaluator stresses the importance of statistical analysis and consistency with respect to time and dose–response relationship.]

Haematology revealed a statistically dose-related increase in platelet (PLT) count in treated females at 12 and 18 months, which achieved significance ($p<0.05$ vs control) at the 7000 ppm feeding level at termination (see Table). Males were unaffected. Although the study authors attributed this finding to the heparin anticoagulant used in the blood sampling procedure, the elevated PLT count in 7000 ppm females is considered to be biologically significant because it was consistent with a trend observed at both the interim and terminal sacrifices.

[Assuming there was no difference between the treatment of blood samples from the 7000 ppm females and the other groups. A statement regarding methods would have been helpful here.]

Treated groups varied considerably from controls with respect to white blood cell and neutrophil counts (WBC, NEU; depressed in males and elevated in females), lymphocyte and monocyte counts (LYM, MON; elevated in females), and large unstained cell counts (LUC; depressed in males, no consistent trend in females). At interim sacrifice, there was a statistical dose-related increase in

basophil count (BAS) among treated females, achieving significance ($p<0.05$ vs control) in the 7000 ppm group. However, the finding was not repeated at termination (see Table). Eosinophil (EOS) counts were elevated in both sexes at interim sacrifice but the trend did not persist at termination (see Table). Consistent with the 90-day study performed at the same facility (Study No. ML-93-111), the group standard deviations were large and these findings are attributed to individual variation.

[For some parameters, the text does not report the absolute and percentage changes relative to control, contrary to these guidance notes (Section 5.2, point 12), but this information is provided in the table where needed to enhance detail and transparency. Repetition is therefore reduced. The text mentions some "noise", or "random" perturbations, in haematology (see Subsection 3.1.4), but this is not given in the table, as clearly negative data would have added little to the evaluation. The evaluator uses between-study consistency when discussing the biological significance of some parameters; however, inclusion of standard deviations in the table would have improved interpretability for the reader.]

The principal finding from clinical chemistry was a statistically dose-related elevation in blood urea nitrogen (BUN) in treated males at termination, which was significant ($p<0.01$ vs control) at 7000 ppm (see Table). A similar but non-significant trend was observed in treated females. An exceptionally high alanine aminotransferase (ALT) activity was detected in a single female from the 700 ppm group at termination. The affected mouse also had elevated aspartate aminotransferase (AST) activity and BUN. This was sufficient to cause a statistically dose-related trend in ALT activity among treated females (see Table), but the trend is probably an artifact because similar peturbations did not occur at higher doses. At interim sacrifice, significant depression in alkaline phosphatase (AP) activity occurred in 30 and 7000 ppm females, and 700 ppm males showed significantly increased AST activity. These findings are not considered to be biologically significant, given that they were not repeated at termination.

[Again, the biological significance of some experimental observations is examined and dismissed because of a lack of temporal consistency and dose-relatedness. This paragraph also illustrates the potential impact of a single "outlying" observation, in this case probably arising from spontaneous disease processes.]

At termination, absolute testis weight and testis:body and testis:brain weight ratios were slightly depressed in the 7000 ppm male group, with a significant dose–response relationship with respect to the absolute and relative (to brain) values (see Table). In the 3000 ppm male group, liver:body weight ratio was increased (significantly) to 112% of control at interim sacrifice, but the finding is not ascribed to treatment because no statistically significant effect was seen at 7000 ppm, or among treated males at termination. Similarly, a significant increase in kidney:brain weight ratio in 30 ppm males at termination is not considered treatment-related in the absence of a response among higher dose groups.

[In this paragraph, the same criteria are applied to exclude some observations from treatment-relatedness, while the decreased testis weights cannot be excluded because both absolute and relative weights were affected and there was a dose–response relationship (see Subsection 3.1.6)]

At interim and terminal sacrifices and in premature decedents, calculi were found in the urinary bladder of many 3000 and 7000 ppm males, in addition to an increased incidence of enlargement/distension of the ureter, bladder and/or renal pelvis and renal atrophy at 7000 ppm (see Table). Lymph node enlargement was seen more frequently in 7000 ppm females than other groups.

[As most of the detail is tabulated, this text is sufficient; duplication by text and table is kept to a minimum. Note the division of the text into paragraphs that deal with separate aspects of the study.]

Histopathological examination yielded further evidence of bladder and renal injury in 3000 and 7000 ppm males at both sacrifices. The renal changes may have been associated with obstruction caused by the presence of calculi in the lower urinary system (see Table). The study pathologist suggested the calculi may have been composed at least in part of the test material and/or its metabolites. Histological abnormalities consisted principally of renal atrophy, dilation of the renal pelvis, inflammation of the bladder, bladder epithelial mucosal hyperplasia and squamous metaplasia. The incidences of these findings were dose-related and statistically significant. Hyperplasia of the mucosal epithelium was characterised by a generally diffuse, increased thickness of the transitional cell epithelium from the normal 1–3 cell layers to 5–8 or even 10 layers. Occasionally, it was accompanied by focal ulceration of the mucosal surface. Squamous metaplasia of the transitional cell epithelium was characterised by the replacement of the normal mucosal surface with foci of well-differentiated keratinising squamous epithelium. Female mice were not affected by calculus formation, other than a possible slight increase in the rate of renal tubule mineralisation (see Table).

[This paragraph opens by correlating the gross and histological observations on the urinary system. Note that the evaluator makes it clear when he/she is repeating a statement made by the study author. The description of renal and bladder abnormalities is detailed but concise, and flows well because numerical data have been placed in the table.

The US EPA and the Canadian PMRA require additional tables of organ weight changes and tumour incidences; these should include measurements that do not indicate a treatment-related effect.]

The incidence of mesenchymal tumours of the urinary bladder was increased from zero among male controls and the lower dose groups, to 1/60 and 5/60 in the 3000 and 7000 ppm male groups, respectively. One control and one 7000 ppm female were also found with the tumour. The mesenchymal tumours consisted of clusters of large pleomorphic cells within the submucosa, surrounded by inflammatory cells, oedema and fibrosis. They did not penetrate the epithelial basal lamina or serosal surface of the bladder and metastases were not observed. All 5 tumour-bearing 7000 ppm males had bladder calculi and hyperplasia of the transitional cell epithelium, and 2 of these mice also displayed squamous mucosal metaplasia.

[Neoplastic findings are described succinctly and correlated with non-neoplastic findings. This helps the reader establish a causal association between the calculi, injury to the urinary tract epithelium and development of bladder tumours.]

The 700 ppm female group had a statistically significantly increased incidence of lymphoma and/or lymphoma or mononuclear cell infiltrate in the spleen, urinary bladder, lung and pancreas. However, the incidences at 3000 and 7000 ppm were lower than at 700 ppm, and the overall number of lymphoma-bearing mice was not affected by treatment (see Table). All other neoplastic and non-neoplastic histological abnormalities were neither statistically nor biologically significant.

[Historical control data, assuming any were available, may have assisted in the interpretation of the apparently high incidence of lymphoma in female mice at termination. However, from the mortality data (in the second paragraph under “Results”), it is clear shortened life span is not responsible for the diminished incidence of lymphoma among 7000 ppm females relative to the 700 and 3000 ppm groups (see Section 5.7).]

Based on renal and bladder lesions associated with urinary calculi at 3000 ppm and above in males, the NOEL was 700 ppm (93 mg/kg bw/d).

[The term "LOEL" is not used, but the LOEL itself is apparent in the concluding sentence, and the evaluator makes the basis on which the NOEL is established perfectly clear. Evaluators from some agencies would also state whether they considered the study adequate or inadequate.]

SELECTED FINDINGS AMONG MICE, 18-MONTH DIETARY STUDY (MJ99117)

FEEDING LEVEL (ppm)	0	30	700	3000	7000	0	30	700	3000	7000
			MALES					FEMALES		
NUMBER IN GROUP 60	60	60	60	60	60	60	60	60	60	
CLINICAL SIGNS (no. affected)										
Urine stained fur 14	2	3	12	1	0	1	0	0		
Abnormal penile erection	2	2	1	1	6	–	–	–	–	–
Abdominal swelling 2	1	1	3	7	3	2	3	4	3	
NUMBER IN GROUP 39	44	44	41	38	42	38	40	41	42	
TERMINAL BODYWT (g)	45	46	47	47	45	38	39	40	39	40
BODYWT. GAIN (g)	15.1	15.4	16.2	16.3	14.4	14.3	15.1	16.4	15.3	15.6
NUMBER IN GROUP 10	10	10	10	10	10	10	10	10	10	
HAEMATOLOGY (mean values)										
BAS (thousand / mm^3), interim*	0.013	0.016	0.010	0.012	0.160	0.008	0.012	0.032	0.013	0.021
BAS (thousand / mm^3), terminal	0.021	0.020	0.011	0.019	0.010	0.017	0.018	0.017	0.027	0.019
EOS (thousand / mm^3), interim	0.104	0.133	0.070	0.129~	0.191	0.062	0.079	0.153	0.065	0.120
EOS (thousand / mm^3), terminal	0.121	0.181	0.122	0.108	0.076	0.106	0.127	0.114	0.131	0.101
PLT (thousand / mm^3), interim@	230	312	235	154	344	195	320	305	379	499
PLT (thousand / mm^3), terminal@@	481	601	349	572	549	248	437	402	507	684~
CLINICAL BIOCHEMISTRY (mean values)										
ALT (IU/L) , terminal @	22	27	27	24	32	45	26	124	67	45
BUN (mg/dL), terminal ##	23	28	23	25	37~~	15	16	17	17	19
NUMBER IN GROUP 39	41	43	40	38						
ORGAN WEIGHTS (at termination)										
Testis (absolute, g)# 0.25	0.25	0.25	0.25	0.23	–	–	–	–	–	
Testis (rel. to bodywt., %)	0.64	0.64	0.62	0.63	0.60	–	–	–	–	–
Testis (rel. to brain wt., %)#	45	46	44	46	42	–	–	–	–	–

FEEDING LEVEL (ppm)	0	30	700	3000	7000	0	30	700	3000	7000
			MALES					FEMALES		
NUMBER IN GROUP 60	60	60	60	60	60	60	60	60	60	
GROSS PATHOLOGY (incidence)										
Kidney - mass / nodule	1	0	0	2	3	1	0	0	0	0
- dilated pelvis	3	6	2	4	21	0	0	1	1	1
- atrophy / small	0	1	0	1	5	0	0	0	0	0
Urinary bladder - calculus	1	0	0	21	41	0	0	0	0	1
- enlarged	4	3	3	5	21	0	0	0	0	0
- thick walled	0	1	1	3	4	0	0	0	0	0
Ureter - calculus 0	0	0	0	1	0	0	0	0	0	
- dilated/distended	0	2	0	2	8	0	0	0	0	0
HISTOPATHOLOGY (incidence)										
Kidney - tubular adenoma	0	0	0	0	1	0	0	0	0	1
- atrophy 2	0	0	3	11~	5	1	2	2	2	
- dilated pelvis	5	0	2	4	27~~	0	0	1	1	1
- necrosis of papilla	0	0	0	2	2	0	0	0	0	0

- tub. mineralisation	38	7	39	34	33	2	1	1	5	6
Urinary bladder - calculus	1	0	0	4	10˜	0	0	0	0	0
- dilatation	5	4	3	8	20˜˜	0	0	0	0	0
- hyperplasia	4	1	2	25˜˜	41˜˜	0	0	0	0	1
- inflammation	2	3	2	23˜˜	41˜˜	0	0	0	0	1
- metaplasia	0	0	0	1	8˜	0	0	0	0	0
- ulceration	0	1	0	3	4	0	0	0	0	0
- lymphoma infil.	1	0	0	1	0	1	0	8˜	4	3
- mesenchymal tumour	0	0	0	1	5	1	0	0	0	1
Lymphoma 1 –1	3	1	9	–	12	15	10			

*significant ($p<0.05$) for 7000 ppm females vs control, and linear dose-related trend ($p<0.05$)
˜significant ($p<0.05$) vs control and ˜˜significant ($p<0.01$) vs control
@significant ($p<0.05$) or @@significant ($p<0.01$) linear dose-related trend for females
#significant ($p<0.05$) or ##significant ($p<0.01$) linear dose-related trend for males

[Presentation of all significant findings in the same table has the advantage of compactness, but some agencies may prefer to present the various parameters in separate tables that accompany the text in which the results are described. This table would be more informative if a measure of variability (*i.e.* SD or SE) had been included with the body/organ weight, haematology and clinical biochemistry data. This may have helped to explain why statistical significance was found in interim EOS count in 3000 ppm males but not at 30 or 7000 ppm. Has the Bonferroni inequality procedure been applied correctly in the case of lymphoma infiltration of the urinary bladder?]

PARAMETERS EXAMINED IN 18-MONTH DIETARY STUDY IN MICE (MJ99117)

Haematology: (10 mice/sex at months 12 and 18): erythrocyte count (RBC), leukocyte count (WBC) and differential, neutrophil count, lymphocyte count, platelet count, haematocrit (HCT), haemoglobin (HGB), RBC indices (MCV, MCH, MCHC)

Clinical chemistry: (10 mice/sex at months 12 and 18): blood urea nitrogen (BUN), alanine aminotransferase activity (ALT), aspartate aminotransferase activity (AST), alkaline phosphatase (ALP), gamma glutamyl transferase (GGT)

Histopathology: Tissues retained — adrenals, aorta, brain, caecum, colon, duodenum, epididymides, oesophagus, eyes, femur with tibio-femoral joint, gross lesions, heart, ileum, jejunum, kidneys, liver with gallbladder, lung with mainstream bronchi, lymph nodes (mesenteric and submandibular), muscle (quadriceps femoris), nerve (sciatic), ovaries, pancreas, pituitary, prostate, rectum, salivary gland, seminal vesicles, skin (with mammary tissue), spinal cord (cervical, thorax and lumbar), spleen, sternum (with marrow), stomach, testes, thymus, thyroids/parathyroids, trachea, urinary bladder, uterus (corpus and cervix).

Tissues examined — all retained tissues from all controls and animals treated at 700, 3000 and 7000 ppm at terminal sacrifice; all retained tissues from all controls and 7000 ppm mice at interim sacrifice; lungs, kidneys, liver and urinary bladder from animals treated at 30, 700 and 3000 ppm at interim sacrifice; lungs, kidneys, liver and urinary bladder (males only) from 30 ppm mice at terminal sacrifice; urinary bladder from all 30 ppm males at terminal sacrifice; special staining and examination of urinary bladder from selected control females and 7000 ppm males.

Control and high dose animals only — spleen, testes, thyroids.

APPENDIX VII

Sample evaluation 2

The following is an actual assessment report, although the names of the chemical, the sponsor, the conducting laboratory and the study, as well as the laboratory codes, have been altered to ensure confidentiality.

Bartholomaeus A & Roberts G (1992) Chemical Z toxicity to dogs by repeated dietary administration for 52 weeks. Study No. F/920532 Lab: Churchill Research Centre Ltd., Oxfordshire, England Sponsor: Kemichefabrikwerk, Darmstadt, Germany Expt Date 17 Dec 1990 – 8 Jun 1992 Unpublished Report Date 7 July 1992. (QA: Yes GLP: OECD Test Guidelines: EPA FIFRA 83-1)

Conduct of the Study

Male and female beagle dogs in groups of 4 were treated with Chemical Z in the diet at 0, 50, 185, 700/500 ppm for 52 weeks. Following a treatment related death of a female at 700 ppm in week 11 this dose was reduced to 500 ppm for the remainder of the study. The food for the 50 ppm group was prepared to contain 57.5 ppm to adjust for an observed Chemical Z loss of 15.8% during and shortly after preparation. Approximate achieved intakes were (m/f): 0, 1.5/2, 6.6/6.7, 17/21 mg/kg bw/d. Toxicity was evaluated through: daily clinical observation, neurological examination, body weight and food consumption measurements, ophthalmology, haematology, serum biochemistry (excluding LDH, including alpha-Hydroxy Butyrate Dehydrogenase, T3, T4), urinalysis, organ weights (except salivary glands), gross pathology, histopathology (except lachrymal gland) and bone marrow examination. Food, water and the housing environment were controlled and monitored.

[The evaluator omits the source, initial age and body weight of the dogs, and does not state the batch number and purity of the test chemical, or describe the analytical methods used to check its homogeneity and stability. Details of the study observations are tabulated at the end of the evaluation, but the times at which the observations were made are not reported. Although minimal detail is provided regarding housing conditions, this is appropriate for a GLP study (see Section 5.2, point 6). However, no information is provided on statistical methods, which makes it difficult for the reader to judge their appropriateness.]

Results

<u>Mortality, Food Consumption and Body Weight Gain</u>

Two females died at 700 ppm. One female was sacrificed on the last day of week 11 after it was found prostrate with excessive salivation and with jerking limbs. These signs were considered to be treatment related and led to the reduction of the high dose from 700 to 500 ppm for the remainder of the study period for both sexes. Another female at 700 ppm with signs of polyarteritis, not considered treatment related, was sacrificed on the first day of week 6. A replacement dog was included at this treatment level for a complete 12 months. Weight gains in males were unaffected by treatment but in

females, animals treated at 185 and 500 ppm lost an average of 0.6 and 0.4 kg in the final 26 weeks of the study compared to a control weight gain of 0.5 kg. This weight loss was seen in all individual females at these two treatment levels. Food consumption was not significantly affected by treatment although a very slight reduction was noted at the two higher doses. On a per dog/d basis the intake at the highest treatment level was 384 g compared to a control value of 399 g. The reduced weight gains do not appear attributable to reduced food intake and presumably therefore reflect a reduction in food utilisation efficiency.

Clinical Signs and Ophthalmoscopy

Other than in the single female humanely sacrificed early due to treatment related prostration there were no clinical signs observed which related to treatment. No treatment related effects on the eyes were noted. No neurological effects were noted.

Urinalysis, Haematology and Clinical Biochemistry

Urinalysis did not reveal any treatment related changes. Minor variations in some haematological parameters were observed but these did not follow a dose response pattern, were slight in magnitude, and were generally within historical control ranges. MCV values were elevated in all animals treated at 500 ppm, significant in weeks 13, 26 and 52 in males but week 26 only in females. The MCV elevation did not exceed 6% in any group at any time point however, and was not considered to be toxicologically significant. Clinical biochemistry revealed elevated ALT and AP levels in males at 185 and 500 ppm in weeks 26 and 52, slightly elevated AP levels in all females at 500 ppm at weeks 13, 26 and 52, increased cholesterol in 500 ppm males at weeks 13, 26 and 52, and a decreased albumin level in 500 ppm males and females in weeks 26 and 52. Elevated AP and ALT in males was due to effects seen in one male at 185 ppm and in two at 500 ppm with the remaining animals at control levels. Liver histology confirmed hepatic damage in these animals that was more marked than in the remaining animals of their groups, and the elevated AP and ALT levels, although modest, are consequently considered to reflect treatment related damage. The magnitude of an observed decrease in albumin levels was small and individual animal levels were generally within the normal range for this parameter; nevertheless, the effect was statistically significant in both sexes at 500 ppm and in females at 185 ppm, and the difference to control was greater than 10% in each case. Increased cholesterol levels in males at 500 ppm were evident in individuals as well as the group value.

T3 and T4 levels in males at 500 ppm were increased but only the T3 levels were consistently increased significantly. No effects on thyroid gross or microscopic morphology were observed however.

[The evaluator uses transparent reasoning to dismiss the treatment-relatedness of inter-group differences in haematology and yet ascribe biological significance to the modest effect on serum albumin levels. Most numerical data have been placed in the table, making~~s~~ the text easier to read. Correlations — and the absence of correlations — between clinical biochemistry and organ histology are discussed (see Section 3.1.4).]

Altered Biochemical Parameters

		Males				Females			
		0	50	185	500 ppm	0	50	185	500 ppm
ALT ***mU/mL***	Week 13	23	–	–	–	–	–	–	–
	Week 26	27	33	70	86	–	–	–	–
	Week 52	29	23	69	51	–	–	–	–
AP *mU/mL*	Week 13	126	–	133	161	135	–	–	175
	Week 26	104	–	143	148	120	–	–	144
	Week 52	99	–	175	160	119	–	–	151
Albumin *g/dL*	Week 13		–	–	–	3.0	–	–	2.6*
	Week 26	2.7	2.8	2.7	2.4*	2.9	2.6	2.6	2.4*
	Week 52	3.0	–	–	2.6*	3.2	2.8	2.7*	2.7*
Cholesterol *mg/dL*	Week 13	140	–	150	205**	141	–	–	162
	Week 26	141	–	–	211**	152	126	171	174
	Week 52	142	–	–	214**	189	125	153	183
T3 *ng/dL*	Week 13	38	–	–	54*	–	–	–	–
	Week 26	44	58	47	67*	–	–	–	–
	Week 52	57	66	53	76*	–	–	–	–
T4 *μg/mL*	Week 13	1.4	1.8	<1.2	2.0*	–	–	–	–
	Week 26	1.4	1.7	1.5	1.9	–	–	–	–
	Week 52	1.5	1.6	1.4	2.1	–	–	–	–

* P<0.05, ** P<0.01, – comparable to control values or not toxicologically relevant

[Inclusion of repeated observations effectively conveys whether there are any time-related trends. Although it is not always necessary to show data from both sexes at every dose, evaluators have to exercise care in omitting data so as to avoid limiting the reader's ability to make judgements about the LOEL and the existence of dose–response relationships.]

Aberrant Individual Male Biochemistry Values (IU/L)

Animal #	Treatment	ALT Wk 26	ALT Wk 52	AP Wk 26	AP Wk 52
359	185 ppm	207	204	262	396
remainder (3)	185 ppm	24	24	103	102
367	500 ppm	196	23	171	112
371	500 ppm	98	145	186	289
remainder (2)	500 ppm	24	19	118	116

[*This table illustrates the contribution of "outlying" values to the group mean data; it makes clear the correlation between elevated serum ALT/AP activity in those individuals with more severe liver injury.*]

Gross Pathology and Organ Weights

Gross pathology revealed significant variations in only three animals. One female animal at 185 ppm had multiple red discoloured linear depressions in the mucosal surface of the body and antrum of the stomach, histologically confirmed as focal erosion of the superficial gastric mucosa. Two males at 185 ppm were found to have single flat, oval choleliths in the gall bladder (animals 359 & 363). Dog 359 had elevated AP and ALT in week 52; however, other dogs with elevated ALT and AP did not have choleliths and the finding was not repeated at 500 ppm. Given the isolated incidences of these effects and their absence at the highest dose, they are considered unlikely to be treatment related.

Differences in relative organ weights between treatment groups were generally small and did not reach statistical significance. Heart weights were slightly reduced in males and females at 500 ppm, and liver weights were increased. Decreased heart weights and, on an individual animal basis, higher liver weights did not correlate with histological alterations in these tissues. In males a slight increase in thyroid weights and a substantially reduced prostate weight was apparent at 500 ppm but these did not correlate with any histological alterations. Relative testicle weights were slightly decreased at 500 ppm. In females an apparent treatment related decrease in uterus and ovary weights was attributable to histologically observed variations in the oestrus cycle of these animals and was not attributed to treatment by the study author. The reduced prostate and testicle weights in males and the reduced ovary and uterus weights in females may however be a reflection of endocrine disruption, particularly at 500 ppm.

[The evaluator makes further correlations between different observations, and the text gives a highly transparent view of the findings when read in conjunction with the table below. Differences between the evaluator's and the study author's opinions are clearly stated. The possibility of endocrine disruption would be covered in greater detail in the discussion section of the evaluation report.]

Relative Organ Weights

Treatment (ppm)		**Heart**	**Liver**	**Prostate/ uterus**	**Thyroids**	**Gonads**
Males	**0**	0.94	3.25	0.086	0.0087	0.242
	50	0.90	3.03	0.073	0.0081	0.233
	185	0.92	3.67	0.074	0.0086	0.237
	500	0.82	3.57	0.045	0.0099	0.212
Females	**0**	0.88	3.4	0.260	–	0.0187
	50	0.93	3.31	0.088	–	0.0103
	185	0.9	3.53	0.055	–	0.0081
	500	0.83	4.08	0.076	–	0.0079

Histopathology and Bone Marrow Examination

Bone marrow, on microscopic examination, was normal in all animals. Treatment related effects were confined to the liver and spleen. A moderate degree of "foci of degenerate hepatocytes" was seen in 1/4 males and 3/4 females at 500 ppm and 1/4 females at 185 ppm. At 50 ppm and in control animals this effect was observed at only trace or minimal degrees. In only one dog, a male at 500 ppm, single cell necrosis was observed. Inflammatory cell infiltration around the central veins and occasionally the branches of the hepatic vein was seen in some of the dogs displaying foci of degenerate hepatocytes. Apparent increases in centrilobular fibrocytes were observed in 1/4 males at 185 ppm and in 3/4 males at 500 ppm. In males the incidence of aggregates of pigmented Kupffer cells and macrophages (0/4, 1/4, 2/4, 4/4, control to highest dose) in the liver was significantly increased and followed a dose response pattern. A similar pattern of increased pigmented macrophages was observed in the male spleen (1/4, 1/4, 3/4, 3/4). The same pattern was not observed in female liver (0/0, 1/4, 4/4, 2/3) although the incidence was significant at 185 ppm, and in the female spleen the pattern was inconclusive (2/4, 2/4, 2/4, 3/3). Increased pigmentation of Kupffer cells, and macrophages with Perls' positive material (*i.e.* containing ferric iron), may reflect increased iron turnover.

[*Here the evaluator has chosen to present data in the text rather than a table.*]

At 50 ppm no clinical or pathological evidence for treatment related effects were observed. The single incidence in one male and one female of aggregates of pigmented Kupffer cells and macrophages in the liver did not correlate with any other biochemical, histological or pathological parameter. As the

degree of pigmentation was designated as minimal the low incidence of this effect in isolation was not sufficient evidence to conclude a treatment-related effect at this treatment level.

The NOEL was 50 ppm (1.5 mg/kg bw/d) based on body weight loss in females and alterations in clinical chemistry and histology at the next higher dose.

Standard Test Parameters

Clinical Chemistry	Haematology	Urinalysis
albumin	clotting parameters	appearance
AP (alkaline phosphatase)	erythrocyte count	specific gravity
ALT (serum alanine aminotransferase – SGPT)	Hct (haematocrit)	glucose
AST (serum aspartate aminotransferase – SGOT)	Hb (haemoglobin)	ketones
bilirubin	leucocyte differential count	sediment
calcium	leucocyte total count	occult blood
chloride	platelet count	pH
cholesterol (total)	reticulocyte count	protein
ChE (cholinesterase activity)	MCH (mean corpuscular haemoglobin)	volume
creatinine (blood)	MCHC (mean corpuscular haemoglobin concentration)	bilirubin
CPK (creatine phosphokinase)	MCV (mean corpuscular volume)	urobilinogen
GGT (gamma glutamyl transferase)	blood smear	
globulin		
glucose		
phosphorous		
potassium		
protein (total)		
sodium		
triglycerides		
BUN (blood urea nitrogen)		

Organs Weighed		Tissues Examined	
adrenals	adrenals	heart	rectum
brain	aorta	ileum	salivary gland
gonads	blood smear	jejunum	seminal vesicle
heart	bone	kidneys	skin
kidneys	bone marrow	lacrimal gland	spinal cord (cervical,
liver	brain (3 levels)	liver	thoracic, lumbar)
lungs	caecum	lungs	spleen
pituitary	colon	lymph nodes	sternum
pancreas	duodenum	mammary gland	stomach
prostate	epididymides	muscle (smooth)	testes
spleen	eyes	muscle (skeletal)	thymus
thyroid	eyes (optic nerve)	nerve (peripheral)	thyroid
thymus	gall bladder	oesophagus	trachea
uterus	Harderian glands	ovaries	urinary bladder
	head (3 sections)	pancreas	uterus
		pituitary	vagina
		prostate	Zymbal's gland

Neurological Assessment (specific to this study and not part of the basic data requirements)

Behaviour and Gait **—** General examination of the animal whilst standing and moving with particular reference to strength and co-ordination.

Cranial Nerve Function		
head tilt	palpebral reflex L&R	facial muscle
ear movement	muscles of mastication	position of philtrum
blink reflex R&L	commissure of lips	pupillary light reflex R&L
jaw closure	eye position	open jaw resistance
strabismus	tongue	abnormal nystagmus
gag reflex	corneal reflex R&L	
Spinal Reflexes		
Muscle tone	flexor reflex	patellar reflex
crossed extensor reflex	triceps reflex	perineal reflex
Postural and attitudinal reactions		
Wheelbarrowing	tactile placing	thoracic hopping
visual placing	pelvic hopping	tonic neck reaction
Extensor postural thrust	righting reaction	

APPENDIX VIII

The IPCS Conceptual Framework for Cancer Risk Assessment

Framework Guidelines: Suggested Section Headings

1. Introduction

This section describes the cancer endpoint or endpoints that have been observed and identifies which of these is addressed in the analysis. (The nature of the framework is such that only one mode of action is analysed at a time; hence, for example, tumour types associated with a different mode of action, even if recorded in the same animals, will require separate framework analyses). However, where different tumours are induced by related mode of action, they are best addressed in a single analysis. It should also be noted that some modes of action will involve multiple contributing components.

2. Postulated mode of action (theory of the case)

This section comprises a brief description of the sequence of events on the path to cancer for the postulated mode of action of the test substance. This explanation of the sequence of events leads into the next section which identifies the events considered "key" (*i.e.* measurable) given the data base available for the analysis.

3. Key events

This section briefly describes the "key events" — *i.e.* measurable events that are critical to the induction of tumours as hypothesised in the postulated mode of action. To support an association, a body of experiments needs to define and measure an event consistently. Pertinent observations: *e.g.* tumour response and key events in same cell type, sites of action logically relate to event(s), increased cell growth, specific biochemical events, organ weight, histology, proliferation assays, hormone or other protein perturbations, receptor-ligand changes, DNA or chromosome effects, and cell cycle effects. For example, key events for tumours hypothesised to be associated with prolonged regenerative proliferation might be cytotoxicity as measured histopathologically and an increase in labelling index. As another example, key events for induction of urinary bladder tumours hypothesised to be due to formation of bladder stones composed primarily of calcium phosphate might include elevated urinary calcium, phosphate and pH and formation of bladder stones followed by irritation and regenerative hyperplasia of the urothelium.

4. Dose–response relationship

This section should detail the observed dose–response relationships and discuss whether the dose–response for the key events parallels the dose–response relationship for tumours. Ideally, one should

be able to correlate increases in incidence of a key event with increases in incidence or severity (*e.g.* lesion progression) of other key events occurring later in the process, and with the ultimate tumour incidence. Comparative tabular presentation of incidence of key events and tumours is often helpful in examining dose–response.

5. Temporal association

This section should detail the observed temporal relationships or sequence of events and discuss whether the key events precede the tumour response. One should see the key events before tumour appearance; this is essential in deciding whether the data support the postulated mode of action. Observations of key events at the same time as the tumours (*e.g.* at the end of a bioassay) do not contribute to temporal association, but can contribute to analysis in the next section. Most often, complete data sets to address the criterion of temporality are not available.

6. Strength, consistency and specificity of association of tumour response with key events

This section should discuss the weight of evidence linking the key events, precursor lesions and the tumour response. Stop/recovery studies showing absence or reduction of subsequent events or tumour when a key event is blocked or diminished are particularly important tests of the association. Consistent observations in a number of such studies with differing experimental designs, increases that support since different designs may reduce unknown biases or confounding. Consistency, which addresses repeatability of key events in the postulated mode of action for cancer in different studies is distinguished from coherence, however, which addresses relation of the postulated mode of action with observations in the broader database (see point 7). Pertinent observations are, *e.g.*, tumour response and key events in same cell type, sites of action logically relate to event(s), initiation–promotion studies, and stop/recovery studies.

7. Biological plausibility and coherence

The postulated mode of action and the events that are part of it need to be based on current understanding of the biology of cancer to be accepted, though the extent to which biological plausibility as a criterion against which weight of evidence is assessed is necessarily limited, due to considerable gaps in our knowledge in this regard. One should consider whether the mode of action is consistent with what is known about carcinogenesis in general (biological plausibility) and in relation to what is also known for the substance specifically (coherence). For the former, likeness of the case to others for structural analogues may be informative (*i.e.* structure–activity analysis). Additionally, this section should consider whether the database on the agent is internally consistent in supporting the purported mode of action, including that for relevant non-cancer toxicities. Some modes of action can be anticipated to evoke effects other than cancer, *e.g.* reproductive effects of certain hormonal disturbances that are carcinogenic. Moreover, some modes of action are consistent with observed lack of genotoxicity. Coherence, which addresses relation of the postulated mode of action with observations in the broader database — for example, association of mode of action for tumours with that for other endpoints — needs to be distinguished from consistency (addressed in Point 6 above) which addresses repeatability of key events in the postulated mode of action for cancer in different studies.

8. Other modes of action

This section discusses alternative modes of action that logically present themselves in the case. If alternative modes of action are supported, they need their own framework analysis. These should be distinguished from additional components of a single mode of action which likely contribute to the observed effect, since these would be addressed in the analysis of the principal mode of action.

9. Assessment of postulated mode of action

This section should include a clear statement of the outcome with an indication of the level of confidence in the postulated mode of action **—** *e.g.* high, moderate or low.

10. Uncertainties, Inconsistencies, and Data Gaps

Uncertainties should include those related to both the biology of tumour development and those for the database on the compound of interest. Inconsistencies should be flagged and data gaps identified. For the identified data gaps, there should be some indication of whether they are critical as support for the postulated mode of action or simply serve to increase confidence therein.