Regulatory Toxicology and Pharmacology 60 (2011) S1–S34

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

A review of mammalian carcinogenicity study design and potential effects of alternate test procedures on the safety evaluation of food ingredients

A.W. Hayes [b,f], A.D. Dayan [d], W.C. Hall [a], R.L. Kodell [c], G.M. Williams [e], W.D. Waddell [g], R.S. Slesinski [f], C.L. Kruger [f,*]

[a] Hall Consulting Services, 12337 Sherwood Forest Drive C6, Mt. Airy, MD 21771, USA
[b] Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
[c] University of Arkansas for Medical Sciences, 4301 W. Markham St., Little Rock, AR72205, USA
[d] University of London, Senate House, Malet Street, WC1E 7HU, London
[e] New York Medical College, 40 Sunshine Cottage Road, Valhalla, New York 10595, USA
[f] Spherix Consulting, Inc., 6430 Rockledge Drive #503, Bethesda, MD 20817, USA
[g] University of Louisville, Louisville, KY 40292, USA

ARTICLE INFO

Article history:
Received 24 August 2010
Available online 19 November 2010

Keywords:
Carcinogenicity
Study design
Food ingredients

ABSTRACT

Extensive experience in conducting long term cancer bioassays has been gained over the past 50 years of animal testing on drugs, pesticides, industrial chemicals, food additives and consumer products. Testing protocols for the conduct of carcinogenicity studies in rodents have been developed in Guidelines promulgated by regulatory agencies, including the US EPA (Environmental Protection Agency), the US FDA (Food and Drug Administration), the OECD (Organization for Economic Co-operation and Development) for the EU member states and the MAFF (Ministries of Agriculture, Forestries and Fisheries) and MHW (Ministry of Health and Welfare) in Japan. The basis of critical elements of the study design that lead to an accepted identification of the carcinogenic hazard of substances in food and beverages is the focus of this review. The approaches used by entities well-known for carcinogenicity testing and/or guideline development are discussed. Particular focus is placed on comparison of testing programs used by the US National Toxicology Program (NTP) and advocated in OECD guidelines to the testing programs of the European Ramazzini Foundation (ERF), an organization with numerous published carcinogenicity studies. This focus allows for a good comparison of differences in approaches to carcinogenicity testing and allows for a critical consideration of elements important to appropriate carcinogenicity study designs and practices. OECD protocols serve as good standard models for carcinogenicity testing protocol design. Additionally, the detailed design of any protocol should include attention to the rationale for inclusion of particular elements, including the impact of those elements on study interpretations. Appropriate interpretation of study results is dependent on rigorous evaluation of the study design and conduct, including differences from standard practices. Important considerations are differences in the strain of animal used, diet and housing practices, rigorousness of test procedures, dose selection, histopathology procedures, application of historical control data, statistical evaluations and whether statistical extrapolations are supported by, or are beyond the limits of, the data generated. Without due consideration, there can be result conflicting data interpretations and uncertainty about the relevance of a study's results to human risk. This paper discusses the critical elements of rodent (rat) carcinogenicity studies, particularly with respect to the study of food ingredients. It also highlights study practices and procedures that can detract from the appropriate evaluation of human relevance of results, indicating the importance of adherence to international consensus protocols, such as those detailed by OECD.

1. Introduction

Extensive experience has been gained over decades of animal tests on drugs, pesticides, industrial chemicals and numerous types of food additives and consumer products to define essential components in animal tests to determine potential carcinogenicity of chemical substances under experimental circumstances. All human carcinogens adequately tested in animals have produced positive results in at least one animal model (Faustman and Omenn, 2008). The International Agency for Research on Cancer (IARC) concludes that, "Although this association cannot establish that all agents and mixtures that cause cancer in experimental

* Corresponding author. Fax: + 1 978 409 1154.
E-mail address: ckruger@spherix.com (C.L. Kruger).

0273-2300/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.yrtph.2010.10.002

animals also cause cancer in humans, nevertheless, in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents and mixtures for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans" (IARC, 2000). The US EPA Cancer Guidelines (EPA, 2005b) also note the absence of a full scientific link, but assume the relevance of animal bioassays unless lack of relevance for humans is specifically determined. Information about potential carcinogenic hazard from such bioassays is then evaluated by risk assessment/management methodologies to reach pragmatic decisions about potential risk to human health (IARC, 2006; NRC, 1994; Cogliano et al., 2004; Chapin et al., 1993; Williams, 2008).

All chronic bioassays, whether conducted by NTP, EPA, FDA, ERF, EU (OECD), industry, or any other entity, are invariably used in the risk assessment/management process by regulatory decision makers. Because of the public role of published science, it is incumbent on regulators and companies that produce or market potentially hazardous products to put bioassay information in perspective in order to assess consumer health risk, make plausible risk management decisions, and implement effective risk communication strategies. Published carcinogenicity studies must be executed in a manner that is scientifically rigorous and appropriate. If study methodology does not conform to guidelines that have been validated and have the benefit of scientific and regulatory approval, then an evaluation of the deviations from standard protocol design must be completed in order to determine whether the methodology has compromised the robustness of conclusions. While rigorous testing guidances have been provided and utilized around the world, differences in approach to carcinogenicity studies still exist. ERF, as an example, has published numerous carcinogenicity studies under protocols and practices that differ significantly from those commonly used to adhere to established regulatory guidance and toxicology testing standards in the evaluation of food ingredients. The ERF, also described as "The European Foundation of Oncology and Environmental Sciences 'B. Ramazzini'", is an independent, non-profit foundation based in Bologna, Italy (Maltoni et al., 1999; Soffritti et al., 2002). Its research studies on food ingredients have, to date, not been performed for the purpose of supporting a pre-market regulatory safety assessment. As such, they have not been part of formal food additive petitions, GRAS affirmation documents, or other like documents, submitted to regulatory agencies when an entity is seeking permission for use of a food ingredient in foods. Similarly, carcinogenicity studies conducted or funded by the National Center for Toxicological Research (NCTR), in cooperation with the US FDA, are not conducted for regulatory purposes. Nevertheless, such studies can impact the overall public perception of safety of food chemicals and raise questions about the overall reliability of safety assessments made by regulatory agencies, when results obtained and/or conclusions reached conflict with existing research and/or published safety reviews and assessments. Despite many common features in carcinogenicity testing protocols, differences in design and practices can influence study results. Understanding the differences is critical to understanding differences in measures related to neoplastic potential and therefore critical for evaluating the relevance of the findings in subsequent human risk assessments. The focus of this paper is to highlight the numerous parameters in animal carcinogenicity testing programs that can affect interpretation of results when studies are conducted with unique testing protocols. In this respect, testing protocols that diverge from standards agreed to by international consensus, such as the extensive carcinogen screening studies conducted at the ERF laboratories, are considered both for their proposed merits (Maltoni et al., 1999; Soffritti et al., 2002) as well as for potential for confounding comparisons with studies performed for submission to regulatory authorities.

The comparison allows for clarification of best practices and procedures in carcinogenicity testing. The topics of carcinogen risk assessment methodology, mechanisms of carcinogenicity and protocols using *in vitro* tests and genetically derived animals have been the subject of numerous recent reviews (Robinson and MacDonald, 2001; Goodman, 2001; MacDonald et al., 2004; Jena et al., 2005; OECD, 2006; Williams et al., 2008; Wells and Spencer-Williams, 2009). Consideration of the merits and deficiencies of different protocols and approaches for these type of tests are beyond the scope of the current review, however, where relevant, alternative testing procedures are briefly discussed.

1.1. Carcinogenicity protocol design

Testing protocols for conduct of carcinogenicity studies in rodents have been developed in Guidelines promulgated by national and international regulatory agencies. In the US, these include, for example, the EPA (Environmental Protection Agency), which evaluates pesticides (FIFRA – Federal Fungicide, Insecticide and Rodenticide Act) (EPA-FIFRA; EPA, 1984) and industrial chemicals (TSCA – Toxic Substances Compliance Act) (EPA-TSCA; EPA, 1983) and the FDA , which evaluates drugs (CDER – Center for Drug Evaluation and Research) (Contrera et al., 1997) and food ingredients (CFSAN – Center for Food Safety and Applied Nutrition) (FDA, 2000a–f). Outside the US, protocol-defining authorities and regulatory agencies include, for example, the OECD, based in the European Union (EU) (OECD, 1981, 2009a,b), Japan's MAFF, and the Food Standards regulatory agency of Australia and New Zealand (FSANZ). Williams et al. (2008), gives a detailed list of regulating agencies and authorities around the world (see Table 25.1). In addition to regulatory agencies, the International Conference for Harmonization (ICH, 1997) has also developed and regularly updates internationally accepted ("harmonized") testing protocols for carcinogenicity tests.

The 1981 OECD guidelines for carcinogenicity tests (Guideline 451) (OECD, 1981) have been revised recently (OECD, 2009a,b). The new guidelines are essentially the same as the ones in force since 1981, but they include additional details that are generally harmonized internationally with EPA's Office of Prevention, Pesticides and Toxic Substances and FDA (CFSAN) regulatory protocols. The technical bases for scientific standards encompassed in all of the testing protocols for foods, pharmaceuticals and other chemicals are similar. They also include specific guidences for modifications to protocol design for the testing of food ingredients, which relates to the types of protocols that are the focus of this review.

In a recent paper, (Melnick et al., 2008) noted that conflicting views on results of different animal carcinogenicity tests have arisen from differences and deficiencies in the design of studies, rigorousness of test procedures, dose selection, histopathology, different methods for statistical evaluations and extensions of interpretations beyond the limits supported by the testing data. These authors considered that deficiencies in design or evaluation of animal carcinogenicity studies used to make public health decisions can be the result of many different aspects of test design and performance and, for a relevant extrapolation of results to humans, test protocols must, at a minimum:

- employ animal models sensitive to the study endpoints,
- thoroughly characterize both the test chemical and administered dose,
- use challenging doses and durations of exposure,
- use sufficient numbers of animals per dose group,
- use multiple dose groups to detect dosage effects,
- employ complete and peer reviewed histological evaluations, and
- evaluate data using pairwise comparisons and analyses of trends that rely on survival-adjusted tumor incidence.

The importance of international standards that have been developed for guidance in the design and execution of carcinogenicity bioassays to achieve valid results was highlighted by the recent controversy over publications of a long-term carcinogenicity study by Soffritti and co-workers (Soffritti et al., 2005, 2006). Results from their studies indicated that aspartame is a 'multipotential carcinogenic agent', even at a dose (20 mg/kg bw/day) corresponding to half of the current acceptable daily intake (ADI). The investigators suggested that the results necessitate the updating of current scientific advice on the safety of aspartame. It should be noted that the use of the term "multipotential carcinogen", in this instance, is meant to convey that tumorigenic potential was found with multiple sites (Tannenbaum et al., 1962). The term was originally coined by Maltoni, in a report on urethane carcinogenicity, however, at that time, its use was meant to convey that tumorigenic (carcinogenic) potential was found in multiple species.

A review of the methodology and results of the studies on aspartame at ERF was conducted in June 2005 by the European Food Safety Authority (EFSA). The evaluation considered the study methodology in relation to the results obtained and the interpretation of those results. After its evaluation, the EFSA Panel concluded that the study had numerous methodologic flaws that brought into question the validity of the findings, as interpreted by ERF. In particular, the EFSA Panel discussed the disadvantages of lifetime treatment to natural death, compared to termination of the study at 104 or 110 weeks. These disadvantages included an increase in background pathology and higher probability of postmortem autolytic change. The EFSA Panel noted that the results from the ERF studies could not be compared with results from testing done by scientifically valid, internationally accepted protocols. This opinion has led to the conclusion by EFSA that the Ramazzini studies gave no reason to revise the ADI for aspartame.

The US FDA Center for Food Safety and Applied Nutrition (CFSAN) (FDA, 2000c) independently evaluated the results of the same studies conducted at ERF and their conclusions were similar to those reached by EFSA:

> "*Based on the available data, however, we have identified significant shortcomings in the design, conduct, reporting, and interpretation of this study. FDA finds that the reliability and interpretation of the study outcome is compromised by these shortcomings and uncontrolled variables, such as the presence of infection in the test animals*",
>
> "*Based on our review, pathological changes were incidental and appeared spontaneously in the study animals, and none of the histopathological changes reported appear to be related to treatment with aspartame*".

This most recent controversy highlights the need to examine the rationale behind key components of typical carcinogenicity testing protocols for food ingredients in order to evaluate the strengths and weaknesses of particular protocol alternatives and to ensure that results from such studies can be appropriately interpreted to identify potential human hazard(s). It is only after a hazard has been identified by valid internationally accepted testing protocols conducted according to Good Laboratory Practices (GLP) that the risk to the human population can then be assessed. Regarding the work completed at ERF, there is uncertainty with respect to several major aspects of the testing protocols utilized, as well as the study practices employed, including the status of GLP compliance.

As noted previously, the scope of this paper is limited to the evaluation of rodent carcinogenicity testing protocols commonly used for compliance with regulatory requirements for pre-market food ingredient testing, or and those used by major laboratories conducting carcinogenicity screening studies of chemicals and food components, specifically the ERF and the NTP testing programs. For testing of food ingredients, protocols focused on testing by the oral route (gavage, diet or drinking water) are the most relevant. A comparison of the major regulatory protocol specifications for oral carcinogenicity studies included in guidelines of the OECD, US EPA (OPPTS–TSCA/FIFRA) and US FDA (CFSAN) is presented in Table 1. A summary of the principal components of carcinogenicity testing protocols employed for NTP studies and descriptions of ERF procedures as presented in recent publications is shown in Table 2. The procedures used by the two most extensive screening programs for chemical substances, namely those conducted by the US NTP and by ERF, are compared and contrasted below, with respect to requirements of major regulatory guidelines and to discern the possible impact of fundamental differences in the interpretation of study results (see Table 3).

2. Testing parameters for animal screening tests to determine potential for chemical carcinogenicity

Rodent carcinogenicity studies are performed to evaluate the safety of chemicals with potential for human exposure, such as substances added intentionally or inadvertently to food (i.e., direct and indirect additives), water impurities and incidental contaminants that can pose a risk to humans from exposure (Pastoor and Stevens, 2005). Regulatory agencies world-wide typically expect study protocols to meet recognized testing guidelines, adhere to requirements of Good Laboratory Practice Standards (GLPs) and incorporate testing features that maximize the probability of detecting weak oncogenic effects so that negative results can be considered reliable and results can be used for human risk assessments. Typical components that must be considered in protocols for testing chemical substances and different approaches in screening versus regulatory purposes of test conduct are summarized in the following sections.

2.1. Selection of test animals

Animal tests of chemicals for potential to pose a carcinogenic hazard to humans have evolved from basic testing practices, principles and protocols developed for subchronic and chronic rodent tests that determine effects of toxicity on survival, growth, hematological, clinical and organ and tissue effects following repeated exposures over a significant portion of the animals' life span (OECD, 1981, 2009a,b; FDA, 2000b). Because of possible differences in mammalian metabolism, physiology, pharmacokinetics and species sensitivity seen in chronic toxicity studies, evaluation of chemicals for carcinogenicity using two species has evolved as a recommended practice in regulatory agency guidelines. Although other species are sometimes used in specific situations, rats and mice have been used for physiological, behavioral and other scientific studies for well over a century (Suckhow et al., 2006) and have become the species of choice in carcinogenicity and other toxicity tests because of their small size and short life span and the extent of our detailed knowledge of their physiology and biochemistry. Some scientists have questioned the need for carcinogenicity studies with mice in addition to rats (Rushton, 1994), but current practice still demands the use of two species in the absence of suitable alternatives.

Outbred (also designated as randomly bred or non-inbred) stocks of rats (e.g., Sprague–Dawley, Wistar and Long Evans) are relatively large in size as test animals, but they are considered to have high levels of disease resistance, rapid growth, high fertility, long life spans and a genetic diversity that is analogous to the human population (NRC, 1996). The Han Wistar rat strain has re-

Table 1
Comparison of study design variables in US and OECD guidelines for rodent dietary carcinogenicity studies.

Protocol parameter	OECD (Guideline 451)	FDA Redbook (Section IV. C. 6 Rev. January 2006)	US Environmental Protection Agency (OPPTS Guideline 870.4200)
Purpose of study	Identify oncogenic effects for a major proportion of rodent normal lifespan	Identify oncogenic effects for a major proportion of rodent normal lifespan	Identify oncogenic effects for a major proportion of rodent normal lifespan
Duration of study	Rats 24 mo; Mice 18 mo. (may be extended to 30 mo. and 24 mo., respectively, if permitted by longevity and low spontaneous tumor rate in strain employed	Rats/mice 24 mo.	Rats ⩾ 24 mo.; Mice ⩾ 18 mo. Studies should not exceed 30 mo. for rats and 24 mo. for mice without Agency consultation
Animals			
Species/strain	Rats and mice; strains not specified	Rats and mice, strains not specified (“consult Agency”)	Rats and mice, strains not specified
Sexes	Males and females	Males and females	Males and females
Age at initiation	Weanlings < 6 weeks	Weanlings <6–8 weeks old	⩽8-weeks old; neonates may be recommended under special conditions
Animal numbers	50/sex/group	50/sex/group (⩾25/sex/group desired at termination and initial number should be adjusted if toxicity expected)	50/sex/group; group size should reflect animals needed if interim kills are scheduled
Study design			
No. dose groups	⩾3 Dose groups	⩾3 Dose groups	⩾3 Dose groups
Number of control groups	1 Concurrent control group/sex	1 Concurrent control group/sex; additional control group(s) if nutritional status may be altered by dose incorporation	1 Concurrent control group/sex; if vehicle of uncertain toxicity is used then untreated and vehicle control groups are required
Sentinel animals (for periodic blood assessment)	Not specified	Not specified	Not specified
Dosage intervals	Highest dose 5% in diet or MTD; lowest dose 10% of high dose	MTD, intermediate and low dose; low dose determined by NOAEL in subchronic study	Highest dose should elicit minimal signs of toxicity seen in 90-day study but need not exceed 1000 mg/kg/day; low dose should be expected to produce no evidence of toxicity (NOEL) and inter-mediate dose should be spaced between high and low dose
Duration and frequency of dosing	Mice: 18–24 mo.; Rats: 24–30 mo.; 7 days/week	24 months – 7 days/week	
Randomization	“Adequate randomization”; no specific method stated	Randomization should be based on body weights or weight ranges; other method must be justified	“Adequate randomization” required for animal allocation to groups to avoid bias.
Determination of individual body weights	At initiation; 1x/week for 13 weeks and 4 week intervals to termination	At initiation; 1x/week for 13 weeks and 4 week intervals to termination	At initiation; 1x/week for 13 weeks and 4 week intervals to termination
Determination of feed consumption	Measured/group weekly for 13 weeks and 3 mo. intervals to termination	Measured/animal at same intervals after initiation as body weights	Measured/group weekly for 13 weeks and at 4-week intervals to termination
Dose selection criteria			
MTD (& basis) Other (describe)	Max. 5% in diet or MTD based on minimal toxicity (<10% bw gain decrement)	MTD based on subchronic toxicity; “Highest dose *should not* be selected based upon margin of safety for human exposure”	Highest dose should elicit minimal signs of toxicity seen in 90 day study but need not exceed 1000 mg/kg/day; low dose should be expected to produce no evidence of toxicity (NOEL) and inter-mediate dose should be spaced between high and low dose
Analytical verification of test substance			
Method(s)	Prerequisite for study initiation but methodology not specified	Analyses must provide information on % purity and % of identified and unidentified impurities	Prerequisite for study initiation purity and names and quantities of contaminants should be provided when possible
Stability of test substance (Y/N)	Y – conducted prestudy	Y– composition should be known	Y – conducted prestudy
Stability in diet (Y/N)	Y – conducted prestudy and defined intervals	Y – required by GLP and in report to FDAb	Y – conducted prestudy and periodically during study
Homogeneity in diet (Y/N)	Y – conducted prestudy and defined intervals	Y – required by GLP and in report to FDAb	Y – conducted prestudy and periodically during study
Verification of concentration in diet	Not specified	Y – required by GLP and in report to FDAb	Y – conducted prestudy and periodically during study

Animals and housing			
Acclimation period	Not specified	⩾5 days	⩾5 days
Animal identification	Not specified	Ear tag, tattoo or chip	Each animal assigned unique I.D. number; method not specified
Animals/cage	Not specified	1 animal/cage (“single housing; “crucial to determining feed intake and efficiency, avoid cannibalism and minimize effects of palatability on bw gain”)	Animals may be group caged/sex, numbers should not interfere with animal observations; single housing may be required in cases of toxic effects (i.e., excitability, morbidity)
Cage/rack type	Not specified	Per NRC/NAS[a]; plastic or metal hard bottom, 40 $in.^2$/animal × 7 in height	Not specified but tests in US subject to NRC/NAS guidelines[a] for animal care
Bedding and QC	Type and changes not specified but “should be sterilized”	Per NRC/NASa; details required in report to FDAb	Not specified but tests in US subject to NRC/NAS guidelines1 for animal care
Environmental Conditions (specify which monitored)	Air changes not specified “should be well ventilated with controlled lighting, temp. and humidity”	Per NRC/NAS[a]; 10–15 air changes/hr; 50% replenishment with fresh air; 64–79 oF; %RH range not specified but ranges required in report to FDAb	Air changes not specified but subject to NRC/NAS guidelines1; 22 ± 3 °C and RH range 50 ± 20%
Type of feed	Adequately nutritious but type not specified; provided ad libitum; replaced weekly	Per NRC/NAS[a]; “should meet nutritional requirements”; *ad libitum*	Per NRC/NAS[a]; “should meet nutritional requirements”; ad libitum
Nutrient and Contaminant QC (Y/N)	Y – details not provided	[1]Per NRC/NAS; Results must be reported per FDA (CFSAN)reporting guidelines[2]	Y – analysis should determine adequacy of nutritional components and impurities of potential concern for study
Water/water consumption	Specifies only ad libitum	[1]Per NRC/NAS; water bottle or automatic source *ad libitum*	[1]Per NRC/NAS; water bottle or automatic source ad libitum
Necropsy evaluations:	Dead, moribund; Interim if scheduled at 12 mos.; terminal at 24 mos. Rats	Dead, moribund and terminal at 24 months	Dead, moribund and terminal at 24 months
Termination method	Not specified	Not specified	Not specified
Gross pathology (Y/N)	Y – detailed procedure	Y – detailed procedure	Y – detailed procedure
Organ weights (Y/N)	Not specified	Y – ∼12 organs at termination	Y – ∼10 organs weighed at interim kills (if scheduled) and at termination from at least 10 animals/group
Hematology	Not specified but blood smear for all animals and diff. blood count for high dose and control groups at 12, 18 and 24 mos.; lower dose groups should be sampled and examined if findings suggest health effects of dosing	Y – 10 animals/sex all groups at 2 weeks, 3, 6, and 12 mo.; + 18 mo. to confirm any findings; ∼10–12 standard hematological parameters and clotting factors evaluated	Not specified but blood smear for all animals and diff. blood count for high dose and control groups at 12, 18 and 24 mos.; lower dose groups should be sampled and examined if findings suggest health effects of dosing
Clinical chemistry	Not specified	Y – same sampling schedule as hematology; ∼23 typical plasma parameters	Not specified
Urinalysis	Not specified	Y – at initiation, and 3, 6, 12 mo. + 18 mo. if confirmation needed for any finding; ∼5–6 typical urinary parameters	Not specified
Histopathology			
Number of tissues/organs	∼34	∼43	Not specified
Fixative(s)	Preservative not specified; lungs and bladder should be inflated with fixative	10% neutral buffered formalin; lungs should be inflated with fixative	10% neutral buffered formalin
Open (not coded) vs. Blinded observations	Not specified	Open	Not specified
Peer review (Y/N; comments)	Not specified	Y – FDA (CFSAN) will assign review pathologist to each study	Not specified
Good Laboratory Practices:	Y	Y	Y
Protocol (Y/N/not specified)	Not specified	Y – required by GLP	Y – required by GLP
SOPs (Y/N/not specified)	Not specified	Y – required by GLP	Y – required by GLP
QA review (Y/N/not specified)	Not specified	Y – required by GLP	Y– required by GLP
Reporting			
Organ/tissue toxicity findings (Y/N)	Not specified	Y – required by reporting guideline[b]	Y – “Findings ... should be evaluated in conjunction with the findings of previous studies and considered in terms of the toxicological effects, necropsy and histopathological finding”

(*continued on next page*)

Table 1 (*continued*)

Protocol parameter	OECD (Guideline 451)	FDA Redbook (Section IV. C. 6 Rev. January 2006)	US Environmental Protection Agency (OPPTS Guideline 870.4200)
Time to tumor recorded (Y/N)	Not specified	Y	Not specified
Incidence/Severity of tumors	Not specified	Y – required by FDA reporting guideline[b]	Y – relationship between dose and severity should be reported
Relation of tumors to other toxicity findings	Not specified	Y – Required by FDA reporting guideline[b]	Y – "Attempts should be made to correlate gross observations with histopathological findings"
Body weight plots/tables	Not specified	Y – required by FDA reporting guideline[b]	Y – Including discussion of relationship to dose
Survival plots/tables	Not specified	Y – required by FDA reporting guideline[b]	Y – mortality effects should be discussed relative to dose
Feed consumption	Y/group	Y - Required/animal for feed incorporation studies by FDA reporting guideline[b]	Y/group
Evaluation and summary of clinical signs	Y	Y – required by FDA reporting guideline[b]	Y – clinical signs of toxicity should be related to incidence and severity of other reported parameters

[a] Specified by NRC/NAS, Institute of Laboratory Animal Resources, Guide for the Care and Use of Laboratory Animals, 1996.
[b] US FDA/CFSAN Redbook 2000 Section IV.B.2. Guidelines for reporting of toxicity studies. November 2003.

cently shown a low rate of mortality in long-term toxicity and carcinogenicity studies and may prove to be a preferable alternative to Sprague–Dawley rats. However, outbred strains in general, can show considerable and undefined genetic variability (Suckhow et al., 2006) and the possibility that differences in test parameters noted in a specific study may be the result of genetic rather than treatment-related factors is a potential confounding factor that must be considered. The same strain from different sources can vary considerably as discussed below in this section. Despite the goal of maximizing genetic diversity in outbred strains, it is surprisingly very difficult to prevent inbreeding in outbred colonies maintained over a prolonged period. Several factors can contribute to inbreeding, when rigorous controls are not in place, and inbreeding can result in loss of desired heterozygosity. In comments abstracted from White and Lee (1998), examples of factors that must be considered with outbred strains include: (1) genetic limitations that stem from use of only a small number of animals to start the non-inbred "founder" colony, since loss of heterozygosity is increased in subsequent populations when low numbers of founder animals are used at the initiation of the initial colony; (2) sampling errors that may occur in the selection of breeders chosen to start the founder colony and that may limit the genetic diversity in the colony relative to animals from which they were derived, which is particularly important if only one colony of animals from several available colonies is used to propagate the ongoing colony. A procedure to ensure that sufficient breeders are taken from all available colonies must be enforced, so that genetic diversity is sustained; and (3) mutations in the breeding population that can become fixed in the colony and may sometimes result in significant physiological or metabolic differences compared to the original population. These differences may be difficult to discern phenotypically without an adequate surveillance system in place. White and Lee (1998) illustrated the possibility of fixed mutations in the breeding colony by examining sample colonies of non-inbred CD-1 mice that were established simultaneously at four separate locations, with all locations using the same number of breeders for each successive population. After 3 years, the isozymes in blood and tissues of 100 animals selected at random showed marked differences in number and types of isozymes despite similar handling and breeding in the individual laboratories. The authors concluded that divergence in genetic features can be caused by normal mutation rates, when a population of outbred animals is maintained over the course of several years, and that this can cause many important, but difficult to detect, differences in the population.

*C*harles *R*iver *L*aboratories (CRL) was the first major animal supplier to attempt to minimize genetic drift by developing the *C*aesarian-derived (CD) *S*prague *D*awley (SD) founder strain in 1950. The SD strain used was originated in 1925 by Robert W. Dawley, by his mating of a hybrid hooded male and a female Wistar rat (CRL, 2009). In 1991, CRL selected eight colonies of the strain to form the "IGS foundation colony". In 1997, the colony was re-derived, with Caesarian delivery, into an isolator (i.e., "barrier raised" [BR]) foundation colony. "IGS" refers to animals bred using the International Genetic Standard system described by White and Lee (1998). This system controls the size of the foundation colony and still allows for application of complex outbreeding schemes to ensure that the tendency toward inbreeding is minimized. Since progeny colony genetics drift from the foundation colony genetics, they are replenished when necessary from the foundation colony. To ensure minimal drift in the foundation animals, they are routinely monitored through the use of microsatellite loci analysis. The worldwide availability of the highly standardized strain of CD® (SD)IGS BR rats from CRL has been instrumental in permitting meaningful comparisons of carcinogenicity testing results on the same or similar chemicals among testing laboratories. The use of

Table 2
Comparison of protocol parameters employed in NTP and ERF programs for chemical carcinogenicity screening.

Study design parameter	NTP testing guidelines[a]	ERF procedures[b]
Purpose of study	Identify oncogenic effects for a major proportion of rodent lifespan (2 years)	Identify oncogenic effects during the entire lifetime of rodents until natural death
Animal species	Fischer 344 rats; B6C3F1 mice	Sprague–Dawley, Outbred strain reared at ERF
During of dosing with test substance	24–30 months; 7 days/week in diet	24–25 months; 7 days/week in diet
Study termination	Animals terminated at cessation of dosing with test substance, typically 24 mo.; mice considered for termination if survival <50% at 18 mo.	Animals allowed to die natural death; dosing with test substance terminated at 24 mo.
Sentinel animals (for periodic hematology tests/serology assessment)	Yes – 15/sex/species	Not used
Numbers of test animals	50–100/sex/group; typically 50/sex/group	50–100/sex/group; 100–150/sex/group "Mega-experiment"
Randomization procedure	Following stratification by weight (and rejection of outlier unhealthy, heavy/light animals), randomization into cage groups based on table of random numbers	Described either as assignment by litter to each treatment and control group (Bucher, 2002) or by randomization on the basis of avoiding >1 M and 1 F from the same litter produced at ERF (Soffritti et al., 2006). All animals apparently used regardless of weight or health status.
Animal housing		
Animals/cage	3 males or 5 females/cage; sentinels and *male mice* for all studies housed 1 animal/cage	5 males or 5 female/cage; no sentinel animals indicated
Cage type and dimensions	Polycarbonate 22 × 12 × 8 in.; 264 $in.^2$ floor space/5 females or 3 males	Polycarbonate 41 × 25 × 15 cm (16 × 10 × 6 in.); 160 $in.^2$ floor space/5 animals
Bedding	Irradiated hardwood changed 2x/weeks	White wood shavings; changes and sterilization not stated
Feed	Irradiated NTP2000 powdered; provided ad libitum	Source of pelleted feed not stated in publications; provided *ad libitum*
Water	Automatic watering system; municipal source (1–2 ppm Cl); provided *ad libitum*	Water source not stated in publications; methods state water consumption measured daily/cage implies water bottle used
No. doses and dose selection	3 Dose levels; maximum tolerated dose (MTD) is highest dose based upon minimal toxic clinical and pathological effects in subchronic study; lower doses typically 1/2 and 1/4 of MTD	2 or more dose levels; highest dose based upon literature information or range finding information if literature not available; highest dose is the "maximum tolerated level" at an "order of magnitude to which humans may be exposed"; ADI used as a reference point in recent studies on food ingredients
Determination of Ingested Dose	Dose calculated based upon individual body weights and feed consumption/group determined at initiation; 1x/week for 13 weeks and 4 week intervals to termination	Dose calculated based on the assumption of an average weight of a rat as 400 g during the entire study and average consumption of 20 g feed/day/animal
Termination/evaluation interval	All moribund animals; at 12 mo. *if* interim scheduled; terminal at 24 mo.	Not applicable; All animals (incl. moribund) allowed to die natural death only
Pathology review	Multistage review: initial review by NTP pathologist; secondary review by external (i.e., independent) pathologist and final blinded review by Pathology Working Group (PWG)	Secondary review by internal pathologist only
Availability of test data	Data publicly available in electronic form on internet	Data must be requested from ERF
Statistical analysis of tumor incidence	Survival adjusted quantal-response analysis by poly *K* testing that assigns a weighted risk to individual animals based on survival time and expected lesion prevalence rate	Prevalence analysis for non-lethal tumors and a log rank test of Mantel and Cox; poly *K* analysis used in some recent studies (Soffritti et al., 2006)
Tumor types used in statistical analyses	Significant effects restricted to tissue-specific tumor responses	Significant effects include differences in total benign and total malignant tumor-bearing animals and total tumors
Good laboratory practices	Y – required by GLP	Not stated in publications "design followed in previous publications"
Protocol (Y/N/not specified)	Y – required by GLP	Not specified "standard methods employed"
Standard operating procedures (Y/N/not specified) Quality Assurance Unit review (Y/N/not specified)	Y – required by GLP	Not stated in publications
Reporting		
Organ/tissue toxicity findings (Y/N)	Y	Not stated in publications
Time to tumor (Y/N)	Y – observed in moribund animals and at interim kill if scheduled	N – tumors observed only at natural death
Incidence/severity of tumors	Y – required	Incidence only
Relation of tumors to other toxicity findings	Y – required	Not stated in publications
Body weight plots/tables	Y – required	Y – in some publications
Survival plots/tables	Y – required	Y – in some publications
Feed consumption	Y/group	Y/group
Procedures to verify homogeneity and concentration of dose in feed	Y	Not described in publications
Evaluation and summary of clinical signs	Y – required	Not reported in publications nor discussed in relation to tumors

[a] Based on: Bucher (2002); NTP (2006).
[b] Soffritti et al. (1999, 2002, 2006).

Table 3
Comparison of study design variables in rodent in utero treatment protocols.

Protocol parameter	FDA Redbook (Section IV. C. 8 Rev. July 2007)	ERF protocol
Purpose of study	Identify oncogenic effects of direct/indirect food ingredients in rodents exposed *in utero* and for subsequent 24 mo. of lifespan	Identify oncogenic effects in rodents exposed *in utero* and during the entire lifetime until natural death
Animals		
Species/strain	Rats and mice, strains not specified (“consult Agency”)	Sprague–Dawley, ERF Inbred strain
Sexes	Males and females	Males and females
Age at initiation	See duration and frequency of dosing section below	P1 parental animals: 4–5 weeks and females through lactation/weaning F1 pups dosed after weaning
Number of animals	⩾70 P1/sex/group (⩾25 F1 animals/sex/group should be available at termination); initial number of P1 parents should be adjusted to provide minimum of 70 F1 males and females with a maximum of 1M/1F/litter for subsequent chronic exposures	100–150/sex/group
Study design		
No. dose groups	⩾3 Dose groups	Various ⩾3 and up to 7 dose groups depending on test compound
Number of control groups	1 Concurrent control; additional control group(s) if nutritional status of diet may be affected by test substance	1 concurrent control group/sex
Sentinel animals (for periodic disease assessments	N	N
Dosage intervals	MTD, intermediate and low dose; low dose determined by NOAEL in subchronic study	Various; Up to 100% of ADI
Duration and frequency of dosing	Parental (P1) males exposed for 10 weeks and parental (P1) females for 4 weeks prior to mating. F1 animals: dosed 24 months – 7 days/week at same dose level as parents	P1 pregnant dams dosed from gestation day 12 through lactation; F1 weanlings dosed with same parental dose level for entire lifetime
Randomization	Randomization of P1 animals based on body weights or weight ranges; F1 weanlings in each dose group assigned for continuing dosing to avoid >2 males or females from same litter	Randomization of P1 not stated in publications; F1 weanlings in each dose group assigned for continued dosing to avoid >2 pups from same litter
Determination of body weights of P1 and F1 animals	All P1 animals should be weighed at initiation of dosing and females weekly through gestation and lactation; F1 animals should be weighed weekly for 13 weeks after weaning and monthly for duration of study	Weighing P1 animals not described. F1 animals weighed “beginning at 6 weeks of age and continuing once each week for the first 13 weeks, then every 2 weeks until the animals reached 110 weeks of age”
Determination of feed consumption	Measured/animal at same intervals after initiation as body weights	Measured/group at same intervals after initiation as body weights
Dose selection criteria: MTD (& basis) Other (describe)	MTD based on subchronic toxicity; “Highest dose *should not* be selected based upon margin of safety for human exposure”	MTD not used; Doses based upon estimated human exposure or percentages of ADI up to 100%
Analyses of test substance		
Method(s)	Analyses must provide information on % purity and % of identified and unidentified impurities	Infrared spectral analysis
Stability of test substance (Y/N)	Y “composition should be known”	Lab determination of chem. stability not stated
Stability in diet (Y/N)	Y – required by GLP and in report to FDA	Y – At predosing and periodically during study
Homogeneity in Diet (Y/N)	Y – required by GLP and in report to FDA	Determination not stated
Verification of concentration in diet	Y – required by GLP and in report to FDA	Not reported in publications
Animals and housing		
Acclimation period	⩾ 5 days	NA – animals reared in-house
Animal identification	Ear tag, tattoo or chip	Ear punch
Animals/cage	1 F1 animal/cage (“single housing”) ; “crucial to determining feed efficiency, avoid cannibalism and minimize effects of palatability on bw gain”)	M/F 5/cage
Cage/rack type	Per NRC/NAS[b]; plastic or metal hard bottom, minimum 40 $in.^2$/animal × 7 in height	Polycarbonate 41 × 25 × 15 cm (16 × 10 × 6 in.); rack not specified
Bedding and QC	Per NRC/NAS[b]; details required in report to FDA	White wood shavings; changes and sterilization not stated
Environmental conditions (specify which monitored)	[a]Per NRC/NAS; 10–15 air changes/h; 50% replenishment with fresh air; 64–79 F/RH not specified but ranges required in report to FDA	Air changes not stated; temp/humidity ranges: 23 ± 2C; 50–60 RH
Type of feed (diet studies)	Per NRC/NAS[b]; “should meet nutritional requirements”; *ad libitum*	Source and type of feed not stated; provided *ad libitum*; daily feed consumption/cage
Nutrient and contaminant QC (Y/N)	Per NRC/NAS[b]	Determination in feed not stated
Water	Per NRC/NAS[b]; water bottle or automatic source *ad libitum*	Water source not stated; water consumption measured daily/cage implies water bottle used
Necropsy evaluations	Dead, moribund and terminal at 24 months	At natural death only
Termination method	Not specified	Not applicable animals allowed to die natural death
Gross pathology (Y/N) Organ weights (Y/N)	Y – detailed procedure Y –∼12 organs at termination	Y – stated as “complete” Not stated in publications
Hematology	Y – 10 animals/sex all groups at 2 weeks, 3, 6, and 12 mo.; + 18 mo. to confirm any findings; ∼10–12 std. parameters and clotting	Not stated in publications
Clinical chemistry	Y – same sampling schedule as hematology; ∼23 typical plasma parameters	Not stated in publications
Urinalysis	Y – at initiation, and 3, 6, 12 mo. + 18 mo.. if confirmation needed for any finding; ∼5–6 parameters	Not stated in publications
Histopathology		
Number of tissues[a]	∼42–44[a]	∼34–36[a]

Table 3 (*continued*)

Protocol parameter	FDA Redbook (Section IV. C. 8 Rev. July 2007)	ERF protocol
Fixative(s)	10% neutral buffered formalin; lungs should be inflated with fixative	70% ethanol; bones in 10% formalin
Open (not coded)/blinded observations	Open	Not specified
Peer review (Y/N; comments)	Y – FDA will assign review pathologist to each study	External peer review not stated but tumors and lesions of interest "evaluated by senior pathologist" at ERF
Statistical analyses		
Test procedure (s)	According to FDA Redbook 2000; Section IV.B.4	Cochran-Armitage trend test; poly *K* test; total tumors/group also evaluated in recent publications
Combining tumors (Y/N)	Not typical	Y
Historical control (s) (Y/N)	Not specified	Y
Good Laboratory Practices	Y – required by FDA reporting guideline[2]	Uncertain; mentioned in some but not all publications and compliance program not described
Protocol (Y/N/not specified)	Y– required by GLP	Not stated in publications
SOPs (Y/N/not specified)	Y – required by GLP	Not specified "standard methods employed"
QA review (Y/N/not specified)	Y – required by GLP	Not stated in publications
Reporting		
Organ/tissue toxicity findings (Y/N)	Y – required by FDA reporting guideline2	Not stated in publications
Time to tumor (Y/N)	Y – required by FDA reporting guideline2	N – tumors only examined at natural death
Incidence/severity of tumors	Y – required by FDA reporting guideline2	Y/N – incidence but not severity reported in publications
Relation of tumors to other toxicity findings	Y – required by FDA reporting guideline2	N – only incidence reported
Body weight plots/tables	Y – required by FDA reporting guideline2	Y – in some publications
Survival plots/tables	Y – required by FDA reporting guideline2deline	Y – in some publications
Feed consumption	Y – required by FDA reporting guideline2	Y – in some publications
Summary of clinical signs	Y – required by FDA reporting guideline2	Not reported in publications

[a] Approximation.
[b] Specified by NRC/NAS, Institute of Laboratory Animal Resources, Guide for the Care and Use of Laboratory animals, 1996.
[2] US FDA/CFSAN Redbook 2000 Section IV.B.2. Guidelines for reporting of toxicity studies. November 2003.

internationally available animals is important for comparison of test results because carcinogenicity tests can be performed with highly similar outbred animals regardless of the location of the test facility.

Because the intentional goal is to maximize genetic diversity in outbred strains, it is "critical that concurrent and historical controls be acquired when conducting studies in order to characterize the population both at the time of the study as well as changes in the population over time" (White and Lee, 1998). Also, because of the greater degree of genetic diversity in outbred populations of test animals, large groups of animals must be used to discern differences and to determine statistical differences in comparisons of test and control groups.

Inbred strains of test animals for carcinogenicity tests are intentionally employed for reducing problems inherent in breeding schemes required to ensure genetic heterogeneity in outbred colonies. Inbred strains are derived from animals bred for 20 or more successive generations following consecutive brother-sister matings. This breeding regimen results in a high degree (>98%) of genetic uniformity (isogenicity). With inbred strains, the probability for genetic differences to produce confounding effects on test parameters is minimized in comparisons of differences in responses seen in test vs. control animals. For example, Haseman and Hoel (1979) concluded that the significance of tumor incidences in tests of DDT in various strains of mice could be more readily discerned using inbred strains than in outbred strains, likely due to decreased biological variability in tumor rates and biological variability in the control groups which facilitates statistical comparisons (Gad and Weil, 1986).

An essential requirement for conduct of meaningful carcinogenicity studies is the use of healthy, disease free animals at initiation of the study. This is a fundamental component of all accepted carcinogenicity test guidelines (EPA, 1998; NTP, 2006; FDA, 2000a,b; OECD, 1981, 2009a,b). Survival of adequate numbers of animals during the prolonged 1.5- to 2-year dosing period until termination is crucial to provide sufficient statistical power in comparisons of treated and control groups. Test animals selected for contemporary carcinogenicity studies in the US and in most European testing laboratories are, in almost all cases, no longer produced in-house, as was typical of studies prior to the 1970s. Animals are typically obtained from commercial suppliers that have extensive historical information on animal disease profiles, health status of breeders and records of typical clinical, laboratory and pathological indices in the animal colony. Also, established suppliers provide animals to multiple testing facilities, allowing for a very large series of historical control data, based on untreated animals from both the supplier facility and at the testing facilities. This situation provides an opportunity for sharing information and for inter-laboratory comparisons of clinical, laboratory and tumor indices in control groups. The common animal source reduces variability in health parameters, while use of caesarian derived (CD) and specific pathogen free (SPF) colonies results in consequent increase in lifespan and facilitates comparisons of growth parameters, clinical, hematological and urinalysis assessments, and incidence of tumors seen in groups treated with similar test substance(s). Although variability among studies and laboratories confounds attempts to use historical control data in a quantitative manner (Haseman et al., 1984, 1998), qualitative comparisons can reveal trends in variables over time and whether concurrent controls are outside the limits seen in recent studies.

ERF is unique among established testing facilities in that it has, and continues to maintain, a breeding colony of Sprague–Dawley rats (strain CRC/RF) for use in its carcinogenicity screening program – a colony program sustained for over a period of more than 30 years. ERF has evaluated numerous test substances using either CRC/RF Sprague–Dawley rats and/or Swiss mice ("as well as other strains") (Maltoni et al., 1999; Soffritti et al., 1999; Soffritti et al., 2002; Belpoggi et al., 2006a,b). It reports that the basal tumor incidences in the CRC/RF rat strain have been well established in control groups over the time span of their numerous studies and that spontaneous cancer susceptibility in this strain "is not very different than in humans" (Maltoni et al., 1999). Similarity of cancer susceptibility of this strain relative to humans is difficult to assess independently, but the latter claim implies that the breeding of

the ERF rat colony, although not specified in their publications, is most likely through outbred matings (i.e., similar to humans). Thus, despite the long history of this strain, the animals would still have a level of genetic diversity similar to other outbred rat strains. The possibility that this strain has undergone genetic drift, however, cannot be eliminated because of multiple technical issues discussed previously. The lack of information on breeding procedures used at ERF to produce the CRC/RF SD strain, in comparison to the IGS breeding scheme used to produce the SD strain of rats used in the majority of testing laboratories in the world, makes meaningful comparisons of carcinogenicity results from ERF with those from other testing programs problematic to impossible.

Although genetic diversity in outbred animals can be envisioned to represent heterogeneity found in the human population, it can also lead to increased variation in study results when results in outbred animals are compared to results in inbred strains that are highly genetically homogenous. ERF researchers have contended that the long-term reliance on a single unique outbred animal colony has resulted in a high degree of standardization in the historical reference data for evaluating significance of tumor incidences (Soffritti et al., 1999), which can be considered a basis for meaningful within-study comparisons of treatment groups. However, Magnuson et al. (2007) and Magnuson and Williams (2008) observed that, in recent ERF studies of aspartame, there were tumor incidences that were reported to be increased relative to controls but that were also within the range of historical control incidences. Soffritti (2008) responded that the toxicological importance of comparisons to concurrent control groups, rather than the historical ranges, was more relevant, effectively negating previous claims of the value of historical data at ERF gained from using the unique CRC/RF strain of rats because of their well-defined tumor incidences over a long history of use (Maltoni et al., 1999; Soffritti et al., 1999; Soffritti et al., 2002).

ERF studies do not indicate use of their breed of rats in comparable studies conducted at other laboratories (Maltoni et al., 1999; Soffritti et al., 1999). Because apparently only ERF is using its internally-established strain of CRC/RF rats, inter-laboratory comparisons of data for this CRC/RF strain are not possible, nor can the possibility of genetic drift effects be determined. Possible differences in genetic constitution in the CRC/RF outbred colony could result in unique patterns of metabolism and pharmacokinetic handling of test substances and pharmacodynamic responses. Without comparable cross-laboratory data for this colony of rats, the magnitude of such effects is not able to be discerned by peer or regulatory reviews. In addition, reviews of recent ERF studies indicate the possibility of compromised health status in ERF's CRC/RF rat colony (EFSA, 2006). Compromised health status in the breeding colony can also impact genetic drift, through effects on survival.

In contrast to the ERF, the US National Toxicology Program (NTP) has used inbred Fischer 344/N (F344) rats and hybrid B6C3F1/N (N = NIH repository) mice for most of their published carcinogenicity studies. NTP controls shipment of the same strain(s) of animals to various contractor facilities. The rationale for choosing these species was based on the availability of good baseline data for growth, longevity and tumor incidences in control animals. This data was derived from studies in the early National Cancer Institute (NCI) carcinogenicity testing program in the 1960s; a program that was merged gradually with the NTP program, in the late 1970s and early 1980s (Weisburger, 1983). F344 rats were originally chosen because their small size would minimize cage sizes needed for housing, low rate of spontaneous tumors, good survival at 2 years and their high fertility and reliable breeding characteristics (Weisburger, 1983; Cameron et al., 1985). F344 rats used in toxicity and carcinogenicity studies over many decades provided substantial comparative data on age-associated lesions and tumor incidences as well as hematology, clinical chemistry, growth parameters and lifespan (Goodman et al., 1994). Sprague–Dawley rats were used briefly by the NCI both in exploratory projects and some early carcinogenicity bioassays, but the "large size and numerous spontaneous tumors in this strain" were considered detrimental characteristics compared to perceived advantages with F344 rats (Weisburger, 1983). Some characteristics of F344/N rats have changed over the decades and increases have been seen in body weight, incidence of mononuclear cell leukemia and incidence of testicular interstitial cell tumors, with an incidence approaching 100% at the termination of studies (Solleveld et al., 1984; Haseman et al., 1997; Brix et al., 2005). In contrast to the dedication to a single animal colony at ERF, NTP has a dynamic process in place to periodically evaluate whether chosen test strains should be continued or replaced, following reviews by qualified scientific forums that evaluate results for potential effects of genetic drift. Evaluation by the NTP of potential issues with use of F344/N rats in carcinogenicity screening tests is discussed below in this section.

Hybrid B6C3F1/N mice were originally developed at the NIH as the F1 generation from a cross between the inbred C57BL/6 females and C3H males for use in carcinogenicity screening tests of 130 chemicals at the NCI (Weisburger, 1983). Hybrid mice were chosen for these studies because of their increased hardiness, acceptable tumor incidence and relatively greater longevity compared to other widely used inbred mice such as the CD-1 HaM/ICR a potential test strain rejected because of a high spontaneous tumor rate and relatively short survival (Weisburger, 1983; Cameron et al., 1985). Hybrid mice also served as a compromise to proponents of inbred vs. outbred strains in that they offered both controlled genetic uniformity and greater hybrid vigor (Goodman et al., 1994).

The animal models employed at NTP are evaluated periodically for acceptability and relevance to defining potential human carcinogenicity. A workshop held in June, 2005 (King-Herbert and Thayer, 2006) resulted in suggestions for possible modification of the animals used for testing because of "drift" in various parameters in the F344/N. Suggestions for a new rat model included: (1) re-establish the F344/N from another source, although such an approach would not address the general issues confronting the strain; (2) create an F1 hybrid such as the F344/Brown Norway cross (FBNF1) to increase genetic heterogeneity similar to that in humans but in a controlled manner; and (3) consider using an alternative outbred strain or stock such as the Wistar Han or Sprague Dawley rat to increase the range of responses to chemical agents. The workshop consensus proposed use of a FBNF1 hybrid because it may have a lower incidence of testicular interstitial cell tumors than the F344 and may also have lower rates of mononuclear cell leukemia. King-Herbert and Thayer (2006) reported that following the workshop, "the NTP discontinued use of the NTP F344/N rat in all new studies and began using a commercial source [Taconic Farms, Inc.] of the F344 (F344/NTac)". Also, "The NTP intends to continue to use an isogenic rat strain to maximize reproducibility in tumorigenic response over time and facilitate genetic monitoring and interpretation of subsequent mechanistic studies". However, in 2010, King-Herbert Taylor et al. (2010) reported that use of the F344/NTac strain in a limited number of scheduled studies during evolution of new programmatic initiatives at NTP to use a single strain for both reproductive and cancer endpoints as well as beginning exposures *in utero* rather than in the young adult period. Use of F344 rats was not considered appropriate for these multiple goals using a common rat model because of their poor reproductive performance. The Wistar Han outbred rat could not be used in studies evaluating dioxin or aromatic hydrocarbons because of genetic properties that confer resistance to such types of chemicals and this strain proved to have unexpectedly poor repro-

ductive performance in timed-mating protocol compatible with the proposed use of the *in utero* dosing protocol. After considerable deliberation, NTP chose the Harlan Sprague–Dawley rat (Hsd:SD) as the primary rat model for future NTP rat studies. Limited historical data from previous carcinogenicity studies with the HSD rat are available from 8 previous studies and data from females indicate favorable outcomes for litter size, sex ratio and body weights for use as the standard protocol with *in utero* dosing followed by a 2-year in-life dosing for future carcinogenicity studies (King-Herbert et al., 2010).

The use of the B6C3F1 mouse strain was not altered by NTP following discussions at the workshop and workshop participants were strongly supportive that continued use of two species in the NTP bioassay program has merit both for cancer hazard identification and in subsequent human risk assessments (King-Herbert et al., 2010).

Neither the OECD nor the US FDA (CFSAN) carcinogenicity testing guidelines (OECD, 1981, 2009a,b; FDA, 2000a–f) specify the use of specific strains of rats or mice. FDA (CFSAN) allows for scientific justification of selected test species in meetings with FDA (CFSAN) scientists responsible for monitoring tests on the chemical to be tested to meet regulatory requirements. OECD (1981) guidance indicated that, the use of inbred strains has the advantage of known characteristics including the average life span and predictable spontaneous tumor incidences. Similarly, F1 hybrid mice derived from mating two inbred strains are preferable because they are more robust, have a controlled genetic constitution and have longer life spans than commonly used inbred mice.

2.2. Age of animals at study initiation

Pioneer studies in screening chemicals for carcinogenicity supported by the NCI in the early 1960s showed that young animals were often more sensitive to effects of known carcinogenic agents as evidenced by earlier appearance of tumors and at lower doses than in older animals (Vesselinovitch and Mihailovich, 1966, 1967). The age of initiation of treatment with a total of 6 injections of urethan given at 3-day intervals (Vesselinovitch and Mihailovich, 1966, 1967) showed that there was an inverse relationship in increasing age and incidence of tumors quantified at 12 months. Incidence of tumors from a limited number of doses was greater when treatment was started in newborn and infant rats, versus older rats, but acute toxicity was a confounding factor in the young animals that could be minimized in mice treated after weaning. The greater sensitivity of young animals to chemical carcinogenicity appears to be due, at least in part, to greater rates of cell proliferation (Alexandrov et al., 1990). However, possible advantages of greater sensitivity to neoplastic induction in mice dosed from birth could also be reproduced in mice post-weaning by using higher doses than tolerated by neonates, but which did not produce excessive numbers of mortalities (Weisburger et al., 1970). In addition, some chemicals have been found to show a higher degree or a more diverse pattern of neoplastic activity when given during a period of 6 weeks to 104 weeks of age (i.e., "standard bioassay") versus dosing only during the gestation period, or in neonates or up to weaning (e.g., aflatoxin, ethylene thiourea) (McConnell, 1992).

The reduced response at early age periods is likely attributable to a requirement for metabolic conversion of some compounds that are not metabolized efficiently in neonates and weanlings (Rice, 1979; Hattis et al., 2005). McConnell (1992) summarized the effects of age of the test animal on carcinogenicity findings from examination of the scientific literature, resulting in the following conclusions: (1) perinatal exposure alone (suggested as a highly sensitive period) does not detect some carcinogens (14% of 22 chemicals not detected); (2) perinatal exposure only rarely identified carcinogens not identified in a standard bioassay (e.g., Eugenol was the only chemical found to be carcinogenic following perinatal, but not adult, exposure; however, effects in adults were possibly confounded by reduced survival in the top dose group); (3) perinatal exposure in conjunction with adult exposure can increase the incidence of a given neoplasm; (4) both adult and perinatal exposure protocols, in most cases, reveal the same types of neoplasms; (5) perinatal exposure only rarely produces neoplasms not seen in a standard bioassay, but there are many examples where tumor multiplicity is greater in adult vs. perinatal exposure regimens (e.g., anethole, benzidine, diethylstilbesterol, estragole and vinyl chloride); (6) perinatal exposure, together with exposure through the lifespan, can reduce the normal latency period of tumor formation; and (7) genotoxicity of the test substance did not alter the conclusions noted previously. Because of the absence of clear advantages of perinatal dosing protocols, the use of an adequately high range of doses and continuous dosing for a significant part of the animals' lifespan were agreed as the compromise approach to testing. This allows detection of carcinogenic effects during different stages of the life span without incurring excessive study costs or significant increases in study complexity, design and conduct for marginal scientific gains.

The NTP carcinogen screening program has employed weanling rats and mice at the initiation of dosing in carcinogenicity studies for years. Periodic reviews by international safety assessment experts give no indication that this practice should change to use younger animals or to dose during the perinatal period (e.g., Bucher, 2002). OECD guidance (1981) indicates that fetal, neonatal and weanling animal differ from adults in numerous factors including biochemical, physiological, hormonal and immunological characteristics, but that, "at present, there is only limited evidence that perinatal exposure may reveal the carcinogenic potential of a chemical that would not have been revealed had the treatment started at a later age". This conclusion is supported by a review of perinatal carcinogenicity results (Rice, 1979). Additional discussion of perinatal dosing protocols and results is presented in the final section.

2.3. Randomization and stratification of test animals

Most regulatory agencies require that carcinogenicity testing protocols incorporate a scheme for randomization of the test animals for allocation of animals to specific test groups in order to permit statistically valid inferences about the results. Randomization using a neutral selection scheme for assignment of animals into test and control groups is considered important to minimize differences between treatment and control groups at the beginning of the study so that differences between groups at the end of the study can be ascribed to the treatment and not animal variability. In studies that evaluate differences in body weight gains in groups of animals (e.g., OECD, FDA, NTP and ERF protocols), it is important that the individual groups have a mean and variance that is as homogenous as possible to permit meaningful statistical comparisons (Robens et al., 1994).

Stratification, the sorting of animals by sex, weight or weight ranges, age, or litter, often precedes the randomization step, and is required by FDA (2000b) before animals are assigned to dose and control groups. Body weight is most typically used for stratification because the animals obtained from commercial sources are generally already sorted by age, sex and weight range and supplied without regard to litter (Hamm, 1994). The FDA (2000b) uses a stratification scheme based upon body weights. Heavy and light animals outside the range of ±20% from the mean and any animals exhibiting an unhealthy appearance are discarded and the remaining animals are randomly assigned to groups using a table of random numbers (NTP, 2006). The OECD (1981, 2009a) carcinogenicity

testing guidelines indicate that randomization is important and should be performed, although no particular stratification variable or method of randomization is suggested. FDA (2000b) specifies that stratification by body weight or weight ranges is acceptable preceding randomization and that any other procedure(s) should be described and supported.

The procedure used by ERF for animal randomization in its carcinogenicity tests is unique relative to typical randomization procedures. Because ERF has its own in-house production of animals, this permits tracking the history of individual animals to parents and litters. Bucher, who reviewed ERF randomization procedures, noted that ERF "assigns litters to the same dose group and uses all animals while keeping track of litter identification information". This description is consistent with comments from ERF investigators about animal assignment in a recent ERF study (Belpoggi et al., 2006a). The investigators state, "at weaning (at 4–5 weeks), offspring were identified by ear punch, weighed, separated by sex, and assigned sequentially litter by litter to the exposed and control group, respectively". However, it is not clear whether entire litters are assigned to dose groups in all ERF studies. For example, in a different ERF report of a recent carcinogenicity study, the investigators state that "animals were randomized in order to have no more than one male and one female/litter in the same group". (Soffritti et al., 2006) It is possible that the randomization procedure used by ERF in its studies of foods and food ingredients may be tailored to specific types of studies, but this is not clear from published studies (e.g., Soffritti et al., 2002, 2006, 2007; Belpoggi et al., 2006a,b). Intentionally assigning multiple animals from the same litter into the same dose group would not provide an optimum randomization of outbred animals. Inadequate randomization could result in clusters of individuals in the test group with genetically unique properties that could bias the independence of the test groups and exacerbate spurious indications of significant increases of rare tumors. Inadvertent assignment of littermates to the same study may occur because of the way animals are supplied from commercial sources. Such an event can be the cause of clusters of rare tumors in a study.

Animals also need to be proportionately distributed, or randomized, to study rooms, when multiple study rooms are used. This is to minimize possible confounding effects related to potential environmental differences, such as humidity, room and air cleanliness, and potential exposure to infectious agents. For example, in a review of an ERF study of aspartame, EFSA reviewers noted, "The large number of animals used in studies conducted by the Ramazzini Institute often necessitate that different dose groups are in different rooms, each having its own microclimate and rate of disease. When animals from various treatment groups are not randomized among the various rooms, different background pathologies per group can occur, including tumor incidence and survival" (EFSA, 2006). Magnuson et al. (2007) reviewed the EFSA findings and indicated that lack of randomization may have led to the low survival rate of the female control group at 104 weeks (27.3%) in comparison to the 45% survival rate in the group given the highest dose of aspartame.

2.4. Route of dosing

The oral route of administration should be used in tests of food substances to represent the normal route of human exposure in the diet or in beverages. The typical oral routes of dosing include feeding in the diet, gavage, or incorporation in drinking water. The dietary and gavage routes are the most frequent choices of administration of test substances in long-term studies. For dosing in the diet, the test chemical can be mixed in ground feed and the ground feed can be used directly or may be pelleted. The test substance may also be added directly to pelleted feed by use of a solvent carrier for the test substance or added to feed by thorough mechanical mixing. An issue with administration in the diet is cross contamination of groups (EMEA, 2005). Volatile chemicals may be added to feed following encapsulation or dosing may be performed by gavage of the chemical dissolved or suspended in a non-toxic solvent (e.g., water, methylcellulose, etc.). OCED (2001, 2009a) and FDA (2000a,b) guidance indicates that when substances are given via the diet or drinking water in long-term studies, it is important to ensure that the quantities of the test substance do not interfere with normal nutrition or water balance. The concentration of the chemical in the feed should not normally exceed an upper limit of 5% w/w of the total diet in order to avoid nutritional imbalances.

For both dietary and gavage dosing, it is extremely important to ensure that the test chemical is stable either in the feed or in any solvent used for gavage or feed preparation. Routine, periodic homogeneity and stability analyses and reports of analytical findings are required during the course of each study by OECD, FDA and EPA regulatory guidelines to ensure that animals receive the reported dose levels. In addition, stability data can be used to determine acceptable intervals for storage of dosing solutions or preparation of new lots of dosed feed during study conduct. Further discussion on evaluation of test substance purity, stability and homogeneity analyses as essential component of both dietary and gavage dosing studies is provided in Section 6.

Administration by gavage is technically more demanding than dietary dosing because this procedure involves daily handling of the animals which can produce stress and the possibility of rupture of the esophagus or introduction into the trachea during the dosing procedure leading to death of animals. A significant issue with gavage dosing is that the animals receive the dose in a bolus, which is not typical of human exposures via the diet. A gavage dose compared to dietary administration provides a known amount of the test substance at a fixed point in time (Johnson, 2006), but a bolus dose can have significantly different rates of absorption than exposures via diet and can produce atypical metabolic and physiologic effects in the test animal (Hamm, 1994; Renwick, 2008). A more significant issue with gavage dosing lies in the production of dosing artifacts caused by stress of handling and the dosing procedure. Õkva et al. (2006) evaluated stress produced by gavage dosing and reported significant effects on systolic and diastolic blood pressures as well as heart rates. Balcombe et al. (2004) summarized findings from five gavage studies that reported significant effects on corticosterone levels (+58% to 596%), weight loss (3.5% to 9%) and other parameters including body temperature (−2.8%), ambulation (−39%), rearing (−53%) and liver apoptosis (+355%). The effects of gavage on corticosterone levels are particularly troubling because of the potential for effects on metabolism, absorption and toxicity of the test substance. Gavage dosing can also produce physical effects in the forestomach which can affect tumorigenesis in this organ (IARC, 2003).

Following dietary administration, the absorption of a test substance from the intestinal tract is gradual and steady state blood levels occur over an extended time period (Krishnan and Andersen, 2007). With gavage dosing, the bolus dose leads to more rapid bioavailability and peak blood levels may be higher and steady state blood levels may be briefer than with dosing in the diet. Because of the possibility of atypical effects from gavage doses, the US EPA, that regulates chemicals in raw foods and the environment (EPA, 1998), requires that studies performed by gavage must provide experimental evidence that this dose route is equivalent to administration in the diet or drinking water or "Alternatively, metabolic information on both modes of administration should be provided so that appropriate interpretation of data can be accomplished" for relevance to human exposures.

2.5. Numbers of test animals/group

The purpose of animal carcinogenicity tests is to determine whether exposure to a chemical substance for a substantial portion of a lifetime produces a significant increase in the incidence of tumors in the test animals compared to concurrent controls. Thus, high numbers of test animals per dose group would be desirable for statistical reasons, but the goal of attaining a high level of statistical power must be counterbalanced with animal numbers that are ethically, technically and economically feasible. The NTP, OECD and FDA (CFSAN) indicate that a minimum of 50 animals/sex/group should be used. Additional animals are recommended to be added if interim kills are desired or a subset of the animals will be used for other purposes, such as pharmacokinetic studies or periodic disease monitoring that would require these animals to be handled differently from the core group of animals used in the carcinogenicity bioassay. Even with the minimum number of 50 animals/sex/group, a carcinogenicity assay of a single substance evaluated in two species at 3 dose levels requires 800 animals [50 animals/sex/group/species $\times$ 2 sexes/group $\times$ 4 groups (3 dose groups plus control group) $\times$ 2 species]. The FDA indicates that at least 25 animals/sex/group should be available at the end of the study to provide a sufficient number of animals to allow meaningful statistical comparisons.

The minimum number of animals required for a study has been evaluated on statistical grounds in comparisons to tumor indices in control animals as illustrated in sample size tables, such as those provided as examples in Gad and Weil (1986) and Robens et al. (1994). In cases where spontaneous tumor incidence at termination of the study is low (e.g., 5%), 35 animals per group would give 90% power to show a significant difference ($p \leqslant 0.05$) if the induced tumor incidence in treated animals was $\geqslant$30%. However, if incidence of the same tumors in control animals was 30%, then 389 animals per group would be required to achieve 90% power to obtain a significant difference in treated animals with a tumor incidence of 40%. Gad and Weil (1986) dramatically illustrate the problem in detecting significant differences relative to background levels by showing that with a control incidence of 1% in tumors, 570 animals would be needed to consider a 3-fold increase in tumors as significantly above control and 5100 animals would be required to detect a doubling as statistically significant. If the incidence of tumors in control animals is high (e.g., 10%), then small increases in induced tumors, possibly caused by weak carcinogens, would currently be impossible to detect on a practical basis as 1100 or 10,000 animals would be needed to obtain significant ($p \leqslant 0.05$) increases above tumor incidences in controls of 1% or 3%, respectively.

In recent years, ERF has advocated the use of large numbers of animals in carcinogenicity testing, "Mega-experiments", and has published results of studies with 150 animals/group. Research indicates that use of such large numbers of animals has possibly increased potential to detect weak carcinogens; i.e., statistically significant differences in dose group tumor incidence from control group tumor incidence when the background incidence in tumors is low (1–5%) and there is an increase of between 5% and 10% above control levels (Gad and Weil, 1986). However, this potentially desirable effect may be negated when the duration of the study is significantly prolonged, as is typical of studies reported by ERF, (see also section, titled Study Duration). With increased study duration, there can be increases in age-related spontaneous tumors/animal that can decrease statistical power (Gad and Weil, 1986). The statistical procedures used by ERF that ascribe significance to increases in *total* tumors rather than comparisons of site/tissue specific tumors or numbers of tumor-bearing animals (Bucher, 2002) are discussed in the section on Statistical Analyses of Tumor Data. Analyses of cumulative tumor incidences in aging animals poses a high hurdle both for pathologists and statisticians, who must make comparisons in relation to an increasing tumor background in untreated control animals, as well as risk assessors who cannot discern contributions from multiple tumors/animal when only totals observed at death are compared.

2.6. Number, selection and spacing of dosage levels

Maximizing the sensitivity for detecting weak carcinogens was a key goal in early screening protocols at the NCI (Weisburger, 1983), but the use of high doses to accomplish this goal sometimes compromised the statistical power to detect small differences from controls in the incidence of spontaneous tumors that increase with age, due to the dose-related early mortality that caused a reduction in the observed number of induced tumors. Thus, the highest dose, estimated from toxicity data to define a dose that did not produce excessive toxicity and decreased survival, except from tumor induction, was termed the *M*aximum *T*olerated *D*ose (MTD) or *E*stimated Maximum *T*olerated *D*ose (EMTD). The MTD was employed as an attempt to maximize any potential tumor effects from the test substance to ensure that weak carcinogens would have a higher probability for detection (Sontag et al., 1976). Haseman and Lockhart (1994) reviewed rodent carcinogenicity results for 216 chemical found to be rodent carcinogens in studies conducted by the NCI and NTP; they concluded approximately two-thirds of the chemicals would have been classified as carcinogens even without the results from the animals given the EMTD. However, the remaining one-third of carcinogens were detected only in studies that included a dose group given the EMTD, as a result of attaining tumor incidences significantly different from that seen in controls. In light of such findings, regulatory and advisory groups (NRC, 1993) have cautioned that conclusions of lack of carcinogenic activity of test substances is defensible only if there is experimental evidence that an MTD or EMTD was evaluated based on detectable physiologic or histopathological indications of toxicity in the highest dose group.

Although there is no universal scientific agreement on the characteristics of an MTD, the ideal MTD is often characterized as the highest dose that is estimated to produce no more than a 10% reduction in body weight gain and/or a quantifiable toxic effect that is not considered life threatening (Sontag et al., 1976; Haseman, 1985). Swenburg (1995) pointed out that the term MTD is defined as the *minimal toxic dose* in the EU and by the OECD (1981) and indicated that "the highest dose level should be sufficiently high to elicit signs of minimal toxicity without substantially altering the normal life span due to effects other than tumors". OECD (2009a) revised guidelines are more specific and indicate that the "highest dose level should normally be chosen to elicit evidence of toxicity as evidenced by, for example, depression of body weight gain (approximately 10%)". These newer guidelines also include guidance on spacing dose levels at 2- to 4-fold intervals and recommends against spacing on factors >10. The subtle differences in the US and EU in the MTD definition and in setting the maximum test dose should be considered in comparing results performed in compliance with EU testing procedures versus those citing FDA, EPA or NTP procedures.

Counts and Goodman (1995) cautioned that a bioassay screening program should not be focused merely on finding chemicals that are carcinogens but should include components to allow assessment of possible hazards under realistic conditions of exposure avoiding abnormal conditions that overwhelm normal detoxification and absorption mechanisms. Early NCI carcinogenicity screening tests were designed to simply detect carcinogenic activity and used only 2 dose groups of males and female mice and rats exposed at the MTD, 1/2 MTD and a third group given a control substance (Sontag et al., 1976). The NCI protocol was designed to

provide information about the carcinogenicity of a compound at the MTD, but it could not reliably indicate the dose response relationship or the minimal or no observed adverse effect concentrations. Refinements in the NCI program were incorporated into the NTP bioassay to test additional doses to provide information more suitable for risk evaluation in addition to the screening function of the bioassay. Hamm (1994) summarized the MTD as the dose (e.g., dietary concentration) that can be given that does not cause a significant decrease in survival from effects other than carcinogenicity and that is chosen based on published information – preferably acute, 14 day, and 90 day studies at the same laboratory using the same species, strain, and sex of animals dosed by the route of administration proposed for the carcinogenicity study.

In addition to use of dose group employing a MTD, three dose groups and a control group using males and females at each dose level are standard features of most current regulatory protocols. Additional groups may be used in cases when the dose–response curve is steep or for tests of highly important substances (either for commercial or exposure reasons) for which additional dose–response information is critical, often to most reliably understand the maximum no effect level (Portier and Hoel, 1983, 1984). A minimum of 3 dose groups are needed for trend tests as well as to help define a no-effect level for use in risk assessment (Portier and Hoel, 1984; Rhomberg et al., 2007). Gad and Weil (1986) suggested that selection of the intermediate doses is relatively easy compared to the complexity involved in selecting the highest dose, which must be sufficient to challenge the animals as evidenced by body weight or clinical effects and not so excessive that it will produce mortalities, with consequent loss of statistical power in a 2-year bioassay. The intermediate and lowest doses can be chosen on a statistical basis (Portier and Hoel, 1983, 1984; Gaylor et al., 1985; Krewski et al., 1984) or empirically (NTP, 2006; OECD, 1981). Plotting doses on a logarithmic scale to follow thermodynamics that govern observed effects has also been proposed and could be used to estimate the toxicity or carcinogenicity threshold (Waddell, 2004a,b, 2008).

FDA (2000b) and OECD (1981) carcinogenicity testing protocols specified three treatment groups of mice and rats exposed at the MTD, 1/2 MTD and 1/4 MTD (lower fractional doses tested as determined empirically from range finding results) and use of a control group. OECD (2009a) revised guidelines specify that dose levels should be spaced at approximately 2- to 4-fold intervals below the maximum dose that is based on toxicological considerations – such as a 10% decrement in body weight gain. The advantage of using both mid- and low-doses close to the high dose in carcinogenicity tests can enhance the acceptability of test results as adequate when the high dose produces excessive (unanticipated) toxicity and the mid dose suffices as an adequate MTD for regulatory purposes (Swenburg, 1995). OECD (1981) specified that the MTD should be used as the highest dose, but that substances incorporated into feed should not exceed 5% by weight, with the exception of dietary nutrients that can be tested at very high percentages when nutritional effects on the feed must be carefully evaluated. Current OECD (2009a) carcinogenicity guidelines base the highest dose on toxicological findings such as body weight gain or target organ effects observed in shorter term or dose range finding studies. The typical procedure used by laboratories to define an MTD is based primarily from results of 90-day toxicity studies in the same strain and route of dosing to be used in the carcinogenicity test. The traditional factors for estimating the MTD, intermediate and low doses have been body weight gain and survival, but more recent tests have evaluated histopathological changes, clinical and pharmacologic effects and metabolism/pharmacokinetic data (Haseman, 1985; Haseman and Lockhart, 1994). Non-neoplastic pathological changes in preliminary studies are also considered for setting the MTD when the effect has the potential to produce life-threatening effects that could produce excessive mortality (e.g., kidney lesions seen in a 90-day toxicity study) (Haseman, 1985).

In 1985, Haseman (1985) reported that, of 52 chemicals tested in the NCI/NTP program, more than 66% of the positive carcinogenicity findings would have been missed if the MTD had not been used as the highest dose. In contrast, in 1994 Haseman and Lockhart (1994) evaluated a larger data base of NTP studies conducted since 1985 and revised the estimate of incorrect findings when the MTD was not used. They noted that "Approximately two-thirds of [216 chemicals found to be] rodent carcinogens would have been detected even without the top dose (estimated MTD), but in many of these studies, some site-specific carcinogenic effects would have been missed". "Among the remaining one-third of rodent carcinogens that required the top [MTD] dose for statistical significance [for detection of a positive effect], approximately 80% had numerically elevated rates of the same site-specific tumors at lower doses as well". Thus, although the use of the MTD has been questioned because of possible tissue damage, metabolic overload, inflammation, cellular proliferation and tumor formation not encountered from real life exposures to much lower concentrations (Contrera et al., 1995, 1997; Foran, 1997), the MTD appears to be the best empirical method to choose the highest test dose for maximizing sensitivity of rodent carcinogenicity tests, bearing in mind that effects at high doses may not be relevant to the consequences of low level human intake. Gaylor and Swirsky Gold (1998) have noted from an analysis of results from 2-year bioassays that the MTD established in a 90-day study divided by 7 provides a relatively precise determination of the dose estimated to produce an excess of tumors in 10% of animals (LTD_{10}). They concluded that for chemicals with anticipated human exposure that is estimated to be small relative to the MTD/7 dose in animals, "there may be little value in conducting a chronic 2-year study in rodents because the estimate of cancer risk [to humans] would be low regardless of the results of a 2-year bioassay".

For test of chemical substances with no caloric value in food, the FDA (2000a,b) requires testing of 3–5 doses up to the MTD defined by minimum toxic responses in preliminary tests or up to a limit dose of 5% in the diet. It should be noted that at these dietary concentrations, rats and mice are required to ingest up to 5–10 g/kg body weight/day of substances that may only be ingested at infinitesimally lower dietary levels by humans, such as dietary supplements. FDA (CFSAN, 2007) is explicit that the MTD should *not* be based on published information that is unrelated to the toxicity of the test substance in the same species and by the same route and should not be selected based upon a selected margin of safety from expected human exposure to the substance. OECD (1981, 2009a,b) procedures for setting an MTD as the highest exposure dose are similar to NTP and FDA (CFSAN, 2007). OECD (1981, 2009a) guidelines indicate that the highest dose should not exceed 5% in the diet and that the lowest dose should be approximately 10% of the high dose. The OECD (2009a) revised guidelines include additional guidance for dose selection including use of toxicokinetic and metabolic information, precursor lesions seen in range-finding studies, mode of action (e.g., hormonal effects) and consideration of human exposure levels. Both FDA (CFSAN) and OECD treat tests on dietary nutrients or food additives with negligible or no observable toxicity as dietary components that have the ability to significantly alter nutritional status of the animals when tested at very high percentages of the diet. Testing features pertinent to assessing carcinogenicity of dietary substances are considered in the section on Animal Diets.

Methods used to determine the highest test dose in carcinogenicity studies performed at ERF are difficult to understand from descriptions provided in publications. Bucher (2002) described ERF dose selection procedures based on "extensive discussions

with Drs. Soffritti and Belpoggi [of ERF] under an agreement for scientific collaboration with the NTP" as based upon "efforts to provide a minimally toxic challenge to the animals, but doses are usually based on toxicity information gleaned from the literature". The concepts of "minimal toxic challenge" at ERF and the "minimal toxic dose" definition of the MTD in the EU present subtle differences that necessitate clear explanations for dose selection in studies performed under each definition of the highest test dose. In recent carcinogenicity studies at ERF, rationale for selection of the highest tested dose appears contradictory. For one study (Soffritti et al., 2002), it was noted that typically 3 dose levels and a control are employed in tests and that the "dose levels are the maximum tolerated level, the order of magnitude to which humans may be exposed, and one intermediate level". The authors also indicate that range-finding experiments are conducted when literature information on toxicity is not available. No discussion or data from preliminary tests , however, are typically mentioned in their publications. In a study on Coca Cola® (Belpoggi et al., 2006a), a total of 1999 male and female rats were used with dosing initiated at different times for different sub-groups; i.e., beginning at *in utero*, or at 7 weeks, 30, 39, or 50 weeks. One-hundred percent Coca Cola® completely took the place of drinking water and water served as the control substance in the study. The rationale for both the use of a staggered dosing initiation and a complete replacement of drinking water in the test groups without compensating for caloric effects from the test substance was not explained. Although fluid consumption and body weights were reported to be measured during the study, no data were presented to illustrate either the effects on the conclusions concerning increases in fluid consumption and/or body weights. In addition, no information was presented on whether there were tests to determine changes in hematological, clinical chemistry and/or urinalysis parameters in test animals, changes in which can have downstream effects in target organs. These kind of deliberations are important for consideration of the reliability of conclusions regarding carcinogenicity of the test material.

In a separate ERF study, the basis for the highest and lower test doses was again difficult to understand from information provided (Soffritti et al., 2006; Belpoggi et al., 2006b). The authors indicated that the test doses were supported by survey data for consumer consumption, which they referenced and that "were remarkably consistent across studies and [consumption was well below the acceptable daily intake (ADI) both in the United States (50 mg/kg bw) and Europe (40 mg/kg bw)". The same report, however, says that the doses were based on "An assumed daily intake by humans of 5,000, 2,500, 500, 200, 20, 4 or 0 mg/kg bw". Statistically significant increases in tumors were reported at doses as low as 20 mg/kg bw, which was described as "much less than the current acceptable daily intake". The data were used to suggest that the ADI for aspartame was not protective of human health and that "a reevaluation of the present guidelines on the use and consumption of [aspartame] is urgent and cannot be delayed". Important to note is that the ADI is not representative of current intake or EDI. With respect to FDA (CFSAN) guidelines in setting an MTD for acceptable dietary exposure studies (FDA, 2000a,b), a dosing rationale based upon "An assumed daily intake by humans" as used by ERF would not be an acceptable rationale. The FDA/CFSAN "Total Diet" surveys (FDA, 2008b) clearly show that dietary intake of various foods, food ingredients and their potential contaminants can vary significantly among individuals and groups and are confounded by many factors including age, nationality, region of the country, season of the year and dietary preferences among many other factors which precludes a simple statement on "assumed daily intake by humans" of any dietary component. ERF describes its procedure for determining the actual ingested dose as "daily consumption [sic] in milligrams per kilogram body weight was calculated considering the average weight of a rat for the duration of the experiment as 400 g, and the average consumption of feed as 20 g/day for both males and females". Thus, the dosages/group assumed all animals of both sexes were equal in weight and all consumed the same amount of feed/day throughout the 2 plus years of testing until death. These assumptions are highly unlikely for actual feed consumption that can change as animals age. The use of such assumptions makes it difficult to compare the study results to the results of other studies and understand carcinogenicity thresholds. Internationally recognized protocols and publications also specify that data on both feed consumption and body weights of individual animals be collected regularly throughout the study. It is not clear why ERF would not utilize this data for calculation of the ingested dose instead of using assumptions of constant weight and feed consumption "during the duration of the experiment". Such practices are unusual and both atypical and unlike standardly accepted toxicological practices advocated by major international testing protocols (FDA, 2000a; NTP, 2006).

2.7. Use of ADME (Absorption, Distribution, Metabolism and Excretion) and PK (Pharmacokinetic) data for dose selection

Alternative procedures to set an appropriate high (MTD) dose in carcinogenicity tests have been proposed because of concerns that the MTD can pose excessively high doses not based on "sound toxicologic principles" (Foran, 1997). ADME and PK studies can be valuable to assess the potential of test doses to cause metabolic "overload" and "saturation" of normal physiological/metabolic pathways and lead to carcinogenicity effects that are "secondary" or unrelated findings relative to assessing human risk (Gehring and Blau, 1977; Gehring et al., 1978; Haseman, 1985; Foran, 1997). The NTP now incorporates routine metabolism and pharmacokinetic (PK) tests as part of the decision process to choose the doses for long-term rodent carcinogenicity studies (Buchanan et al., 1997).

To avoid using test concentrations with questionable relevance to anticipated human exposures, the International Congress for Harmonization (ICH) (ICH, 2008) and European Medicines Evaluation Agency (EMEA, 2005) have recommended that the MTD for carcinogenicity studies could alternatively be based on pharmacokinetic (PK) parameters. The ICH recommended that the MTD used in rodent tests should be 25 times the measured AUC (area under the curve representing peak circulating dose in blood) of the parent compound and/or metabolites determined in human PK studies. FDA (2008a) adopted the ICH guideline S1C (R2) (ICH, 2008) definition for an MTD. A minimum systemic exposure ratio is considered useful for the high dose selection for rodent carcinogenicity studies on candidate drugs and biologics. Contrera et al. (1995) reported that, for 35 of 84 drugs in FDA files with available information on both rodent carcinogenicity tests and PK analyses, all 35 had human–rat systemic ratios of $\leqslant$10 so that an MTD at a ratio of 25 would be sufficient to detect all carcinogenic substances in the FDA dataset and IARC Type 1 and 2A carcinogens as well. Swenburg (1995) noted that, for most chemicals evaluated in rodent and human PK studies, the AUC in rodents is typically 1–3 times the AUC found in humans, meaning that this approach for setting the MTD for carcinogenicity studies should be considered.

A PK approach may be useful for setting the MTD for tests involving human and animal drugs and biologics, because PK studies are performed in rodents and humans as part of the drug characterization process. In addition, exposure to drugs is based on a risk:benefit evaluation such that effects seen in animal tests are subject to risk-benefit considerations in which administration to humans for potential benefits may be controlled by application of safety factors to effects seen at particular dose levels in animal tests. In contrast, risk-benefit considerations are not applicable

for tests on proposed new food substances or food additives. Similarly, the use of ADME and PK data may have more limited applications for safety assessments of food ingredients compared to safety assessments of pesticides or safety and efficacy tests of drugs. ADME and PK data for food substances as finished products for consumers are typically limited or not available. Generally, safety of such substances is based on historical uses and/or proven safety of the compositional ingredients, such that safety concerns are not expected with their consumption. The expense and difficulties in performing studies with such substances, that often contain complex ingredients, is also significant. In addition, doses for long-term toxicity and carcinogenicity studies of food substances are based on regulatory requirements for attaining an MTD (or limit dose) and a NOAEL (no adverse effect level) for the highest and lowest doses in an acceptable test as specified in most testing guidelines as discussed previously. When ADME and PK information are available, however, these can be used for guiding selection of the intermediate dose(s) in carcinogenicity studies and for evaluating whether the MTD is excessive with respect to saturation of ADME parameters (Sumner and Stevens, 1994). Although beyond the scope of the present review, when such data are available, ADME and PK can also provide crucial information for interpretation of carcinogenicity study data and for qualitative and quantitative risk assessment of the relevance of the findings for human exposures (Sumner and Stevens, 1994; Barton et al., 2006).

2.8. Control groups

Carcinogenicity tests conducted by NTP (2006) and ERF (Soffritti et al., 2006) use equal numbers of male and female animals in control groups as in treatment groups. In most cases, the control groups in feeding studies receive untreated feed, although a second vehicle-treated feed control group may also be used if dosing mixtures in feed required use of a vehicle. FDA (2000b) indicates that when the addition of the test substance to feed may affect diet consumption due to palatability and/or overall dietary nutrient quality, which could result in reduced feed intake and/or altered nutritional status of test animals, two control groups can be used to help understand the effect of the dietary-related change. For example, in testing a non-nutritive test substance through dietary consumption, one control group can be fed basal diet and a second control group can be fed the basal diet supplemented with inert filler (e.g., methylcellulose) at a percentage equal to the highest percentage of the test substance in the diet. In studies of aspartame conducted by ERF, such additional control groups were not employed. These studies do not include hematology, clinical chemistry or urinalysis testing, which could allow for further consideration of differences in histopathologic changes. The combination makes difficult the interpretation of results in test animals compared to controls with regard to measures of tumorigenicity.

Studies employing two (dual) control groups in carcinogenicity tests have also been used to identify the extent of variability in untreated animals (Baldrick, 2005; Baldrick and Reeve, 2007) and to evaluate biological significance of tumor increases in treated groups (Haseman et al., 1986; FDA, 2001). Guidance provided by FDA (2001) indicates that arguments for use of two controls are that the results from the two identical controls can be used as a mechanism for identifying the extent of control variability. The results can be used to help evaluate the biological significance of increases in tumor incidence in the treated groups (i.e., true increases versus random variation). Arguments have been made, however, that the two control groups should be treated as a single control group for statistical purposes.

In considering appropriate tests for drugs, which can be relevant to consideration of appropriate tests for food ingredients and foods, the FDA (2001) provides guidance on the use of historical controls for qualitative purposes perspective on concurrent control data:

> "*The concurrent control group is always the most appropriate and important in testing [substance] related increases in tumor rates in a carcinogenicity experiment. However, if used appropriately, historical control data can be very valuable in the final interpretation of the study results. Large differences between studies can result from differences in nomenclature, pathologists reading slides, the specific animal strain used and laboratory conditions. It is therefore extremely important that the historical control data chosen be from studies comparable to the current study, generally recent studies from the same laboratory using the same strain of rodent. Historical control data are particularly useful in classifying tumors as rare or common. A statistically significant increase in a rare tumor is unlikely as a chance occurrence so that it is critical to decide whether a tumor is rare or not. Rare tumors are generally tested with less stringent statistical decision rules. Historical control data can also be used as a quality control mechanism for a carcinogenicity experiment by assessing the reasonableness of the spontaneous tumor rates in the concurrent control.*"

In cases of atypical frequencies of neoplastic or pre-neoplastic lesions, concurrent controls may present certain limitations. A lower-than-normal tumor frequency in the concurrent control animals may lead to the finding of a statistically significant increase in the incidence of lesions in the dose groups, whereas a higher-than-normal tumor frequency in the concurrent control could mask a carcinogenic response in dosed animals (Deschl et al., 2002). In such cases, historical control data may help in discerning the presence or absence of a carcinogenic hazard (Haseman et al., 1984, 1992; Elmore and Peddada, 2009). A number of laboratories have collected historical information on tumor incidences in untreated animals for several species, including data from rodent strains used to evaluated chemicals for carcinogenic potential, and these data are available as reference points for studies in the same species/strains (Bomhard et al., 1986; Bomhard and Mohr, 1989; Chandra et al., 1992; Chandra and Frith, 1992; Bomhard, 1992, 1993; Bomhard and Rinke, 1994; Brix et al., 2005).

Recent guidance from the Historical Control Data (HCD) Working Group under the auspices of the Society of Toxicologic Pathology (Keenan et al., 2009) produced the following recommendations on the use of HCD summarized in excerpts from their publication:

- the concurrent control group is the most relevant to use for comparison of treatment related effects; however,
- HCD may be useful in the interpretation of rare tumors, marginally greater incidences in treated animals compared to controls and to review trends in tumors that may evolve over time;
- HCD are best when factors including strain, route, vehicle, feed, feeding practices, study duration, necropsy, tissue and slide preparation and diagnostic criteria are standardized for the data collected.
- HCD from studies that had a peer-review process are generally more reliable for reference purposes;
- HCD compiled at the same laboratory than data from multiple laboratories and the difficulty in ensuring quality of published data must be considered for relevance to a particular study;
- HCD should be considered as one source of information that adds to the "weight of evidence" approach to assessing carcinogenic potential.

A major difficulty in using historical control data is the comparability with the study under evaluation. For confidence in the quality of historical control data, standardized procedures in all as-

pects of study conduct and extensive background information on any study entered into a database is needed (Deschl et al., 2002). The Registry of Industrial Toxicology Animal-Data (RITA, 2009) database for historical data of tumors and pre-neoplastic lesions was created to establish a centralized European database providing standardized and valid historical background data for carcinogenic risk assessment. In accordance with goals of the HCD working group, this database provides information about variability on standardized, peer-reviewed spontaneous tumors in the most common animal strains used in the EU and US. The data can be used to track survival rates and occurrence of tumor-bearing animals to assess stability over time. In addition, they can be used to help carcinogenic hazard identification of rare tumors or marginally increased incidences of tumors. Unfortunately, these data cannot be used to evaluate comparability of the numerous studies conducted at ERF, because of the unique strain of rats used by ERF (CRC/RF) employed for its carcinogenicity studies and because of the potential for confounding effects with the different animal care practices and histopathological peer-review procedures at ERF compared to those at laboratories that provide the data for the RITA database.

3. Animal facilities and welfare, study conduct and management

Minimum requirements governing animal facilities in laboratories that perform rodent chronic carcinogenicity tests for submission to regulatory authorities are specified in detailed standards established in international harmonized standards for Good Laboratory Practices (GLP) (FDA, 1978; EPA, 1983, 1984; OECD, 1998). Animal care and use in the US was addressed in 1963 when the National Research Council (NRC), part of the National Academy of Sciences in the US, published the first Guide on the Care and Use of Laboratory Animals (“The Guide”). The Guide has undergone numerous revisions over decades, providing a living document for the incorporation of the latest scientific input from industry, government and academic laboratory scientists. The Guide (NRC, 1996) is an international resource and has become a standard reference in formulating standards for animal treatment in the US. Recommendations in The Guide have been incorporated into Good Laboratory Practices (GLP) standards and used to set policy by regulatory agencies responsible for animal welfare. In addition, the welfare of laboratory animals is regulated by US federal law, as a result of the passage of the Animal Welfare Act in 1966. The Act has been amended 6 times since its passage, and it is enforced by the USDA, APHIS animal care agency. The Act (7 USC, 2131–2159), amendments to the Act, and the USDA–APHIS Animal Care Policy Manual and Final Rules for Animal Welfare can be accessed via the Internet (US Animal Welfare Act, 1966). In Europe, the European Convention for the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes provides guidance on the use of animals in experimental procedures, and the ETS 123 document contains detailed information regarding animal welfare that is similar to information in The Guide (NRC, 1996). Technical Appendix A, from the European Convention document, was since revised, with the revisions adopted in 2006, and it contains comprehensive guidelines for the accommodation and care of animals (Council of Europe, 2006).

3.1. Animal housing

The benefits and pitfalls of housing animals singly or in groups during carcinogenicity studies have been debated for years. Influencing the debate are factors such as animal welfare considerations, comparative economics, and labor and logistical considerations with long-term animal maintenance. (Combes et al., 2002). Group housing requires fewer cages, feed dishes, amounts of bedding, etc., than single housing, but it negates the ability to accurately determine individual animal feed consumption, makes routine animal identification checks more difficult and time consuming, and can engender losses of animals due to fighting and cannibalism. However, rodents are typically social animals, so group housing has been found to have beneficial effects in some studies (Riley, 1981). Hamm (1994) succinctly described the choices between single and group housing as conflicting choices between, respectively, possible “isolation” stress (Riley, 1981) and the even greater stress produced from fighting and/or cannibalism. In feeding studies, only single-housing allows for actual measurement of individual animal feed and dose consumption. In group housing, numerous factors affect feed and dose consumption per-animal estimates. For example, dominance (i.e., hierarchical) patterns are possible in established in groups of rats, which can lead to variations in feed availability/rat. Differences in animal size can influence total food and dosage intake. Stress can influence feed intake, which, in dietary studies, then influences dose intake. In group housing conditions, stress levels can vary from cage to cage, and can vary from animal to animal within a single cage.

FDA (2000b) requires individual housing for feeding studies, to avoid complications with animal competition for feed, make animal identification and feed consumption data more reliable, and eliminate possibilities for fighting and cannibalism. OECD (1981, 2009a) guidelines provide for both single and group-housing in carcinogenicity studies, stating that test “animals may be housed individually, or be caged in small groups [typically $\leqslant$5] of the same sex; individual housing should be considered only if scientifically justified”. NTP (2006) has found that aggression in their testing program is particularly an issue with male animals and deals with this negative aspect of group housing by specifying that NTP contractors should house no more than 3 males/cage and 5 females/cage. Both the NTP and the ERF testing programs employ group housing of animals regardless of whether the test substance is an industrial, environmental or food chemical. Thus, results obtained in tests at NTP rely on estimates of feed consumption/animal and the dose/group from estimates of average consumption by group rather than individual. As discussed, this negates the possibility of actual individual feed and dose intakes. As noted previously, ERF bases the dose/group from default assumptions on animal mean weights and feed consumption for the duration of the lifespan studies, despite the apparent collection of data that could be used to provide specific calculations.

Cage-location -related effects can also alter some physiological responses (Riley, 1981). For example, animals housed constantly near room lights have been found to have a greater incidence of ocular pathology (Reuter and Hobbelen, 1977). As a result, rotation of cage locations on a regular basis is a recommended procedure in long-term rodent studies (Hamm, 1994; Gad and Weil, 1986).

The size of animal caging has been found to be an important variable in carcinogenicity testing. In studies employing group housing, the comparatively decreased floor space/animal has been associated with increased stress. Increased stress can, in turn, alter animals’ responses to toxic insults (Clark et al., 1981; Hurst et al., 1999). Nonetheless, the Guide (NRC, 1996) leaves the choice of single or group-housing open and is relatively flexible with regard to choosing types of caging materials and bedding. The primary concern is, instead, for the size of caging and the ease of cage sanitation, with a focus on eliminating the potential for bedding material contaminants. The latter is discussed further in the section on Animal Care and Housing. For both rats and mice used in carcinogenicity tests, The Guide (NRC, 1996) specifies the amount of space that should be allowed per animal. The specifications are based on animal body weight which, for the sake of convenience, leads most laboratories to use caging that will

accommodate test animals up to the maximum weight attained at maturity.

With F344/N rats used in the NTP screening program fed an NTP2000 diet, typical maximum attained weights for males and females were 496.8 ± 30.8 g and 344.5 ± 28.7 g, respectively (Rao et al., 2001). For SD rats, such as the CRC/RF strain at ERF (type of diet not specified in publications), the maximum average weight/group can be gleaned from study publications. For example, from a published graph on estimated per-animal body weight, (Soffritti et al., 2006), CRC/RF males and females attain an average maximum weight of approximately 500 g and 350 g, respectively. Thus, the maximum weights for these two different strains, maintained at the respective laboratories, are remarkably similar. In contrast, Chapin et al. (1993) reported that SD rats fed an *ad lib* diet attained approximately 650+g and 400+g for males and females, respectively. This indicates that individual body weight estimates for ERF rats are conservative and true weights may be variable from animal to animal within a cage.

According to The Guide (NRC, 1996) specifications for group housing, rats with body weights up to 400 g should be housed in cages with approximately 40 in.^{2} floor area/rat and rats weighing up to 500 g should have a floor area of approximately 60 in.^{2}/rat. At both NTP and ERF, maximum male weights are estimated at about 500 g, so floor space for males should be at least 60 in.^{2}/rat. Both NTP and ERF use polycarbonate cages with floor dimensions of, respectively, 22×12 in (264 in.^{2}) and $\sim 16 \times 10$ in. (160 in,^{2}) (reported as 41 cm × 25 cm) (NTP, 2006; Soffritti et al., 2007; Belpoggi et al., 2006a,b). As previously discussed, NTP allows for a maximum of 3 males per cage, thus caging floor space is for male rats is approximately 88 in.^{2}/animal. At ERF, where group housing permits 5 males per cage, available floor space for males is approximately 32 in.^{2}/animal. This amount of floor space is well below (about 1/2) the minimum specified by The Guide. Floor space for females at NTP and ERF (5 females allowed/cage and maximum estimated body weight not >400 g) is , respectively, approximately 53 in.^{2} and 32 in.^{2}/animal, while the minimum floor space specified by The Guide for animals of this size is 40 in.^{2}/rat. Thus, NTP studies are conducted typically well within recommended guidelines, while ERF studies are not. Guidelines for minimum per-animal floor space are there to help minimize stress and possible disease transfer encountered in group housing. ERF notes that its housing of both male and female rats is in accordance with Italian law that regulates the use of animals for scientific purposes (Decreto Legislativo, 1992). No specifications for the size of animal cages, however, are given in this legislation. In contrast, animal housing at ERF appears to be crowded, based on recommendations and specifications provided by animal experts, NRC guidelines and by the European Community (EC) animal welfare Directive 86/609/EEC (EC, 1986). Crowded housing has been correlated with increased spread of colony diseases, animal aggression and cannibalism that may have an impact on study outcomes.

Both types of caging and number of animals/cage may potentially influence the incidence rates of certain hormone-dependent or stress-related proliferative responses (Keenan et al., 2009). Clark et al. (1981) reported that, in studies of hexachlorobenzene (HCB) toxicity, when singly housed rats (1000 cm^2 floor space) were transferred to smaller cages each with 4 rats (100 cm^2 floor space/rat) and compared with rats that continued single housing, the crowded group had severe loss of body weight, lower liver and kidney weight, higher mortality rates and greater tissue residues of HCB compared to the dosed rats that were singly housed. The authors attributed the differences in toxicity and mortality to stress caused by the crowded housing conditions.

The confined space allotted to the test animals at ERF has not been reported to cause difficulties in study conduct or to affect survival in the lifetime studies. However, the absence of any such reports is not evidence of no effect. The EFSA Panel that reviewed the first aspartame study reported by ERF (Soffritti et al., 2005, 2006) concluded (EFSA, 2006):

> "the size of the cages in which the animals were housed during the study might not sufficiently cover the need for space of adult animals (Directive 86/609/EEC), and considers that this could have had an influence on the outcome of the study in that high-density housing could have contributed to the high incidence of infection seen in the study".

Concerns about infection in these ERF studies were also expressed by FDA (FDA, 2000c).

Questions about the housing of animals at ERF and the consequent health of the animals have also been raised in the EFSA (2006) review of the unpublished portions of raw data from the study (Soffritti and Belpoggi, 2005). This review revealed disease problems in the animals not mentioned in the publications:

> "*The Panel notes that survival of rats in the study at 103 weeks of exposure was relatively poor, ranging from 22% (controls) to 31% in males and 27.3% (controls) to 45% in females. The non-tumour pathological findings in the rats indicated a very high incidence of infection, both in treated and untreated rats, which is likely to be linked to the poor survival. There was no treatment-related trend observable in the incidence of this infection. The incidences in treated and control groups for brain abscesses ranged from 7% to 11% in males and from 4% to 20% in females, for pyelonephritis from 23% to 62% in males and from 31% to 83% in females, for pleuritis from 22% to 71% in males and from 47%–94% in females, and for bronchopneumonia from 81% to 95% in males and from 69% to 97% in females. In addition, a relatively high incidence of peritonitis, liver abscesses and hepatitis, pericarditis and meningitis occurred in all groups. This very high incidence of infections is unusual for toxicology studies according to current standards.*

Animal crowding is also known to have potential effects on both survival and tumor rates. Extrapolation from plots of survival provided in two recent ERF publications (Soffritti et al., 2005, 2007) shows less than 50% survival for control males at 96 weeks in one study (Soffritti et al., 2007) and approximately 40% survival of male and female control animals at 104 weeks in a second study (Soffritti et al., 2005). This relatively low survival percentage would be a cause to terminate the study under US FDA (CFSAN) testing guidelines, in order to preserve statistical power, if the test had been conducted for regulatory purposes. The duration of these ERF studies, however, was not curtailed at these points. The OECD (2009a) revised carcinogenicity test guidelines also indicate that a "study should not be extended beyond the point when the data available from a study are no longer sufficient to enable a statistically valid conclusion to be made".

The EFSA review of the unpublished data provides background information that may explain the low survival percentages from extrapolations from publications discussed above (Soffritti et al., 2005, 2007). Belpoggi et al. (2006b) published a more detailed presentation of tumor data from the same study (Soffritti et al., 2005, 2006) and similarly did not indicate any disease or other problems in the study possibly related to crowding but instead stated "the study proceeded smoothly without unexpected occurrences". Thus, the impact of crowding and disease incidence on the conclusions from this or other studies at ERF cannot be determined. However, Keenan et al. (2009) indicate "The type of caging and number of animals per cage can influence proliferative responses. A decrease in testicular interstitial cell tumors and increase in pituitary tumors in F344 rats was reported in studies where rats were housed individually compared to studies with group housing".

Other parameters at ERF such as environmental conditions (e.g., lighting, humidity, temperature, cage conditions, water source, general sanitation, monitoring potential exposure to other test articles and ventilation) have also not been reported in publications cited previously.

3.2. Animal diet

According to all of the regulatory and scientific guidelines for conduct of carcinogenicity studies, feed and water should be provided *ad libitum* and the diet should contain recommended levels of nutrients (NRC, 1996; OECD, 1981, 2009a; NTP, 2006; FDA, 2000b). Guidelines also recommend tests of diet for nutritional adequacy and demonstration of absence of impurities that could affect study outcomes. Because many substances tested in feeding tests may have little or no nutritional value, they can displace dietary nutrients on a dietary weight basis. At high doses (>5% by weight), both caloric and nutrient content of the diet is diluted in comparison to diets containing lower doses, particularly control diet. OECD (1981, 2009a), but particularly FDA (2000b), guidelines caution that specific attention must be paid to nutritional status of the animals in such studies, and FDA (CFSAN) requires that feed consumption of the animals in the treated groups should receive particular attention. As noted previously, FDA (2000b) indicates that the use of two control groups can help in determining adverse nutritional effects on modifications of the diet received by the high dose group.

Both OECD and FDA (CFSAN) guidelines recognize that studies that evaluate substances that are also nutrients can result in dietary and nutrient modification at the different doses tested in a study. Rhomberg et al. (2007) pointed out that in addition to simpler caloric modifications in diets fortified with the test substance, additional dietary disturbances must be considered, including effects on gut transit times and/or gut microflora, vitamin and nutrient bioavailability, digestive enzyme secretion and regulation, and hormonal status. FDA (2000b) specifically recommends that the diets with different doses of substances that can affect nutritional status may need respective adjustment for caloric and/or nutritional components to ensure that nutritional factors do not cause health effects that could compromise interpretation of test results.

A significant problem in chronic rodent studies with *ad libitum* feeding is significantly increased weight and decreased percentage survival in comparison to historical control ranges (Haseman, 1992; Seilkop, 1995; Christian et al., 1998). Numerous studies have shown that *ad libitum* feeding or inappropriate diet can compromise the health and longevity of animals and concomitantly reduce the sensitivity of the carcinogenicity bioassay by reducing animals that survive to termination (Hart et al., 1995; Newberne and Sotnikov, 1996; Keenan, 1996; Keenan et al., 1994, 1996, 2009). Rodent obesity has also been implicated in an increased incidence of background tumor rates in control animals resulting in decreased sensitivity in statistical comparisons in carcinogenicity bioassays (Seilkop, 1995; Abdo and Kari, 1996; Allaben et al., 1996; Christian et al., 1998; Keenan et al., 1994, 2009; Keenan, 1996; Nold et al., 2001). The influence of dietary restriction on reducing tumor incidence, growth or development has been known since the early 1940s, as discussed by Visscher et al. (1942) and Tannenbaum (1940). These same authors cite earlier studies with similar effects of diet restriction on growth of transplanted tumors.. Incidences of certain site-specific tumors, most notably mammary gland and pituitary gland tumors in F344 rats and liver tumors in mice, were shown to have a strong positive correlation with 52-week body weight (Haseman et al., 1997). Although a comprehensive evaluation of the effect of diet on carcinogenicity results is beyond the scope of the present review, the role of *ad libitum* feeding (as required in EPA, FDA (CFSAN) and OECD testing guidelines) should be kept in mind in evaluating the results of carcinogenicity tests. Imai et al. (1990) and others (Abdo and Kari, 1996; Haseman et al., 1997; Molon-Noblot et al., 2003) demonstrated that reductions in spontaneous tumor rates and increased lifespan could be achieved by dietary restriction. Keenan et al. (1996, 2009) cited the observation of benefits on longevity from moderate caloric restriction or diet modification by reducing protein and/or increasing fiber compared to *ad libitum* feeding of a standard nutritionally balanced rodent diet. The authors noted that "dietary modifications that contribute to a reduction in body weight gain and a consequent reduced incidence and severity of chronic degenerative conditions lead to improved survival [that] allows for the detection of tumors that occur later in life".

Following the growing number of reports on the role of diet and obesity in test animals in affecting tumor incidences, the NTP evaluated the effects of various diets on growth, survival and chronic disease incidence in F344 rats fed a standard NIH-07 diet used in the NTP bioassay program since 1980 into the 1990s (Rao, 1994). This review led to the development of an altered diet formulation termed the NTP-2000 diet (Rao et al., 2001). Since 1994, the NTP-2000 diet has been used in all NTP bioassays, which has resulted in decreased growth rate of F344 rats, significant increases in survival percentages of both males and females, and significant decreases in severity of nephropathy and cardiomyopathy, particularly in males (Rao et al., 2001; Haseman et al., 2003). In the study by Rao et al. (2001), the percentage of survival in rats fed the NTP-2000 diet at 104 weeks was 60.2 ± 2.9 and 73.8 ± 2.1 in males and females, respectively, and, in both sexes, significantly ($p < 0.01$) higher than the survival percentages for F344 rats males and females fed the NIH-07 diet (42% and 58.5%, respectively). Similar to NTP, ERF conducts large numbers of carcinogenicity screening studies and employs *ad libitum* feeding. Lower-than-expected survival rates, however, have been noted in ERF studies. In its review of a recent ERF study (2006), EFSA found that "survival of rats in the study at 103 weeks of exposure was relatively low ranging from 22% (controls) to 31% [treated groups] in males and 27% (controls) to 45% [treated groups] in females". It is possible that the cause of the low survival rate in the ERF study is partially related to *ad libitum* feeding effects associated with the type of diet used in ERF studies.

4. Frequency and duration of dosing

OECD (1981, 2009a) and FDA (2000a,b) testing guidelines as well as NTP (2006) and ERF (Maltoni et al., 1999; Soffritti et al., 2002) protocols for carcinogenicity tests indicate that a test substance incorporated in feed or drinking water should be available *ad libitum* and supplied continuously for 7 days/week. Continuous availability of the test substance is intended to achieve the highest possible intake of the test substance in order to maximize the possibility to detect effects

The duration of dosing required by most regulatory guidelines and carcinogenicity study protocols includes a substantial portion of the animal lifespan. The NTP (2006) and FDA (2000b) carcinogenicity testing protocols specify an exposure duration of 24 months, in studies of rats or mice. OECD (1981, 2009a) guidelines consider longevity of the test animals and set goals of 18 month exposures for mice and 24 month exposures for rats. Termination of the study is typically concomitant with termination of dosing. OECD guidelines indicate that early termination of the study is acceptable when the incidence of deaths in the lower doses or the control group is equal to or greater than 25% (OECD, 1981, 2009a).

The ERF carcinogenicity testing program exposes animals, starting either *in utero* or more typically after weaning, for a period of 24 months, after which dosing is terminated and the surviving animals are allowed to live out their normal life until natural death

(Maltoni et al., 1999; Soffritti et al., 2002; Bucher, 2002). The 24 month treatment interval is described as representative of possible human exposure to a chemical or contaminant (Maltoni et al., 1999). It is not clear from published ERF studies why dosing in studies of foods or food ingredients is not continued until natural death, if the goal is to mirror human exposure. A dosing duration of 18–24 months in mouse and rat studies, however, typically obtains the goal of continuous daily exposure throughout all major phases of life, i.e., from the period of early growth and development through maturation and advanced adulthood. Termination of the study at termination of dosing maximizes the likelihood of detecting effects related to test material administration, since results are based on comparing treated animals to control animals. Extending study duration beyond test material administration is inconsistent with the goal of comparing treatment to non-treatment conditions.

5. Study duration

Most carcinogenicity studies are conducted in accordance with regulatory guidelines that specify termination of the study on the last day of dosing (e.g., EPA, 1998; FDA, 2000b; NTP, 2006). For studies in mice and rats, this is typically 18 months and 24 months after study initiation, respectively. In instances when animal survival percentages are significantly lower than expected prior to termination, early termination may be sometimes recommended to ensure an adequate number of survivors for reliable statistical sensitivity. Other proposals to shorten or lengthen the duration of animal carcinogenicity studies are discussed in the following sections on Limited Carcinogenicity Studies and Life-Span and Extended Dosing. FDA carcinogenicity guidelines (FDA, 2000b) indicate that test material exposure should be daily throughout the study. US EPA carcinogenicity guidelines (EPA, 1998) for study duration are essentially the same as the FDA guidelines (2006b). The duration of dosing in NTP rat studies is specified as 2 years. Bucher (2002) noted that this fixed interval has allowed pathologists at NTP to become adept at understanding and predicting the typical progression of rat lesions in a variety of organs during this exposure period. Bucher also indicates that the generally uniform practice of a 2-year study duration with test material administered in feed available daily considerably enhances the ability to compare different studies and to standardize the historical database collected in the NCI/NTP bioassay program.

In the EU, OECD carcinogenicity guidelines published in 1981 and revised and adopted in 2009 are generally consistent with the FDA guidelines for study duration. They specify a 24 month exposure period for rodents, "representing the majority of the normal life span of the animals", noting, however, that, for various strains of mice, a study "duration of 18 months may be more appropriate". The OECD (2009a) guidelines also note that early termination of a study before the target periods may be considered "when the number of survivors in the lower dose groups or the control group falls below 25%", with the additional guidance that "survival of each sex should be considered separately". The OECD guidelines (2009a) do not consider excessive deaths in only the high dose group as a sufficient reason to trigger early study termination, although the guidelines indicate that a "study should not be extended beyond the point when the data available from the study are no longer sufficient to enable a statistically valid evaluation". No guidance is provided that would be pertinent to analyses of tumor data from lifetime studies such as those conducted at ERF, which allows each animal to live until its "natural death". With this practice, irrespective of continuation or discontinuation of dosing after 24 months (see above section: Dosing Duration), effects with varying ages past termination would make between-group comparisons of tumorigenicity measures difficult to interpret (see also below section: Statistical Analyses).

6. Verification of test substance purity, stability, concentration/homogeneity in the diet

Both OECD (1981, 2009a) and FDA (2000b) regulatory testing guidelines for feeding studies require that, prior to initiation of dosing, the test substance should be chemically analyzed to confirm its identity and purity. Such analytical tests are essential, since impurities can have toxicological effects that can confound interpretation of test study results. Rodent diets containing the test substance should then be evaluated to determine whether the test substance is stable in the diet and that the dietary target concentration has been achieved. Stability tests ensure that the test substance is not altered in the formulated feed mixture and provide information on acceptable time intervals for storage before new formulated diets should be prepared. Routine procedures for preparation of dosed feed incorporate analyses to determine that the test substance is homogenously distributed in each prepared batch of dosed feed to avoid dosing animals with unintended low or high concentrations of the test material. Control diet must also be analyzed to confirm the absence of test material. This analysis is critical, in order to exclude inadvertent administration of the test substance to control animals. This is particularly important, as unintended cross-contamination can sometimes occur during preparation of feed batches or with feed handling practices in the animal room.

Guidelines for NTP feeding studies (NTP, 2006) specify that the test substance and any new lots should be analyzed for purity within 30 days prior to use in the test and reanalyzed at intervals during the study (24 ± 2 weeks). Jameson and Goehl (1994) suggest that the identity of the test substance should be analyzed with a variety of validated analytical methods and that, prior to initiation of dosing, any impurities >1% should be identified, quantified and evaluated for potential effect on study conduct or interpretation of study results. The homogeneity of formulated diet has been found to be improved by mixing the test substance with a small amount of the powdered diet and gradually adding this to the remainder of the diet (Jameson and Goehl, 1994). NTP recommends that test material stability in the diet be tested prior to the preparation of actual test diets. Stability tests are to be conducted weekly, for six weeks, and prior to study initiation, on batches of diet containing test material at the lowest concentration expected to be used in the study. Weekly tests are to be done both on samples of diet that have been stored frozen (−20 and 5 °C) and additional samples that have been stored at elevated room temperature (25 °C). The results are used to determine the acceptable time period for storage of dosed feed and the intervals when new batches must be made, and old batches discarded. These data are essential to have prior to the initiation of the study, to avoid unexpected study dosing errors. Similar to OECD and FDA guidelines, NTP guidelines recommend verification of diet homogeneity. Uniformity of dose distribution in the diet (homogeneity) is to be determined at the highest and lowest dietary concentrations, based on analyses of duplicate samples collected from a minimum of three different points in the feed blender (e.g., top, middle and bottom). Homogeneity tests are repeated whenever a new lot of test material is used. As a precautionary measure, a reserve sample of all batches of prepared diet is to be collected and stored frozen. Test material stability in formulated diets is also to be verified at periodic intervals (10 ± 2 weeks), using an analytical method validated over the range of concentrations to be used in the diet. The diet formulation procedure is acceptable under NTP guidelines if the coefficient of variation for the test material concentration in collected samples is ±5% of the target value. If the variability of

the analyzed sample is outside of the acceptable range, then the reserve samples are analyzed to determine the range of variability of the substance concentration in the diet.

Dietary carcinogenicity studies published by ERF indicate that the relevant test material is administered in a type of pelleted feed routinely used at ERF (“Corticella diet”) (Belpoggi et al., 2006b; Soffritti et al., 2006). These studies do not, however, always reveal the composition of the diet used or the methods used to confirm test material purity, stability and homogeneity in the diet (see, for example, EFSA, 2006). Formulation of diets with pelleted animal feed presents a particular challenge in attaining dietary homogeneity. Feed pellets are coated with a liquid solution or suspension of the test material, typically using a revolving mixer. Uniform distribution of test material to all pellets can be impacted by numerous factors, including uniformity of pellet separation, total exposure time of pellets to the test material solution or suspension (which may differ at different levels of the mixer, due to gravity effects), uniformity of pellet surface features (porosity, surface tension), etc. There can also be solvent effects on test material stability. The ERF studies cited above did not provide information on how the dosed feed was prepared, whether solvents or carriers were used, or how test material homogeneity or concentration in the diet was assessed. EFSA found, in a review of related unpublished information provided by ERF, a declaration that “at the start of the experiment, the various concentrations and the stability of aspartame in feed was evaluated”, however, the report did not clarify how these evaluations were conducted EFSA (2006). This information is critical to assure that planned dietary dose levels are achieved, test material in diet is homogeneously distributed, test material in diets is stable over the time of use, and that potential contaminants in feed, the test substance or carriers are either absent or well-characterized and below levels that would confound test results. Without this information, it is unclear whether results should be accepted as reliable or comparable to results of other studies investigating the same test material.

7. In-life observations and clinical examinations

The extent of animal observations included in NTP and ERF protocols and required by the various US and EU regulatory agencies varies in the amount of detail. NTP protocols require that regular, detailed clinical observations be performed and recorded in the conduct of any carcinogenicity study (NTP, 2006). It is unclear whether ERF includes in its carcinogenicity study protocols a requirement for regular clinical observations, as copies of the detailed ERF carcinogenicity testing protocols and/or standard operating procedures have apparently not been published. Published ERF studies provide insufficient information to determine if clinical observations are a part of the study procedures. Regular clinical observations help in the evaluation of overall test material effects, including possible tissue/organ toxicity and the evaluation of oncological progression.

OECD Guideline 451 (OECD, 1981, 2009a) require clinical observations in the conduct of carcinogenicity tests, with the following general recommendations: clinical examinations of the test animals should be conducted at least once daily and records kept on animal appearance, onset, location, dimensions, appearance, and progression of any visible tumors. Body weights should be recorded individually for all animals at least once weekly, for the first 13 weeks of the study, and at least once every four weeks, thereafter. Feed intake should be determined once during the first 13 weeks, and, thereafter, at approximately 13 week intervals. If deterioration of animal health is noticed during the study, a differential blood count should be performed. Also, differential blood counts are required at 12 and 18 months for all animals in the highest dose and control groups. If questionable results are obtained, or there are pathological signs, then a differential blood count is required for the animals in the low and intermediate groups as well.

FDA (CFSAN) provides relatively detailed guidance on requirements for animal observations and handling during the conduct of long term studies as part of the updated sections of the FDA Redbook (FDA, 2000a,b). The guidelines state that complete records for body weights, feed and water consumption are noted to be important for assessing time-related occurrences of changes that may be related to toxic effects of the test substance. Individual animals should be weighed weekly for the first 13 weeks and monthly thereafter for the duration of the study, similar to OECD (1981, 2009a) guidelines. In contrast to OECD, FDA CSFAN (FDA, 2000b) indicates that animals should be weighed twice per week, when there are (significant) issues with feed palatability and/or marked changes in body weights and/or deaths, and that this enhanced schedule may need to be extended throughout the study. Feed intake is advised to be measured at the same intervals as body weights. Measurement of water consumption is required only in cases where the test substance is given in the drinking water.

Clinical tests in FDA (CFSAN) protocols are relatively extensive (FDA, 2000b), compared to OECD guidelines discussed previously, possibly because the FDA Redbook guidelines are applied to animal tests on food components that have a potential for extensive human exposure. FDA requires ophthalmologic examinations prior to, and at termination of, the study on all animals in the control and high dose groups and on all animals in other groups if ocular changes are seen as possibly related to the test substance. Hematology tests should be performed on at least 10 animals/sex in each group during the first 2 weeks of the study, at 3, 6, and 12 months, and at 18 months if trends in any of the hematological parameters are observed. The parameters to be evaluated include hematocrit, hemoglobin concentration, erythrocyte count, total and differential leukocyte counts, mean corpuscular hemoglobin, mean corpuscular volume, mean corpuscular hemoglobin concentration, platelet count, and a measure of clotting potential (e.g., clotting time, prothrombin time, activated partial thromboplastin time). When significant adverse effects are noted during the study, reticulocyte counts and bone marrow cytology tests should be considered for inclusion at study termination. Clinical chemistry requirements for FDA (CFSAN) studies are also more extensive than in the OECD test specifications (OECD, 1981, 2009a). Clinical chemistry tests are performed at the same intervals and circumstances as for the hematology tests. Table 4 provides a summary of the parameters required by the FDA (CFSAN) guideline (FDA, 2000b).

Hematology and clinical chemistry tests using the same animals at each sampling point is desirable, but not required, according to FDA (CFSAN) guidance. Urinalyses should also be conducted with the same animals employed for the hematology and clinical chemistry determinations, and urine is typically collected by housing the animals in metabolism cages for a period before or after blood sampling. Determination should include urine specific gravity, pH, glucose, and protein, as well as microscopic analysis of urine for sediment and presence of blood and/or blood cells.

NTP (2006) carcinogenicity protocols do not require the same diversity of in-life tests as for studies for submission to the FDA. NTP specifications require determination of feed consumption for a one-week (7-day) period, every 4 weeks during the study. Clinical examination for signs of toxicity “should be performed two times daily (once in the early morning and once in the late afternoon at least 6 h apart, before 10:00 AM and after 2:00 PM, including holidays and weekends), particularly for a check on morbidity/mortality. Any animal whose condition makes it unlikely that it will survive until the next observation, based on the view of the assigned Laboratory Animal Veterinarian or Study Director, shall be terminated immediately, and necropsied. NTP protocols also

require, at four week intervals, that each animal be subjected to a formal, detailed clinical examination, to supplement less extensive daily observations. Hematology, clinical chemistry and urinalysis evaluations are performed as part of subchronic testing at NTP, but are not required for carcinogenicity tests.

At the ERF facility, carcinogenicity study parameters more closely resemble those of NTP protocols than FDA tests that incorporate hematology, clinical chemistry and urinalysis testing as described previously. However, NTP derives clinical chemistry and toxicity data from results of preliminary 90 day studies while ERF appears to rely primarily on published toxicity information from studies that are not conducted using the same life span protocols or rat strain employed at ERF and that may not contain information on clinical changes produced by agents chosen for testing at ERF. Information from recent ERF studies (Soffritti et al., 2006; Belpoggi et al., 2006b; Soffritti et al., 2007) indicates that both drinking water and feed consumption were measured for each group once a week for the first 13 weeks and every 2 weeks for approximately 2 years. This procedure is more detailed than those described for OECD, NTP or FDA (CFSAN) protocols. Individual body weights are measured once/week for 13 weeks and at 8 week intervals thereafter. ERF publications indicate that cage examinations for clinical and behavioral observations are performed 3 times daily during the week and twice on weekends and holidays. ERF publications indicate no disposition of moribund animals, rather, all animals are allowed to die naturally. Necropsy is begun only after an animal is found dead during a cage check, and then placed in storage until such time as the necropsy can be performed. This practice could allow for a high prevalence of tissue/organ autolysis. Neither NTP nor ERF incorporate hematology, clinical chemistry or urinalysis tests in routine long term studies. However, NTP incorporates these parameters in preliminary 90-day dose finding studies that serve both to determine subchronic toxicity and to assist in dose selection for the 2-year oncogenicity study. It is not evident from publications if ERF conducts dose range finding tests that include hematology, clinical chemistry and urinalysis evaluations. The absence of information on clinical chemistry and hematology data from ERF studies prevents utility of this information in tumor diagnosis. This could affect both the reliability of comparisons of tumor incidence data from ERF studies to tumor incidence data from non-ERF studies and the subsequent assessment of potential carcinogenicity risk to humans.

As noted in a previous section, OECD (1981) carcinogenicity protocols require that blood smears be taken at 12 and 18 months and that differential blood counts be recorded. ERF publications indicate that these procedures are not part of the ERF protocols for carcinogenicity testing. EFSA cited this as a deficiency in compliance with EU approved testing procedures, in its review of recent ERF studies of aspartame (EFSA, 2006). It is not clear from ERF publications why such practices are apparently not undertaken, given the value of the data in evaluating animal health status and factors in tumorigenicity.

8. Good Laboratory Practice Standards (GLPs)

GLPs are a quality system for controlling the manner in which nonclinical safety studies are planned, performed, monitored, recorded, verified/audited, reported and archived. The principal value of conducting studies in compliance with GLPs is to allow independent assessment and reconstruction of study events and conclusions using study records. GLP compliance requires that study records be regularly reviewed for compliance to laboratory SOPs and study protocols, by the laboratory's designated, qualified Quality Assurance Unit. Deviations are to be documented and discussed in final reports, with regard to their potential impact on study outcomes. In both the US and the EU, carcinogenicity studies must be conducted in compliance with GLP standards to be accepted for regulatory purposes. Compliance with GLP regulations is also required for acceptance of many other types of safety related studies and/or certain aspects of the safety-related investigations (e.g., hematologic, clinical and analytical chemistry analyses, etc.). The requirements apply to investigations of food of components of pharmaceutical products, cosmetic products, pesticides, veterinary drugs, medical devices and food additives submitted for regulatory approval. In the US, GLP compliance programs and provisions are similar in FDA and EPA programs (EPA, 1983; EPA-FIFRA, 1984; FDA, 1978). Compliance with GLP standards for a submitted study must be stated in reports as a legally enforceable commitment and GLP compliance is usually required for acceptance of non-clinical study reports for consideration in support of product safety assessments.

In the EU, the European Commission (EC) (EC, 2009) has adopted revised principles of GLPs applicable to preclinical studies that support product approvals in two revised Directives: Directive 2004/10/EC and Directive 2004/9/EC. Directive 2004/10/EC relates to the application of GLPs and verification of GLPs for tests on chemical substances. Directive 2004/9/EC governs the harmonization of various provisions of GLP compliance and enforcement in the EU. This is to require compliance with the OECD Revised Guides for Compliance Monitoring Procedures for GLP and the OECD Guidance for the Conduct of Test Facility Inspections and Study Audits during laboratory inspections and study audits (OECD, 1998).

At a minimum, establishment of a GLP compliance program is generally accepted as requiring the following (Ertz and Preu, 2008):

1. a management-appointed Study Director, responsible for study conduct and reporting, for every study conducted;
2. establishment of a Quality Assurance Unit (QAU) and QAU officer who report directly to management and independently determine the compliance of facilities, personnel, practices and record keeping with regulations. The QAU also maintains a master schedule sheet of all studies conducted in the test facility, inspects each non-clinical study at intervals to assure compliance, reports findings to the Study Director and management, reviews the final report to assure that it accurately reflects the raw data, and prepares and signs a QA statement included in the final report;
3. standard Operating Procedures (SOPs), including ones for equipment use and maintenance calibration; laboratory tests and methods; animal use practices, such as identification, care, handling and transfer; necropsy procedures; histopathologic evaluations; test material and feed handling and formulation; and data handling and storage;
4. a written protocol for every study conducted, each prepared prior to study initiation and clearly describing the objectives and methods for the conduct of the study;
5. clear data records, with both original entries and corrections recorded in ink, dated and initialed (or otherwise treated so that the data is preserved intact and the date of each entry and identity of the relevant recorder obvious);
6. separate laboratory and animal facilities; and
7. a final report, for every study conducted, which contains a GLP compliance statement that is signed and dated by the QA officer, indicating that the study has been inspected and that the report accurately reports data as collected, and is reviewed for general accuracy in its entirety, which is to be indicated by a dated signature by the Study Director.

In addition to these basic requirements for GLP compliance, GLPs have compliance provisions for facilities, laboratory equip-

Table 4
Clinical chemistry and histopathology parameters included in typical US and EU carcinogenicity study designs.

Clinical chemistry parameters
Hepatocellular evaluation (at least 3 of the following 5)
Alanine aminotransferase (SGPT, ALT)
Aspartate aminotransferase (SGOT, AST)
Sorbitol dehydrogenase
Glutamate dehydrogenase
Total bile acids
Hepatobiliary evaluation (at least 3 of the following 5)
Alkaline phosphatase
Bilirubin (total)
Gamma-glutamyl transpeptidase (GG transferase)
5′ nucleotidase
Total bile acids
Other clinical markers of cell changes or cellular function
Albumin
Calcium
Chloride
Cholesterol (total)
Cholinesterase
Globulin (calculated)
Glucose (in fasted animals)
Phosphorous
Potassium
Protein (total)
Sodium
Triglycerides (fasting)
Urea nitrogen
Tissue/organ histological evaluations
Adrenals
Aorta
Bone (femur)
Bone marrow (sternum)
Brain (at least 3 levels)
Cecum
Colon
Corpus and cervix uteri
Duodenum
Epididymides
Esophagus
Eyes
Gall bladder (if present)
Harderian gland
Heart
Ileum
Jejunum
Kidneys
Liver
Lung (with main-stem bronchi)
Lymph nodes (1 mandibular and 1 mesenteric)
Mammary glands
Nasal turbinates
Ovaries and fallopian tubes
Pancreas
Pituitary
Prostate
Rectum
Salivary gland
Sciatic nerve
Seminal vesicle (if present)
Skeletal muscle
Skin
Spinal cord (3 locations: cervical, mid-thoracic, and lumbar)
Spleen
Stomach
Testes
Thymus (if present)
Thyroid/parathyroid
Trachea
Urinary bladder
Vagina
Zymbal's gland
All tissues showing abnormality

ment, handling procedures for test and control articles, and requirements for retention of records and reports.

For NTP carcinogenicity studies, contract testing laboratories are required to conduct these and supporting tests in compliance, minimally, with FDA GLP regulations (FDA, 1978). In some instances, NTP may also require adherence to relevant EPA guidelines (EPA, 1983: EPA-FIFRA, 1984). NTP ensures that laboratories are in compliance with its requirements by conducting on-site Quality Assurance (QA) monitoring audits, prior to, and during, study conduct. NTP staff also independently audit reports filed by the contract laboratory's QAU. This would include, minimally, the QAU audits of (a) prestart study activities, (b) the quality and appropriateness of the study protocol and any subsequent amendments, (c) critical study events (some of which require auditing as the event is in progress and/or sufficiently before initiation of a related key next step) , e.g., randomization, identification, study start, necropsy, slide preparation, etc., (d) procedures of other key practices and procedures that have an impact on either study data collection or outcome, e.g., dose formulation calculations and preparation, the process for weighing animals and determining feed consumption, etc., and (e) representative samples of all data generated in support of the major study components.

GLP compliance procedures and programs for carcinogenicity screening tests at ERF are difficult to appreciate from the details available primarily in publications. Maltoni et al. (1999) notes that among the general prerequisites for "protecting this branch of research [carcinogenicity bioassays] from the amateur or anecdotal approach" ... "Following the rules of Good Laboratory Practice [is used at ERF] as a minimum standard in experiment management". However, there is no indication or any details on the program that is used for ERF studies. Soffritti et al. (2002) states that "All studies are performed on Sprague–Dawley rats conducted according to Good Laboratory Practices", but no details of compliance procedures or citation of OECD regulations for GLP are cited. Finally, no citation of GLP compliance has appeared in any of the recent carcinogenicity study publications from ERF (Soffritti et al., 1999, 2002, 2006, 2007; Belpoggi et al., 2006b). In a review of one ERF study (see both Soffritti et al., 2006; Belpoggi et al., 2006b), the European Food Safety Authority (EFSA, 2006) had access to unpublished data in a report submitted by Soffritti and Belpoggi (2005)that cited a statement that the study ("was conducted in accordance with GLP"). However, the EFSA review noted that neither the study (used as the basis for the publications cited above) nor the ERF laboratory has received an inspection by the Italian GLP compliance monitoring authority and the claim of GLP compliance could not be confirmed.

9. Histopathology evaluations

Comparison of histopathological findings from evaluation of tissues from control and treated animals is a critical component of carcinogenicity tests and these findings, in conjunction with appropriate statistical analyses, provide the basis for concluding whether a chemical possesses significant carcinogenic potential. Requirements for numbers of tissues/organs that should be subjected to histopathological examinations are specified by EPA, FDA (CFSAN) (FDA, 2000b) and OECD regulatory guidelines. Both NTP and ERF carcinogenicity screening studies generally conform to the minimum requirements of these respective guidelines, with minor additions, modifications or multiple sections/tissue. The FDA (CFSAN) tissue list shown below (FDA, 2000b) indicates the major tissues to be evaluated in carcinogenicity evaluations and is representative of minimum requirements of similar lists provided by

other regulatory guidelines for carcinogenicity evaluations as summarized in Table 4.

In addition to specific tissues and organs listed in Table 4, FDA CFSAN (FDA, 2000b) guidance indicates:

> "*All gross lesions should be examined microscopically. All tissues from the animals in the control and high dose groups should be examined. If treatment related effects are noted in certain tissues, then those specific tissues in the next lower dose level tested should be examined. Successive examination of the next lower dose level continues until no effects are noted. In addition, all tissues from animals that died prematurely or were killed during the study should be examined microscopically.*"

All of the regulatory agency carcinogenicity guidelines, as well as NTP and ERF carcinogen screening studies, incorporate additional specifications to conduct histopathological examination of all tissues or organs that show pathologic lesions.

Because of the subjective nature of histopathological diagnoses, both independent experts and international pathology advisory groups (Williams et al., 2008; Crissman et al., 2004; FDA, 2000d) have stressed the importance of appropriate qualifications and certification of pathologists, the use of standardized pathology nomenclature, consistent recording procedures and peer review of histopathological evaluations. In the US, the American College of Veterinary Pathology (ACVP) has stringent qualifications for study pathologists who must demonstrate minimal requirements of training and experience for board certification (ACVP, 2009). Similar requirements for training and experience have been implemented for accreditation of toxicologic pathologists in Europe and Asia in harmonization with programs in North America (Ettlin et al., 2007; ESTP, 2007). The International Academy of Toxicologic Pathologists (www.iatpfellows.org) provides global accreditation.

Histopathological evaluations conducted at the NTP and for submission to US regulatory authorities rely on the Standardized System of Nomenclature promulgated by the Society of Toxicologic Pathologists (STP) (STP, 2009) and cited previously in Crissman et al. (2004). The histopathology nomenclature system used by ERF is not cited in ERF publications. Comparison of ERF and non-ERF histopathologic findings may be complicated by differences in the nomenclature used. Animal histopathology evaluations for studies submitted to regulatory agencies in Europe normally employ terminology specified by WHO/IARC (World Health Organization/International Agency for Research on Cancer) (cited in ESTP, 2005; Mohr, 1992-1997, 2001; IARC, 1992–1997). Differences in histopathological nomenclature between North America and Europe/Asia have been recognized, and a revision of the Standardized System of Nomenclature and Diagnostic Criteria used for studies (Crissman et al., 2004) with rats and mice in North America has been initiated by the Society of Toxicologic Pathologists (STP, 2009). This process is being coordinated internationally with comments provided by the European STP (ESTP, 2005).

As discussed earlier, consideration of relevant in-life data, e.g., clinical signs, hematology, clinical chemistry and urinalysis, organ weight changes, macroscopic changes and non-neoplastic findings observed at necropsy are critical parameters that enhance the ability to understand the relevance of microscopic findings (Crissman et al., 2004). FDA, EPA and NTP guidelines all recommend this consideration. The OECD testing guidelines are less specific about the integration of toxicological and histological findings, although the guidelines indicate that information from clinical examination "should be available before microscopic evaluation, since it may give significant guidance to the pathologist".

ERF publications have typically reported tumor incidences without comment on necropsy findings, organ-specific toxicity or non-neoplastic findings (Soffritti et al., 2005, 2006, 2007; Belpoggi et al., 2006b). EFSA (2006) noted that ERF reporting of study results differ from OECD Guideline 451 (1981) recommendations in several ways including lack of information on clinical and macroscopic changes, lack of blood smears and restriction of study reports to tabulation of incidences of tumors and inflammatory events. In an EFSA audit of recent ERF studies (EFSA, 2006), this practice was found to significantly affect the ability to evaluate carcinogenicity test results. For example, EFSA noted that non-neoplastic findings were significant and relevant to the interpretation of reported increases in lymphoimmunoblastic lymphomas and histiocytic sarcomas, particularly in the lung. Supplementary information not reported in publications and later supplied by ERF to EFSA showed high rates of infection in the study animals, including pleuritis (22–71% in males and 47–94% in females) and bronchopneumonia (81–95% in males and 69–79% in females). EFSA reviewers commented that, for the lung tumors reported in the ERF study, " it is well established that this pulmonary type of lymphoreticular tumor may occur as a consequence of severe chronic respiratory disease". This example illustrates how presenting incidence of tumors without information on non-neoplastic effects can result in a misleading picture on causation of tumors.

Independent pathology peer review is also critical for ensuring reliability of histopathology evaluations. In unpublished pathology NTP reviews of an ERF study on aspartame (Soffritti et al., 2005, 2006), an NTP Pathology Working Group (PWG) reported that "a number of hyperplastic and neoplastic lesions were more severely classified by the study pathologists from ERF [when] compared with the diagnoses of the NTP review group" (Hailey, 2004). For example, with a set of three renal pelvis carcinomas reported by ERF, only 1 was confirmed by the PWG. The other two (66%) were diagnosed as hyperplasia. Based on its review, the PWG found that the different ERF evaluations, i.e., hyper-classification, "could have a significant impact on the outcome and interpretation of the study". The PWG review showed similar cases of "overdiagnosis" of "adenocarcinomas of the mammary glands considered as fibroadenomas by the PWG, cases of early squamous cell carcinomas of the ear duct or oral cavity considered as hyperplasias, a case of adenocarcinoma of the pituitary gland as cystic change, and cases of early transitional cell carcinomas of the renal pelvis as hyperplasia". Because of the examples of misdiagnoses noted by the NTP pathology group, EFSA concluded that determination of the significance of carcinogenicity findings by "aggregation of all malignant tumour incidences or all malignant tumour-bearing animals for statistical purposes is not justified" (EFSA, 2006). Similarly, ERF reported that kidney calcification was an important part of hyperplastic and neoplastic renal pelvic lesions in a study of aspartame, however, in an EFSA review (2006) of unreported information supplied by ERF, EFSA found no dose-related trend in calcification. In contrast, EFSA found that there was sufficient evidence "to undermine the hypothesis published by Soffritti et al. (2006) that kidney calcification was an important determinant of hyperplastic and neoplastic renal pelvic lesions". In addition, the EFSA reviewers concluded that the "irritant induced" alterations in the renal pelvic tissues, suggested to be related to neoplasia in the ERF studies, were specific to rats and "the effects are of no relevance for humans". These EFSA findings again point out the importance of reporting all non-neoplastic findings and other findings and learnings in interpreting carcinogenicity testing results.

The importance of independent pathology peer reviews for ensuring reliability of histopathology evaluations is illustrated in the EFSA (2006) reviews and unpublished NTP PWG audit information cited therein as noted above. A system for peer review has been incorporated into the STP "Best Practices Guideline for Toxicologic Histopathology (Crissman et al., 2004) and STP positions on the purposes and documentation of peer review have been pre-

sented previously (Black, 1991; STP, 1997). For drugs, the ICH (S1B) testing guideline requires independent review of 10% of slides. At NTP, pathology reviews are performed in concordance with STP Best Practices by a multi-step process involving an initial review by a NTP pathologist, followed by an independent review by a non-NTP pathologist and a final blinded review by the Pathology Working Group (PWG). The process of "blinded" or "masked" observations without knowledge of the treatment group is employed to reduce potential bias. However, as noted previously, correlation with clinical observations, clinical pathology and macroscopic findings at necropsy and information about the reactivity of the test compound, which can elucidate potential mechanisms of action, all could enhance interpretation of histological findings. Thus, at NTP and in regulatory guidelines specified by OECD, EPA and FDA, blinded histological observations are not required for initial pathology interpretations but may be employed in final PWG reviews for difficult assessments. Iatropoulos (1988) reported the position of the Society of Toxicologic Pathologists that "blinded" histological evaluation of slides (i.e., without information on the animal or treatment) "is not appropriate as a routine procedure". However, "blinded re-examination of selected target organs .... is accepted, recommended and routinely practiced by pathologists in industry to clarify specific diagnostic problems".

The NTP uses a third party review procedure of histological preparations that employs a blinded slide examination procedure at this final review stage. ERF does not indicate in its publications if slide evaluations are "open" or blinded at the stage of initial or secondary internal review. ERF has indicated in publications that pathology reviews are performed in-house by internal ERF pathologists that provide a second opinion of pathology findings. For recent studies (Soffritti et al., 2005, 2006), an NTP PWG performed independent evaluations of ERF data, however, details of these audits are only partially available, limiting conclusions that can be drawn with relationship to previous ERF publications not subjected to independent review.

10. Statistical analyses

Following the termination of the dosing and in-life phase of a chronic carcinogenicity test, incidence of tumors in animal tissues and organs are determined and severity scores or relevance to the cause of death may also be recorded in some protocols. Appropriate statistical tests are then used to evaluate the significance of differences seen in treated groups of animals relative to the untreated control groups to determine carcinogenic potential of the test chemical. In most current studies that employ multiple dose levels, evaluation of trends in the tumor incidence relative to dose are desirable for subsequent use in conduct of risk assessments for extrapolation of the data to humans. A series of papers (Waddell, 2002, 2003a,b,c,d, 2004a,b, 2005, 2006, 2008, 2010 and Rozman et al., 1996 and Rozman, 2003) has suggested that both non-reactive and DNA-reactive carcinogens have thresholds, departing from the generally accepted concept of a linear, no threshold approach used in cancer risk assessment. The notion that no threshold exists for carcinogens results from plotting dose on a linear scale, making it difficult, if not impossible, to put doses received by humans into perspective.

The extent and clarity of regulatory guidance for conduct of appropriate statistical tests is quite variable. IARC (1980, 2006) provides explicit and detailed guidance for statistical analysis of chronic carcinogenicity data. FDA CFSAN (2000f) provides guidance for specific statistical issues that pertain to design of the study, presentation and interpretation of analytical data and the support available from CFSAN statistical reviewers to address particular questions. Although the FDA (CFSAN) does not specify specific statistical tests for evaluating carcinogenicity of substances in food, guidance is provided for use and reporting of parametric versus non-parametric tests, analysis of survival parameters, and use of Life Table and trend tests (FDA, 2000f). For evaluating candidate pharmaceutical products, FDA (2001) does recommend specific tests, favoring the Peto test when reliable cause of death data are available and the poly-3 test otherwise. In addition, Lin (2000) provides additional guidance on evaluation of carcinogenicity data and various procedures considered by FDA (CFSAN) to address complexities in tumor and survival data.

For reporting of carcinogenicity study results, OECD (1981) provides minimal guidance by requiring only a "summary... and analysis of the data" and a "statement of the conclusions drawn from the analysis". OECD (2009a) revised guidelines are similarly non-specific concerning statistical analyses and "summary tables should provide the means and standard deviations" (for continuous test (data) of animal showing toxic effects of lesions), is the only guidance provided. No specific statistical procedures are suggested. Similarly, EPA (1998) is also non-specific about statistical evaluations of the study data and specifies only that "all observed results (quantitative and qualitative) should be evaluated by an appropriate statistical method. Any generally accepted statistical methods may be used". However, EPA does require that "the statistical methods including significance criteria should be selected during the design of the study" and would, therefore, have to be part of the study protocol required for conduct of studies in compliance with Good Laboratory Practice Standards. FDA (CFSAN) (FDA, 2000e), in contrast, has formulated a guidance document for reporting of toxicity studies that provides a generalized structure for data presentation and analysis of carcinogenicity studies. FDA (CFSAN) (FDA, 2000f) specifies the format for data tabulation, requirements for referencing statistical tests used, null and alternative hypotheses, values and degrees of freedom of the test statistic (when appropriate, *p*-values, specification for one vs. two tailed tests and tabulation of the results of all statistical analyses).

Development of appropriate statistical tests for tumor incidence data has been an active topic for decades because of the importance and ramifications of conclusions of carcinogenicity tests of food, drug, industrial and consumer chemicals. Procedures to guide selection and application of specific statistical tests for carcinogenicity studies developed by numerous statisticians have been summarized in excellent reviews by Gad and Weil (1986) and Robens et al. (1994) (see in addition, Gart et al., 1986; Fairweather et al., 1998; and Chow and Liu, 1998). Major points to be considered in statistical evaluations include possible survival differences among groups, extent of "type 1" statistical errors (incorrectly concluding a result represents a significant effect) and the need for evaluations of trends in addition to simple pairwise comparisons.

Evaluating significance of the total incidence of a specific lesion in treated groups versus controls can be performed by simple pairwise tests including chi-square or Fisher's exact test. However, such comparisons are justified only when the test agent does not shorten life span (Robens et al., 1994). The latter can result in underestimating a possible significant effect when there are substantial numbers of animals that do not survive to study termination (i.e., early deaths) (Gad and Weil, 1986). Similarly, when data from multiple dose groups are available, as in most contemporary test protocols, a test for a linear trend in tumor incidence has substantial statistical power and can be determined using the Cochran-Armitage test. But this procedure, like the pairwise tests mentioned previously, also does not adjust for possible survival differences between the groups of animals, and the power of such statistical comparisons of tumor incidences is diminished if there are survival differences among groups (Robens et al., 1994). Gad and Weil (1986) point out that when there are marked differences in survival between groups there is a need to adjust for survival to

avoid temporal bias. They also note that age (actuarial) adjustment can increase the power to detect group differences even when there are no differences in survival. This suggestion is consistent with similar recommendations of Peto et al. (1980) that tumor rates should be adjusted for survival whether or not differences among groups are apparent. Melnick et al. (2008) also noted that survival adjustment of tumor rates is important because animals that died early from causes other than tumors at the site of interest would detract from the complete picture of risk in that dose group and could result in unreliable estimates of cancer risk and the true site-specific effect of the test agent.

Bucher (2002) described differences in statistical analyses of carcinogenicity studies conducted by the NTP and by ERF. The NTP analyzes data from their bioassay screening studies using a survival adjusted, quantal-response procedure, the poly *K* test [a survival adjusted modification of the Cochran-Armitage test by Bailer and Portier (1988)] that assigns a weighted risk to each animal in the study based on its survival time and an assumed shape of lesion prevalence distribution (Bucher, 2002). ERF relies on prevalence analysis for non-lethal tumors and the log rank test of Mantel and Cox for all other tumors (Bucher, 2002). A major difference between statistical tests at NTP and ERF is that significant effects at ERF are considered to include differences in *total* benign and malignant tumor-bearing animals as well as *total* numbers of tumors/group. NTP, in contrast, restricts statistical comparisons and interpretations of significant differences to tissue-specific tumor responses. Although NTP reports list total numbers of tumors, this parameter is not used for statistical purposes as a measure of carcinogenic potential as at ERF (Huff, 2002). Statistical procedures at ERF have not always adjusted tumor rates for survival. Recent publications indicate that analyses were conducted using either the chi-squared test for pairwise comparisons of tumor bearing animals or total tumors/100 animals (Belpoggi et al., 2006b) or an analysis for trend using the Cox regression model (Soffritti et al., 2007). In one recent publication (Soffritti et al., 2006), the poly *K* ($K = 3$) test that adjusts tumor incidences for survival was employed, but analyses consisted of total tumors or tumor bearing animals rather than site specific tumors as evaluated in NTP applications of this statistical procedure. Robens et al. (1994) noted that "the use of data comprising the total number of tumors [such as at ERF], as opposed to analyses with animals with a tumor of specific morphological site or tissue of origin, has many theoretical difficulties". They concluded that using total tumors for analyses can bias the analysis by the weighting provided from a high spontaneous tumor rate at a specific site and by the multiple contribution of a single animal to the total tumor count, while the "true carcinogenic potential of a compound is that proportion of the animals that incur tumors over and above the true spontaneous rate". Gad and Weil (1986) noted that one of several methods to be avoided in analyses of carcinogenicity was "evaluation of number of tumors of all sites as opposed to the number of animals with tumors for specific sites of specific organs". McConnell et al. (1986) described rigorous conditions that could justify combination of tumors including substantial evidence for progression of benign to malignant neoplasia, that most neoplasms are of the same histomorphogenic type even if in different sites or neoplasms are classified with different morphology but have comparable morphogenic origins. Hyperplasia may be used as supporting evidence in their opinion when there are equivocal or arbitrary differences between hyperplasia and benign neoplasia.

Interpretation of the results of statistical analyses of carcinogenicity tests is fraught with risks of both false positive and false negative determinations because there are numerous sources of variability in study conduct that can affect tumor rates (Haseman et al., 1989). Lin (2000) notes that false negatives (failure to identify a true carcinogen) can result from the relatively low number of animals used in studies or from low tumor incidence rates. The issue with incidence rates is confounded by the observations of increasing numbers of spontaneous tumors in test animals as they age. False positive conclusions can result from the cumulative effect of the large number of comparisons involved (species tested, two sexes, 3–4 doses and more than 30 tissues/animal typically evaluated) such that the potential is great for finding statistically significant trends or comparisons to control levels by chance alone (Lin, 2000). In interpretation of statistical results from pairwise comparisons, Lin (2000) notes that FDA (CDEF) had used a rule in early studies that set significance levels at a $p \leqslant 0.05$ for rare tumors and at a 0.01 level for common tumors; where rare tumors were defined by an incidence rate of $\leqslant$1.0%. Use of this significance rule for pairwise comparisons of treated to control values resulted in false positive rates between 7% to 11%. If these same significance levels were applied to results from trend analyses, analyses showed that the false positive rate was roughly double. Based on this finding, FDA (CDER) sets levels of significant trends at a level of 0.025 for rare tumors and 0.005 for common tumors. With these new levels of statistical significance for consideration of a result as positive, the overall false positive rate was reduced to approximately 10%.

Although statistical evaluation of tumor data from carcinogenicity studies serves a valuable purpose for determining quantitative significance of a treatment-related effect, the application of pairwise comparisons to the multiple tissues evaluated can result in a false assessment of at least some of the comparisons as significant as noted previously. Gad and Weil (1986) caution that additional issues must be addressed, including questions such as:

1. Are the study data adequate without interference from low survival in test or control groups, extremely high or low tumor incidence levels, excessive loss of tissues from autolysis or possible effects from infection of animals during the study?
2. If there is a significant increase in tumors in the treated groups vs. controls, is there a trend in the response that corresponds to known biological or toxicity effects from the test agent in the same tissue(s)?
3. Is there supporting evidence for carcinogenicity such as positive genotoxicity studies or evidence of hyperplasia associated with tumor progression?

Robens et al. (1994) also caution on over-reliance on a single statistical rule without corroboration from factors listed above and supporting evidence by related, non neoplastic effects, similar tumorigenic evidence across species or between sexes, similar target organ effects and evaluation of historical control data as a point of reference.

Comparison of tumor data to historical control data within a laboratory where animal care, handling, common source of animals and environmental conditions are routinely conducted on standard operating procedures for different studies can provide a valuable tool to interpret biological significance of carcinogenicity data. Although the most important control group for statistical comparisons remains the concurrent control, the historical control ranges can be used to provide additional resources for assessing occurrence of rare or marginally increased tumor incidences. Historical control data can provide a point of reference for evaluation of typical ranges for incidences of spontaneous tumors (Greim et al., 2003). Because of the typical heterogeneity in historical control data collected over time, comparisons to current studies are typically on a qualitative rather than quantitative basis (Gad and McCord, 2008). In cases where incidence of a rare tumor is not found to be significantly different from the concurrent control, comparisons to the historical control values can assist in determination of the biological relevance of the tumor (Deschl et al., 2002).

Similarly, for tumors with a high spontaneous control incidence, the biological relevance of a significant increase from the concurrent control can be considered to be low when assessed in the framework of the typical range of variability for that tissue and strain of test animal.

11. Alternative animal carcinogenicity testing protocols

Development of standard protocols for conduct of carcinogenicity tests in animals has been accompanied by numerous suggestions for improvements that could increase sensitivity and specificity of bioassays as well as alternative test procedures to reduce the use of animals. Weisburger and Williams (1981) pointed out discrepancies in results from NCI carcinogenicity studies over two decades ago and suggested that a decision point approach that evaluates chemical structure and results from genotoxicity studies and short-term bioassays in animals could provide a mechanistic understanding of chemical actions to justify expending animals and resources for requiring conduct of a long-term carcinogenicity bioassay. The NTP carcinogenicity program that succeeded the NCI effort has evolved to incorporate several features of these same proposals but the conduct of a 2-year bioassay still continues to be a requirement in the regulation of food, pesticide and industrial chemicals in protocol components of regulatory agencies summarized in Table 1.

Alternative carcinogenicity testing procedures employing *in vitro* tests, transgenic mice and non rodent species have been reviewed elsewhere and are beyond the scope of the present review (Milman and Haber, 1994; Williams et al., 2008), and numerous excellent publications and reviews on the topic are available (Spalding et al., 1999, 2000; Tennant et al., 1993, 1996; Thompson et al., 1998; Weaver et al., 1998; French et al., 1999; Pritchard et al., 2003; Wells and Spencer-Williams, 2009). The next sections focus on some of the suggested modifications of procedures for incorporation into standard animal carcinogenicity studies and perceived strengths and weaknesses of each approach.

12. Limited carcinogenicity bioassays

Because of the cost, animal usage and time intervals required to obtain results from standard animal carcinogen bioassays, there have been numerous suggestions for alternatives for simplifying *in vivo* carcinogenicity protocols (IARC, 1980, 1999, 2006; Van der Laan et al., 2002; Cohen et al., 2001). Ward (2007) enumerated and provided citations for many of the proposed alternatives to rat and mouse carcinogenesis tests, including tests using no rodents, such as *in vitro* cell transformation and mutagenicity tests, computer-based structure function approaches, and tests with rodents, including the neonatal mouse assay, 6 month genetically engineered mouse assays, rat only assays, female rats and male mouse assays, multi-mouse strain studies and abbreviated liver or multi-organ rat bioassays. Williams et al. (2008) provided more detailed procedures for using initiation/promotion and accelerated carcinogenicity bioassays, assays using transgenic mice or newborn mice, and *in ovo* assays with avian eggs as potential adjuncts or possible alternatives to animal testing. Discussion of all of these alternatives are beyond the scope of this paper and, at this time, none of these tests have been sufficiently developed to supplant the regulatory requirement for standard two year (or longer) carcinogenicity bioassays in rodents.

The need for a two year exposure period for detection of carcinogenicity has been questioned (Davies et al., 2000) and alternative abbreviated time spans were proposed based on a review of published tests in the International Agency for Cancer Research (IARC) monographs (volumes 1–70). The feasibility of reducing the time frame for rodent carcinogenicity studies remains to be validated by comparative studies using the standard and abbreviated protocols for tests on a selected panel of candidate chemicals. Nevertheless, the IARC (1999) recommended using results from initiation/promotion protocols in carcinogenicity hazard identification.

13. Life span and extended dosing studies

Although EPA, FDA and OECD routine guidelines for carcinogenicity studies indicate that a dosing period of 18 months for mice and 24 months for rats should be employed, there has been interest in extending rodent tests beyond 24 months (Huff, 1999) and up to the end of the animals' natural life (Maltoni et al., 1999). The purpose of allowing animals to die a "natural death" is to putatively increase the sensitivity of the assay to detect weak carcinogens that produce tumors in old animals and/or at lower doses similar to human exposure levels. At ERF, animals are not terminated at the cessation of dosing but allowed to die a natural death because ERF scientists that pioneered this testing protocol believe this permits detection of "diffuse carcinogenic" properties of chemicals and weak carcinogens (Maltoni et al., 1999; Soffritti et al., 2002). According to Magnuson et al. (2007), important disadvantages of prolonging studies beyond 104–110 weeks, and allowing unscheduled deaths can "include an increase in age-related background lesions and higher probability of autolytic tissue changes in animals found dead" The review of one of the ERF studies (Soffritti et al., 2005, 2006) conducted by EFSA confirmed the occurrence of autolytic changes in tissues noted by an NTP Pathology Working Group resulting in histological changes that confounded microscopic evaluations of the ERF histological preparations (Hailey, 2004). It is not unexpected that when animals are allowed to die in an unscheduled manner, incidents that may or may not be related to treatment can arise when deaths are not detected for an extended period until the next daily cage checks are performed (EFSA, 2006).

Keenan et al. (2009) noted that for studies that are extended beyond the typical 24 month period of dosing, historical control data are unavailable from other laboratories for comparison and there are often substantial increases in spontaneous tumor rates in control groups that can interfere with study evaluation. They noted that in an evaluation of respective 24 and 30 month studies by Bomhard and Rinke (1994) and Bomhard (1992), "there were clear trends of higher incidence rates over time in pituitary tumors, adrenal pheochromocytomas, and mammary gland tumors, which also showed a shift toward malignancy from 2/31 malignant tumors after 24 months to 34/101 after 30 months in study durations".

A critical question arises about whether prolonging a study beyond 24 months gains a true benefit for increasing the sensitivity of the bioassay to detect effects from carcinogenic substances that are only seen late in life and beyond the time when typical bioassays are terminated (Monro, 1996; Haseman et al., 2001). Monro (1996) has argued that for chemicals that exhibit significant risk factors (genotoxicity, immune suppression, hormonal activity and chronic irritation/inflammation), the rodent carcinogenicity bioassay is redundant considering that the outcome of the carcinogenicity bioassay has been shown to be predictable for about half of a random selection of chemicals in the US National Toxicology Program (NTP). Gad and McCord (2008) conjectured on why carcinogenicity studies are terminated at fixed intervals while prolonging a study up to the end of an animal's natural life could allow for greater time to tumor development. They concluded, however, that depending upon tumor type the ability to detect a true increase in tumor incidence would *decrease* rather than in-

crease with time because the incidence of spontaneous (non-treatment related) tumors increases with age of the test animals.

The ability to discern the statistical significance of tumor incidence with age is a particular confounding issue because there is a significantly higher incidence of spontaneous tumors seen in control groups as animals age (Solleveld et al., 1984; Haseman et al., 1998). Hollander et al. (1984) evaluated pathological changes in aging rats and mice and cautioned that, when pathology data are gathered from dead and dying animals, questions arise about the applicability of tumor data for use in risk assessment for humans. However, they noted that pathological lesions unrelated to the cause of death (or treatment) can be expected to be present in the living population in a frequency comparable to that found in rats that died or were killed while moribund and provide a better approximation to prevalence in a living cohort. Also, the same authors noted that for those tumors that appeared late in life, shortly before death, or that killed the aged animal within a short period, the prevalence of the tumor in living animals although low, could be calculated from life-table analyses. Because of the age-related tumor increases, some investigators have questioned whether positive increases in carcinogenicity bioassays may in fact represent an acceleration of development rates for tumors that would have appeared even in untreated aged animals. In untreated rats (Hollander et al., 1984), the number of pathological lesions observed in rats increased with survival age, and the differences were most apparent in comparisons of groups that attained a survival of 90% of the estimated lifespan versus incidence of lesions in either 50% or 10% groups. Solleveld et al. (1984) showed that incidence of common lesions found in control F344 rats had a clear upward trend after 110 weeks of age, but, importantly:

> "*The variety of neoplastic lesions did not increase with age and that old age is not characterized by unique neoplasms. This finding means that life-span studies do not appear to have an advantage over 2-year studies. In contrast, life-span studies have a major disadvantage over 2-year studies, namely, a higher background incidence of many types of neoplasms that increases the chance of obtaining false negative results in carcinogenicity testing.*"

Specific examples of increases in tumor incidences cited by Solleveld et al. (1984) in comparisons of incidence of neoplastic lesions at 96–110 weeks versus 124–136 weeks were neurofibroma in subcutis increased from 9% to 21%; fibroadenoma in mammary gland, 4–18%; neoplastic nodules in liver, 5–14%; C-cell adenoma in thyroid, 6–14%; pheochromocytoma in adrenal gland, 16–40% (and 51% at >137 weeks); adenoma and carcinoma in pancreas, 3–4% to 7–10%. With such high background levels of tumors in individual organs, the number of animals required to detect a significant, test substance-related increase in carcinogenic endpoints for a specific tumor would be daunting. Specifically, numbers of test animals needed to detect test article related increases in tumor incidences above levels in control animals is significantly increased in relation to the incidence of tumors in the control group relative to increasing age. With a spontaneous level of 21% neurofibromas in subcutis, for example, 592 animals would be needed to determine that an incidence of 25% neurofibromas in test animals was significantly different from the control (Gad and Weil, 1986). Thus, at high levels of background tumors characteristic of aging animals, the laudatory goal to improve the assay for detection of weak carcinogens by extending the length of the bioassay is thwarted by a steep statistical hurdle regarding sample size required to detect significance above background levels.

Despite the significant problems in extending the time frame of carcinogenicity studies discussed previously, Huff (1999) has supported extending the time span of the bioassay to "30 or more months" or to continue exposures until 10–20% of the animals remain. However, Huff (1999) did not address numerous concerns about dealing with increased background tumor rates in the aged animals and the potential for reduced statistical power of the assay in the presence of high spontaneous rates in aged animals evaluated after "natural death" as noted by many others (Hollander et al., 1984; Gad and Weil, 1986; Solleveld and McConnell, 1985; Williams et al., 2008; Gad and McCord, 2008; Keenan et al., 2009). Despite the significant issues raised by these authors, Huff et al. (2008) have continued to suggest that the sensitivity of chemical carcinogenesis assays would be enhanced by exposing rodents beginning *in utero* and continuing for at least 30 months (130 weeks) or until their natural deaths at "up to about 3 years" by citing results from studies at ERF as examples to support the benefits of their proposals.

The deficiencies in execution and evaluation of those studies found in EFSA (2006, 2009) and FDA (2000b) reviews of ERF studies, and deficiencies found in NTP PWG audits of pathology data (Hailey, 2004), make aspartame studies at ERF a poor example for supporting lifetime protocols since the published "positive" results would likely be significantly modified if deficiencies in pathology evaluation were addressed and the data reanalyzed. In conclusion, the relevance of a lifetime dosing regimen in animal studies for determining carcinogenic risk to humans (where daily-lifetime exposure to a single substance would be unusual and extremely rare) remains unproven. The statistical arguments for large sizes of animal groups required to detect significance of differences from control levels of tumors in aging animals (Gad and Weil, 1986) remain as realistic impediments to implementation, and to feasibility and human relevance of life span rodent carcinogenicity studies as a routine testing protocol for all substances.

14. Perinatal (in utero) dosing

Current protocols for conducting carcinogenicity tests for regulatory submission, as well as in screening tests for oncogenic potential, employ young 5- to 6-week-old animals which are dosed with the test and control substances at or shortly after weaning (FDA, 2000a,b; OCED, 1981, 2008; NTP, 2006; Maltoni et al., 1999). However, for carcinogenicity tests of chemicals either intended for direct addition into food, or for food contact substances that could pose cumulative exposures greater than 1 ppm in the diet, current FDA (CFSAN) guidelines indicate that an *in utero* exposure phase of dosing should be employed in one of the two species used for testing to determine possible effects on incidence of tumors or "chronic disease outcomes" (FDA, 2000c). Although OECD testing guidelines for developmental toxicity include *in utero* dosing, no indications of changes in OECD guidance for dosing prenatally are contained in past or current revised guidelines for carcinogenicity studies (OECD, 1981, 2009a). The NTP carcinogen testing program has not employed a protocol that incorporated perinatal dosing in studies published to date. However, as discussed previously, a recent NTP decision to use Harlan SD rats as the model test animal for all future carcinogenicity and reproductive/developmental studies will allow switching to an *in utero* dosing protocol on a chemical specific basis determined by anticipated human exposure to the test chemical agent.

The carcinogen screening program at ERF has employed perinatal dosing in two published carcinogenicity tests on aspartame (Soffritti et al., 2007, 2008), although this dosing protocol has not been a routine procedure for other studies at this laboratory. Comparisons of the strengths and weaknesses of perinatal exposure tests are complicated by the lack of adequate data on parallel tests of similar or related classes of chemical substances in weanlings versus perinatal dosing regimens. However, perinatal dosing re-

sults at ERF provide an opportunity to evaluate these data in comparison to previous ERF studies on aspartame conducted by dietary administration (Soffritti et al., 2005; Soffritti and Belpoggi, 2005; Belpoggi et al., 2006b).

Soffritti et al. (2007) dosed groups of 70–95 male and female CRC/ERF rats with 0, 400 or 2000 ppm of aspartame (APM) in the diet from "the 12th day of fetal life until natural death". The results of the study reported "(*a*) a significant dose-related increase of malignant tumor–bearing animals in males ($p < 0.01$), particularly in the group treated with 2000 ppm APM ($p < 0.01$); (*b*) a significant increase in incidence of lymphomas/leukemias in males treated with 2000 ppm ($p < 0.05$) and a significant dose-related increase in incidence of lymphomas/leukemias in females($p < 0.01$), particularly in the 2000-ppm group ($p < 0.01$); and (*c*) a significant dose-related increase in incidence of mammary cancer in females ($p < 0.05$), particularly in the 2000-ppm group ($p < 0.05$)". The results were described as corroborative of carcinogenicity effects of aspartame in weanlings and that exposures beginning during fetal life increased tumor incidences.

Because of questions and concerns raised in reviews of previous studies conducted at ERF (EFSA, 2006; FDA, 2000c; Hailey, 2004), EFSA responded to a request from the EC and conducted a review of the perinatal dosing study summarized above to determine the relevance of the findings for human health risk (EFSA, 2009). EFSA noted that histopathological observations were restricted to the incidence and total numbers of malignant tumors and incidence of lymphomas, leukemias and mammary gland carcinomas and that all malignant tumors were aggregated for statistical evaluations. Aggregation of tumors [without justification for common morphogenetic tissue origins] was not considered a sound scientific approach. When the incidence of leukemias and lymphomas in the prenatal study (Soffritti et al., 2007) were compared with results from the post-natal dosing study with respect to the individual concurrent control groups (Soffritti et al., 2006), the 20 mg/kg bw dose produced a 2.3-fold increase in tumors following post-natal dosing but only 1.4-fold following the pre-natal dosing regimen. At the highest dose of 100 mg/kg in both studies, there was a 2.2-fold increase above control levels in the post-natal dosing study and a slightly higher ratio of 2.5-fold relative to the concurrent control. These data indicate that sensitivity of the perinatal dosing protocol at the low dose of aspartame was lower than seen in the previous post-natal study and there was only a slight increase in leukemias/lymphoma over respective controls at the highest dose in the perinatal dosing protocol.

The conclusions of the EFSA review indicated that:

- Evaluation of aggregated malignant tumor incidences as evidence of carcinogenic potential of the test compound can only be performed based on a thorough consideration of all tumor data including onset, and data on non-neoplastic, hyperplastic and preneoplastic lesions, and that these data were not provided by the authors. Only limited information on the presence of inflammatory changes in the lungs of animals with lymphomas and leukemias was provided by ERF and only in data not included in their publication.
- The majority of the lymphomas and leukemias observed appeared to have developed in rats suffering from inflammatory changes in the lungs, which is characteristic for chronic respiratory disease. In accordance with the previous view of the studies of Soffritti and Belpoggi (2005) and Soffritti et al. (2006), these incidences of leukemias and lymphomas were not considered to be related to the treatment with aspartame.
- The increase in incidence of mammary carcinoma is not considered indicative of a carcinogenic potential of aspartame since the incidence of mammary tumors in female rats (15.7%) was only slightly higher than the upper value for historical controls (14.2%) and is known to vary considerably between carcinogenicity studies. In addition, an increased incidence of mammary carcinomas was not reported to suggest this as a target tissue in the previous ERF study with aspartame which used much higher doses of the compound.

The lack of a substantial increase in sensitivity in detecting leukemias and lymphomas in comparisons of perinatal and postnatal protocols indicates that perinatal dosing for carcinogenicity testing may have limited and questionable value. Results from future NTP studies that incorporate perinatal dosing regimes will provide a better basis for assessing the utility of this stringent route of exposure in rodents given doses at and near the MTD and its relevance to humans.

15. Perspectives

Appropriate protocol design should be based upon the application of scientifically established test methods. Carcinogenicity studies that publish results of food ingredient testing must be done in a manner that is scientifically rigorous and appropriate. If a study methodology does not conform to guidelines that have the benefit of scientific and regulatory consensus, then an evaluation of the impact of the deviations from such standard protocol design must be completed to determine whether the methodology has compromised the robustness of conclusions that can be drawn from the data. The review, included in this paper, of study design elements, methodologies and general practices utilized in carcinogenicity studies of food ingredients elucidates the problems that can be inherent with certain deviations from standardly recommended designs, methodologies and practices. It further indicates the critical considerations for development of a carcinogenicity study design that permits, with appropriate relation to human exposure, an acceptable identification of potential carcinogenic hazard. Protocols for testing the carcinogenic potential of industrial, environmental or consumer chemicals has evolved from early studies prior to the 1960s, which evaluated new chemical entities, to the formalized NCI testing regimen developed in the 1960–1980 period, and later to the current NTP testing program and other guidelines, such as OECD guidelines for carcinogenicity testing. Experience in early programs led to improvements and harmonization of all current regulatory and international testing guidelines, although some differences in procedures still exist. Efforts to improve the sensitivity and reliability of tests have been a critical part of the evolution in guidances. Important aspects for test sensitivity and reliability are discussed in this paper. They include, but are not limited to: (1) use of healthy SPF animals to prevent confounding problems of disease and ensure adequate survival of animals up to study termination, (2) appropriate animal husbandry practices to ensure survival of maximum numbers of animals to permit meaningful statistical comparisons, (3) harmonized pathology assessment and appropriate pathology review procedures for consistent evaluation of study pathology related data., (4) continuous exposure for a substantial proportion of a rodent lifespan, typically 18–24 months, to allow for assessment of potential effects from youth to advanced age and to compare with teratology data, for younger ages, (5) an adequate number of animals, typically 50 per sex per group, to ensure adequate numbers of survivors and permit meaningful statistical comparisons, and (6) employment of high doses, often very high multiples of possible human exposure, up to an MTD found to produce some toxicological effects in preliminary studies. The use of such high doses, while an unlikely scenario for human exposure, is an important element to allow for a reasonable estimation of the margin of safety for human risk.

In recent years, there have been suggestions to begin exposure earlier than at weaning or to extend study duration beyond 24 months. These suggestions include exposures beginning in the perinatal period, studies for the entire lifespan of the animals, or both. Perinatal exposures can be appropriate when there is a risk for human exposure during gestation and/or there is an indication, based on related structural and mechanistic studies, that gestational and/or gestational-related effects are possible. Inclusion of perinatal exposure is, however, tempered by current scientific evidence that indicates that few if any new carcinogens are detected by this procedure. Similarly, proponents of extending the span of testing until the natural death of animals have not shown convincing evidence from data generated from independently audited studies that would support such a change in study design. In addition, the number of animals needed to detect a significant increase in tumor incidence can become an insurmountable problem due to the high background incidence of spontaneous tumors in the control groups from animals in lifetime exposure studies.

The current OECD protocols serve as a good basis for carcinogenicity tests on the majority of test chemicals because of the availability of their large comparative database on chemicals to provide perspective on results on new and related agents. They recommend protocols that include elements important to test sensitivity and reliability, such as those summarized above. Alternative protocols are likely best used selectively, and only when justified by good scientific rationale.

Because published science becomes public knowledge, it is important that carcinogenicity testing of food ingredients is placed into perspective, particularly when public health is involved. Chronic/carcinogenicity studies that publish results of ingredient testing in the open literature must be done in a manner that is scientifically rigorous. If study methodology does not conform to guidelines that have the benefit of scientific and regulatory consensus, then an evaluation of the impact of the deviations from the standard protocol design must determine whether the methodology has compromised the robustness of the conclusions.

Acknowledgment

Supported by a contract to Spherix Consulting, Inc. from Tate and Lyle and McNeil Nutritionals, LLC for independent research and Expert Panel deliberation.

References

Abdo, K.M., Kari, F.W., 1996. The sensitivity of the NTP bioassay for carcinogen hazard evaluation can be modulated by dietary restriction. Exp. Toxicol. Pathol. 48, 129–137.

ACVP (American College of Veterinary Pathologists), 2009. Certifying examinations and qualifications. Available from: <http://www.acvp.org/exam/> (accessed October, 2009).

Alexandrov, V., Aiello, C., Rossi, L., 1990. Modifying factors in prenatal carcinogenesis (review). In Vivo 5, 327–335.

Allaben, W.T., Turturro, A., Leakey, J.E.A., Seng, J.E., Hart, R.W., 1996. FDA points-to-consider documents: the need for dietary control for the reduction of experimental variability within animal assays and the use of dietary restriction to achieve dietary control. Toxicol. Pathol. 24, 776–781.

Bailer, A.J., Portier, C.J., 1988. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. Biometrics 44, 417–431.

Balcombe, J.P., Barnard, N.D., Sandusky, C., 2004. Laboratory routines cause animal stress. Contemp. Topics Lab. Animal Sci. 43, 42–51.

Baldrick, P., 2005. Carcinogenicity evaluation: comparison of tumor data from dual control groups in the Sprague–Dawley Rat. Toxicol. Pathol. 33, 283–291.

Baldrick, P., Reeve, L., 2007. Carcinogenicity evaluation: comparison of tumor data from dual control groups in the CD-1 mouse. Toxicol. Pathol. 35, 562–569.

Barton, H.A., Pastoor, T.P., Baetcke, K., Chambers, J.E., Diliberto, J., Doerrer, N., Driver, J.H., Hastings, C.E., Iyengar, S., Krieger, R., Stahl, B., Timchalk, C., 2006. The acquisition and application of absorption, distribution, metabolism, and excretion (ADME) data in agricultural chemical safety assessments. Crit. Rev. Toxicol. 36, 9–35.

Belpoggi, F., Soffritti, M., Tibaldi, E., Falcioni, L., Bua, L., Trabucco, F., 2006a. Results of long-term carcinogenicity bioassays on Coca-Cola administered to Sprague–Dawley rats. Ann. NY Acad. Sci. 1076, 736–752.

Belpoggi, F., Soffritti, M., Padovani, M., Degli Esposti, D., Lauriola, M., Minardi, F., 2006b. Results of long-term carcinogenicity bioassay on Sprague–Dawley rats exposed to aspartame administered in feed. Ann. NY Acad. Sci. 1076, 559–577.

Black, H.E., 1991. Peer review in toxicologic pathology: some recommendations. Toxicol. Pathol. 19, 290–292.

Bomhard, E., 1992. Frequency of spontaneous tumors in Wistar rats in 30-month studies. Exp. Toxicol. Pathol. 44, 381–392.

Bomhard, E., 1993. Frequency of spontaneous tumours in NMRI mice in 21-month studies. Exp. Toxicol. Pathol. 45, 269–289.

Bomhard, E., Mohr, U., 1989. Spontaneous tumors in NMRI mice from carcinogenicity studies. Exp. Pathol. 36, 129–145.

Bomhard, E., Rinke, M., 1994. Frequency of spontaneous tumours in Wistar rats in 2-year studies. Exp. Toxicol. Pathol. 46, 17–29.

Bomhard, E., Karbe, E., Loeser, E., 1986. Spontaneous tumors of 2000 Wistar TNO/W.70 rats in two-year carcinogenicity studies. J. Environ. Pathol. Toxicol. Oncol. 7, 35–52.

Brix, A.E., Nyska, A., Haseman, J.K., Sells, D.M., Jokinen, M.P., Walker, N.J., 2005. Incidences of selected lesions in control female Harlan Sprague–Dawley rats from two-year studies performed by the National Toxicology Program. Toxicol. Pathol. 33, 477–483.

Buchanan, J.R., Burka, L.T., Melnick, R.L., 1997. Purpose and guidelines for toxicokinetic studies within the National Toxicology Program. Environ. Health Perspect. 105, 468–471.

Bucher, J.R., 2002. The national toxicology program rodent bioassay, designs, interpretations, and scientific contributions. Ann. NY Acad. Sci. 982, 198–207.

Cameron, T.P., Hickman, R.L., Kornreich, M.R., Tarone, R.E., 1985. History, survival and growth patterns of B6C3F1 mice and F344 rats in the National Cancer Institute carcinogenesis testing program. Fund. Appl. Toxicol. 5, 526–538.

Chandra, M., Frith, C.F., 1992. Spontaneous neoplasms in aged control Fischer 344 rats. Cancer Lett. 62, 49–56.

Chandra, M., Riley, M.G.I., Johnson, D.E., 1992. Spontaneous neoplasms in aged Sprague–Dawley rats. Arch. Toxicol. 66, 496–502.

Chapin, R.E., DK, Gulati, Barnes, L.H., Teague, J.L., 1993. The effects of feed restriction on reproductive function in Sprague–Dawley rats. Fund. Appl. Toxicol. 20, 23–29.

Chow, S.C., Liu, J.P., 1998. Design and Analysis of Animal Studies in Pharmaceutical Development. Marcel Dekker, NY, 406 pp.

Christian, M.S., Hoberman, A.M., Johnson, M.D., Ray Brown, W., Bucci, T.J., 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.

Clark, D.E., Elissalde, M.H., Doyle, J.J., Mollenhauer, H.H., 1981. Interaction of environmental stress and hexachlorobenzene in the laboratory rat. J. Toxicol. Environ. Health 8, 297–306.

Cogliano, V.J., Baan, R.A., Straif, K., Grosse, Y., Secretan, M.B., El Ghissassi, F., Kleihues, P., 2004. The science and practice of carcinogen identification and evaluation. Environ. Health Perspect. 112, 1269–1274. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1247515/?tool=pubmed>.

Cohen, S., Robinson, D., MacDonald, J., 2001. Alternative models for carcinogen testing. Toxicol. Sci. 64, 14–19. Available from: <http://toxsci.oxfordjournals.org/cgi/reprint/64/1/14>.

Combes, R., Schechtman, L., Stokes, W.S., Blakey, D., 2002. The International Symposium on Regulatory Testing and Animal Welfare: Recommedations on best scientific practices for subchronic/chronic toxicity and carcinogenicity testing. ILAR J. 43 (Suppl.), S112–S117. Available from: <http://dels.nas.edu/ilar_n/ilarjournal/43_supp/v43supCombes.shtml>.

Contrera, J.F., Jacobs, A.C., Prasanna, H.R., Mehta, M., Schmidt, W.J., DeGeorge, J., 1995. A systemic exposure-based-alternative to the maximum tolerated dose for carcinogenicity studies of human therapeutics. J Amer. College Toxicol. 14, 1–10.

Contrera, J.F., Jacobs, A.C., DeGeorge, J.J., 1997. Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals. Regul. Toxicol. Pharmacol. 25, 130–145.

Council of Europe, 2006. Appendix A of the European Convention for the protection of vertebrate animals used for experimental and other scientific purposes (ETS NO. 123): Guidelines for Accomodation and Care of Animals (Article 5 of the Convention) Approved by the Multilateral Consultation. Strassburg. Available from: <http://www.coe.int/t/e/legal_affairs/legal_co-operation/biological_safety%2C_use_of_animals/laboratory_animals/2_006/Cons123%282006%293AppendixA_en.pdf> (accessed December 2009).

Counts, J.L., Goodman, J.I., 1995. Principles underlying dose selection for, and extrapolation from, the carcinogen bioassay: dose influences mechanism. Reg. Toxicol. Pharmacol. 21, 418–421.

Crissman, J.W., Goodman, D.G., Hildebrandt, P.K., Maronpot, R.R., Prater, D.A., Riley, J.H., Seaman, W.J., Thake, D.C., 2004. Society of Toxicologic Pathologists best practices guideline: toxicologic histopathology. Toxicol. Pathol. 32, 126–131.

CRL (Charles River Laboratories, Inc.), 2009. CRL® IGS (SD) rat model. Available from: <http://www.criver.com/en-US/ProdServ/ByType/ResModOver/ResMod/Pages/CDRat.aspx> (accessed October, 2009).

Davies, T.S., Lynch, B.S., Monro, A.M., Munro, I.C., Nestmann, E.R., 2000. Rodent carcinogenicity tests need be no longer than 18 months: an analysis based on 210 chemicals in the IARC monographs. Food Chem. Toxicol. 38, 219–235.

Decreto Legislativo 116, 1992. Attuazione della direttiva. 86/609/CEE in materia di protezione degli animali utilizzati a fini sperimentali o ad altri fini scientifici [in Italian]. Supplemento ordinario alla Gazzetta Ufficiale 40, 5–25. Cited in: Soffritti et al. (2006).

Deschl, U., Kittel, B., Rittinghausen, S., Morawietz, G., Kohler, M., Mohr, U., Keenan, C., 2002. The value of historical control data – scientific advantages for pathologists, industry and agencies. Toxicol. Pathol. 30, 80–87.

EC (European Commission), 1986. Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. Available from: <http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=EN&numdoc=31986L0609&model=guichett> (accessed October, 2009).

EC (European Commission), 2009. Implementation of GLP directives in the European states: Directives 2004/9/EC and 2004/10/EC. Available from: <http://ec.europa.eu/enterprise/chemicals/legislation/glp/index_en.htm> (accessed November, 2009).

EFSA (European Food Safety Authority), 2006. Opinion of the Scientific Panel on Food Additives, Flavourings, Processing Aids and Materials in Vontact with Food (AFC) on a request from the Commission related to a new long-term carcinogenicity study on aspartame. Question number EFSA-Q-2005-122. Adopted May 3, 2006. The EFSA J. 356, 1–44.

EFSA, 2009. Updated Scientific Opinion of the Panel on Food Additives and Nutrient Sources added to Food on a request from the European Commission related to the 2nd ERF carcinogenicity study on aspartame taking into consideration study data submitted by the Ramazzini Foundation in February 2009. EFSA-Q-2009-00474. Adopted March 19, 2009. The EFSA J. 1015, 1–18.

Elmore, S.A., Peddada, S.D., 2009. Points to consider on the statistical analysis of rodent cancer bioassay data when incorporating historical control data. Toxicol. Pathol. 37, 672–676.

EMEA (European Medicines Evaluation Agency), 2005. Guideline on the evaluation of control samples in nonclinical safety studies: checking for contamination with the test substance. Committee for Medicinal Products for Human Use. CPMP/SWP/1094/04. Available from: <http://www.emea.europa.eu/pdfs/human/swp/109404en.pdf> (accessed December 2009).

EPA, 1983. Environmental Protection Agency Good laboratory practice standards. (Source: 48 FR 52922, Nov.29, 1983). 40 CFR Part 160.1-160.195. 7/1/2007 Ed. Available from: <http://www.access.gpo.gov/nara/cfr/waisidx_07/40cfr160_07.html> (accessed October, 2009).

EPA, 1984. Environmental Protection Agency – Fungicides, Insecticides and Rodenticides Act. 48 Federal Register 53946. Effective May 2, 1984. Available from: <http://www.epa.gov/oecaerth/monitoring/programs/fifra/glp.html> (accessed October, 2009).

EPA, 1998. Environmental Protection Agency, Office of Prevention, Pesticide and Toxic Substances. Health Effect Test Guidelines OPPTS 870-4200: Carcinogenicity. EPA 712-C-98-211. Available from: <http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/> (accessed October, 2009).

EPA, 2005. Guidelines for Carcinogen Risk Assessment (Cancer Guidelines). Federal Register Notice. Available from: <http://cfpub.epa.gov/ncea/cfm/rcprdos[;au/cf,?deod=116283>.

Ertz, K., Preu, M., 2008. International GLP: a critical reflection on the harmonized global standard from a test facility viewpoint. Annali del Instituto Superiore di Sanita 44, 390–394.

ESTP (European Society of Toxicologic Pathology), 2005. Regarding Society of Toxicology Pathology (STP) "Revision of standardized nomenclature for lesions in the rat and mouse". Unpublished Letter from W. Kauffmann, President of ESTP to Nancy Evards, President STP, September 19, 2005.

ESTP (European Society of Toxicologic Pathology), 2007. Accreditation as "IFSTP/IATP Accredited Toxicologic Pathologist". October 2007 Draft SOP. Available from: <http://www.eurotoxpath.org/IFSTP_IATP_accreditation.pdf> (accessed October, 2009).

Ettlin, R.A., Bolon, B., Pyrah, I., Konishi, Y., Black, H.E., 2007. Global recognition of toxicologic pathologists: where we are now and where we need to go. J. Toxicol. Pathol. 20, 267–272.

Fairweather, W.R., Bhattacharyya, A., Ceuppens, P.P., Heimann, G., Hothorn, L.A., Kodell, R.L., Lin, K.K., Mager, H., Middleton, B.J., Slob, W., Soper, K.A., Stallard, N., Ventre, J., Wright, J., 1998. Biostatistical methods in carcinogenicity studies. Drug Inform. J. 32, 401–421.

Faustman, E.L., Omenn, G.S., 2008. In: Klaassen, C.D. (Ed.), Risk Assessment, in Casarett & Doull;s Toxicology, seventh ed., p. 112.

FDA, 1978. Good laboratory practice for non-clinical laboratory studies. (Source: 43 FR 60013, December 22, 1978; as amended at 52 FR 33779, September 4, 1987). 21 CFR Parts 58.1 – 58.219. 4/1/2007 Ed. Available from: <http://www.access.gpo.gov/nara/cfr/waisidx_07/21cfr58_07.html> (accessed October, 2009).

FDA, 2000a. Toxicological principles for the safety assessment of food ingredients: FDA Redbook 2000: Section IV.B.1. General guidelines for designing and conducting toxicity studies. November 2003, 18 pp. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/ucm078315.htm> (accessed October, 2009).

FDA, 2000b. Toxicological principles for the safety assessment of food ingredients: FDA Redbook 2000: Section IV.C.6. Carcinogenicity studies with rodents. January 2006, 21 p. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/ucm078388.htm> (accessed October, 2009).

FDA, 2000c. Toxicological principles for the safety assessment of food ingredients: FDA Redbook 2000: Section IV.C.8. *In utero* exposure phase for addition to carcinogenicity studies or chronic toxicity studies with rodents. July 2007, 26 pp. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/ucm078394.htm> (accessed October, 2009).

FDA, 2000d. Toxicological principles for the safety assessment of food ingredients: FDA Redbook 2000: Section IV.B.3. Pathology considerations in toxicity studies. July 2000, 8 pp. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/ucm078318.htm> (accessed October, 2009).

FDA, 2000e. Toxicological principles for the safety assessment of food ingredients: FDA Redbook 2000: Section IV.B.4. Statistical considerations in toxicity studies. July 2000, 9 pp. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/ucm078320.htm> (accessed October, 2009).

FDA, 2000f. Toxicological principles for the safety assessment of food ingredients: FDA Redbook 2000: Section IV.B.2. Guidelines on reporting the results of toxicity studies. November 2003, 8 pp. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/ucm078409.htm> (accessed October, 2009).

FDA, 2001. Center for Drug and Evaluation and Research. Guidance for industry: Statistical aspects of the design, analysis and interpretation of chronic rodent carcinogenicity studies of pharmaceuticals. Draft guidance, KKK Lin, FDA. Available from: <http://www.fda.gov/downloads/Drugs/. . ./Guidances/UCM079272.pdf> (accessed October, 2009).

FDA, 2007. Center for Food Safety and Applied Nutrition. FDA Statement on European Aspartame Study. April 20, 2007. Available from: <http://www.cfsan.fda.gov/~lrd/fpaspar2.html> (accessed November, 2009).

FDA, 2008a. Guidance for Industry. S1C(R2) Dose selection for carcinogenicity studies. Center for Drug Evaluation and Research (CDER); Center for Biologics Evaluation and Research (CBER). September 2008: ICH Revision 1. Fed. Reg. 60, 11278–11281; Addendum published in the Fed. Reg. 62, 64260.

FDA, 2008b. CFSAN/Office of Food Safety: Total Diet Study, Study Design. Available from: <http://www.fda.gov/Food/FoodSafety/FoodContaminantsAdulteration/TotalDietStudy/ucm184232.htm> (accessed November, 2009).

Foran, J.A., 1997. Principles for the selection of doses in chronic rodent bioassays. ILSI Risk Science Working Group on Dose Selection. Environ. Health Perspect. 105, 18–20.

French, J.E., Spalding, J.W., Dunnick, J.K., Tice, R.R., Furedi-Machacek, M., Tennant, R.W., 1999. The use of transgenic animals in cancer testing. Inhal. Toxicol. 11, 541–544.

Gad, S.C., McCord, M.G., 2008. Carcinogenicity. In: Safety Evaluation in the Development of Medical Devices and Combination Products, third ed. Informa Healthcare, USA, NY, pp. 197–202.

Gad, S.C., Weil, C.S., 1986. Statistics and Experimental Design for Toxicologists. The Telford Press, NJ. 380 pp.

Gart, J.J., Krewski, D., Lee, P.N., Tarone, R.E., Wahrendorf, J., 1986. Statistical Methods in Cancer Research, vol. III, The Design and Analysis of Long-term Animal Experiments. International Agency for Research on Cancer, Lyon, 219 pp.

Gaylor, D.W., Swirsky Gold, L., 1998. Regulatory cancer risk assessment based on a quick estimate of a benchmark dose derived from the maximum tolerated dose. Reg. Tox. Pharmacol. 2S, 222–225.

Gaylor, D.W., Chen, J.J., Kodel, R.L., 1985. Experimental design of bioassays for screening and low dose extrapolation. Risk Anal. 5, 9–16. Cited in: Rhomberg et al. (2007).

Gehring, P.J., Blau, G.E., 1977. Mechanisms of carcinogenesis: dose response. J Environ. Path. Toxicol. 1, 163–179.

Gehring, P.J., Watanabe, P.G., Park, C.N., 1978. Resolution of dose–response toxicity data for chemicals requiring metabolic activation: example vinyl chloride. Toxicol. Appl. Pharmacol. 44, 581–591. Cited in Haseman (1985) referenced herein.

Goodman, J.I., 2001. A perspective on current and future uses of alternative models for carcinogenicity testing. Toxicol. Pathol. 29 (Suppl.), 173–176.

Goodman, D.G., Boorman, G.A., Strandberg, J.D., 1994. Selection and use of the B6C3HF1 mice and F344 rats in long-term assays for carcinogenicity. In: Handbook of Carcinogen Testing, second ed., 904 pp. William Andrew Publ., NY, pp. 347–390.

Greim, H., Gelbke, H.-P., Reuter, U., Thielmann, H.W., Edler, L., 2003. Evaluation of historical control data in carcinogenicity studies. Hum. Exper.Toxicol. 22, 541–549.

Hailey, J.R., 2004. Pathology Working Group Chairperson's Report. Lifetime study in rats conducted by the Ramazzini Foundation. Prepared by James R. Hailey, Pathology Working Group Chair. National Institute of Environmental Health Sciences, Research Triangle Park, USA. Submitted to: Dr. Fiorella Belpoggi, Ramazzini Foundation, Bologna, Italy. November 30, 2004: Cited In Report from: Committee on Carcinogenicty of Chemicals in Food, Consumer Products and the Environment: Statement on a Carcinogenicity Study of Aspartame by the European Ramazzini Foundation; Statement COC/06/S2 – December 2006. Available from: <http://www.iacoc.org.uk/statements/documents/COC06S2AspartamestatementDec2006_000.pdf> (accessed November 2009).

Hamm Jr., T.E., 1994. Design of a long-term animal bioassay for carcinogenicity. In: Milman, H.A., Weisburger, E.K. (Eds.), Handbook of Carcinogen Testing, second ed. Noyes Publications, NJ, pp. 270–285. 904 pp.

Hart, R.W., Kennan, K., Turturro, A., Abdo, K.M., Leakey, J., Lyn-Cook, B., 1995. Caloric restriction and toxicity. Fund. Appl. Toxicol. 25, 184–195.
Haseman, J.K., 1985. Issues in carcinogenicity testing: dose selection. Fund. Appl. Toxicol. 5, 66–78.
Haseman, J.K., 1992. Value of historical controls in the interpretation of rodent tumor data. Drug Info. 26, 191–200.
Haseman, J.K., Hoel, D.G., 1979. Statistical design of toxicity assays: role of genetic structure of test animal population. J. Environ. Pathol. Toxicol. 2, 1313–1327.
Haseman, J.K., Lockhart, A., 1994. The relationship between use of the maximum tolerated dose and study sensitivity for detecting rodent carcinogenicity. Fund. Appl. Toxicol. 22, 382–391.
Haseman, J.K., Huff, J., Boorman, G.A., 1984. Use of historical control data in carcinogenicity studies in rodents. Toxicol. Pathol. 12, 126–135.
Haseman, J.K., Winbush, J.S., O'Donnell Jr., M.W., 1986. Use of dual control groups to estimate false positive rates in laboratory animal carcinogenicity studies. Fund. Appl. Toxicol. 7, 573–584.
Haseman, J.K., Huff, J.E., Rao, G.N., Eustis, S.L., 1989. Sources of variability in rodent carcinogenicity studies. Fund. Appl. Toxicol. 12, 793–804.
Haseman, J.K., Young, E., Eustis, S.L., Hailey, J.R., 1997. Body weight-tumor incidence correlations in long-term rodent carcinogenicity studies. Toxicol. Pathol. 25, 256–263.
Haseman, J.K., Hailey, J.R., Morris, R.W., 1998. Spontaneous neoplasm incidences in Fischer 344 rats and B6C3F1 mice in two-year carcinogenicity studies: National Toxicology Program update. Toxicol. Pathol. 26, 428–441.
Haseman, J., Melnick, R., Tomatis, L., Huff, J., 2001. Carcinogenesis bioassays: study duration and biological relevance. Food Chem. Toxicol. 39, 739–744.
Haseman, J.K., Ney, E., Nyska, A., Rao, G.N., 2003. Effect of diet and animal care/housing protocols on body weight, survival, tumor incidences, and nephropathy severity of F344 rats in chronic studies. Toxicol. Pathol. 31, 674–681.
Hattis, D., Goble, R., Chu, M., 2005. Age-related differences in susceptibility to carcinogenesis. II. Approaches for applicatiuon and uncertainty analyses for individual genetically acting carcinogens. Environ. Health Perspect. 113, 509–516.
Hollander, C.F., Solleveld, H.A., Zurcher, C., Nooteboom, A.L., Van Zwieten, M.J., 1984. Biological and clinical consequences of longitudinal studies in rodents: their possibilities and limitations. An overview. Mech. Aging Develop. 28, 249–260.
Huff, J., 1999. Long-term chemical carcinogenesis bioassays predict human cancer hazards. Ann. NY Acad. Sci. 895, 56–79.
Huff, J., 2002. Chemicals studied and evaluated in long-term carcinogenesis bioassays by both the Ramazzini foundation and the National Toxicology program. In tribute to Cesare Maltoni and David Rall. Ann. NY Acad. Sci. 982, 208–230.
Huff, J., Jacobson, M.F., Davis, D.L., 2008. The limits of two-year bioassay exposure regimens for identifying chemical carcinogens. Environ. Health Perspect. 116, 1439–1442.
Hurst, J.L., Barnard, C.J., Tolladay, U., Nevision, C.M., West, C.D., 1999. Housing and welfare in laboratory rats: effects of cage stocking density and behavioural predictors of welfare. Animal Behav. 58, 563–586.
IARC (International Agency for Research on Cancer), 1980. Long-term and short-term screening assays for carcinogens: a critical appraisal. IARC Monographs Suppl. 2, Lyon.
IARC, 1999. The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. In: McGregor, D.B., Rice, J.M., Venit, S. (Ed.), IARC Scientific Publications No. 146. Lyon, France. Available from: <http://monographs.iarc.fr/ENG/Publications/index1.php>.
IARC, 2000: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon, France, World Health Organization.
IARC, 2006. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans: Preamble to the IARC monographs (amended January 2006). Lyon, France. Available from: <http://monographs.iarc.fr/ENG/Classification/index.php> (accessed November, 2009).
Iatropoulos, M.J., 1988. Society of toxicologic pathologists position paper: "Blinded" microscopic examination of tissues from toxicologic or oncogenic studies. In: Grice, H.C., Ciminera, J.L. (Eds.), Carcinogenicity. The Design, Analysis and Interpretation of Long-term Animal Studies. Proceedings of the Interdisciplinary Discussion Group on Carcinogenicity Studies held in Chapel Hill, NC, June 1986 . ILSI Monographs. Springer-Verlag, NY.
ICH (International Conference for Harmonization). 1997. Testing for carcinogenicity of pharmaceuticals. Guideline S1B. US Fed. Reg. 63: 8983; EU: CPMP/ICH/299/95. Available from: <http://www.ich.org/cache/compo/502-272-1.html#S1B> (accessed October, 2009).
ICH, 2008. Dose selection for carcinogenicity studies of pharmaceuticals. Guideline S1C(R2). US Fed. Reg. 60, 11278–11281; EU CHMP/ICH/383/1995. Available from: <http://www.ich.org/cache/compo/502-272-1.html#S1B> (accessed November, 2009).
Imai, D., Yoshimura, S., Yamaguchi, K., Matsui, E., Isaka, H., Hashimoto, K., Boorman, G.A., 1990. Effects of dietary restriction on age-associated pathologic changes in F-344 rats. J. Toxicol. Pathol. 3, 209–221.
Jameson, C.S., Goehl, T.J., 1994. Chemistry requirements for toxicological and carcinogenicity evaluation of chemicals. In: Milman, H.A., Weisburger, E.K. (Eds.), Handbook of Carcinogen Testing, second ed. Noyes Publications, NJ, pp. 286–297. 904 pp.
Jena, G.B., Kaul, C.L., Ramarao, P., 2005. Regulatory requirements and ICH guidelines on carcinogenicity testing of pharmaceuticals: a review on current status. Indian J. Pharmacol. 37, 209–222.
Johnson, M.D., 2006. The rat: toxicology. In: Gad, S.C. (Ed.), Animal Models in Toxicology, second ed. Taylor & Francis Group, LLC Publ., NY, NY, pp. 150–192. 933 pp.
Keenan, K.P., 1996. The uncontrolled variable in risk assessment: ad libitum overfed rodents – fat, facts and fiction. Toxicol. Pathol. 24, 376–383.
Keenan, K.P., Smith, P.F., Hertzog, P., Soper, K., Ballam, G.C., Clark, R.L., 1994. The effects of overfeeding and dietary restriction on Sprague Dawley rat survival and early pathology biomarker of aging. Toxicol. Pathol. 22, 300–315.
Keenan, K.P., Laroque, P., Ballam, G.C., Soper, K.A., Dixit, R., Mattson, B.A., Adams, S.P., Coleman, J.B., 1996. The effects of diet, *ad libitum* overfeeding, and moderate dietary restriction on the rodent bioassay: the uncontrolled variable in safety assessment. Toxicol. Pathol. 24, 757–768.
Keenan, K.P., Ballam, G.C., Soper, K.A., Laroque, P., Colemen, J.B., Dixit, R., 2009. Diet, caloric restriction, and the rodent bioassay. Toxicol. Sci. 52 (2Suppl), 24–34.
King-Herbert, A., Thayer, K., 2006. NTP Workshop: animal models for the NTP rodent cancer bioassay: stocks and strains should we switch? Toxicol. Pathol. 34, 802–805.
King-Herbert, A.P., Sills, R.C., Bucher, J.R., 2010. Commentary: update on animal models for NTP studies. Toxicol. Pathol. 38, 180–181.
Krewski, D., Brown, C., Murdoch, D., 1984. Determining "safe" levels of exposure: safety factors or mathematical models? Fund. Appl. Toxicol. 4, S383–S394.
Krishnan, K., Andersen, M.F., 2007. Physiologically-based pharmacokinetic and toxicokinetic models. In: Hayes, A.W. (Ed.), Principles and Methods of Toxicology, fifth ed. CRC Press, Taylor & Francis, NY, pp. 179–292. 2270 pp.
Lin, K.K., 2000. Progress report on the guidance for industry for statistical aspects of the design, analysis, and interpretation of chronic rodent carcinogenicity studies of pharmaceuticals. J. Biopharm. Stat. 10, 481–501.
MacDonald, J., French, J.E., Gerson, R.J., Goodman, J., Inoue, T., Jacobs, A., Kasper, P., Keller, D., Lavin, A., Long, G., McCullough, B., Sistare, F.D., Storer, R., van der Laan, J.W., 2004. The utility of genetically modified mouse assays for identifying human carcinogens. The Alternatives to Carcinogenicity Testing Committee ILSI-HESI. Toxicol. Sci. 77, 188–194.
Magnuson, B., Williams, G.M., 2008. Carcinogenicity of apartame in rats not proven. Environ. Health Perspect. 116, 239–240. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2430246/?tool=pubmed> (accessed November, 2009).
Magnuson, B.A., Burdock, G.A., Doull, J., Kroes, R.M., Marsh, G.M., Pariza, M.W., Spencer, P.S., Waddell, W.J., Walker, R., Williams, G.M., 2007. Aspartame: a safety evaluation based on current use levels, regulations, and toxicological and epidemiological studies. Crit. Rev. Toxicol. 37, 629–727.
Maltoni, C., Soffritti, M., Belpoggi, F., 1999. The scientific and methodological bases of experimental studies for detecting and quantifying carcinogenic risks. Eur. J. Oncol. 3, 5–10.
McConnell, E.E., 1992. Comparative responses in carcinogenesis bioassays as a function of age at first exposure. In: Guzelian, P.S., Henry, C.J., Olin, S.S. (Eds.), Similarities & Differences Between Children & Adults. ILSI Press, Washington, DC, pp. 66–78. 285 pp.
McConnell, E.E., Solleveld, H.A., Swenberg, J.A., Boorman, G.A., 1986. Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies. J. Nat. Cancer Inst. 76, 283–289.
Melnick, R.L., Thayer, K.A., Bucher, J.R., 2008. Conflicting views arising from the design and evaluation of rodent carcinogenicity studies. Environ. Health Persp. 116, 130–135.
, second ed.Milman, H.A., Haber, E.K. (Eds.), 1994. Handbook of Carcinogen Testing William Andrew Publ., NY. 904 pp.
Mohr, U., 1992–1997. International Classification of Rodent Tumours, Part I: The Rat. 10 Volumes, IARC Scientific Publ. Series 122.
Mohr, U., 2001. International Classification of Rodent Tumours: The Mouse. Springer Publ.. 474 pp.
Molon-Noblot, S., Laroque, P., Coleman, J.B., Hoe, C.M., Keenan, K.P., 2003. The effects of ad libitum overfeeding and moderate and marked dietary restriction on age-related spontaneous pituitary gland pathology in Sprague–Dawley rats. Toxicol. Pathol. 31, 310–320.
Monro, A., 1996. Are lifespan rodent carcinogenicity studies defensible for pharmaceutical agents? Exp. Toxicol. Pathol. 48, 155–166.
Newberne, P.M., Sotnikov, A.V., 1996. Diet: the neglected variable in chemical safety evaluations. Toxicol. Pathol. 24, 746–756.
Nold, J.B., Keenan, K.P., Nyska, A., Cartwright, M.E., 2001. Society of toxicology pathology position paper: diet as a variable in rodent toxicology and carcinogenicity studies. Toxicol. Pathol. 29, 585–586.
NRC (National Research Council), 1993. Use of the maximum tolerated dose in animal bioassays for carcinogenicity. In: Goldstein, B. et al. (Eds.), Issues in Risk Assessment, pp. 15–83 (Cited in Haseman and Lockhart (1994)).
NRC, 1994. Science and Judgement in Risk Assessment. National Academy Press, Washington, DC, 651 pp. Available from: <http://www.nap.edu/openbook.php?isbn=030904894X> (accessed November, 2009).
NRC, 1996. Guide to the Care and Use of Laboratory Animals. Institute of Laboratory Animal Resources (ILAR). National Academy Press, Washington, 125 pp. Available from: <http://www.nap.edu/openbook.php?record_id=5140> (accessed November, 2009).
NTP (National Toxicology Program), 2006. Specifications for the conduct of studies to evaluate the toxic and carcinogenic potential of chemical, biological and physical agents in laboratory animals for the National Toxicology Program (NTP). October 2006. Available from: <http://ntp.niehs.nih.gov/?objectid=72015DAF-BDB7-CEBA- F9A7F9CAA57DD7F5> (accessed October, 2009).

OECD (Organization for Economic Cooperation and Development), 1981. OECD Guideline for Testing of Chemicals. Guideline 451: Carcinogenicity Studies, 15 pp. Adopted May 12, 1991.

OECD, 1998. OECD Series on Application of Principles of Good Laboratory Practice and Compliance Monitoring. OECD Principles on Good Laboratory Practice (as revised in 1997). Document: ENV/MC/CHEM (98)17. Available from: <http://www.olis.oecd.org/olis/1998doc.nsf/LinkTo/env-mc-chem(98)17> (accessed October, 2009).

OECD, 2006. Draft Revised Version of the DRP No.31 on Cell Transformation Assays for Detection of Chemical Carcinogens. Available from: <http://www.oecd.org/LongAbstract/0,3425, en_2649_34377_37863758_1_1_1_1,00.htm> (accessed October, 2009).

OECD, 2009a. OECD Guideline for the Testing of Chemicals. Test Guideline 451: Carcinogenicity Studies. 15 pp. Available from: <http://lysander.sourceoecd.org/vl=1999285/cl=28/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1n4/contp1-1.htm> (accessed October, 2009).

OECD, 2009b. Draft Guidance Document on the Design and Conduct of Chronic Toxicity and Carcinogenicity Studies, Series on Testing and Assessment No. 116. Available from: <http://www.oecd.org/document/55/0,3343,en_2649_34377_2349687_1_1_1_1,00.html> (accessed October, 2009).

Õkva, K., Tamoševičiūtė, E., Čižiūtė, A., Pokk, P., Rukšėnas, O., Nevalainen, T., 2006. Refinements for intragastric gavage in rats. Scand. J. Lab. Animal Sci. 33, 243–252.

Pastoor, T., Stevens, J., 2005. Historical perspective of the cancer bioassay. Scand. J. Environ. Health. 31 (Suppl. 1), 129–140.

Peto, R., Pike, M.C., Day, N.E., Gray, R.G., Lee, P.N., 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. World Health Organization. Cited in Lin (2000) referenced herein.

Portier, C., Hoel, D., 1983. Optimal design of the chronic animal bioassay. J. Toxicol. Environ. Health 12, 1–19.

Portier, C.J., Hoel, D.G., 1984. Design of animal carcinogenicity studies for goodness-of-fit of multistage models. Fund. Appl. Toxicol. 4, 949–959. Cited in Rhomberg et al. (2007) referenced herein.

Pritchard, J.B., French, J.E., Davis, B.J., Haseman, J.K., 2003. The role of transgenic mouse models in carcinogen identification. Environ. Health Persp. 111, 444–454.

Rao, G.N., 1994. Diet for Fischer-344 rats in long-term studies. Environ. Health Perspect. 102, 314–315.

Rao, G.N., Morris, R.W., Seely, J.C., 2001. Beneficial effects of NTP-2000 diet on growth, survival, and kidney and heart diseases of Fischer 344 rats in chronic studies. Toxicol. Sci. 63, 245–255.

Renwick, A.G. Toxicokinetics, p. 214. In: Hayes, A.W. (Ed.), Principles and Methods of Toxicology, fifth ed. CRC Press, Taylor & Francis, NY, p. 2270.

Reuter, J., Hobbelen, H., 1977. The effect of continuous light exposure on the retina in albino and pigmented rats. Physiol. Behav. 18, 939–944.

Rhomberg, L.R., Baetcke, K., Blancato, J., Bus, J., Cohen, S., Conolly, R., Dixit, R., Doe, J., Ekelman, K., Fenner-Crisp, P., Harvey, P., Hattis, D., Jacobs, A., Jacobson-Kram, D., Lewandowski, T., Liteplo, R., Pelkonen, O., Rice, J., Somers, D., Turturro, A., West, W., Olin, S., 2007. Issues in the design and interpretation of chronic toxicity and carcinogenicity studies in rodents: Approaches to dose selection. Crit. Rev. Toxicol. 37, 729–837.

Rice, J.M., 1979. Problems and perspectives in perinatal carcinogenesis: a summary of the conference. Nat. Cancer Inst. Monogr. 51, 271–278.

Riley, V., 1981. Psychoneuroendocrine influences on immunocompetence and neoplasia. Science 212, 1100–1109.

RITA (Registry of Industrial Toxicology Animal data), 2009. Continuous advancement of rodent tumor data acquisition and interpretation. Available from: <http://reni.item.fraunhofer.de/reni/public/rita/> (accessed December, 2009).

Robens, J.F., Calabrese, E.J., Piegorsch, W.W., Schueler, R.L., Hayes, A.W., 1994. Principles of testing for carcinogenicity. In: Hayes, A.W. (Ed.), Principles and Methods of Toxicology, third ed. Raven Press, NY, pp. 697–728 (Chapter 19)

Robinson, D.E., MacDonald, J.S., 2001. Background and framework for ILSI's collaborative evaluation program on alternative models for carcinogenicity assessment. International Life Sciences Institute. Toxicol. Pathol. 29, 13–19. Available from: <http://tpx.sagepub.com/cgi/reprint/29/1_suppl/13> (accessed November, 2009).

Rozman, K.K., 2003. Letter to the editor. Toxicol. Sci. 75, 224–225.

Rozman, K.K., Kerecsen, L., Viluksela, M.K., Österle, D., Deml, E., Viluksela, M., Stahl, B.U., Greim, H., Doull, J., 1996. A toxicologist's view of cancer risk assessment. Drug Metab. Rev. 28, 29–52.

Rushton, B., 1994. Oncogenicity/carcinogenicity. In: Cartwright, A.C., Matthews, B.R. (Eds.), International Pharmaceutical Product Registration. Aspects of Quality, Safety and Efficacy. Ellis Horwood Ltd, NY, pp. 553–568 (Chapter 15).

Seilkop, S.K., 1995. The effect of body weight on tumor incidence and carcinogenicity testing in B6C3F1 mice and F344 rats. Fund. Appl. Toxicol. 24, 247–259.

Soffritti, M., 2008. Carcinogenicity of aspartame: soffritti responds. Environ. Health Perspect. 116, A240–A241. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2430255/> (accessed November, 2009).

Soffritti, M., Belpoggi, F., 2005. Long-term carcinogenicity bioassay to evaluate the potential biological effects, in particular carcinogenic, of aspartame administered in feed to Sprague–Dawley rats (Protocol No.: BT 6008). Unpublished report of the European Foundation of Oncology and Environmental Sciences "B. Ramazzini", December 2005, Bologna, Submitted to EFSA. In EFSA (2006) cited separately herein.

Soffritti, M., Belpoggi, F., Minardi, F., Maltoni, C., 1999. Mega-experiments to identify and assess diffuse carcinogenic risks. Ann. NY Acad. Sci. 895, 34–55.

Soffritti, M., Belpoggi, F., Minardi, F., Maltoni, C., 2002. Ramazzini Foundation Cancer Program. History and major projects, life-span carcinogenicity bioassay design, chemicals studied, and results. Ann. NY Acad. Sci. 982, 26–45.

Soffritti, M., Belpoggi, F., Esposti, D.D., Lambertini, L., 2005. Aspartame induces lymphomas and leukaemias in rats. Eur. J. Oncol. 10, 107–116.

Soffritti, M., Soffritti, M., Belpoggi, F., Tibaldi, E., Esposti, D.D., Lauriola, M., 2006. First experimental demonstration of the multipotential carcinogenic effects of aspartame administered in the feed to Sprague–Dawley rats. Environ. Health Perspect. 114, 379–385.

Soffritti, M., Belpoggi, F., Tibaldi, E., Esposti, D.D., Lauriola, M., 2007. Life-span exposure to low doses of aspartame beginning during fetal life. Environ. Health Perspect. 115, 1293–1297.

Soffritti, M., Belpoggi, F., Esposti, D.D., Falcioni, L., Bua, L., 2008. Consequences of exposure to carcinogens beginning during developmental life. Basic Clin. Pharmacol. Toxicol. 102, 118–124.

Solleveld, H.A., McConnell, E.E., 1985. The value and significance of life span and scheduled termination data in long-term toxicity and carcinogenesis studies. Toxicol. Pathol. 13, 128–134.

Solleveld, H.A., Haseman, J.K., McConnell, E.E., 1984. Natural history of body weight gain, survival, and neoplasia in the F344 rat. J. Natl. Cancer Inst. 72, 929–940.

Sontag, J.M., Page, N.P., Safiotti, U., 1976. Guidelines for carcinogen bioassay in small rodents. DHSS Publication (NIH) 76-801. National Cancer Institute, Bethesda, MD.

Spalding, J.W., French, J.E., Tice, R.R., Furedi-Machacek, M., Haseman, J.K., Tennant, R.W., 1999. Development of a transgenic mouse model for carcinogenesis bioassays: evaluation of chemically induced skin turmor in Tg.AC mice. Toxicol. Sci. 49, 241–254.

Spalding, J.W., French, J.E., Stasiewicz, S., Furedi-Machacek, M., Conner, F., Tice, R.R., Tennant, R.W., 2000. Responses of transgenic mouse lines p53(=/-) and Tg.AC to agents tested in conventional carcinogenicity bioassays. Toxicol. Sci. 53, 213–223.

STP (Society of Toxicologic Pathologists), 1997. Documentation of pathology peer review: position of the Society of Toxicologic Pathologists. Toxicol. Pathol. 6, 655.

STP (Society of Toxicologic Pathologists), 2009. Standardized System of Nomenclature and Diagnostic Criteria (SSNDC) Guides: Series of 31 online publications for guidance on evaluation of animal tumors. Available from: <http://www.toxpath.org/ssndc.asp> (accessed October, 2009).

Suckhow, M.A., Weisbroth, S.H., Franklin, C.L., 2006. The Laboratory Rat, second ed. Academic Press, NY. 912 pp.

Sumner, D.D., Stevens, J.T., 1994. Pharmacokinetic factors influencing risk assessment: saturation of biochemical processes and cofactor depletion. Environ. Health Perspect. 102 (Suppl1), 13–22.

Swenburg, J.A., 1995. Bioassay design and MTD setting: old methods and new approaches. Reg. Toxicol. Pharmacol. 21, 44–51.

Tannenbaum, A., 1940. The initiation and growth of tumors. Introduction. I. Effects of underfeeding. Am. J. Cancer 38, 335–350.

Tannenbaum, A., Vesselinovitch, S.D., Maltoni, C., Mitchell, D.S., 1962. Multipotential carcinogenicity of urethan in the Sprague–Dawley rat. Cancer Res. 22, 1362–1371. Available from: <http://cancerres.aacrjournals.org/cgi/content/abstract/22/11_Part_1/1362> (accessed November, 2009)..

Tennant, R.W., French, J.E., Spalding, J.W., 1993. Identifying chemical carcinogens and assessing potential risk in short-term bioassays using transgenic mouse models. Environ. Health Perspect. 103, 942–950.

Tennant, R.W., Spalding, J., French, J.E., 1996. Evaluation of transgenic mouse bioassays for identifying carcinogens and non carcinogens. Mut. Res. 365, 119–127.

Thompson, K.L., Rosenzweig, B.A., Sistare, F.D., 1998. An evaluation of the hemizygous transgenic Tg. AC mouse for carcinogenicity testing of pharmaceutical. II. A genotypic marker that predicts tumorigenic responsiveness. Toxicol. Pathol. 26, 548–555.

US Animal Welfare Act, 1966. Animal Welfare Act 7 USC, 2131-2159, Animal Welfare Regulations and Final Rules Animal Welfare. Available from: <http://awic.nal.usda.gov/nal_display/index.php?info_center=3&tax_level=3&tax_subject=182&topic_id=1118& level3_id= 6735& level4_id=0& level5_id=0& placement_default=0> (accessed December, 2009).

Van der Laan, J.W., Lima, B.S., Snodin, D., 2002. Alternative models in carcinogenicity testing: a European perspective. Toxicol. Pathol. 30, 157–159. Available from: <http://tpx.sagepub.com/cgi/content/abstract/30/1/157> (accessed November, 2009)..

Vesselinovitch, S.D., Mihailovich, N., 1966. Significance of newborn age and dose of urethane in leukemogenesis. Cancer Res. 26, 1633–1637.

Vesselinovitch, S.D., Mihailovich, N., 1967. The neonatal and infant age period as biologic factors which modify multicarcinogenesis by urethane. Cancer Res. 27, 1422–1429.

Visscher, M.B., Ball, Z.B., Barnes, R.H., Siversten, I., 1942. The influence of caloric restriction upon the incidence of spontaneous mammary carcinoma in mice. Surgery 11, 48–55.

Waddell, W.J., 2002. Thresholds of carcinogenicity of flavors. Toxicol. Sci. 68, 275–279.

Waddell, W.J., 2003a. Thresholds of carcinogenicity in the ED01 study. Toxicol. Sci. 72, 158–163.

Waddell, W.J., 2003b. Thresholds in chemical carcinogenesis: what are animal experiments telling us? Toxicol. Pathol. 31, 260–262.
Waddell, W.J., 2003c. Threshold for carcinogenicity of N-nitrosodiethylamine for esophageal tumors in rats. Food Chem. Toxicol. 41, 739–741.
Waddell, W.J., 2003d. Comparison of human exposures to selected chemicals with thresholds from NTP carcinogenicity studies in rodents. Human Exp. Toxicol. 22, 501–506.
Waddell, W.J., 2004a. Dose–response curves in chemical carcinogenesis. Nonlinearity in Biology, Toxicology, and Medicine 2, 11–20.
Waddell, W.J., 2004b. Analysis of thresholds for carcinogenicity. Toxicol. Lett. 149, 415–419.
Waddell, W.J., 2005. Comparisons of thresholds for carcinogenicity on linear and logarithmic dosage scales. Human Exp. Toxicol. 24, 325–332.
Waddell, W.J., 2006. Critique of dose response in carcinogenesis. Human Exp. Toxicol. 25, 413–436.
Waddell, W.J., 2008. Thermodynamic basis for expressing dose logarithmically. Toxicol. Appl. Pharmacol. 228, 156–157.
Waddell, W.J., 2010. History of dose response. J. Toxicol. Sci. 35, 1–8.
Ward, J.M., 2007. The two-year rodent carcinogenesis bioassay – Will it survive? J. Toxicol. Pathol. 20, 13–19.
Weaver, J.L., Contrera, J.F., Rosenzweig, B.A., Thompson, K.L., Faustino, P.J., Strong, J.M., Ellison, C.D., Anderson, L.W., Prasanna, H.R., Long-Bradley, P.E., Lin, K.K., Zhang, Sistare, F.D., 1998. An evaluation of the hemizigous transgenenic Tg. AC mouse for carcinogenicity testing of pharmaceuticals. I. Evidence for a confounding nonresponder phenotype. Toxicol. Pathol. 26, 532–540.
Weisburger, E.K., 1983. History of the bioassay program of the National Cancer Institute. Prog. Exp. Tumor Res. 26, 187–201.
Weisburger, J.H., Williams, G.M., 1981. Carcinogen testing: current problems and new approaches. Science 214, 401–407.
Weisburger, J.H., Klein, M., Weisburger, E.K., Glass, R.M., Woodward, G., Cronin, M.T.I., 1970. A comparison of the effect of the carcinogen N-hydroxy-N-fluorenylacetamide in infant and weanling rats. J. Nat. Cancer Inst. 45, 29–35.
Wells, M.Y., Spencer Williams, E., 2009. The transgenic mouse assay as an alternative test method for regulatory carcinogenicity studies–implications for REACH. Reg. Toxicol. Pharmacol. 53, 150–155.
White, W.J., Lee, C.S., 1998. The development and maintenance of the Crl:CD®(SD)IGS BR rat breeding system. Charles River Laboratories: CD(SD) IGS 1998; 8–14. Available from: <http://paicriver.com/research_models_and_services/research_models/CD_IGS.html>.
Williams, G.M., 2008. Applicaation of mode-of-action considerations in human cancer risk assessment. Toxicol. Lett. 180, 75–80.
Williams, G.M., Iatropoulos, J.M., Enzmann, H.G., 2008. Principles of testing for carcinogenic activity, pp. 1265–1316. In: Hayes, A.W. (Ed.), Principles and Methods of Toxicology, fifth ed. CRC Press, Taylor & Francis, NY, 2270 pp.