COMMENTS OF THE INTERNATIONAL ANIMAL PROTECTION COMMUNITY ON OPP'S PROPOSED DATA REQUIREMENTS FOR CONVENTIONAL PESTICIDE CHEMICALS

DOCKET NUMBER OPP-2004-0387

OVERVIEW OF POSITIONS

Proposal to Eliminate Requirements

The parties to this submission strongly **support** OPP's proposal to eliminate the following existing data requirements from Part 158:

- Chronic (1-year) toxicity in dogs: This proposal is fully supported by the analysis and conclusions of OPP's *Comparison of the Results of Studies on Pesticides from 12- or 24-Month Dog Studies with Dog Studies of Shorter Duration* (Baetcke et al., 2005) and similar studies carried out by regulatory authorities in Europe (Box and Spielmann, 2005; Spielmann and Gerbacht, 2001).
- Wild mammal acute toxicity testing for aquatic nonfood residential uses.

In addition to deletions proposed by OPP, we support a number of recommendations presented in draft white papers prepared by the ILSI Health and Environmental Science Institute's Technical Committee on Agricultural Chemical Safety Assessment (ACSA). These include:

- Elimination of the 2-year mouse carcinogenicity study (ACSA, 2005a).
- Amalgamation of several reproduction-type studies into a single, more efficient protocol (i.e., F1-extended one-generation study), which could obviate the conduct of separate 2-generation reproduction, rat prenatal developmental, and developmental neuro/immunotoxicity studies (ACSA, 2005b).
- Making better and earlier use of ADME studies and using these data to inform subsequent testing decisions (ACSA, 2005c).

Newly Imposed or Codified Requirements

For the reasons outlined below, we are strongly **opposed** to OPP's proposal to codify the following endpoints, conditionally or otherwise, under Part 158:

- Developmental neurotoxicity
- Acute neurotoxicity – rat
- Immunotoxicity
- Avian acute oral toxicity testing with a passerine species

Proposed Expansion of Existing Requirements

For the reasons outlined below, we are strongly **opposed** to OPP's proposal to expand the application of the following existing data requirements:

- 2-generation reproduction – from CR to R for nonfood uses
- Prenatal developmental toxicity – from a single-species study in rats to a two-species study in rats and rabbits
- 90-day rodent feeding study – from a single-species study in rats to a two-species study in rats and mice
- Primary ocular and dermal irritation and dermal sensitization – from testing of only end-use products to testing of EP and TGAI
- 21- and 90-day dermal – from CR to R for food- and nonfood uses, respectively
- 90-day neurotoxicity – from a CR to R
- Avian reproduction – from CR to R for terrestrial, aquatic, forestry and residential outdoor uses
- Avian acute oral – from testing of TGAI only to TGAI (R) and EP (CR)
- Avian subacute dietary – from CR to R for terrestrial, aquatic, forestry and residential outdoor uses; new CR for EP; new CR for second species when LC_{50} is <500
- Simulated or actual field testing in birds and mammals – from CR to R for all outdoor uses
- Acute toxicity to freshwater fish – from single-species study to a two-species study for greenhouse and indoor uses
- Acute toxicity to estuarine and marine fish – from CR to R for outdoor (except residential nonfood aquatic) uses
- Chronic toxicity to fish – from CR to R for outdoor (except residential nonfood aquatic) uses

STATUTORY BACKGROUND

Mandate to Protect Human Health & the Environment

The sale and use of pesticides is closely regulated by the detailed requirements of two federal statutes and the regulations promulgated there under. First, in order to be sold or distributed in the United States, a pesticide must be registered pursuant to the provisions of the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), 7 U.S.C. §§ 136-136y. In order to register a pesticide, FIFRA requires the EPA Administrator to determine that "it will perform its intended function without unreasonable adverse effects on the environment." 7 U.S.C. § 136a(c)(5)(C). FIFRA defines "unreasonable adverse effects on the environment" as including "a human dietary risk from residues that result from a use of a pesticide in or on any food inconsistent with the standard under section 346a of Title 21 [the Food Quality Protection Act (FQPA)]." 7 U.S.C. § 136(bb). Thus, to be registered under FIFRA, a pesticide use must meet the safety standard of the FQPA.

The FQPA provides that "the Administrator may establish or leave in effect a tolerance for a pesticide chemical residue in or on a food only if the Administrator determines that the tolerance is safe." 21 U.S.C. § 346a(b)(2)(A)(i). The FQPA further provides that "[t]he term 'safe,' with respect to a tolerance for a pesticide chemical residue, means that the Administrator has determined that there is a reasonable certainty that no harm will result from aggregate exposure to the pesticide chemical residue, including all anticipated dietary exposures and all other exposures for which there is reliable information." 21 U.S.C. § 346a(b)(2)(A)(ii).

The FQPA further provides that, "[i]n establishing, modifying, leaving in effect or revoking a tolerance or exemption for a pesticide chemical residue, the Administrator shall consider, among other relevant factors . . . (v) available information concerning the cumulative effects of such residues and other substances that have a common mechanism of toxicity; (vi) . . . exposure from other non-occupational sources; . . .[and] (vii) such information as the Administrator may require on whether the pesticide chemical may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen or other endocrine effects." 21 U.S.C. § 346a(b)(2)(D).

The FQPA also includes specific requirements for EPA to consider the special susceptibility of infants and children when establishing, modifying, revoking or leaving in effect a pesticide tolerance. The Administrator "shall assess the risk of the pesticide chemical residue based on – (I) **available information** about consumption patterns among infants and children that are likely to result in disproportionately high consumption of foods containing or bearing such residue among infants and children in comparison to the general population; (II) **available information** concerning the special susceptibility of infants and children to the pesticide chemical residues, including neurological differences between infants and children and adults . . . ; and (III) **available information** concerning the cumulative effects on infants and children of such residues and other substances that have a common mechanism of toxicity." 21 U.S.C. § 346a(b)(2)(C) (emphasis supplied). The Administrator "shall ensure that there is a **reasonable certainty** that no harm will result to infants and children from aggregate exposure to the pesticide chemical residue." Id. (emphasis supplied).

To further account for the potential special toxicity of a chemical to infants and children, the FQPA provides that "an additional tenfold margin of safety for the pesticide chemical residue and other sources of exposure shall be applied for infants and children to take into account potential pre- and post-natal toxicity and completeness of the data with respect to exposure and toxicity to infants and children. Notwithstanding such requirement for an additional margin of safety, the Administrator may use a different margin of safety for the pesticide chemical residue only if, on the basis of **reliable data**, such margin will be safe for infants and children." 21 U.S.C. § 346a(b)(2)(C) (emphasis supplied).

Mandate to Develop & Utilize Alternatives to Animal Testing

The 1993 National Institutes of Health (NIH) Revitalization Act, 42 U.S.C.A. §283e, directs the NIH, through an Interagency Coordinating Committee on the Use of Animals in Research, to prepare a plan to conduct or support research into methods of research that "do not require the use of animals," that "reduce the number of animals used in such research," that encourage the "acceptance by the scientific community" of alternative methods, and that trains "scientists in the use of such methods." 42 U.S.C. §283e. It is clear from the language of the statute that Congress intended for the EPA to be an active contributor to development and implementation of the above-mentioned plan, as involvement by "representatives of the Environmental Protection Agency…" on the interagency committee is a specific requirement under the NIH Revitalization Act.

Similarly, the central aim of the ICCVAM Authorization Act of 2000, 42 U.S.C. §2851 *et seq*, is "to establish, wherever feasible, guidelines, recommendations, and regulations that promote the regulatory acceptance of new or revised scientifically valid toxicological tests that protect human and animal health and the environment while reducing, refining, or replacing animal tests and ensuring human safety and product effectiveness." In establishing ICCVAM as a permanent Committee, Congress signaled its firm commitment to the promotion and advancement of alternatives to animal-based toxicity testing.

Mandate to Ensure Test Method Validation & Data Quality

The ICCVAM Authorization Act also reflects Congress' intent that "each federal agency … shall ensure that any new or revised acute or chronic toxicity test method, including animal test methods and alternatives, is determined to be valid for its proposed use prior to requiring, recommending, or encouraging the application of such test method." 42 U.S.C. §2851-4(c). "Validation" is defined by ICCVAM and its international counterparts as "the process by which the reliability and relevance of a procedure are established for a particular purpose" (ICCVAM, 2003). The EPA's own Scientific Advisory Panel (SAP) has likewise emphasized, "that any new test guideline adopted by the Agency requires validation. This entails producing reproducible results among laboratories and selection of endpoints in test animals that are applicable to humans" (SAP, 1998).

The Data Quality Act (DQA), 44 U.S.C. §3516 reflects Congress' further requirement that governmental agencies ensure the "quality, objectivity, utility, and integrity of information disseminated by the agency…." The DQA's Objectivity Standard, for example, requires the EPA to ensure that information it disseminates is "accurate, reliable, and unbiased" (OMB, 2002; EPA, 2002a). The EPA and most other federal agencies have established a government-wide data quality standard that requires proper validation of toxicity test methods before data from such tests may be considered reliable:

> Before a new or revised test method is used to generate information to support regulatory decisions, it must be…validated to determine its reliability and relevance for its proposed use…. (ICCVAM, 2003).

Moreover, for Influential Scientific Information, such as data from regulatory toxicity studies, the EPA's Data Quality Guidelines require that the agency "ensure reproducibility for disseminated original and supporting data according to commonly accepted scientific, financial, or statistical methods" (EPA, 2002a). However, the EPA cannot ensure reproducibility of original and supporting data from toxicity studies until and unless these data are generated by reliable tests, and validation is necessary to ensure reliability (ICCVAM, 2003).

The EPA is further obligated to adhere to the DQA's Utility Standard, which requires that information disseminated by a federal agency be useful to its intended users, including the public (OMB, 2002; EPA, 2002a). Information from toxicity tests is not useful when it is generated from non-validated, and therefore potentially unreliable and irrelevant tests.

Commitment to the Mutual Acceptance of Data

The testing of chemicals is labor-intensive and expensive. Often the same chemical is tested and assessed in several countries. Because of the need to relieve some of this burden, the Organization for Economic Cooperation and Development (OECD)—an economic alliance of 30 member countries, including the United States—adopted a Decision in 1981 stating that:

> Data generated in a member country in accordance with OECD Test Guidelines and Principles of Good Laboratory Practice (GLP) shall be accepted in other member countries for assessment purposes and other uses relating to the protection of human health and the environment.

The OECD is not a supranational organization. However, if member countries consider it appropriate, an accord can be embodied in a formal OECD Council Act, which is agreed to at the highest level of OECD, the Council. In general, there are two types of Council Act. A Council Decision, which is legally binding on OECD member countries, and a Council Recommendation, which is a strong expression of political will.

The statement above on Mutual Acceptance of Data (MAD) is a legally binding OECD Council Decision (OECD, 1981). Thus, all OECD member countries—including the US—are obligated to accept data generated in accordance with an OECD Test Guideline. For the purposes of these comments, refusal by the EPA to accept non-animal data generated in accordance with an OECD Test Guideline would constitute a breach of the US' commitments and obligations pursuant to MAD.

FACTUAL BACKGROUND

Costs of Current & Proposed Data Requirements

Upwards of 10,000 rats, mice, fish, birds, rabbits, dogs and other vertebrate animals may be consumed in toxicity studies in order to satisfy *current* OPP registration data requirements for a single food-use chemical active ingredient (Mattsson et al., 2003). In addition, an undetermined number of additional animals are used in "special studies" requested by OPP or conducted voluntarily by registrants (e.g., dose-setting studies, etc.). Thus, in practical terms, OPP's current paradigm "seeks to generate all data that might be required without necessarily taking into consideration the use profile or toxicity" and "it predefines the studies without detailed consideration of the risk assessments needed or effects of concern" (ACSA, 2005b).

Despite the recognized redundancies and overall inefficiency of the current paradigm, OPP's proposed amendments do more to exacerbate these problems than to remedy them. Table 1 outlines the numerous toxicity endpoints considered in pesticide hazard assessments and the average cost (economic and animal welfare) concomitant to a standard guideline study in the context of current and proposed Part 158 data requirements for food- and nonfood-use pesticides. Many of the studies associated with proposed new or additional data requirements each cost hundreds of thousands of dollars and consume hundreds or thousands of animals. Thus, the cost of requiring a 2-generation reproduction study for all nonfood-use chemicals, a developmental neurotoxicity study on a conditional basis, and rabbit developmental, mouse subchronic, acute neurotoxicity and immunotoxicity studies for all pesticide chemicals—to say nothing of the laundry list of other proposed additions—will soar to unprecedented highs, which we believe are grossly underestimated in OPP's proposal. These proposals stand in stark contrast to the EPA's mandate to minimize its reliance on animal testing pursuant to the NIH Revitalization Act and the ICCVAM Authorization Act.

Table 1—Technical Grade of the Active Ingredient (TGAI)

Endpoint	OPPTS guideline	Avg. # of animals/test[1]	Avg. cost/ test[2-3]	Part 158 req's[4]	
				current[3]	proposed
Acute systemic toxicity—oral	870.1100	7 rats	$4,200	R/R	R/R
Acute systemic toxicity—inhalation	870.1300	20 rats	$16,000	R/R	R/R
Acute systemic toxicity—dermal	870.1200	20 rabbits	$6,500	R/R	R/R
Acute eye irritation	870.2400	3 rabbits	$1,800	–/–	R/R
Acute skin irritation	870.2500	3 rabbits	$1,800	–/–	R/R
Skin sensitization	870.2600	16 mice or 32 guinea pigs	$6,000	–/–	R/R
Acute neurotoxicity	870.6200	80 rats	$89,600	–/–	R/R
Acute delayed OP neurotoxicity	870.6100	24 hens	$79,375	CR/CR	CR/CR

Table 1, cont'd

Subacute (21/28 d) toxicity—dermal	870.3200	40 rats	$83,250	CR/–	R/–
Subchronic (90 d) feeding—rat	870.3100	120 rats	$95,000	R/R	R/R
Subchronic (90 d) feeding—mouse	870.3100	120 mice	$95,000	–/–	R/R
Subchronic (90 d) feeding—non-rodent	870.3150	32 dogs	$130,000	R/R	R/R
Subchronic (90 d) toxicity rodent—dermal	870.3250	120 rats	$137,000	–/CR	–/R
Subchronic (90 d) toxicity rodent—inhalation	870.3465	120 rats	$205,000	CR/CR	CR/CR
Delayed (28 d) OP neurotoxicity	870.6100	40 hens	$79,375	CR/CR	CR/CR
Subchronic (90 d) neurotoxicity	870.6200	80 rats	$184,000	CR/CR	R/R
Chronic toxicity/ carcinogenicity—rat	870.4300	400 rats	$700,000	R/CR	R/CR
Chronic (1-year) toxicity—non-rodent	870.4100	32 dogs	$350,000	R/CR	–/–
Carcinogenicity—mouse	870.4200	400 mice	$600,000	R/CR	R/CR
Reproductive toxicity in 2 generations	870.3800	2,640 rats	$378,500	R/CR	R/R
Developmental toxicity—rodent	870.3700	1,300 rats	$77,000	R/R	R/R
Developmental toxicity—non-rodent	870.3700	900 rabbits	$77,000	CR/CR	R/R
Developmental neurotoxicity	870.6300	1,300 rats	$407,000	–/–	CR/CR
Mutagenicity—bacterial reverse mutation (Ames)	870.5100	0	$3,500	R/R	R/R
Mutagenicity—cell gene mutation or *in vitro* chromosomal aberration	870.5300 870.5375	0	$18,000	R/R	R/R

Table 1, cont'd

Mutagenicity—*in vivo* chromosomal aberration or mouse micronucleus	870.5385 870.5395	80 rodents	$15,000[4]	R/R	R/R
Mutagenicity—UDS *in vitro*	870.5550	0	$11,000	CR/CR	CR/CR
Mutagenicity—SCE *in vitro*	870.5900	0	$8,000	CR/CR	CR/CR
Mutagenicity—dominant lethal	870.5450	80 rodents	$32,000	CR/CR	CR/CR

General metabolism	870.7485	4-32 rats	$31,650[4]	R/R	R/CR
Dermal penetration	870.7600	96 rats	$	CR/CR	CR/CR
Immunotoxicity	870.7800	32 rats	$56,700	–/–	R/R
Scheduled controlled operant behavior	870.6500	40 rats	$164,000	CR/CR	CR/CR
Peripheral nerve function	870.6850	40 rats	$110,000	CR/CR	CR/CR
Neurophysiology: sensory evoked pot.	870.6855	40 rats	$110,000	CR/CR	CR/CR
Companion animal safety	870.7200	variable	variable	CR/CR	CR/CR

Avian oral LC_{50}—mallard	850.2100	60 birds	$10,000	R/R	R/R
Avian oral LC_{50}—bobwhite	850.2100	60 birds	$10,000	R/R	R/R
Avian oral LC_{50}—redwing blackbird	850.2100	60 birds	$10,000	–/–	R/R
Avian dietary LC_{50}—mallard	850.2200	80 birds	$6,500	R/R	R/R
Avian dietary LC_{50}—bobwhite	850.2200	80 birds	$6,500	R/R	R/R
Avian reproduction––mallard	850.2300	1,450 birds	$168,250	CR/CR	R/R
Avian reproduction––bobwhite	850.2300	1,450 birds	$168,250	CR/CR	R/R
Wild mammal toxicity	850.2400	variable	variable	CR/CR	CR/CR
Simulated or actual field testing	850.2500	variable	$650,000	CR/CR	CR/CR

Table 1, cont'd

Freshwater fish LC_{50}—rainbow trout	850.1075	30-120 fish	$17,000	R/R	R/R
Freshwater fish LC_{50}—bluegill	850.1075	30-120 fish	$17,000	R/R	R/R
Estuarine fish LC_{50}	850.1075	30-120 fish	$17,000	R/R	R/R
Fish early life stage—freshwater	850.1400	360 fish	$37,250	CR/CR	CR/CR
Fish early life stage—saltwater	850.1400	360 fish	$37,250	CR/CR	CR/CR
Fish life-cycle	850.1500	360 fish	$37,250	CR/CR	CR/CR
Fish bio-[ ]	850.1730	12 fish	$39,000	CR/CR	CR/CR
Simulated or actual field testing for aquatic organisms	850.1950	variable	$512,500	CR/CR	CR/CR

[1] Calculations based on OPPTS (2005)
[2] BEAD (2004)
[3] Derelanko and Hollinger (2002)
[4] R = unconditional requirement; CR = conditional requirement. The two entries within each cell in the latter two columns reflect data requirements for food- and nonfood-use chemicals, respectively (i.e., R/CR = required for food-use and conditionally required for nonfood-use)
[5] EPA (2002d)

Table 2—Formulated End-Use Product (EP)

Endpoint	OPPTS guideline	Avg. # of animals/test[1]	Avg. cost/ test[2]	Part 158 req's[3] current	proposed
Acute systemic toxicity—oral	870.1100	7 rats	$4,200	R	R
Acute systemic toxicity—inhalation	870.1300	20 rats	$16,000	R	R
Acute systemic toxicity—dermal	870.1200	20 rabbits	$6,500	R	R
Acute eye irritation	870.2400	3 rabbits	$1,800	R	R
Acute skin irritation	870.2500	3 rabbits	$1,800	R	R
Skin sensitization	870.2600	16 mice or 32 guinea pigs	$6,000	R	R

[1] Calculations based on OPPTS (2005)
[2] Derelanko and Hollinger (2002)
[3] R = unconditional requirement; CR = conditional requirement

Criticisms of Some Animal-Based Toxicity Studies

The relevance to humans of laboratory tests on rodents and other animals is the subject of much controversy, due to the myriad of biological differences that exist between animal species, as well as methodological issues such as chemical dosing, behavioral measures, etc. (Derelanko & Hollinger, 2002). Despite the requirements of the ICCVAM Authorization Act and DQA regarding the assurance of test method validation and data quality, virtually none of the animal-based toxicity tests referenced in the current Part 158 regulations—nor those in the proposed amendments—have been properly validated to demonstrate their reliability and relevance as predictors of human health or population-level ecological effects. The following are but a handful of the many available examples that highlight the dubious relevance and reliability of animal-based toxicity tests, the results of which the EPA uses as a basis for making statutorily mandated safety findings for registered pesticides (i.e., "reasonable certainty of no harm"):

- **Acute Systemic Toxicity:** Acute toxicity, or "lethal-dose," studies date back to the World War I era (Trevan, 1927), yet have never been formally or adequately validated according to modern standards (OECD, 1996; ICCVAM, 2003). Ekwall and colleagues (1998) examined the results of rat and mouse acute toxicity tests for 50 chemicals and found a relatively poor concordance between LD_{50} data for rats and mice ($R^2 = 0.61$)—two closely related species—as well as between rodent LD_{50} data and peak human lethal blood concentrations ($R^2 = 0.65$). Similar examples of species and strain variations in the toxic effects of chemicals have been present in the toxicological literature for decades. In 1945, a study showed that the lethal dose for the chemical thiourea in the Hopkins strain of rat was 4 mg/kg, but in the wild Norway rat was 1,830 mg/kg (Dieke SH and Richter, 1945).

  A report published in 1948 found large differences in chemical toxicities to be typical, with human sensitivities to some chemicals being up to 2,000 times greater than in animals (Müller, 1948). Similar examples abound (Morrison et al., 1968; Goldenthal, 1971; Jollow et al., 1974; Paxton, 1995; Himmelstein et al., 1996; Quick and Shuler, 1999; Olson et al., 2000; Rosenkranz and Cunningham, 2005). Zbinden and Flury-Roversi (1981) therefore concluded: "For the recognition of symptomatology of acute poisoning in man, and for the determination of the human lethal dose, the LD_{50} in animals is of very little value." Lorke (1983) added: "… even if the LD_{50} could be measured exactly and reproducibly, the knowledge of its precise numerical value would barely be of practical importance, because an extrapolation from the experimental animals to man is hardly possible." In addition to their marginal scientific value, acute lethality studies have been uniformly condemned for the suffering they inflict upon animals (Zbinden et al., 1981; Balls, 1991).

- **Acute Dermal Irritation:** Despite 60 years of use in regulatory toxicology (Draize et al., 1944), animal-based skin irritation studies have likewise never been properly validated by modern standards. For example, a comparison of data from rabbit tests and four-hour human skin patch tests for 65 substances found that 45 percent of classifications of chemical irritation potential based on animal tests were incorrect (Robinson et al., 2002). Studies reporting similarly of poor inter-species comparability of skin responses have been available in the scientific literature for

decades (Carter and Griffith, 1965; Weil and Scala, 1971; Phillips et al., 1972; Davies et al.,1972; Nixon et al., 1975).

- **Developmental Toxicity:** A single rodent developmental toxicity study consumes an average of 1,300 animals at a cost of US$50,000, while the same study in rabbits consumes approximately 900 animals at a cost of $77,000 (OPPTS, 2005; BEAD, 2004). Since the thalidomide incident in the early 1960s, testing for developmental defects has been a routine component of hazard assessments of chemicals to which humans are highly exposed. By the end of the 1990s, more than 1,200 substances had been identified as developmental toxicants in one or more animal species (Shepard, 1998). However, only about 50 (4.2%) of these can be clearly linked to developmental defects in humans (Schardein, 2000; Shepard, 1998; Friedman and Polifka, 1994).

  The precise reason(s) for this apparent lack of specificity and high false-positive rate are unclear. The National Research Council (NRC) Committee on Developmental Toxicology noted that "all too frequently the focus of developmental toxicity testing has been to study the effects of an agent only at high doses that are most likely irrelevant to environmental and occupational exposures. For industrial and environmental chemicals, the dosing regimens at or even above MTDs, as applied in hazard identification studies, typically contrast sharply with anticipated human exposures that are commonly much lower in extent or magnitude, often uncertain, or even entirely unknown" (CDT, 2000). However, even at identical exposure levels, evidence exists that developmental toxicity studies in rats and rabbits may falsely implicate substances as developmental hazards to humans (Table 3).

Table 3—Substances with No Positive Report of Developmental Toxicity in Humans: A Comparison of Human Usage Levels and Doses Found Teratogenic in Animals (Derelanko and Hollinger, 2002)

Substance	Human exposure	Minimum embryotoxic dose		Approx. ratio
		Rat	Rabbit	
1-Asparaginase	50-200 IU	100 IU	500 IU	1:1
Actinomycin D	0.015 mg/kg	0.025-0.1 mg/kg	–	1:2
Caffeine	20-25 mg/kg	75-100 mg/kg	100 mg/kg	1:3
Colchicine	0.06 mg/kg	–	0.1-0.5 mg/kg	1:1.5
5-Fluorouracil	12 mg/kg	12-37 mg/kg	–	1:1
Hydroxyurea	20-30 mg/kg	137 mg/kg	–	1:2.5
Procarbazine	2-6 mg/kg	5-10 mg/kg	–	1:1
Prochlorperazine	0.6-1.2 mg/kg	2.5-10 mg/kg	–	1:2
Sodium salicylate	2 mg/kg	10 mg/kg	–	2:1

Griffin (1997) cautioned that rodents and rabbits may be insensitive indicators of certain other developmental hazards to humans, possibly leading to false-negative results. The EPA clearly shares this concern, as evidenced by the fact that OPP data requirements specify that developmental toxicity studies are to be carried out in two species: a rodent and a non-rodent. However, the inability to settle on a single "model" only underscores the profound differences that exist between animal species and the fact that no amount of animal testing will ever provide a clear and accurate prediction of toxicity to humans. In fact, the more species that are tested, the greater the discordance and the disarray in the data. Tables 4–6 below summarize the developmental responses of 8 animal species to a range of pesticides, painkillers and chemotherapeutic drugs, respectively.

Table 4—Discordance in Developmental Toxicity Classifications Among Animal Species for Pesticides (Schardein, 2000)

Substance	Species							
	Rat	Mouse	Rabbit	Hamster	NHP	Dog	Cat	GP
Captan	–	–	±	+		±		
Imidazolidine-thione	+	–	–	+			+	–
Thiram	–	±	+	+				
2,4,5-T	±	+	–	±	–			
TCDD (dioxin)	±	+	+	+	–			
Butiphos	–		+					
Chloridazon	–			+				
Dichlorprop	–	+						
Ethephon	–		–	+				
Linuron	+		–					
Carbaryl	–	±	±	–	–	+		±
Carbofuran	–	+	–			–		
Diazinon	+	–	–	–		+		
Dieldrin	±	+	–	+		–		
Dimethoate	+	–						+
Malathion	+		–					
Pentachloro-phenol	±	–		–				
Phosmet	+	+	–					

Legend: (+) = positive finding; (-) = negative finding; (±) = equivocal finding

Table 5—Discordance in Developmental Toxicity Classifications Among Animal Species for Anesthetic and Analgesic Drugs (Schardein, 2000)

Substance	Species							
	Rat	Rabbit	Mouse	NHP	Hamster	GP	Cat	Dog
Acetylsalicylic acid	+	±	+	±	–	–	+	+
Aminopyrine	±	–	+					
Codeine	–	–	±		±			
Dimethyl Sulfoxide	±	–	+		+			
Halothane	±	–	+					
Methyl salicylate	±	+	+		+			
Morphine	–	–	+		+			
Nitrous oxide	+	–	–		–			
Propanidid	–							+
Sodium salicylate	+	–	+			–		+

Legend: (+) = positive finding; (-) = negative finding; (±) = equivocal finding

Table 6—Discordance in Developmental Toxicity Classifications Among Animal Species for Chemotherapeutic Drugs (Schardein, 2000)

Substance	Species							
	Rat	Dog	Mouse	Rabbit	Pig	NHP	Cat	Hamster
Alloxan	+	–	+	+	+			
Aminopterin	+	+	–		+	–	–	
Azaserine	+	–						+
Azauridine	±		±	–				
Bleomycin	+		+	–				
Cactinomycin	+		–	–				
Daunorubicin	+		–	–				–
Demecolcine	–		±	±		–		
Diazo oxo norleucine	+	+	–	–				
Doxorubicin	+		–	–				

Methotrexate	+	–	+	+		±	+	
Mitotane	–	–	+					
Pentostatin	+		–	–				
Thiamiprine	–			+		–		
Vinblastine	+		±	+				+
Vincristine	+		+	+		±		+

Legend: (+) = positive finding; (-) = negative finding; (±) = equivocal finding

Bailey et al. (2005) summarized the state of the science for developmental toxicity as follows: "Mean positive and negative predictivities barely exceed 50%; discordance among the species used is substantial; reliable extrapolation from animal data to humans is impossible, and virtually all known human teratogens have so far been identified in spite of, rather than because of, animal-based methods."

- **Avian Reproductive Toxicity:** A one-generation avian reproduction study consumes at least 1,450 ducks and/or quail (OPPTS, 2005), while a two-generation study could consume as many as 5,500 birds (Fischer, 2003). Despite the exorbitant cost of these studies—both economic and in terms of animal utilization—their use tends to be limited to "risk characterization" rather than as a basis for regulatory restrictions on environmental exposure. Moreover, avian reproduction studies have never been properly validated by modern standards (OECD, 1996; ICCVAM, 2003). In fact, there is ample evidence in the published literature to suggest that the species commonly used in avian reproduction studies may not be representative of the thousands of species of wild birds who may be adversely impacted by environmental chemicals (Baert and De Backer, 2003; Mineau et al., 1994; Schaeffer and Beasley, 1989; Hill and Murray, 1987; Sefarin, 1984).

- **Carcinogenicity:** Following the announcement of the "war on cancer" in the 1970s, testing began in earnest to detect carcinogenic and mutagenic properties of pesticides, drugs and other synthetic substances. Since that time, the US National Toxicology Program (NTP) alone has conducted more than 500 such studies (Gaylor, 2005; Huff et al, 1991; Haseman, 2000), each consuming approximately 400 rats and 400 mice (OPPTS, 2005) at a cost of $1.3 million (Derelanko and Hollinger, 2002). Yet despite more than three decades of use in regulatory toxicology, rodent cancer "bioassays" have never been properly validated according to modern standards (OECD, 1996; ICCVAM, 2003); thus, interpretation of study results is often fraught with controversy. A case in point: saccharine; however, similar examples abound (e.g., Trosko and Brad, 2005; Ennever and Lave, 2003).

  A valid test is one which produces results that are reproducible and relevant to the biological effect of interest in the species of interest—in this case, humans (ICCVAM, 2003). Pritchard and colleagues (2002) evaluated the predictivity of a number of conventional and proposed "alternative" carcinogenicity assays using a database of 99 chemicals that had been tested by the NTP in each of these models.

For comparison, results from conventional rodent bioassays were analyzed against NTP classifications, yielding consistent (reliable) results in only 69% of cases (Table 7). Similar findings were reported by Gottmann et al. (2001), who compared 121 replicate rodent cancer bioassays to estimate the reliability of these experiments, reporting:

> We estimated a concordance of 57% between the overall rodent carcinogenicity classifications from both sources [NTP/National Cancer Institute and the open literature]. This value did not improve substantially when additional biological information (species, sex, strain, target organs) was considered. These results indicate that rodent carcinogenicity assays are much less reproducible than previously expected…

Table 7 (reproduced from Pritchard et al., 2002)

Strategy	Positive for carcinogens	Negative for noncarcinogens	Positive for noncarcinogens	Negative for carcinogens	Overall accuracy
Trp53+/–	21	27	1	10	81% (48/59)
Trp53+/– (genotoxic)	16	6	0	4	85% (22/26)
Tg.AC	17	29	10	6	74% (44/62)
RasH2	21	18	5	7	76% (39/51)
Trp53+/– (genotoxic); RasH2 (nongenotoxic)	17	18	3	6	80% (35/44)
Trp53+/– (genotoxic); RasH2 (all)	30	14	5	4	83% (44/53)
Trp53+/– (genotoxic); Tg.AC (nongenotoxic)	21	23	3	6	83% (44/53)
Trp53+/– (genotoxic); Tg.AC for all	25	22	10	4	77% (47/61)
NTP rodent bioassay	23	17	18	0	69% (40/58)
NTP rat bioassay; Tg.AC (nongenotoxic); Trp53+/– (genotoxic)	35	13	9	0	84% (48/57)
NTP rat bioassay; RasH2 (nongenotoxic); Trp53+/– (genotoxic)	33	12	8	0	85% (45/53)
NTP rat bioassay; genotoxicity	36	7	23	0	65% (43/66)

Definitions: positive for carcinogens, positive assay results for IARC/ROC carcinogens; negative for noncarcinogens, negative assay results for IARC/ROC noncarcinogens; positive for noncarcinogens, positive assay results for IARC/ROC noncarcinogens; negative for carcinogens, negative assay results for IARC/ROC carcinogens.

In their article "Chemical carcinogenesis: too many rodent carcinogens," Ames and Gold (1990) note:

> The administration of chemicals at the maximum tolerated dose (MTD) in standard animal cancer tests is postulated to increase cell division (mitogenesis), which in turn increases rates of mutagenesis and thus carcinogenesis. The animal data are consistent with this mechanism, because a high proportion—about half—of all chemicals tested (whether natural or synthetic) are indeed rodent carcinogens. We conclude that at the low doses of most human exposures, where cell killing does not occur, the hazards to humans of rodent carcinogens may be much lower than is commonly assumed.

Ennever and Lave (2003) likewise contend that rodent cancer bioassays produce an unacceptably large number of false-positive results. One literature review found that 19 out of 20 probable human non-carcinogens tested positive in rodent bioassays, implying that the specificity of these assays may be as low as 0.05 (Ennever et al., 1987). Similarly, Gold et al. (1984; 1986) found that rodent studies identified nearly two-thirds of the 800 chemicals tested as carcinogens. Even the NTP's own analyses reveal an implausibly high proportion of positive results (Haseman, 1983). For example, of 471 NTP studies 250 (53%) produced carcinogenic effects in at least one sex/species group (Haseman, 2000). Similar findings were reported by Fung et al. (1995), who examined 400 chemicals tested by the NTP, noting that:

> Of these chemicals, 210 (52%) induced carcinogenicity in at least one organ of one sex of one species of the four sex/species groups typically used by NCI/NTP. Only 92 of the 400 chemicals (23%) were positive in two species and thus by international criteria are considered likely to pose a carcinogenic hazard to humans. A total of 267 chemicals (67%) were selected as suspect carcinogens, and 187 (68%) of these were carcinogenic. Suspect chemicals account for 86% of chemicals with at least one positive result and account for 90% of chemicals considered positive in two species. The International Agency for Research on Cancer (IARC) lists only 5 of the 400 chemicals as carcinogenic to humans (group 1) and 10 as probably carcinogenic to humans (group 2A). The majority (80%) of the 133 chemicals selected only on production/ exposure considerations were not carcinogenic in animals, even when tested at the maximum tolerated (or minimally toxic) dose. Only 9 (6.8%) were positive in two species, and none is listed in IARC groups 1 or 2A.

Such findings led Ashby and Purchase (1993) to question: "Will all chemicals be carcinogenic to rodents if adequately evaluated?" The short answer appears to be "no." Despite generating an implausibly high proportion of false-positive results, rat and mouse cancer bioassays produce concordant results (both positive and negative) generally less than 70 percent of the time (Byrd et al., 1990; Zeiger, 1987). Given that rats and mice are more biologically similar to one another than either is to humans, it is reasonable to assume that rodent-human concordance would be less than 70 percent. For example, rodents used in these studies are well known to have high spontaneous incidence rates for many of the tumor types commonly observed in a bioassay (Haseman, 2000; Kuschner, 1995; Huff et al., 1991; Kodell et al., 1989; Lave et al., 1988). In addition to the potential confounding effects of this background "noise," many of the mechanisms of action (e.g., peroxisome proliferation) and tumor types (e.g., kidney, forestomach, gastric neuroendocrine, thyroid and urinary bladder) observed in rodents are now openly recognized as being irrelevant to humans, according to the international consensus reports of the International Agency for Research on Cancer (IARC, 1995; 1999; 2003).

Notwithstanding the difficulties associated with false-positive results, the problem of false negatives is of greater concern from a public health perspective. Salsburg (1983) reported that rodent bioassays were capable of identifying only 37 percent of known human carcinogens, which is a worrisome underestimate. More recently,

Johnson (2001) identified only 10 known human carcinogens among the hundreds of chemicals tested by the NTP. Of these, only three of the six substances tested in both species caused cancer in both, and one caused cancer in neither (Table 8).

Table 8 (reproduced from Johnson, 2001)

Known human carcinogen	Carcinogenicity in rats	Carcinogenicity in mice
Thiotepa	+	+
Benzene	+	+
Benzidine and dyes	+	+
1,3-Butadiene	Not tested	+
Ethylene oxide	Not tested	+
8-Methoxypsoralin	+	Not tested
Nickel compounds	+	–
Asbestos	+	Not tested
Talc	+	–
Aspirin/phenacetin/ caffeine (APC)	–	–

Given the serious limitations associated with rodent cancer assays, data from these studies on their own are seldom regarded as sufficient evidence of carcinogenic hazard to humans (Bucher and Portier, 2004; Festing, 1997). This is clearly demonstrated in an analysis by Knight (unplublished data) of EPA carcinogenicity classifications for160 chemicals listed in the IRIS database. In the majority (58.1%) of instances, the Agency considered animal data inadequate to support a classification of probable human carcinogen or non-carcinogen (Table 9). Similarly, for the 128 chemicals with human or animal data also assessed by IARC, Knight reports that human carcinogenicity classifications were compatible with EPA classifications only for those 17 substances possessing at least limited human data (Table 10).

Table 9 (reproduced from Knight, unpublished)

EPA human carcinogenicity classification	**No. of chemicals**	**% of total (235)**
A: Human carcinogen *(convincing human data)*	11	4.7
B1: Probable human carcinogen *(limited human data)*	6	2.6
B2: Probable human carcinogen *(sufficient animal data)*	64	27.2
C: Possible human carcinogen *(animal data inadequate for stronger classification)*	40	17.0
D: Unclassifiable *(animal data inadequate for stronger classification)*	53	22.6
D: Unclassifiable *(no animal or human data)*	58	24.7
E: Probable human non-carcinogen *(sufficient animal data)*	3	1.3
Total	**235**	**100.1**

Table 10 (reproduced from Knight, unpublished)

IARC human carcinogenicity classification		No. of chemicals	% of total (885)
1: Definite Human Carcinogen		88	9.9
2A: Probable Human Carcinogen		64	7.2
2B: Possible Human Carcinogen		237	26.8
3: Unclassifiable		495	55.9
4: Probable Human Non-Carcinogen		1	0.1
	Total	**885**	**99.9**

As Cunningham (2002) succinctly concluded: "A mouse is not a rat is not a human."

- **Developmental Neurotoxicity:** Test guidelines for developmental neurotoxicity (DNT) studies have existed in draft and final form since the 1980s. However, as discussed below, no DNT protocol, past or present, has ever been properly validated to confirm its relevance to neurodevelopmental effects in humans. For example, an EPA cross-species comparison study of the effects of PCBs on the developing nervous system (Tilson et al., 1990) found humans to be an order of magnitude more sensitive than monkeys and four orders of magnitude more sensitive than rodents to similar neonatal exposure levels. Moreover, EPA researchers (Goldey et al., 1994) studying the effects of the known human developmental neurotoxicant, methylmercury (MeHg) stated:

  > It is troubling that existing test methods fail to readily detect CNS-specific effects of MeHg, a known human developmental neurotoxicant. It is clear that a search for more sensitive test methods or test species should continue. The current findings support the view that humans are significantly more sensitive than laboratory animals to known teratogens.

  The magnitude and confounding effect of species differences in DNT studies was specifically addressed by an EPA Workshop on the Qualitative and Quantitative Comparability of Human and Animal Developmental Neurotoxicity. Based on an examination of no-observed-adverse-effect-levels (NOAEL) obtained in DNT studies conducted in a variety of species, the workshop acknowledged that there was "a wide range of differences across species (up to a 10,000-fold difference)" (Francis et al., 1990). This workshop further concluded that, "[i]n many cases—for example, lead, PCBs, and radiation—the proposed [DNT testing] battery probably would have underestimated human risk. This is true even when uncertainty factors are taken into account" (Stanton and Spear, 1990).

  In addition to species differences in toxicokinetics, the interpretation of DNT studies is confounded by the fact that animal species are born at developmentally different stages and mature at markedly different rates (Miller, 2003). The NRC examined this issue in its 1993 report, *Pesticides in the Diets of Infants and Children*, noting: "The newborn rabbit, rat, mouse and hamster can double their birth weights in less than one week, much faster than the human infant can. These different

growth velocities may alter the toxicity of pesticides and other chemicals among different species of infant animals" (NRC, 1993). The NRC further noted that, "[t]he age period in which specific organs or tissues undergo their most rapid rate of development and the age at which development is completed have major implications for studies of toxicity to those organs in growing animals… Thus, the impact of toxic products can produce quite different outcomes that vary both with time and with species" (NRC, 1993).

This issue was further explored during the EPA's Workshop on the Qualitative and Quantitative Comparability of Human and Animal Developmental Neurotoxicity, which reported that, "[t]he full range of critical periods of development of nervous system in humans and experimental animals and sensitive periods of exposure to toxic agents are not well characterized" (Francis et al., 1990). This limitation raises serious doubt as to whether "windows of exposure" in the DNT are representative of all periods of neurodevelopmental vulnerability in infants and children (Claudio et al., 2000). In fact, the EPA's SAP concluded, following its review of an EPA *Retrospective Analysis of Twelve DNT Studies* (Makris et al., 1998), that: "Exposure of rat fetus/pup were not shown to be equivalent to human fetus/infant during equivalent stages of brain development, both with respect to third trimester equivalent exposure (lactational exposure in rat/transplacental in human), and in length of total exposure (does not cover extensive postnatal period of brain development in humans)" (SAP, 1999).

Rats, the recommended species in the EPA's DNT test guidelines, also differ from humans in other developmentally relevant ways. Dorman and colleagues (2001) have cautioned that "[t]here are marked interspecies differences in types of placenta, orientation of exchanging vessels, and number of exchanging layers. … [L]arge species differences have been shown for placental permeability of hydrophilic molecules. … The observed species differences in placental transfer of hydrophilic xenobiotics are caused predominantly by structural differences among placenta." Such differences may have a profound effect on the extent to which unborn rat and human fetuses are exposed to a chemical in the womb.

With respect to dosing, the EPA's DNT test guidelines specify that, "the test substance … should be administered orally," which can include the addition of a test substance to an animal's diet, drinking water, or pumping it directly into an animal's stomach (known as "gavage" administration). It is well established that chemical exposure via oral gavage has the potential to deliver chemical to a target site at a rate that far exceeds anything that would occur in the real world (Conolly et al., 1999). The EPA itself (Makris et al., 1998) has acknowledged the dubious relevance of laboratory dosing procedures to realistic human exposure scenarios:

> [A]n advantage of gavage administration is that the exact measurement of the administered chemical is known and can be adjusted to body weight throughout the study. However, gavage dosing may be more irritating to the stomach. Additionally, gavage administration provides a discrete bolus dose of test substance while rats on a dietary study eat throughout each night and receive lower doses over multiple hours; it

> can be argued that neither of these scenarios is similar to human exposure to a test substance. Likewise, protocols that utilize dermal or inhalation administration (which generally expose animals to discrete 4- to 6-hour daily exposure periods that are designed to mimic worker exposure scenarios) may not have a direct corollary to human exposure in a residential setting either.

It has also been demonstrated that dosing-induced stress can produce massive damage in the liver of rats, and that prenatal stress induced by handling procedures can significantly impact the nervous system of developing pups (Claudio et al., 2000).

A further issue of concern is the EPA's specification that, "…the highest dose level should be chosen with the aim to induce some maternal toxicity (e.g., clinical signs, decreased body weight… and/or evidence of toxicity in a target organ." However, as EPA's Tilson (1992) acknowledges: "agent-induced maternal toxicity can contribute to behavioral indicators of neurotoxicity in the offspring, confounding interpretation of the data." More specifically, in the words of the EPA's *Retrospective Analysis of Twelve DNT Studies* (Makris et al., 1998):

> When developmental neurotoxicity is observed in the presence of maternal toxicity, it is often difficult to determine if the findings in young pups are secondary to the maternal toxicity. For example, decreased pup survival during early lactation and/or perinatal alterations in behavioral findings may be related to other events in either the offspring or the dam, such as an increase in the amount of the test substance in the milk at a critical time of development, inability of the offspring to suckle, general toxicity to the offspring, or insufficient maternal care of the litter.

Factors such as impaired maternal care behavior (e.g., nurturing and grooming), while critical to the growth and development of offspring, are not assessed in the DNT (Sheets, 2003). Nutrition is likewise critical to pups' growth and development, yet milk quantity and quality are not evaluated, nor are measurements recorded of components in the milk (Sheets, 2003). Maternal toxicity during the *in utero* phase may be equally problematic, as a result of reduced maternal food consumption, maternal pulmonary damage, maternal renal damage, effects on the maternal central nervous system, etc. (Tyl and Sette, 1990). Thus, "under circumstances of severe toxicity, you cannot distinguish whether effects are due to developmental neurotoxicity or secondary to maternal toxicity" (Sheets, 2003).

Because developing animals are exposed to a test substance via the dam—which relies on transport of the substance through the placenta or via lactation—the actual level(s) and route(s) of exposure levels in the pups are highly speculative (Claudio et al., 2000; Dorman et al., 2001). The EPA's Makris and colleagues (1998) clearly acknowledge this problem, noting:

> For developmental neurotoxicity studies reviewed by the Agency, there is generally a lack of knowledge regarding actual exposure of the chemical to the offspring *in utero* or via the milk. Pharmacokinetic data, which might assist in this determination, are not addressed in the standard developmental neurotoxicity study guideline… This suggests that a developmental neurotoxicity protocol which includes direct postnatal exposure may be necessary to adequately evaluate the developmental neurotoxic potential of some chemicals… although no standardized testing guideline for postnatal dosing has been developed by the Agency….

This finding was echoed by the Agency's Risk Assessment Forum, which noted: "Under the DNT protocol, there is currently no requirement to perform kinetic studies to ascertain either *in utero* or postnatal exposure. There is no mechanism to guarantee exposure postnatally because the compound may not be excreted into breast milk or it may be excreted only at very low concentrations" (RfD/RfC Panel, 2002). Direct dosing of pups has been proposed as a means of resolving questions surrounding route of exposure, dose levels, etc. However, according to Miller (2003) of the Centers for Disease Control: "Direct dosing requires a lot of manipulation of the animals [and] is itself a problem. The continued handling of pups required in direct dosing can affect the endpoints of the developmental neurotoxicity study, especially the behavioral endpoints that are being evaluated at the same time." The unavoidable stress and potential injury to pre-weaning pups by oral force-feeding has been documented to cause a host of non-specific functional and/or behavioral changes that have nothing to do with a chemical's toxic mode of action, thus confounding interpretation of the results of certain DNT studies (Conolly et al., 1999). Sheets (2003) has therefore questioned:

> Does this provide a relevant circumstance for use in risk assessment? I think we all agree that such a dose does not model dietary, dermal, or inhalation exposure that children may experience in the real world. … If the pesticide does not pass through the milk, then it would be inappropriate to dose pups by gavage because it would not be relevant to a human circumstance of exposure. It is important to remember that this is not a research project to satisfy scientific curiosity; this is a study that should model realistic human exposure circumstances.

Behavioral testing has become a central component of DNT studies due to the perceived sensitivity of behavioral endpoints in detecting subtle chemical insults to the central nervous system. However, current measures of an animal's cognition, sensory-motor function, and other behavioral parameters are numerous and diverse, often differing widely in terms of the subjectivity of observations, breadth and specificity of results, quantity and quality of available validation data, and extrapolation of results among species (Cory-Slechta et al., 2001). For example, Elsner (1992) has commented:

> Experimental psycho-teratology, with the purpose of determining potential effects of chemical exposures during brain development on

> human behavior, assumes that behavior of experimental animals is a valid model for human behavior. This practice implicitly takes for granted that changes in animal behavior, which are detectable in artificial laboratory settings, relate directly to sometimes poorly understood and highly complex psychopathologies in the diversified human environment. It is understandable that this bold assumption is doubted by many neurotoxicologists…

As detailed by Anger (1990), laboratory tests employed to assess the effects of chemicals in animals "differ markedly" from those used to assess neurotoxic effects in humans. Anger's comparison of laboratory and human clinical and occupational assessment techniques "reveals a lack of parallelism between the screening tests that will be employed to test animals and those used to assess humans, suggesting that the respective tests do not test similar functions." For example, sensory tests in human volunteers include sophisticated measures of visual perception and memory, whereas the animal-based tests for sensory function include simplistic "finger snap response," "tail pinch response," and "pupil response" (Anger, 1990). A similar picture emerges for assessments of affect, cognition, memory/learning, and psychomotor function. In the words of the head of Health Canada's Pest Management Regulatory Agency (Franklin, 1999): "Few if any standardized/ validated protocols have yet been established which specifically address cognition, memory and higher brain function." Thus, "few studies of higher neurobehavioral function in humans and animals are directly comparable" (Davis et al., 1990).

Exacerbating this problem is the question of whether current behavioral measures of learning and memory can accurately be applied early in a rodent's life span, and whether they are sufficiently sensitive to detect subtle effects on the central nervous system (SAP, 1999; Claudio et al., 2000). Indeed, the agency's SAP "was divided whether the methods identified [in the EPA's DNT guidelines] are reasonable for assessment of toxicity to offspring."

Also widely recognized is the fact that reliance on often subjective behavioral observations, versus more objective physiological measures, has the potential to introduce major uncontrolled variability into conduct and interpretation of DNT studies (Claudio et al., 2000; Tilson, 1995; Gerber and O'Shaughnessy, 1986). Studies that have relied upon neurobehavioural tests have suffered from poor inter-laboratory reproducibility, even where rigorous control has been maintained over variables such as test apparatus, testing protocols, animal husbandry, acclimation times, order of administration of tests, etc. (Crabbe et al., 1999). This situation is particularly worrisome in view of the fact that EPA DNT test guidelines allow investigators considerable "flexibility" in their choice of methods of behavioral testing.

Another question relates to the specificity of behavioral tests for neurotoxicity, or their ability to distinguish between "true" neurotoxicity as distinct from more generalized systemic toxicity. This question was examined by Gerber and O'Shaughnessy (1986), who concluded:

> These results further illustrate the point that none of the standard behavioral tests are unique indicators of impaired central nervous system function. All the behavioral tests used in this study were affected by impairment of critical systemic organs as well as by restriction of food and water intake. Before it can be concluded that a compound is neurotoxic on the basis of behavioral test results, it must be ascertained that non-neural organs have not been damaged by the test compound, and that food and water consumption have not been severely decreased. Unless these factors are considered, it is possible to obtain a behavioral test result that suggests a neurotoxic effect, but which is actually the consequence of systemic organ toxicity.

In terms of reliability, he DNT has not been subject to a level of scrutiny even remotely approximating that of a prospective, inter-laboratory validation study. Moreover, existing data and published assertions by EPA personnel do not inspire confidence in the reliability of data from DNT studies. Perhaps not surprisingly, a study of behavioral tests (which represent a core component of the EPA's DNT test guidelines), "found extreme variability in the results [of replicate neurobehavioral studies] obtained in different laboratories" (Claudio et al., 2000). In this study, three different laboratories conducted a battery of six neurobehavioral tests in inbred strains of (adult) mice. The results varied widely among the laboratories, in spite of rigorous controls of methodological variables, including test apparatus, testing protocols, animal husbandry, acclimation times, order of administration of tests, etc. The results demonstrated that confounding influences in the laboratory environment produced widely different neurobehavioral outcomes (Crabbe et al., 1999).

Variability among DNT studies may be even greater than in the above example, due the high degree of "flexibility" the EPA permits in the choice of behavioral tests for learning and memory, as well as strain and species of animals used in DNT studies (Cooper Rees et al., 1990). In the words of one high-ranking EPA neurotoxicologist: "the outcome of a [DNT] study can depend on the inherent variability of a test measure" (Tilson, 2000). This was clearly demonstrated by a recent EPA retrospective analysis of "positive control" data in DNT studies (Crofton et al., 2004). The authors note that: "A necessary property of a good positive control chemical is that the effects on the endpoint of concern are well characterized and accepted by the general scientific community." Thus, positive controls in the context of DNT studies are substances that have been well established to be toxic to neurodevelopment. Significantly, however, the EPA's retrospective analysis found that "[l]ack of effect of the positive control chemical was a problem that occurred at least once in over 50% of the [16] test laboratories" (Crofton et al., 2004). This suggests that DNT results often cannot even be duplicated from one laboratory to another.

The EPA itself has acknowledged the litany of problems associated with the conduct and interpretation of DNT studies, which was summarized by Makris and colleagues (1998) as follows:

> It is recognized that the conclusions drawn from this initial retrospective survey of developmental neurotoxicity data must be examined in light of the many confounding factors that may have contributed to the study results and conclusions. Some of these factors are common to many or all of the studies, such as the influence of dose selection on determination of the NOEL [no-observed-effect-level], inaccuracies or inconsistencies in the conversion of dietary or inhalation dose levels to mg/kg/day values, a lack of knowledge regarding actual exposure of the chemical to offspring *in utero* or via the milk (pharmacokinetic data), or differences in the endpoints examined for the various protocols (for example, the timing of measurements, variations in laboratory procedures, missing or inadequate assessments of any particular endpoint). Some factors are specific to a chemical or a particular study protocol. These might include utilization of knowledge on the chemical to aid in the selection of tests to assess learning and memory or of the most appropriate species for testing. It is also acknowledged that the conclusions of the studies, as well as the endpoints selected for risk assessment, are often issues of contention between the Agency and the regulated community. There are ongoing, unresolved controversies regarding some of the studies presented in this paper as well as some of the Agency decisions cited in this analysis.

In the absence of proper validation, EPA cannot conclude that the results of DNT studies have any bearing on the potential threat of the investigated substance to humans, particularly infants and children. This conclusion was also expressed by a panel of the National Academy of Sciences, which reported: "the subcommittee finds that the developmental neurotoxicity test, as it is currently described in the U.S. Environmental Protection Agency (EPA) guidelines (EPA 1991), might be inadequate to identify and characterize specific developmental neurotoxicants (CLS, 2000). Similarly, in the words of one EPA scientist (Rice et al., 1996):

> It is clear from comparison of the human and rodent data that the results from rodent studies often vastly underestimate intakes at which neurotoxicity was observed in humans. For lead, deficits were revealed on activity and simple learning tests at doses that would also result in allowable intakes much higher than those at which cognitive impairment has been demonstrated for children. One conclusion that may be drawn from this analysis is that current methods of calculating acceptable intakes based on animal data … are insufficient to protect the human population against behavioral toxicity.

Likewise, when asked by the EPA for its opinion as to the sensitivity of the DNT relative to other developmental or reproductive toxicity studies, the SAP stated: "The DNT is not more sensitive in its current form, given what is known in the broader neuroscience and pediatric community… Therefore, the current form of the DNT guidelines … is not a sensitive indicator of toxicity to the offspring" (SAP, 1999).

- **Immunotoxicity:** The divergence in immunological structure and response between species is extensive. A recent paper concluded that "...animal studies, even those conducted in non-human primates, have limited predictive power for immunogenicity in humans..." (Bugelski and Treacy, 2004). Whether a xenobiotic is deemed immunomodulatory depends heavily on which species and which parameters are evaluated, and may not be representative of the human response.

  For example, regarding organ composition and function, Haley (2003) reports that mouse spleens are major sites for hematopoietic activity throughout their lives, rat spleens to a much lesser extent, and humans have little hematopoietic activity in the embryonic spleen and virtually none in normal adult spleens. Given the functional differences between human and rodent spleens, the author questions the relevance of using rodent spleen cells as target cells in immunotoxicity assays. With respect to the lymph nodes, most species have significantly fewer nodes organized into less complex networks as compared to humans: the rat lung is drained by two lymph nodes, the dog lung by three to five, and the human lung by 35 nodes classified into five separate groups (Haley, 2003). Lastly, the thymus of aged female rats varies considerably by strain: for Brown Norway rats, the thymus consists mainly of epithelial structures (cords and tubules) with few lymphocytes, but Wistar and WAG rats show lymphocyte masses with limited epithelial components (Haley, 2003). The significance of changes to these organs' weights and histopathology (evaluated in the STS) and their relevance to humans is obfuscated by these important structural differences.

  With regard to cell populations, the composition of circulating leukocytes differs markedly by species. For humans, the circulating leukocyte profile is strongly neutrophilic with 50-70% neutrophils but for rodents, it is strongly lymphocytic with 50-100% lymphocytes (Table 11; Haley, 2003). Thus, since the initial composition of animal leukocyte populations is drastically different from that of humans, it would be unclear what any alterations might mean for humans. Furthermore, there is some evidence that the absence of subset changes in lymphocyte populations does not correlate with an absence of functional immunosuppression (Kerkvliet and Brauner, 1990). The *in vitro* clonogenic assays outlined later in these comments (e.g., CFU-GM assay) can much more meaningfully measure adverse effects on subsets of human leukocytes, and therefore help predict myelosuppression.

Table 11

Approximate total white blood cells, percent neutrophils and lymphocytes in blood of different species

Species	WBC (10^3/μl)	Neutrophils		Lymphocytes	
		%	10^3/μl	%	10^3/μl
Human	7.0–10.0	50–70	3.5–7.0	20–40	1.4–4.0
Cynomolgus monkey	5.0–18.0	10–42	0.5–7.5	40–67	2.0–12.0
Beagle dog	6.4–14.6	47–65	3.0–9.5	16–41	1.0–6.0
CD rat	3.0–14.5	10–21	0.3–3.0	83–100	3.0–12.0
Hamster	4.0–10.0	13–35	0.5–3.5	63–80	2.5–3.5
Guinea pig	2.0–12.0	25–30	0.5–3.5	75–83	1.5–10.0
CD-1 mouse	2.0–10.0	15–20	0.3–2.0	50–70	1.0–7.0
New Zealand white rabbit	4.0–13.0	25–46	1.0–6.0	50–70	2.0–9.0

Data derived from Payne et al. (1976), Williams et al. (1995) and Hall (1992).

Turning to cellular responses in humoral immunity, species vary considerably in T-cell dependent antibody response (TDAR) studies. For example, Smialowicz and colleagues (1994) reported that 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) causes a dose-dependent suppression of the plaque-forming cell (PFC) response to sheep red blood cells (SRBC) in adult female B6C3F1 mice, but enhances the PFC response to SRBC in F344 and Long-Evans rats at doses as high as 30 μg/kg. Smialowicz (1992, 1994) also investigated T-independent antibody responses and found that 2-methoxyethanol (ME) and its principal metabolite 2-methoxyacetic acid (MAA) are immunosuppressants in rats (though sensitivity varies significantly by strain) because they suppress the PFC and lymphoproliferative (LP) responses (and cause thymic involution) but none of these effects were seen in the mice strains studied.

Significant species differences have also been identified in cell-mediated immunity (Trinchieri and Perussia, 1984). For example, natural killer (NK) cell activity in mice is high in the lungs and tends to peak early in life in the spleen and blood while in humans, NK cell activity is low in the lungs and relatively stable in the spleen and blood throughout life. Also, when B6C3F1 mouse and Fischer 344 rats are dosed with cyclosporine A, NK cell activity is significantly depressed in the mouse but not in the rat, however, cytotoxic T-lymphocyte activity and mixed lymphocyte response are significantly inhibited in the rat but not in the mouse (Blot et al., 1994). Additionally, there is a considerable degree of variation of NK cell activity across mouse strains, which could clearly interfere with accurate determination of xenobiotic effects on this parameter (Itoh et al., 1982).

Similarly, different species' macrophages have been found to have vastly differing efficiencies in bacterial phagocytosis and killing (e.g., human pulmonary alveolar macrophages (PAMs) are much more effective than those of rats'), which leads to differing sensitivities to certain immunotoxicants (Nyugen et al., 1982). Other basic differences include the fact that human PAMs are cytocidal whereas PAMs from mice, rats and guinea pigs require activation before they become tumoricidal (Lemarbre et al., 1980). These problems can be sidestepped by the use of *in vitro* systems based on human macrophages and neutrophils (Hartung, 1996; Tsuchiya, 1980).

At the molecular level, the classes, regulation, and functions of antibodies and cytokines also differ by species. For example, despite the fact that the molecules are relatively well conserved, the functions of IL-5 are quite dissimilar between humans and rodents (House, 1999). IgG1 binds complement and is reagenic in mice, but not in rats or rabbits. Callard and Turner (1990) reviewed cytokine control of immunoglobulin regulation in mice and humans, concluding that the mechanisms diverge significantly and cautioning against broad application from murine studies to humans.

In addition to the problems associated with species extrapolation, animal-based immunotoxicity studies are limited by differences between strains, some of which have been mentioned earlier in this section. Strain differences within a species can be significant, especially for inbred strains. For example, dosing three common mouse strains with 2-deoxy-D-glucose (2-DG) results in highly variable changes in splenic cell populations, blood glucose levels, and corticosterone levels, and different effects on the *in vitro* proliferation of mature T splenocytes (Dreau et al., 2000). Other mouse studies revealed variation in susceptibility to various immunotoxicants between different strains/ages ranging from a factor of 5 to 18 (Selgrade et al., 1995).

Artifacts in laboratory methodology also render some animal data suspect. For example, one of the parameters evaluated in the initial STS screen is a change in the weight of immune system organs after dosing with the drug candidate. However, the variation around the median thymic weight of a Sprague-Dawley rat may be 70% or more, and for a beagle dog, 100-170% (Haley, 2003). Collection techniques also influence organ or tissue weight. It is difficult to obtain meaningful data around this type of highly variable endpoint.

Further and importantly, stress can negatively impact the immune system, thus confounding the results of *in vivo* experiments testing potential immunotoxicants. Procedures that cause pain clearly set off a series of undesirable physiological changes, but it is less well known that routine procedures and manipulations of animals, such as animal handling, blood collection, and exposure by gavage cause significant physiological changes associated with stress (e.g., changes in serum or plasma concentrations of corticosterone, glucose, growth hormone or prolactin, heart rate, blood pressure, and behavior) (Balcombe et al., 2004). Changes from baseline or control measures ranged from 20% to 100% and more, with duration of 30 minutes or longer. Furthermore, laboratory conditions induce boredom and distress, sometimes resulting in abnormal stereotypic or compulsive behaviors such as self-mutilation. These stress-related problems confound all animal experiments, but are especially problematic in the field of immunotoxicology.

Dealing With Uncertainty: Additional Testing v. Safety Factors

The FQPA amended the FFDCA to allow EPA to "establish or leave in effect a tolerance for a pesticide chemical residue in or on a food **only if** the Administrator determines that the tolerance is safe." FFDCA § 408(b)(2)(A)(i), 21 U.S.C. § 346a(b)(2)(A)(i)(emphasis supplied). A tolerance is only considered safe if the Administrator determines, to a reasonable certainty, that no harm will result from aggregate exposure to the pesticide chemical residue, including all anticipated exposures. FFDCA § 408(b)(2)(A)(ii), 21 U.S.C. § 346a(b)(2)(A)(ii). To further account for the potential special toxicity of a chemical to infants and children, the FQPA provides that "an additional tenfold margin of safety for the pesticide chemical residue and other sources of exposure shall be applied for infants and children to take into account potential pre- and post-natal toxicity and completeness of the data with respect to exposure and toxicity to infants and children. Notwithstanding such requirement for an additional margin of safety, the Administrator may use a different margin of safety for the pesticide chemical residue only if, on the basis of **reliable data**, such margin will be safe for infants and children." 21 U.S.C. § 346a(b)(2)(C) (emphasis supplied).

Despite clear Congressional direction that infants and children were to receive stronger protection than they had at the time of enactment of the FQPA, OPP's retention of the full 10x children's safety factor has become the exception rather than the rule (HIARC/FQPA SFC, 1998; EPA, 2002b; 2005). Instead, OPP has moved to ostensibly "modernize" its safety standard through the marked expansion of its data and testing requirements. For example, testing for DNT and immunotoxicity (both adult and developmental) have become conditional requirements for the registration and re-registration of conventional pesticide chemicals, and the absence of such information is being characterized as "data gaps" (HIARC, 1998; 2001; 2003a; 2003b). Yet as documented above, these studies are of unproven reliability and relevance to humans, and thus cannot provide the requisite "reasonable certainty of no harm" that OPP requires in order to set lawful tolerances according to its Congressional mandate. Moreover, studies such as the DNT fail to provide any demonstrable "value added" in pesticide risk assessments.

For example, DNT data have yet to be used as the basis for setting even a single chronic dietary reference dose (RfD) (Makris et al., 1998), which means "there were other studies that were more sensitive than the NOELs [no-observed-effect-levels] generated from the DNT. For acute RfD, the DNT study was used three times. In two of these cases, the RfDs were based on maternal toxicity, not pup toxicity. One acute RfD was based on pup toxicity. In evaluating this one acute RfD, there were also NOELs from other reproduction/developmental studies that were really hovering right around that same NOEL as the DNT study, if not lower like the two gen repro study" (Li, 2003). Nor has the DNT lead to the lowering of any pre-existing reference doses for pesticides. As stated by the American Industrial Health Council (AIHC, 1999):

> EPA ignores the key outcome [of the Retrospective Analysis by Makris et al. (1998)] that the developmental neurotoxicity tests have not caused a single chemical on the list of 12 to be regulated at lower exposure levels than the levels that can be determined by other more traditional

> toxicology tests…. EPA's level of confidence in the ability of extensive developmental neurotoxicity testing to lead to greater protection of children's health is simply overstated and unsupported by the evidence.

Indeed, evidence to date suggests that the conduct of DNT and other animal studies to fill "data gaps" functions more as a mechanism to undermine children's health than to enhance it. For example, OPP tolerance decisions in which the full 10x safety factor is retained almost invariably point to the toxicity database being "incomplete" due to the "lack of a developmental neurotoxicity study" as the basis for retention. Conversely, OPP is much more likely to reduce or remove the FQPA safety factor when the toxicity database includes a DNT study. Thus, infants and children are much more likely to be afforded the maximum level of protection from toxic pesticides (i.e., lower tolerances) if DNT and other additional animal tests are *not* carried out. Simply put: more testing almost always equals less protection, which neither serves the interests of children's health nor reflects the will of Congress in enacting the FQPA.

In summary, OPP's proposed additions to Part 158 data requirements are predicated on the simplistic assumption that "more is better" (i.e., "The addition of some data requirements is likely to further communicate to domestic and world-wide marketplaces that pesticide products and items treated with them are safer, thus enhancing the reputation of American agricultural products and registered pesticides as tools for public health, etc." (70 Fed. Reg. 12281)). Notwithstanding the political basis for these proposed amendments, the analysis above clearly invalidates the blanket assumption that increased data and testing requirements will contribute in any way to improved protection for either human health or the environment.

ALTERNATIVE APPROACHES

The revision of Part 158 provides an optimal opportunity for OPP to streamline and modernize its pesticide data requirements to eliminate wasteful redundancies and open the door to emerging technologies and testing strategies. The parties to this submission strongly urge OPP to abandon the rigid "checklist" approach that is currently in place in favor of a flexible tiered testing paradigm. One such approach has recently been proposed by the ILSI Health and Environmental Science Institute's Technical Committee on Agricultural Chemical Safety Assessment (ACSA). For the sake of efficiency, we have structured our comments and recommendations based on the ACSA proposal: identifying specific elements that we do and do not support, and outlining opportunities for further refinement which we believe strikes an optimal balance between the EPA's Congressional and other mandates to protect human health and the environment; to develop and utilize alternatives to animal testing; and to ensure test method validation, data quality, and the mutual acceptance of data.

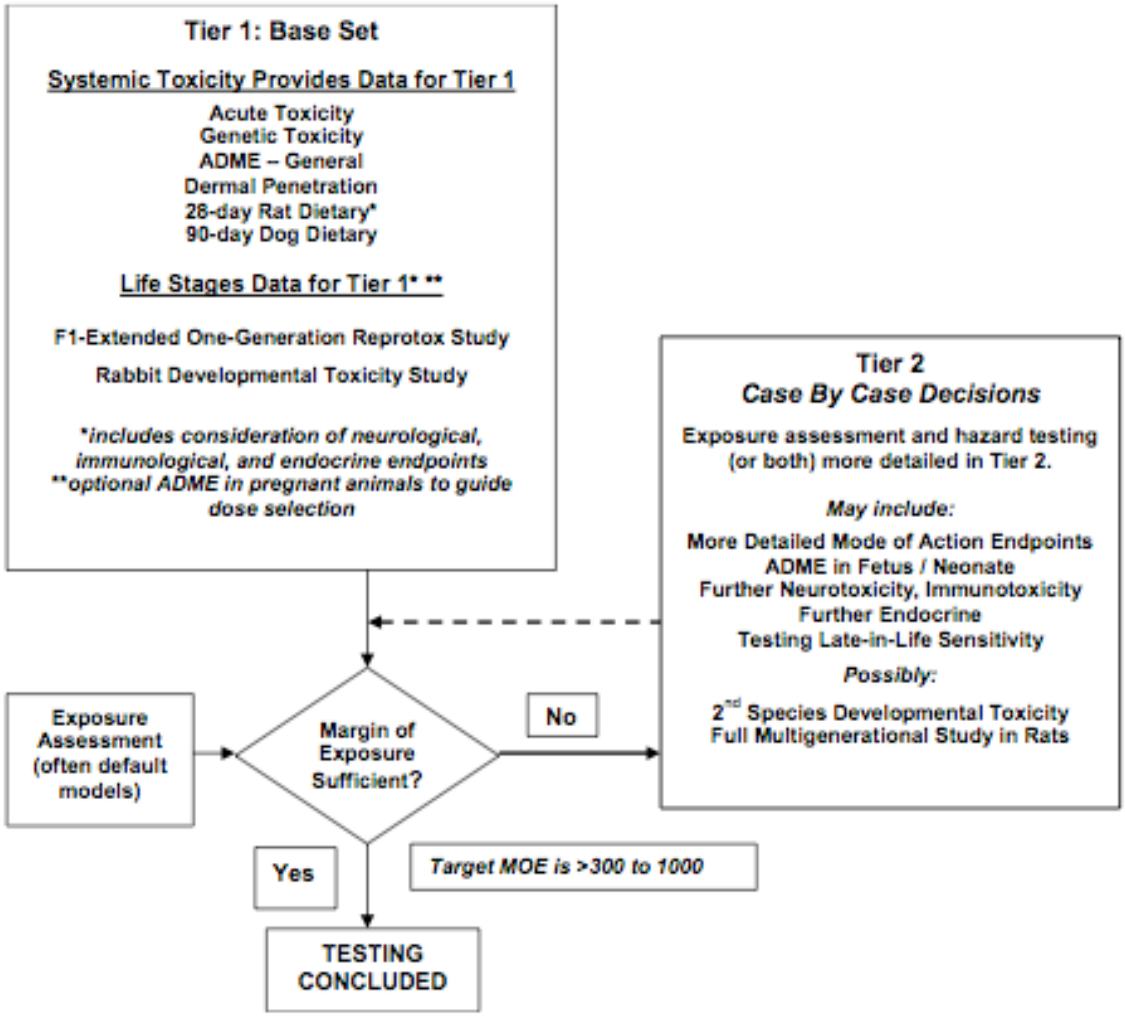

Figure 1—ACSA Tiered Testing Proposal (ACSA, 2005b)

Points of Agreement

ADME White Paper (ACSA, 2005c):

- Making better and earlier use of ADME studies and using these data to inform subsequent testing decisions, as is the case with pharmaceutical evaluations.

- Utilizing pharmacokinetic data to extrapolate across route, frequency and duration of exposure. In particular:

    - "Use of route-to-route extrapolation based on pharmacokinetics can provide a sufficiently comprehensive database for systemic toxicity without conducting each toxicity study by multiple routes of exposure." This approach could reduce the acute toxicity "6-pack" to a 4-pack, and obviate the need for subacute and subchronic dermal and inhalation studies.
    - "*In vitro* dermal studies may be used for route-to-route extrapolation in the risk assessment, but a registrant may want this information earlier in the development process as an important decision-making criteria for selecting among development candidates." Moreover, "[s]ince dermal absorption is a major concern for risk assessment, early evaluation of dermal absorption *in vitro* could be included in the first tier of studies during product development." This approach makes appropriate use of an available, valid and internationally accepted non-animal test method, now codified as OECD TG 428, and subject to the requirements pursuant to MAD (OECD, 1981).

- Utilizing *in vitro* metabolism studies, which "may provide one of the most valuable ways to limit animal use, by assisting in pharmacokinetic modeling for toxicity study design, interpretation, and use in risk assessment" and "… might appropriately be included in the first tier." This approach likewise takes advantage of available and emerging non-animal approaches for evaluating xenobiotic metabolism.

Systemic Toxicity White Paper (ACSA, 2005a):

- Elimination of the chronic (1- or 2-year) dog study, on the grounds that they "neither result in appreciably lower NOAELs nor identify new effects, for the majority of chemicals, when compared to the shorter duration study (90 day) in this species." This finding is consistent with the preponderance of literature in this area (i.e., Spielmann and Gerbacht, 2001; Baetcke et al., 2005; Box and Spielmann, 2005).

- Elimination of the 2-year mouse carcinogenicity study, given that "additional information provided by [this] study is of limited value in risk assessment." This finding is consistent with the preponderance of literature in this area.

Life Stages White Paper (ACSA, 2005b):

- Amalgamation of several reproduction-type studies into a single, more efficient Tier 1 protocol (i.e., F1-extended one-generation study). This approach would reduce animal use by more than 70%—consuming approximately 1,400 rats instead of the 5,200 killed at present by conducting 2-generation reproduction, prenatal developmental toxicity and DNT studies separately.

Opportunities for Further Refinement & Animal Reduction

ADME: Chemical absorption, distribution, metabolism and elimination (ADME) should be the first parameters studied in a pesticide hazard assessment, as is the case in the pharmaceutical sector. A comprehensive overview of available *in vitro* and computational methods for the evaluation of toxicokinetics and metabolism has recently been published by Coecke et al. (2005). The use of cell lines that have been genetically engineered to express various phase I and phase II enzymes (Bernauer et al., 2003; Macé et al., 1998; Combes, 1992) should also be explored in the context of *in vitro* metabolism studies. While these non-animal methods cannot yet fully model all of the parameters examined in the OPPTS 870.7485 study, they have immediate applicability in informing testing decisions. For example:

- Species-specific metabolism and kinetics should be evaluated for commonly used species (i.e., humans, rats, dogs and rabbits on the human health side, and mallard ducks, bobwhite quail, red-winged blackbirds, rainbow trout, bluegill sunfish, fathead minnow, sheepshead minnow, etc., on the ecotoxicity side) using a combination of *in vitro* assays and PBPK modeling. Knowledge of various species' sensitivity to a chemical agent would enable registrants to discontinue the current, wasteful practice of running the same study in two or more species in a blind search for the lowest NOAEL. Instead, ADME data would support the conduct of a single study in the most sensitive species.

- The evaluation of dermal penetration should be moved from a top-tier study to a first-tier investigation based on the internationally accepted *in vitro* protocol codified in OECD TG 428 (OECD, 2004a). *In vitro* data generated in accordance with this guideline are subject to the provisions of MAD (OECD, 1981), and must therefore be accepted by OPP as a stand-alone measure of dermal absorption and penetration. These data can be put to further use in estimating the likelihood of systemic toxicity following exposure via the dermal route, and whether there would be any appreciable "value added" in conducting a stand-alone 21- or 90-day dermal study (which OPP is proposing to make an unconditional requirement, which we oppose). In addition, these *in vitro* dermal penetration data can be used to support in route-to-route extrapolations together with physiologically-based pharmacokinetic (PBPK) modeling, as discussed in greater depth in ACSA (2005c) and Coecke et al. (2005).

Acute Toxicity—Lethality: The parties to this submission cannot overstate our opposition to the continued reliance on acute lethality studies (i.e., oral, dermal and inhalation LD_{50}/LC_{50} studies), whether for the testing of a TGAI or for the labeling of a

pesticide EP. We do not accept the conclusion of the ACSA Systemic Toxicity Task Force (ACSA, 2005a) that the Up-and-Down Procedure (OECD TG 425) is an adequate alternative to the classical oral LD_{50} test (i.e., the now deleted OECD TG 401), although animal use is reduced somewhat.

To the extent that acute systemic toxicity is assessed for the purpose of making labeling decisions for formulated pesticide product containing ingredients that have already been subjected to extensive study and toxicological characterization, we urge OPP to permit registrants to submit non-animal data for this endpoint, consisting of:

- Computerized expert system/(Q)SAR modeling (e.g., TOPKAT Rat Oral LD_{50})
- *In vitro* basal cytotoxicity assay (e.g., NHK cell line)
- *In vitro* metabolism studies (e.g., human microsomal enzymes)

Based on the premise that the actions of chemicals that produce toxicity do so at the cellular level (Grisham and Smith, 1984), many non-specific cell toxicity tests have been developed as potential replacements to acute lethality tests on animals (Paul, 1975; Ekwall and Sandström, 1978; Kangas et al., 1984; Dierickx, 1989; Wakuri et al., 1993; Clemendson et al., 1996). A battery of three such tests and a simple mathematical algorithm was found to be significantly more predictive ($R^2 = 0.83$) of human lethal doses for 50 chemicals (derived from forensic medical reference books) than predictions based on rat and mouse LD_{50} values ($R^2 = 0.65$) (Ekwall et al., 1998). Regulatory guidance and recommended *in vitro* study protocols have since been published (NICEATM, 2001a; EPA, 2001) for use with "normal human keratinocytes" (NHK) and other standardized cell lines (Heimann and Rice, 1983; Borenfreund and Puerner, 1984; Borenfreund and Puerner, 1985; Spielmann and Liebsch, 1992; Liebsch and Spielmann, 1995; Spielmann et al., 1999). Although these methods assess only basal cytotoxicity—as distinct from multiple organ toxicity—the *Report of the International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity* (NICEATM, 2001b) concluded that it is most likely not necessary to routinely test for all possible specific organ effects in various *in vitro* models. Instead, *in vitro* metabolism assays and tests on the ability of a compound to disrupt epithelial barrier function may be sufficient.

In addition to *in vitro* assays, there are several computer software packages for predicting acute toxicity from chemical structure. One example is TOPKAT, a statistically based package comprising several (Q)SAR models, including TOPKAT Model Rat Oral LD_{50} (Anon, 2004). This has been developed using experimentally derived LD_{50} values for about 4,000 chemicals, and assesses acute toxicity for a range of chemical classes. Extrapolation from one exposure route to another (e.g., from oral to inhalation or skin) can be accomplished with the aid of computerised biokinetic modelling (ACSA, 2005c; Paxton, 1995). It was the view of the European Centre for the Validation of Alternative Methods (ECVAM) in 2002 that a combined in vitro/computational testing strategy could be used immediately for priority setting among chemicals, which would permit the most hazardous chemicals to be controlled at an early stage (Worth and Balls, 2002). This strategy should be deemed sufficient for pesticide labelling purposes.

Acute Reference Dose: We note that OPP's amendments do not propose the addition of a further acute study for the purpose of calculating a reference dose (RfD) for a human exposure duration of up to one day, as was recommended by ACSA (2005b) and EPA's Reference Dose/Reference Concentration Technical Panel (2002). We support this "omission," as we believe that it is neither necessary nor appropriate to conduct a separate animal-based study in either dogs or rats for this purpose. Instead, we recommend that if OPP wishes to calculate an acute RfD, this value should be derived by extrapolation from either the 28- or 90-day study (whichever yields the lowest NOAEL).

Acute Ocular Irritation: The rationale behind OPP's proposal to expand dermal irritation testing requirements to both EPs and TGAIs is neither persuasive nor supportable. Testing of EPs alone is quite sufficient for labeling purposes. With respect to testing methodology, Table 12 identifies some of the most advanced alternative methods currently available, including their validation and regulatory acceptance status:

Table 12 (adapted from Eskes et al., 2005)

Alternative method	Endpoint(s) measured	Area(s) of application	Validation status	Regulatory acceptance
BCOP	Opacity, permeability, depth of injury	Moderate to severe irritants	Optimized	A positive outcome is acceptable in the EU for classification + labeling of severe irritants
IRE	Opacity, swelling	Severe irritants	Optimized	A positive outcome is acceptable in the EU for classification + labeling of severe irritants
CEET	Corneal swelling, opacity, peameability + gross morphological lesions	Mild, moderate + severe irritants	Optimized	A positive outcome is acceptable in the EU for classification + labeling of severe irritants
HET-CAM	Haemorrhage, lysis, coagulation	Severe irritants *or* mild/non-irritants for surfactants	Optimized	A positive outcome is acceptable in the EU for classification + labeling of severe irritants

EpiOcular	Cytotoxicity	Moderate to mild irritants	Under validation	
HCE (SkinEthic)	Cytotoxicity, histology	Discrimination of irritant v. non-irritant	Under prevalidation	
HCE-TEP (Gillette)	Permeability	Moderate to mild irritants (surfactants)	Under validation	

Methods accepted as "positive screens" by European regulators (i.e., the BCOP, IRE, CEET and HET-CAM) should likewise be accepted by OPP for identifying and labeling severely irritating substances. We understand that OPP has already accepted BCOP data for this purpose on at least one occasion (Eskes et al., 2005).

Acute Dermal Irritation: The rationale behind OPP's proposal to expand dermal irritation testing requirements to both EPs and TGAIs is arbitrary and is neither persuasive nor supportable. Testing of EPs alone is quite sufficient for labeling purposes. With respect to testing methodology, we urge OPP to accept the results of ethically conducted clinical skin-patch tests in human volunteers (Basketter et al., 1997; OECD, 1997; Robinson et al., 2001) as a full replacement for rabbit skin irritation studies. In addition to eliminating animal use, clinical patch tests offer the benefit of being directly relevant to humans, thus obviating the questionable practice of cross-species extrapolation. Human data are regarded as being so superior to animal data that the Canadian government's testing guidelines (EC/HC, 2001) stipulate: "In most cases, properly conducted human patch tests (positive or negative response) are an acceptable alternative to animal testing for skin irritation or skin sensitization." However, before a formulation is considered for a clinical patch test, it must first be determined to be non-corrosive and free of other hazardous properties using the non-animal strategy prescribed by OECD TG 431 (OECD, 2004b). Only chemicals that appear to be non-irritating and free of other health hazards would move on to a clinical patch test to confirm their non-irritancy. Under no circumstances should negative results *in vitro* (i.e., indicative of non-corrosivity) generated according to OECD TG 431 be subject to "confirmatory" testing in animals, per MAD (OECD, 1981).

Dermal Sensitization: The rationale behind OPP's proposal to expand dermal irritation testing requirements to both EPs and TGAIs is neither persuasive nor supportable. Testing of EPs alone is quite sufficient for labeling purposes. With respect to testing methodology, although the Local Lymph Node Assay (LLNA) is often referred to as an "alternative" method because animal use is "refined" and "reduced" relative to the Guinea Pig Maximization and Buehler tests, it is an *in vivo* study nonetheless, and one of dubious relevance to humans (Basketter and Chamberlain, 1995). We therefore urge OPP to permit registrants to carry out a non-animal weight-of-evidence assessment for this endpoint, consisting of:

- Computerized expert system/(Q)SAR modelling (e.g., DEREK)
- *In vitro* dermal penetration study (OECD TG 428)
- *In vitro* protein binding study (e.g., human serum albumin)

The approach outlined above involves identifying "structural alerts" for sensitization potential using computerized (Q)SAR modeling and expert systems such as DEREK (Deductive Estimation of Risk from Existing Knowledge; LHASA, 2005; Barratt et al., 1994; Sanderson and Earnshaw, 1991), followed by a combination of *in vitro* dermal penetration (OECD TG 428) and protein binding studies (Basketter et al., 1995;). Any chemical scoring positives in all three methods is highly likely to be a skin sensitizer and can be classified accordingly. Three negatives would rule out the likelihood of sensitizing hazard, since if a chemical has no structural alerts, does not penetrate the skin, and cannot bind to protein, it lacks the necessary properties to induce an allergic skin reaction. Expert judgment would be required for mixed positive and negative results.

Subchronic Toxicity: We strongly urge OPP to abandon its longstanding practice of requiring duplicative testing in more than one species (i.e., 90-day feeding studies in rats, dogs, and potentially now mice as well), and by more than one exposure route (i.e, oral, dermal and inhalation). Such obvious and unnecessary redundancy is completely unacceptable. We therefore support in principle the ACSA (2005a) proposal to limit subchronic feeding studies to a single species,* to the extent that it would reduce animal use by over 80% (from approximately 272 to 48 animals). At the same time, however, we do *not* support the ACSA (2005a) proposal to increase group sizes to 6/sex/group, and call OPP's attention to opportunities to further refine the subchronic testing paradigm identified by Appleman and Feron (1986).

We likewise oppose OPP's proposal to change 21- and 90-day dermal studies from conditional to unconditional requirements, as questions of systemic toxicity via the dermal route can be amply addressed by means of route-to-route extrapolation using a combination of subchronic oral data, *in vitro* dermal penetration data generated in the context of ADME investigations, and PBPK modeling to bridge any perceived data gaps (ACSA, 2005c).

Organ/System-Specific Toxicity: We strongly oppose OPP's proposal to create new data requirements for acute and subchronic neurotoxicity and immunotoxicity, as screening for these endpoints can be undertaken more efficiently and appropriately using the ACSA (2005a) paradigm. This approach incorporates functional and extensive neuropathological evaluations into the design of 28- and 90-day studies, and should be further extended to include a set of pathology studies similar to what as is undertaken in OPPTS 870.7800 and the draft "enhanced" OECD 407 study (Freyberger et al., 2003; Cho et al., 2003). Further hazard characterization for these endpoints should focus on improving the mechanistic understanding of effects identified *in vivo*, making maximum use of available *in vitro* studies. A major advantage of *in vitro* methods is that they bypass systemic effects imposed by other organs, which can interact with and confound a chemical's primary effect. This offers the investigator a unique opportunity to assess inherent toxicity to the organ and/or system of interest.

* Endorsement a more efficient study design does not reflect a diminished ethical or scientific opposition to animal testing by the parties to this submission.

Neuroblastoma cell lines have been used most extensively for neurotoxicity studies (Lendahl and McKay, 1990), as these cells form relatively homogeneous populations, proliferate rapidly in chemically defined media, extend neuritic-like processes, and contain precursors for neurotransmitter enzymes (e.g., choline acetyltransferase and tyrosine hydroxylase). Neuroblastomas of human (IMR-31, SK-N-MC) and mouse (Neuro-2a, NB41A3, N1E-115) origin are commercially available for screening neurotoxicants. Both the human SH-SY5Y and the N1E-115 mouse cell lines have been used in insecticide research, and have been proposed as an *in vitro* surrogate for the "hen test" for delayed organophosphate neurotoxicity (50 Fed. Reg. 8540; Claudio, 1992). Specific neural endpoints to consider include changes in neurotransmitter synthesis, uptake, reception or degradation, ion channel alterations, loss of capability for differentiation (neurite outgrowth) and changes in proteins unique to neuronal cells (e.g., neurofilaments, neural-specific enolase) or non-neuronal cells of the nervous system (e.g., glial fibrillary acid protein, fibronectin, S100; Atterwill et al., 1991, 1994; Heijink et al., 1995; Huang et al., 1993; Nister and Westermark, 1994; Oortgiesen et al., 1993; Reuveny and Narahashi, 1991; Rowles et al., 1995; Shea et al., 1989; Taylor et al., 1995; Veronesi, 1992a, 1992b).

Biochemical endpoints are also suitable for *in vitro* neurotoxicity testing given that they can be assayed and collected rapidly, are relatively inexpensive, and are amenable to routine use. Neurotoxicants can target neurotransmitter systems at the level of presynaptic release and/or postsynaptic receptor. Thus, assays for natural enzymes such as acetylcholinesterase, choline acetyltransferase, dopamine □-hydroxylase, glutamic acid decarboxylase, neuron-specific enolase, neurotoxic esterase, tyrosine hydroxylase, and other synaptically relevant enzymes, provide reliable information and are relatively easy to perform. Other endpoints commonly examined in *in vitro* studies have been reviewed by Freshney (1991) and Lendahl and McKay (1990).

With respect to immunotoxicity, human-biology based studies can be conducted on freshly donated blood or tissue samples, primary cell or tissue/organ cultures, or immortalized cell lines. Most types of human immune cells/tissues have been studied *in vitro*, including mixed peripheral blood mononuclear cells (PBMC), NK cells, macrophages, lymphocytes, splenocytes, antigen presenting cells (APC), bone marrow (from human cadaver donors), thymus tissue, and lymph node tissue (e.g., from tonsils). Experiments are performed by adding the test chemical to the immune system cells *in vitro* (human metabolic enzymes can be added in order to ensure that all reactive metabolites of the chemical are tested). Studies have shown high concordance between *in vitro* and *in vivo* findings for this endpoint (Lebrec et al., 1995). Immunotoxicity can be further assessed by measuring endpoints such as cytotoxicity, proliferation, and molecular indicators of immune function; the effects on various immune system cell types can then be separately assessed.

There are numerous *in vitro* systems that have been used for assessing immunotoxicity, but some of the most advanced methods currently in use or development are based on cytokine release. The following three methods hold particular promise:

- A medium throughput system based on human whole blood has been developed that measures the endpoints of cytokine release and cytotoxicity (Langezaal et

al, 2001; 2002). Blood is collected from volunteers and exposed to a xenobiotic. Standard antigen preparations (e.g., LPS, SEB) are used to evoke an immune response, and changes in cytokine release levels can be assayed for a variety of cells present in the sample. This method has the following advantages: human cells are used, so that species extrapolation is avoided; primary cells are used, so that cultivation artifacts are avoided; the amount of test substance required is much less than is necessary for animal testing, so that testing can take place much earlier in the process; the results are quantitative; and effects on different blood cell populations can be tested in a single model. This test has undergone pre-validation studies utilizing certain cytokines as endpoints (IL-1beta and IL-4, produced by monocytes and Th2 lymphocytes respectively), but can be expanded to other cytokines and cell types.

- The "Cell Chip" is an *in vitro* system based on cell lines that have been genetically modified to express enhanced green fluorescent protein (EGFP) under various cytokine promoters (e.g., IL-2, TNF-alpha) (Ringerike 2005; Ulleras, 2005). The system enables chemical exposure effects leading to changes in cytokine expression, and thus immune system perturbation, to be easily assessed in a high throughput manner. Flow cytometry is used to measure fluorescence as well as viability for large numbers of individual cells. This system has undergone method development and evaluation studies. It currently utilizes mouse T-cell, mast cell, and macrophage cell lines, but would be relatively straightforward to migrate to human cells, which would likely be more predictive.

- *In vitro* clonogenic assays have been widely used for years to investigate the viability, proliferation and differentiation of blood cell lineages, their progenitors, and stem cells for preclinical safety screening of drug compounds for myelosuppression. Most notably, the colony-forming unit-granulocyte/macrophage (CFU-GM) *in vitro* assay to predict neutropenia has been integrated into the US National Cancer Institute's drug discovery pathway (Gribaldo, 2002) and has been scientifically validated by international inter-laboratory ECVAM-sponsored studies which found the test to be highly predictive and reproducible (Pessina et al., 2003; 2005). In this assay, human myeloid progenitor cells from donated umbilical cord blood are mixed with a drug candidate and media containing human cytokine GM-CSF in order to determine the effect of the xenobiotic on colony formation. This assay has recently been transferred to a 96-well plate microassay format, dramatically reducing reagent use, work, and cost (Pessina et al., 2004).

A comprehensive overview of available *in vitro* methods for the evaluation of mechanistic and other effects on the endocrine system has recently been published by Bremer et al. (2005).

Reproductive and Developmental Toxicity: We strongly oppose OPP's proposal to require all nonfood-use pesticide chemicals to undergo reproductive toxicity testing in 2-generations of rats. This proposal epitomizes the glaring disregard we have come to expect from OPP's Health Effects Division for both the economic and animal welfare

implications of its testing demands—which in this instance total more than 2,500 animals and $300,000 per substance (OPPTS, 2005; Derelanko and Hollinger, 2002). Moreover, this proposal fails to acknowledge both the tremendous inefficiency and waste in the design of OPPTS 870.3800 (i.e., most of the 1,200 pups from the second generation are not examined and are simply killed and discarded), as well as the existence of combination protocols, which amalgamate several reproduction- and developmental-type toxicity endpoints into a single study. A good example of this is the F1-extended 1-generation study proposed by ACSA (2005b). To the extent that animal use could be reduced by up to 70% by evaluating reproductive toxicity, rat prenatal developmental toxicity, and developmental neuro- and immunotoxicity in a single study rather than separately, we urge OPP to delete and replace the current data requirement and test guideline for the 2-generation reproduction study with this combined study design.*

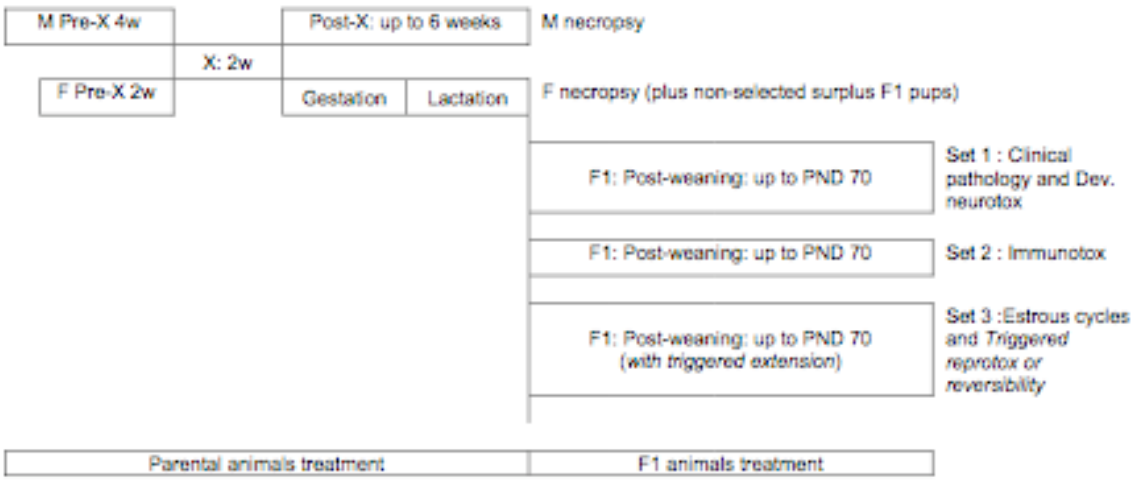

Figure 2—ACSA Proposed F1-Extended 1-Generation Study (ACSA, 2005b)

As documented above and in Schardein (2000), multi-species studies accomplish little more than the wasteful and inefficient consumption of resources without any guarantee of "value added" (i.e., a lower NOAEL). We also strongly oppose OPP's proposal to require all pesticide chemicals to undergo a second developmental toxicity study in rabbits—at a cost of 900 animals and $62,000 (OPPTS, 2005; Derelanko and Hollinger, 2002). We also cannot support the ACSA (2005b) proposal for a stand-alone rabbit developmental study over an above the F1-extended 1-generation study. The question of species sensitivity should be addressed explicitly at the outset of a pesticide hazard assessment using a variety of *in vitro* ADME studies (ACSA, 2005c; Eskes et al., 2005).

One of the most critical parameters in a developmental toxicity study—a chemical's potential to disrupt normal embryonic development—can be assessed *in vitro* using an ECVAM-validated test method (Spielmann et al., 1997; Genschow et al., 2002; ESAC, 2002). In the Embryonic Stem Cell Test (EST), the capacity of the stem cells (rodent-

* Endorsement a more efficient study design does not reflect a diminished ethical or scientific opposition to animal testing by the parties to this submission.

derived ES cell line D3) to differentiate into contracting myocardial cells within 10 days is used to assess the embryotoxic potential of test compounds by light microscopic evaluation. In addition, the effects on the viability of ES cells and differentiated mouse fibroblasts (3T3 cells) are compared. Although the EST does not purport to fully model all of the parameters examined in the OPPTS 870.3700 study, it has immediate applicability as a mechanistic screening tool, and OPP should encourage submission of *in vitro* EST data, which would not only enhance the robustness of the database for developmental toxicity and possibly aid in the interpretation of other study data.

Chronic Toxicity & Carcinogenicity: As stated above, the parties to this submission strongly support OPP's proposal to eliminate the requirement for a chronic (1-year) toxicity study in dogs. We commend Drs. Baetcke, Phang and Dellarco for their rigorous and thoughtful analysis presented in *A Comparison of the Results of Studies on Pesticides from 12- or 24-Month Dog Studies with Dog Studies of Shorter Duration*, and concur with their conclusion that data from 13-week studies are sufficient for establishing a chronic reference dose (RfD) without the application of an additional modifying factor. This review by Baetcke et al. (2005) revealed that of the 304 chronic RfD's established by the Agency as of October 2004, only 116 (38%) have been based on data from dog studies, meaning that six out of every 10 OPP-mandated testing on dogs has had no impact on the determination of chronic RfD. In cost/benefit terms, at least 188 1-year dog studies have been conducted—at a cost of more than $65 million and 6,000 dogs—with *no* commensurate benefit to public health.

In addition to eliminating this data requirement from Part 158, we strongly urge the deletion of the guideline for stand-alone chronic toxicity studies (OPPTS 870.4100), given that chronic toxicity in rodents can be assessed using the combined chronic toxicity/carcinogenicity protocol outlined in OPPTS 870.4300 or, preferably, as part of the F1-extended 1-generation study proposed by ACSA (2005b).* Upon analysis of the study design and group sizes for the extended 1-generation, it appears that approximately 400 F1 pups could fall into the "non-selected surplus" category and go directly to necropsy (assuming standardization of 100 F1 litters to 10 pups/litter and subsequent assignment of 200 pups each to Sets 1-3, leaving approximately 400 "surplus" pups). If this is indeed the case, every effort should be made to assign these animals to a chronic/cancer protocol rather than conducting this study using a separate "batch" of animals. Doing so would reduce animal use by a further 5% relative to carring out separate reproductive, rat developmental, DNT and rat cancer studies and exposes animals to the test agent from gestation on.

As stated above, we strongly support the ACSA (2005a) proposal to eliminate the 2-year mouse carcinogenicity study, on the grounds that "additional information provided by [this] study is of limited value in risk assessment." This finding is consistent with the preponderance of literature in this area (i.e., Pritchard et al., 2002; Gold and Slone, 1993).

* Endorsement a more efficient study design does not reflect a diminished ethical or scientific opposition to animal testing by the parties to this submission.

Genetic Toxicity: We urge OPP to limit its data requirements for genetic toxicity to *in vitro* studies. In particular, we call OPP's attention to draft OECD TG 487 – *In Vitro* Micronucleus Test (OECD, 2004c).

Toxicity to Fish: The parties to this submission are strongly opposed to OPP's proposed expansion of existing aquatic toxicity testing requirements in fish, including:

- Expanding the current CR to require two fish species for greenhouse and indoor use;
- Dropping conditions from the requirement for acute testing on estuarine and marine organisms for outdoor uses; and
- Unconditionally requiring chronic testing on freshwater fish for outdoor uses.

Acute fish studies conducted according to OPPTS (2005) guidelines consume several hundred animals (30-120 fish x 2-3 freshwater + possibly 1 or more saltwater species), with each individual study costing upwards of $3,000 (Derelanko and Hollinger, 2002). The rationale behind OPP's proposal to expand acute fish testing to additional species and outdoor uses (i.e., because "species sensitivity can vary by an order of magnitude or greater") is neither persuasive nor supportable. The extent of species differences can and should be determined through *in vitro* ADME investigations, as discussed above and in ACSA (2005c). Moreover, recent studies call into question the value of acute fish studies in general as a biomarker for aquatic toxicity. For example, Jeram and colleagues (2005) have reported that, when fish LC_{50} values were compared to EC_{50} values in Daphnia and algae for a dataset of 1,439 substances, fish were the most sensitive species in fewer than 15% of instances. Given the marginal "value added" by acute studies in fish, these authors proposed a testing strategy that could markedly reduce animal use for this endpoint. In addition to the limit-test approach proposed by Jeram et al. (2005), we urge OPP to be more flexible in general regarding the data it will accept for this endpoint, and request that the following two *in vitro* systems in particular be recognized as surrogates for acute toxicity studies in fish:

- TETRATOX: In this assay, the protozoan *Tetrahymena pyriformis* is used as a biomarker for acute lethality in fish (Schultz, 1997). The biochemistry and physiology of *T. pyriformis* have been thoroughly investigated since the 1950s, and this assay has been used, in various forms, for aquatic toxicity testing since the 1970s (Sinks and Schultz, 2001). The TETRATOX protocol has been standardized and has undergone a preliminary ring test (Larsen et al.,1997). The German EPA is currently funding a second, more elaborate ring test, with the goal of establishing an OECD TG. In the interim, data generated by TETRATOX demonstrate a consistently high degree of concordance to data from *in vivo* acute studies in fish, which supports the use of this assay as a prospective replacement for toxicity studies in fish (Seward et al., 2001).

- *Dar*T Fish Egg Test: In this assay, fertilised zebrafish (*Danio rerio*) eggs are used as a surrogate for living fish (Schulte and Nagel, 1994; Nagel, 1998, 2002). The reliability and relevance of the *Dar*T test have been confirmed through an international, multi-laboratory validation study run by the German EPA; predictions of acute toxicity from the *Dar*T test were highly concordant with *in*

vivo reference data (Schulte et al., 1996). This *in vitro* test has been accepted in Germany as a replacement for the use of fish in the assessment of wastewater effluent (Friccius et al.,1995), and has recently been proposed as a new OECD TG.

Chronic fish toxicity studies can cost up to $38,000 apiece (Derelanko and Hollinger, 2002) and consume an average of 360 animals per chemical tested (OPPTS, 2005). This, despite a recognition by the OECD (2002) that, "… the actual classification for chronic aquatic hazards is based on acute toxicity data, in combination with data on degradability and bioaccumulation, together being considered as surrogates to chronic toxicity values. The GHS only uses chronic toxicity data as additional information, which may remove the need for classification. It is currently not possible within the GHS to assign a positive identification of chronic hazard based directly on a chronic toxicity value. This is because chronic toxicity data are expensive to generate and less well standardised and are thus not readily available for most substances." We urge OPP to follow this international precedent and delete all data requirements for chronic fish testing from Part 158.

Toxicity to Birds: We strongly oppose OPP's proposed expansion of existing data requirements for toxicity studies in birds, including:

- The addition of a third species in acute studies (red-winged blackbirds or other passerine species) as an unconditional requirement, as well as conditionally requiring the testing of end-use products;
- Making subacute dietary toxicity studies an unconditional requirement for terrestrial, forestry, and residential outdoor uses;
- Conditionally requiring subacute dietary testing of end-use products;
- Conditionally requiring subacute dietary testing on a second species for nonfood outdoor residential uses; and
- Making 1-generation reproduction studies an unconditional requirement for terrestrial, aquatic food crop, forestry, residential, and other nonfood outdoor uses.

Notwithstanding the aforementioned animal welfare and economic implications of avian studies, the rationale behind these proposed expansions (i.e., to "address interspecies range of sensitivity," "concerns for potential greater sensitivity of passerines," etc.) is neither persuasive nor supportable. The extent of species differences can and should be determined through *in vitro* ADME investigations, as discussed above and in ACSA (2005c).

* * * * *

LITERATURE CITED

ACSA [ILSI Health and Environmental Sciences Institute Technical Committee on Agricultural Chemical Safety Assessment]. *Systemic Toxicity White Paper*, 67 pp. Washington, DC: HESI (2005a).

ACSA [ILSI Health and Environmental Sciences Institute Technical Committee on Agricultural Chemical Safety Assessment]. *Life Stages White Paper*, 64 pp. Washington, DC: HESI (2005b).

ACSA [ILSI Health and Environmental Sciences Institute Technical Committee on Agricultural Chemical Safety Assessment]. *ADME White Paper*, 59 pp. Washington, DC: HESI (2005c).

AIHC [American Industrial Health Council]. Comments on Draft Toxicology Data Requirements for Assessing Risks of Pesticides to Children's Health – EPA Docket No. OPP-00610. October 7, 1999.

Ames BN and Gold LS. Chemical carcinogenesis: too many rodent carcinogens." *Proceedings of the US National Academy of Sciences* 87, 7772-7776 (1990).

Anger WK. Worksite behavioral research: Results, sensitive methods, test batteries and the transition from laboratory data to human health. *NeuroToxicology* 11, 629-720 (1990).

Anon. *TOPKAT® QSTR Module Summary: Rat Oral LD_{50}*. Website http://www.accelrys.com/products/topkat/ratoral.html. San Diego, CA: Accelrys (2004).

Appleman LM and Feron VJ. Significance of the dog as "second animal species" in toxicity testing for establishing the lowest "no-toxic-effect level." *Journal of Applied Toxicology* 6, 271-279 (1986).

Ashby J and Purchase IF. Will all chemicals be carcinogenic to rodents when adequately evaluated? *Mutagenesis* 8, 489-493 (1993).

Atterwill CK, Simpson MG, Evan RJ, Allen SL and Ray D. Alternative methods and their application to neurotoxicity testing. In: Balls, M., Bridges, J. and Southee, J. (Eds) *Animals and Alternatives in Toxicology, Present Status and Future Prospects*, pp. 121-152, New York, Macmillan Press (1991).

Atterwill CK, Bruinink A, Drejer J, Cuarte E, Abdulla EM, Meredith C, Nicotera P, Regan C, Rodriguez-Farre E, Simpson MG, Smith R, Veronesi B, Vijverberg H, Walum E, and Williams DC. *In vitro* Neurotoxicity Testing. The Report and Recommendations of ECVAM Workshop 3. *Alternatives to Laboratory Animals* 22, 350-362 (1994).

Baert K and De Backer P. Comparative pharmacokinetics of three non-steroidal anti-inflammatory drugs in five bird species. *Comp Biochem Physiol C Toxicol Pharmacol* 134,

25-33 (2003).

Baetcke KP, Phang W and Dellarco V. A Comparison of the Results of Studies on Pesticides from 12- or 24-Month Dog Studies with Dog Studies of Shorter Duration, 45 pp. Washington, DC: EPA Office of Pesticide Programs (2005).

Bailey J, Knight A and Balcombe J. The future of teratology research is *in vitro*. *Biogenic Amines* 19, 97-145 (2005).

Balcombe JP, Barnard ND and Sandusky C. Laboratory routines cause animal stress. *Contemporary Topics in Laboratory Animal Science* 43, 42-51 (2004).

Balls M. Why modification of the LD50 test will not be enough. *Laboratory Animals* 25, 198-206 (1991).

Barratt MD, Basketter DA, Chamberlain M, Admans GD and Langowski JJ. An expert system rulebase for identifying contact allergens. *Toxicology In Vitro* 8, 1053-1060 (1994).

Basketter DA, Scholes EW, Chamberlain M and Barratt MD. An alternative strategy to the use of guinea pigs for the identification of skin sensitization hazard. *Food and Chemical Toxicology* 33, 1051-1056 (1995).

Basketter DA and Chamberlain M. Validation of skin sensitization assays. *Food and Chemical Toxicology* 33, 1057-1059 (1995).

Basketter DA, Chamberlain M, Griffiths HA, Rowson M, Whittle E and York M. The classification of skin irritants by human patch test. *Food and Chemical Toxicology* 35, 845-852 (1997).

BEAD [Biological and Economic Analysis Division, US EPA Office of Pesticide Programs]. Economic Analysis of the Proposed Change in Data Requirements Rule for Conventional Pesticides, 78 pp. Washington, DC: EPA (2004).

Bernauer U, Glatt H-R, Heinrich-Hirsh B, Liu Y, Muckel E, Vieth B and Gundert-Remy U. (2003). Heterologous expression of mouse cytochrome P450 2e1 in V79 cells: construction and characterization of the cell line and comparison with V79 cell lines stably expressing rat P450 2E1 and human P450 2E1. *Alternatives to Laboratory Animals* 31, 21-30 (2003).

Blot C, Lebrec H, Roger R, Bohuon R and Pallardy M. Immune parameters are affected differently after cyclosporine A exposure in Fischer 344 rats and B6C3F1 mice: implications for immunotoxicology. *Toxicology* 94, 231-245 (1994).
Borenfreund E and Puerner J. A simple quantitative procedure using monolayer cultures for cytotoxicity assays (HTD/NR-90). *Journal of Tissue Culture Methods* 9, 7-9 (1984).

Borenfreund E and Puerner J. Toxicity determination *in vitro* by morphological alternations and neutral red absorption. *Toxicology Letters* 24, 119-124 (1985).

Box RJ and Spielmann H. Use of the dog as non-rodent test species in the safety testing schedule associated with the registration of crop and plant protection products (pesticides): present status. *Archives of Toxicology* [e-pub ahead of print] (2005).

Bremer S, Cortvrindt R, Datson G, Eletti B, Mantovani A et al. 3.11. Reproductive and Developmental Toxicity. In: Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Prospects. *Alternatives to Laboratory Animals* 33 (Suppl. 1), 183-209 (2005).

Bucher JR and Portier C. Human carcinogenic risk evaluation, Part V: The national toxicology program vision for assessing the human carcinogenic hazard of chemicals. *Toxicological Sciences* 82, 363-366 (2004).

Buelke-Sam J, Kimmel CJ and Adams J. Design considerations in screening for behavioral teratogens: Results of the collaborative behavioral teratology study. *Neurobehavioral Toxicology and Teratology* 7, 537-789 (1985).

Bugelski PJ and Treacy G. Predictive power of preclinical studies in animals for the immunogenicity of recombinant therapeutic proteins in humans. *Current Opinions in Molecular Therapy* 6, 10-16 (2004).

Byrd III DM, Crouch EAC and Wilson R. Do Mouse Liver Tumors Predict Rat Tumors? A Study of Concordance Between Tumors Induced at Different Sites in Rats and Mice. In: *Mouse Liver Carcinogenesis: Mechanisms and Species Comparisons*, pp 19-41. Alan R. Liss, Inc. (1990).

Callard RE and Turner MW. Cytokines and Ig switching: evolutionary divergence between mice and humans. *Immunolology Today*. 11, 200-203 (1990).

Carter RO and Griffith JF. Experimental bases for the realistic assessment of safety of topical agents. *Toxicology and Applied Pharmacology* 7 (Suppl. 2), 60-73 (1965).

CDT [National Research Council Committee on Developmental Toxicology]. *Scientific Frontiers in Developmental Toxicology and Risk Assessment*, 327 pp. Washington, DC: National Academy Press (2000).

Cho SD, Kim JH, Kim DY, Lee YS and Kang KS. Pre-validation study for OECD enhanced guideline 407 protocol by gavage for 4 weeks for propylthiouracil and tamoxifen. *Toxicology Letters* 144, 195-204 (2003).

Claudio L. An analysis of the U.S. Environmental Protection Agency of *in vitro* testing in neurotoxicology *Regulatory Toxicology and Pharmacology* 16, 202-212 (1992).

Claudio L, Kwa WC, Russell AL and Wallinga D. Testing methods for developmental neurotoxicity of environmental chemicals. *Toxicology and Applied Pharmacology* 164, 1-14 (2000).

Clemendson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC et al. MEIC evaluation of acute systemic toxicity: Part I. Methodology of 68 *in vitro* toxicity assays used to test the first 30 reference chemicals. *Alternatives to Laboratory Animals* 24, 251-272 (1996).

CLS [National Academy of Sciences, Commission on Life Sciences]. Methyl Bromide Risk Characterization in California. Washington, DC: National Academy Press (2000), p. 60.

Coecke S, Blaauboer BJ, Elaut G, Freeman S, Freidig A et al. 3.9. Toxicokinetics and Metabolism. In: Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Prospects. *Alternatives to Laboratory Animals* 33 (Suppl. 1), 147-175 (2005).

Combes RD. The *in vivo* relevance of *in vitro* genotoxicity assays incorporating enzyme activation systems. In: GG Gibson (Ed.) *Progress in Drug Metabolism* Vol. 13, pp. 295-321. London, UK: Taylor and Francis Ltd. (1992).

Conolly, RB, Beck, BD and Goodman, JI. Forum: Stimulating research to improve the scientific basis of risk assessment. *Toxicological Sciences* 49, 1-4 (1999).

Cooper Rees D, Francis EZ and Kimmel CA. Scientific and regulatory issues relevant to assessing risk for developmental neurotoxicity: An overview. *Neurotoxicology and Teratology* 12, 171-181 (1990).

Cory-Slechta DA, Crofton KM, Foran JA, Ross JF, Sheets LP, Weiss B and Mileson B. Methods to Identify and Characterize Developmental Neurotoxicity for Human Health Risk Assessment: Behavioral Effects. *Environmental Health Perspectives* 109 (Suppl. 1), 79-91 (2001).

Crabbe JC, Wahlesten D and Dudek BC. Genetics of mouse behavior: Interactions within laboratory environment. *Science* 284, 1670-1672 (1999).

Crofton KM, Makris SL, Sette WF, Mendez E and Raffaele KC. A qualitative retrospective analysis of positive control data in developmental neurotoxicity studies. *Neurotoxicology and Teratology* 26, 345-352 (2004).

Cunningham ML. A mouse is not a rat is not a human: Species differences exist. *Toxicological Science* 70, 157-158 (2002).

Davies RE, Harper KH and Kynoch SR. Inter-species variation in dermal reactivity. *J Soc Cosmet Chem* 23, 371-381 (1972).

Davis JM, Otto DA, Weil DE and Grant LD. The comparative developmental neurotoxicity of lead in humans and animals. *Neurotoxicology and Teratology* 12, 215-229 (1990).

Derelanko MJ and Hollinger MA (Eds.). *Handbook of Toxicology*, Second Ed, 1414 pp. Washington, DC: CRC Press (2002).

Dieke SH and Richter CP. Acute toxicity to rats in relation to age, diet, strain and species variation. *Journal of Pharmacology and Experimental Therapeutics* 83, 195-202 (1945).

Dierickx P. Cytotoxicity testing of 114 compounds by determination of the protein content in Hep G2 cell cultures. *Toxicology In Vitro* 3, 189-193 (1989).

Dorman DC, Allen SL, Byczkowski JZ, Claudio L, Fisher, JE Jr, Fisher JW, Harry GJ, Li, AA, Makris SL, Padilla S, Sultatos LG and Mileson BE. Methods to identify and characterize developmental neurotoxicity for human health risk assessment: pharmacokinetic and pharmacodynamic considerations. *Environmental Health Perspectives* 109 (Suppl. 1), 101-11 (2001).

Dreau D, Foster M, Morton DS, Fowler N, Kinney K and Sonnenfeld G. Immune alterations in three mouse strains following 2-deoxy-D-glucose administration. *Physiological Behavior* 70, 513-520 (2000).

EC/HC [Environment Canada/Health Canada]. Guidelines for the Notification and Testing of New Substances: Chemicals and Polymers, 228 pp. Ottawa: Government of Canada (2001).

Ekwall B and Sandström B. Combined toxicity to HeLa cells of 30 drug pairs, studied by a two-dimensional microtitre method. *Toxicology Letters* 2, 285-292 (1978).

Ekwall B, Barile FA, Castano A, Clemendson C, Clothier RH, Dierickx P, Ekwall B, Ferro M, Fiskesjö G, Garza-Ocanas L, Gómez-Lechón MJ, Gülden M, Hall T, Isomaa B, Kahru A, Kerszman G, Kristen U, Junimoto M, Kärenlampi S, Lewan L, Loukianov A, Ohno T, Persoone G, Romert L, Sawyer TW, Shrivastava R, Segner H, Stammati A, Tanaka N, Valentino M, Walum E and Zucco F. MEIC evaluation of acute systemic toxicity. Part VI. The prediction of human toxicity by rodent LD_{50} values and results from 61 *in vitro* methods. *Alternatives to Laboratory Animals* 26 (Suppl. 2), 617-658 (1998).

Elsner J. Animal models for human behavioral deficiencies during development. *Neurotoxicology* 12, 65-76 (1992).

Ennever FK, Noonan TJ and Rosenkranz HS. The predictivity of animal bioassays and short-term genotoxicity tests for carcinogenicity and non-carcinogenicity in humans. *Mutagenisis* 2, 73-78 (1987).

Ennever FK and Lave LB. Implications of the lack of accuracy of the lifetime rodent bioassay for predicting human carcinogenicity. *Regulatory Toxicology and Pharmacology* 38, 52-57 (2003).

EPA [U.S. Environmental Protection Agency]. Supplemental Acute Toxicity Protocol. Website http://www.epa.gov/chemrtk/toxprtcl.htm. Washington, DC: EPA (2001).

EPA [U.S. Environmental Protection Agency]. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency, 57 pp. Washington, DC: EPA (2002a).

EPA [U.S. Environmental Protection Agency]. Organophosphate Pesticides: Revised OP Risk Assessment. http://www.epa.gov/pesticides/cumulative/rra-op. EPA: Washington, DC (2002b).

EPA [U.S. Environmental Protection Agency]. Determination of the Appropriate FQPA Safety Factor(s) in the Organophosphorus Pesticide Cumulative Risk Assessment. Evaluation of Sensitivity and Susceptibility to the Common Mechanism of Toxicity, Acetylcholinerase Inhibition, 55 pp. Washington, DC: EPA (2002c).

EPA [U.S. Environmental Protection Agency]. ICR No. 2055.01 – Data Submissions for the Voluntary Children's Chemical Evaluation Program (VCCEP), 103 pp. Washington, DC: EPA (2002d).

EPA [U.S. Environmental Protection Agency]. 2003-2008 EPA Strategic Plan, 234 pp. Washington, DC: EPA (2003).

EPA [U.S. Environmental Protection Agency]. Order Denying Objections to Issuance of Tolerances; Final Rule. 70 Fed. Reg. 46706, August 10 (2005).

ESAC [ECVAM Scientific Advisory Committee]. *Statement on the Scientific Validity of the Embryonic Stem Cell Test (EST) – An In Vitro Test for Embryotoxicity*. Website: http://ecvam.jrc.it. Ispra, Italy: ECVAM (2002).

Eskes C, Bessou S, Bruner L, Curren R, Harbell J, et al. 3.5. Eye Irritation. In: Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Prospects. *Alternatives to Laboratory Animals* 33 (Suppl. 1), 47-81 (2005).

Festing MFW. Fat rats and carcinogenesis screening. *Nature* 388, 321 (1997).

Fischer R. CropLife America Comments to Avian Two Generation Toxicity Test Detailed Review Paper. Public comments presented to EPA Endocrine Disruptor Methods Validation Subcommittee, Golden, CO, 18-20 Aug 2003.

Francis, EZ, Kimmel CA and Rees DC. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity: Summary and implications. *Neurotoxicology and Teratology* 12, 285-292 (1990).

Franklin, CA. Policy Issues Related to the Food Quality Protection Act – EPA Docket No. OPP-00610. Health Canada Pest Management Regulatory Agency, August 25, 1999.

Freshney RI. *Cultures of Animal Cells: A Manual of Basic Techniques*, 2nd Ed. New York, NY: Wiley-Liss (1991).

Freyberger A, Andrews P, Hartmann E, Elben R, Loof I, Schmidt U, Temerowski M, Folkerts A, Becka M, Stahl B and Kayser M. Testing of endocrine active substances using an enhanced OECD test guideline 407: Experiences from studies on flutamide and ethinylestradiol. *Pure Applied Chemistry* 75, 2483-2489 (2003).

Friccius T, Schulte C, Ensenbach U, Seel P and Nagel R. Der Embryotest mid dem Zebrabärbling – eine neue Möglichkeit zur Prüfung und Bewertung der Toxizität von Abwasserproben. *Vom Wasser* 84, 407-418 (1995).

Friedman JM and Polifka JE. *Teratogenic Effects of Drugs: A Resource for Clinicians: TERIS*, 703 pp. Baltimore, MD: Johns Hopkins University Press (1994).

Fung VA, Barrett JC, et al. The carcinogenesis bioassay in perspective: application in identifying human cancer hazards. *Environmental Health Perspectives* 103, 680-683 (1995).

Gaylor DW. Are tumor incidence rates from chronic bioassays telling us what we need to know about carcinogens? *Regulatory Toxicology and Pharmacology* 41, 128-133 (2005).

Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, et al. The ECVAM international validation study on *in vitro* embryotoxicity tests: Results of the definitive phase and evaluation of prediction models. *Alternatives to Laboratory Animals* 30, 151-176 (2002).

Gerber GJ and O'Shaughnessy DO. Comparison of the behavioral effects of neurotoxic and systemically toxic agents: How discriminatory are behavioral tests of neurotoxicity? *Neurobehavioral Toxicology and Teratology* 8, 703-710 (1986).

Gerbracht U and Spielmann H. The use of dogs as second species in regulatory testing of pesticides. I. Interspecies comparison. *Archives of Toxicology* 72, 319-329 (1998).

Gold LS and Slone TH. Prediction of carcinogenicity from two versus four sex-species groups in the carcinogenic potency database. *Journal of Toxicology and Environmental Health* 39, 143-157 (1993).

Goldenthal E. A compilation of LD_{50} values in newborn and adult animals. *Toxicology and Applied Pharmacology* 18, 185-207 (1971).

Goldey ES, O'Callaghan JP, Stanton ME, Barone Jr S and Crofton KM. Developmental neurotoxicity: Evaluation of testing procedures with methylazoxymethanol and methylmercury. *Fundamental and Applied Toxicology* 23, 447-464 (1994).

Gottmann E, Kramer S, Pfahringer B and Helma C. Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environmental Health Perspectives* 109, 509-514 (2001).

Gribaldo L. Haematotoxicology: scientific basis and regulatory aspects. *Alternatives to Laboratory Animals* 30 (Suppl. 2), 111-113 (2002).

Griffin GK. *Status of Alternative Methods in Toxicity Testing*, 124 pp. Ottawa, ON: Joseph F. Morgan Research Foundation (1997).

Grisham JW and Smith GJ. Predictive and mechanistic evaluation of toxic responses in mammalian cell culture systems. *Pharmacological Reviews* 36(Suppl.), 151S-171S (1984).
Hartung T, Pfaller W, Ryan MP, Morin JP, Witkamp R, Villa P, Van Gompel J and Wendel A. *In vitro* testing for chronic hepatic and renal toxicity. *Alternatives to Laboratory Animals* 27 (Special Issue), 415 (1999).

Haley PJ. Species differences in the structure and function of the immune system. *Toxicology* 188, 49-71 (2003).

Hartung T, Sauer A and Wendel A. Testing of immunomodulatory properties *in vitro*. *Dev Biol Stand.* 86, 5-96 (1996).

Haseman JK. A reexamination of false-positive rates for carcinogenesis studies. *Fundamental and Applied Toxicology* 3, 334-339 (1983).

Haseman JK. Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. *Drug Metabolism Review* 32, 169-186 (2000).

Heimann R and Rice RH. Polycyclic aromatic hydrocarbon toxicity and induction of metabolism in cultivated esophageal and epidermal keratinocytes. *Cancer Research* 43, 48-56 (1983).

Heijink E, Bolhuis PA and de Wolff FA. Sensitivity to 2,5-hexanedione of neurofilaments in neuroblastoma cell line SK-N-SH increases during differentiation. *Journal of Neuropathology and Experimental Neurology* 54, 82-90 (1995).

HIARC [EPA Hazard Identification Assessment Review Committee]. Triphenyltin Hydroxide (TPTH) – 1st Report, 16 pp. Washington, DC: EPA. November 13, 1998.

HIARC [EPA Hazard Identification Assessment Review Committee]. Triphenyltin Hydroxide: Reconsideration on the need for a developmental immunotoxicity study and sustaining the requirement for a developmental neurotoxicity study, 3 pp. Washington, DC: EPA. March 23, 2001.

HIARC [EPA Hazard Identification Assessment Review Committee]. Clothianidin (TI-435) – 1st Report, 35 pp. Washington, DC: EPA. January 6, 2003.

HIARC [EPA Hazard Identification Assessment Review Committee]. Clothianidin (TI-435) – 2nd Report, 40 pp. Washington, DC: EPA. March 31, 2003.

HIARC/FQPA-SFC [EPA Hazard Identification Assessment Review Committee and FQPA Safety Factor Committee]. FQPA Safety Factor Recommendations for the Organophosphates. A Combined Report of the HIARC and FQPA-SFC. Health Effects Division, Office of Pesticide Programs, U.S. EPA, August 6, 1998.

Himmelstein MW, Turner MJ, Asgharian B and Bond JA. Metabolism of 1,3-butadiene: inhalation pharmacokinetics and tissue dosimetry of butadiene epoxides in rats and mice. *Toxicology* 113, 306-309 (1996).

Hill EF and Murray HC. Seasonal variation in diagnostic enzymes and biochemical constituents of captive northern bobwhites and passerines. *Comparative Biochemistry and Physiology* 87, 933-940 (1987).

House RV. Theory and practice of cytokine assessment in immunotoxicology. *Methods* 19, 17-27 (1999).

Huang J, Tanii J, Kato K and Hashimoto K. Neuron and glial cell marker proteins as indicators of heavy metal-induced neurotoxicity in neuroblastoma and glioma cell lines. *Archives of Toxicology* 67, 491-496 (1993).

Huff J, Cirvello J, Haseman J and Bucher J. Chemicals associated with site-specific neoplasia in 1394 long-term carcinogenesis experiments in laboratory rodents. *Environmental Health Perspectives* 93, 247-270 (1991).

IARC [International Agency for Research on Cancer]. IARC Technical Report No. 24 – Peroxisome Proliferation and its Role in Carcinogenesis. Website: http://www-cie.iarc.fr/htdocs/iarcpubs/techrep24/contents.html. Lyon, France: IARC (1995).

IARC [International Agency for Research on Cancer]. IARC Scientific Publications No. 147 – Species Differences in Thyroid, Kidney and Urinary Bladder Carcinogenesis. Website: http://www-cie.iarc.fr/htdocs/iarcpubs/pub147/pub147contents.html. (1999).

IARC [International Agency for Research on Cancer]. IARC Technical Publication No. 39 – Predictive Value of Rodent Forestomach and Gastric Neuroendocrine Tumors in Evaluating Carcinogenic Risks to Humans. Website: http://www-cie.iarc.fr/htdocs/iarcpubs/techpub39/contents.html. Lyon, France: IARC (2003).

ICCVAM [Interagency Coordinating Committee on the Validation of Alternative Methods]. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods – NIH Publication No: 03-4508. Research Triangle Park, NC: ICCVAM/NICEATM. (2003).

Itoh K, Suzuki R, Umezu Y, Hanaumi K and Kumagai K. Studies of murine large granular lymphocytes. II. Tissue, strain, and age distributions of LGL and LAL. *Journal of Immunology* 129, 395-405 (1982).

Jeram S, Riego Sintes JM, Halder M, Bariabar Fentanes J, Sokull-Klüttgen B and Hutchinson TH. A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union. *Regulatory Toxicology and Pharmacology* 42, 218-224 (2005).

Johnson FM. Response to Tennant et al.: Attempts to replace the NTP rodent bioassay with transgenic alternatives are unlikely to succeed. *Environmental Molecular Mutagenesis* 37, 89-92 (2001).

Jollow DJ, Thorgeirsson SS, Potter WZ, Hashimoto M and Mitchell JR. Acetaminophen-induced hepatic necrosis. VI. Metabolic disposition of toxic and non-toxic doses of acetaminophen. *Pharmacology* 12, 251-271 (1974).

Kangas L, Gronroos M and Nieminen AL. Bioluminescence of cellular ATP: a new method for evaluating cytotoxicity agents *in vitro*. *Medical Biology* 62, 338-343 (1984).

Kerkvliet NI and Brauner JA. Flow cytometric analysis of lymphocyte subpopulations in the spleen and thymus of mice exposed to an acute immunosuppressive dose of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). *Environmental Research* 52, 146-154 (1990).

Kodell RL, Chen JJ and Gaylor DW. A note on the role of background tumor incidence in risk assessment for carcinogens. *Regulatory Toxicology and Pharmacology* 9, 141-146 (1989).

Kuschner M. The relevance of rodent tumors in assessing carcinogenicity in human beings. *Regulatory Toxicology and Pharmacology* 21, 250-251 (1995).

Langezaal I, Hoffmann S, Hartung T and Coecke S. Evaluation and prevalidation of an immunotoxicity test based on human whole-blood cytokine release. *Alternatives to Laboratory Animals* 30, 581-595 (2002).

Langezaal I, Coecke S and Hartung T. Whole blood cytokine response as a measure of immunotoxicity. *Toxicology In Vitro* 15, 313-318 (2001).

Larsen J, Schultz TW, Rasmussen L, Hoofman R and Pauli W. Progress in an ecotoxicological standard protocol with protozoa: results from a pilot ring test with *Tetrahymena pyriformis*. *Chemosphere* 35, 1023-1041 (1997).

Lave LB, Ennever FK, Rosenkranz HS and Omenn GS. Information value of the rodent bioassay. *Nature* 336, 631-633 (1988).

Lebrec H, Roger R, Blot C, Burleson GR, Bohuon C and Pallardy M. Immuno-toxicological investigation using pharmaceutical drugs. *In vitro* evaluation of immune effects using rodent or human immune cells. *Toxicology* 96, 147-156 (1995).

Lemarbre PJ, Hoidal J, Vesella R and Rinehart J. Human pulmonary macrophage tumor cell cytotoxicity. *Blood* 55, 612-617 (1980).

Lendahl U and McKay RDG. The use of cell lines in neurobiology. *Trends in Neuroscience* 13, 132-137 (1990).

LHASA Limited. DEREK for Windows. Website: http://www.lhasalimited.org. Leeds, UK: LHASA (2005).

Li A. From Rat to Human: Toxicokinetic and Risk Assessment Considerations. Transcript from a presentation delivered to The Toxicology Forum – Summer Meeting 2003. Aspen, CO: Toxicology Forum (2003).

Liebsch M and Spielmann H. Balb/c 3T3 cytotoxicity test. In S O'Hare (Ed.). *Methods and Molecular Biology, Vol. 43, In Vitro Toxicity Testing Protocols*, pp. 177-187. Totowa, NJ: Humana Press (1995).

Lorke D. A new approach to practical acute toxicity testing. *Archives of Toxicology* 54, 275-287 (1983).

Macé K, Offord EA and Pfeifer AMA. Drug metabolism and carcinogen activation studies with human genetically engineered cells. In: JJ Castell and MJ Gómez-Lechón (Eds). *In vitro Methods in Pharmaceutical Research*,, pp. 433-456. London, UK: Academic Press (1997).

Makris S, Raffaele K, Sette W and Seed J. A Retrospective Analysis of Twelve Developmental Neurotoxicity Studies Submitted to the USEPA Office of Prevention, Pesticides and Toxic Substances (OPPTS). Washington, DC: Environmental Protection Agency (1998).

Mattsson JL, Eisenbrandt DL and Doe JE. More Than 10,000 Animals Are Required for the Registration of a Single Pesticide – This Paradigm Must Be Changed. Poster Presented at 2003 Meeting of the Society of Toxicicology (2003).

Miller D. Study Design and Critical Review of Functional and Morphological Endpoints. Transcript from a presentation delivered to The Toxicology Forum – Summer Meeting 2003. Aspen, CO: Toxicology Forum (2003).

Mineau P, Boersma DC and Collins B. An analysis of avian reproduction studies submitted for pesticide registration. *Ecotoxicology and Environmental Safety* 29, 304-329 (1994).

Morrison JK, Quinton RM and Reinter M. The purpose and value of LD_{50} determinations. In: *Modern Trends in Toxicology* (Eds. Boyland E and Goulding R). London: Butterworths (1968).

Müller R. Vergleich der im Tierexperiment und beim Menschen tödlichen Dosen

wichtiger Pharmaka. Diss. Univ. Frankfurt/Main (1948).

Nagel R. Umweltchemikalien und Fische – Beiträge zu einer Bewertung. Habilitationschrift. Mainz: Johannes Gutenberg-Universität (1998).

Nagel R. *Dar*T: the embryo test with the zebrafish *Danio rerio* – a general model in ecotoxicology and toxicology. *ALTEX* 19 (Suppl. 1), 38-48 (2002).

NICEATM [NTP Interagency Center for the Evaluation of Alternative Toxicological Methods]. *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity*, 102 pp. Website http://iccvam.niehs.nih.gov/methods/invidocs/guidance/iv_guide.htm. Research Triangle Park, NC: NIEHS (2001a).

NICEATM [NTP Interagency Center for the Evaluation of Alternative Toxicological Methods]. *Report of the International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity*, 370 pp. Website http://iccvam.niehs.nih.gov/methods/invidocs/finalrpt/finalcov.htm. Research Triangle Park, NC: NIEHS (2001b).

Nixon GA, Tyson CA and Wertz WC. Interspecies comparisons of skin irritancy. *Toxicology and Applied Pharmacology* 31, 481-490 (1975).

NRC [National Research Council]. *Pesticides in the Diets of Infants and Children.* Washington, DC: National Academy Press (1993).

Nyugen BT, Peterson PK, Verbrugh HA, Quie PG and Hoidal JR. Difference in phagocytosis and killing by alveolar macrophages from humans, rabbits, rats, and hamsters. *Infect. Immunol.* 36, 504-509 (1982).

OECD [Organization for Economic Cooperation and Development]. *Mutual Acceptance of Data (MAD).* Website: http://www.oecd.org/document/41/0,2340,en_2649_34379_1890473_1_1_1_1,00.html. Paris, France: OECD (1981).

OECD [Organization for Economic Cooperation and Development]. *Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*, 49 pp. Paris, France: OECD. (1996).

OECD [Organization for Economic Cooperation and Development]. *Proposal for a Draft New Guideline: Acute Dermal Irritation Study in Human Volunteers*, 7 pp. Paris, France: OECD (1997).

OECD [Organization for Economic Cooperation and Development]. *Guideline 425: Acute Oral Toxicity—Up-and-Down Procedure*, 26 pp. Paris, France: OECD (2001).

OECD [Organization for Economic Cooperation and Development]. *Possible Need for the Development of Test Guidelines for Chronic Aquatic Toxicity in Species Other Than Fish*, 5 pp. Paris, France: OECD (2002).

OECD [Organization for Economic Cooperation and Development]. *Guideline 428: Skin*

Absorption – In Vitro Method, 8 pp. Paris: OECD (2004a).

OECD [Organization for Economic Cooperation and Development]. *Guideline 431: In Vitro Skin Corrosion: Human Skin Model Test*, 8 pp. Paris: OECD (2004b).

OECD [Organization for Economic Cooperation and Development]. Draft Proposal for a New Guideline: *In Vitro* Micronucleus Test, 13 pp. Paris: OECD (2004c).

Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Van Deun K, Smith P, Berger B and Heller A. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regulatory Toxicology and Pharmacology* 32, 56-67 (2000).

OMB [White House Office of Management and Budget]. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies. 67 FR 8452. Website: http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf. Washington, DC: OMB (2002).

Oortgiesen M, Van Kleef RGDM, Zwart R, Van den Beukel I and Vijverberg HPM. Differential neurotoxicological effects of lead on voltage-dependent and receptor-operated ion channels. *Neurotoxicology* 14, 87-96 (1993).

OPPTS [EPA Office of Prevention, Pesticides and Toxic Substances]. Harmonized Test Guidelines. Website http://www.epa.gov/opptsfrs/home/guidelin.htm. Washington, DC: EPA (2005).

Paul J. In: *Cell and Tissue Culture*, 430 pp. Edinburgh, Scotland: Churchill-Livingstone (1975).

Paxton JW. The allometric approach for interspecies scaling of pharmacokinetics and toxicity of anti-cancer drugs. *Clinical and Experimental Pharmacology and Physiology* 22, 851-854 (1995).

Pessina A, Malerba I and Gribaldo L. Hematotoxicity testing by cell clonogenic assay in drug development and preclinical trials. *Curr Pharm Des.* 11, 1055-1065 (2005).

Pessina A, Croera C, Bayo M, Malerba I, Passardi L, Cavicchini L, Neri MG and Gribaldo L. A methylcellulose microculture assay for the in vitro assessment of drug toxicity on granulocyte/macrophage progenitors (CFU-GM). *Alternatives to Laboratory Animals* 32, 17-23 (2004).

Pessina A, Albella B, Bayo M, Bueren J, Brantom P, Casati S, Croera C, Gagliardi G, Foti P, Parchment R, Parent-Massin D, Schoeters G, Sibiril Y, Van Den Heuvel R and Gribaldo L. Application of the CFU-GM assay to predict acute drug-induced neutropenia: an international blind trial to validate a prediction model for the maximum tolerated dose (MTD) of myelosuppressive xenobiotics. *Toxicological Sciences* 75, 355-367 (2003).

Phillips, II L, Steinberg M, Maibach H, Akers WA. A comparison of rabbit and human skin responses to certain irritants. *Toxicology and Applied Pharmacology* 21, 369-382 (1972).

Pritchard JB, French JE, Davis BJ and Haseman JK. The role of transgenic mouse models in carcinogen identification. *Environmental Health Perspectives* 111, 444-454 (2002).

Quick DJ and Shuler ML. Use of *in vitro* data for construction of a physiologically based pharmacokinetic model for naphthalene in rats and mice to probe species differences. *Biotechnology Progress* 15, 540-555 (1999).

Reuveny E and Narahashi T. Potent blocking action of lead on voltage-activated calcium channels in human neuroblastoma cells SH-SY5Y. *Brain Research* 545, 312-314 (1991).

RfD/RfC Panel [U.S. EPA Reference Dose/Reference Concentration Technical Panel]. *A Review of the Reference Dose and Reference Concentration Processes*, 192 pp. Prepared for the Risk Assessment Forum, U.S. EPA, Washington DC (December 2002).

Rice DC, Evangelista de Duffard AM, Duffard R, Iregnen A and Satoh H. Lessons for neurotoxicity from selected model compounds. *Environmental Health Perspectives* 104(Suppl. 12), 205-215 (1996).

Ringerike T, Ulleras E, Volker R, Verlaan B, Eikeset A, Trzaska D, Adamczewska V, Olszewski M, Walczak-Drzewiecka A, Arkusz J, van Loveren H, Nilsson G, Lovik M, Dastych J and Vandebriel RJ. Detection of immunotoxicity using T-cell based cytokine reporter cell lines ("Cell Chip"). *Toxicology* 206, 257-272 (2005).

Robinson MK, McFadden JP and Basketter DA. Validity and ethics of the human 4-h patch test as an alternative method to assess acute skin irritation potential. *Contact Dermatitis* 45, 1-12 (2001).

Robinson MK, Cohen C, de Brugerolle de Fraissinette A, Ponec M, Whittle E, Fentem JH. Non-animal testing strategies for assessment of the skin corrosion and skin irritation potential of ingredients and finished products. *Food and Chemical Toxicology* 40, 573-592 (2002).

Rosenkranz HS and Cunningham AR. Lack of predictivity of the rat lethality (LD_{50}) assay for ecological and human health effects. *Alternatives to Laboratory Animals* 33, 9-19 (2005).

Rowles TK, Song X and Ehrich M. Identification of endpoints affected by exposure of human neuroblastoma cells to neurotoxicants at concentrations below those that affect cell viability. *In Vitro Toxicology* 8, 3-13 (1995).

Salsburg D. The lifetime feeding study in mice and rats--an examination of its validity as a bioassay for human carcinogens. *Fundamental and Applied Toxicology* 3, 63-67

(1983).

Sanderson DM and Earnshaw CG. Computer prediction of possible toxic action from chemical structure: the DEREK system. *Human and Experimental Toxicology* 10, 261-273 (1991).

SAP [U.S. EPA Scientific Advisory Panel]. A set of scientific issues being considered by the agency in connection with the use of FQPA 10X safety factor to address special sensitivity of infants and children to pesticides – Final report. Washington, DC (March 1998).

SAP [U.S. EPA Scientific Advisory Panel]. SAP Report No. 99-01 – Report of the FIFRA Scientific Advisory Panel Meeting, December 8, 1998 – A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: A Retrospective Analysis of Developmental Neurotoxicity Studies. Washington, DC (January 1999).

SAP [U.S. EPA Scientific Advisory Panel]. SAP Report No. 99-03 – Report of the FIFRA Scientific Advisory Panel Meeting, May 25-27, 1999 – Sets of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: Office of Pesticide Programs Policy for the Use of the FQPA 10x Safety Factor. Washington, DC (May 1999).

Schaeffer DJ and Beasley VR. Ecosystem health. II. Quantifying and predicting ecosystem effects of toxic chemicals: can mammalian testing be used for lab-to-fiend and field-to-lab extrapolations? *Regulatory Toxicology and Pharmacology* 9, 296-311 (1989).

Schardein JL. *Chemically Induced Birth Defects*, 3rd Ed. Rev, 1109 pp. New York: Marcel Dekker (2000).

Schulte C and Nagel R. Testing acute toxicity in the embryo in zebrafish, *Brachydanio rerio*, as an alternative to the acute fish test: preliminary results. *Alternatives to Laboratory Animals* 22, 12-19 (1994).

Schulte C, Bachmann J, Fliedner A, Meinelt T and Nagel R. Testing acute toxicity in the embryo of zebrafish (*Brachydanio rerio*) – an alternative to the fish acute toxicity test. Proceedings of the 2nd World Congress on Alternatives and Animal Use in the Life Sciences. Utrecht, The Netherlands (1996).

Schultz TW. TETRATOX: *Tetrahymena pyriformis* population growth impairment endpoint: a surrogate for fish lethality. *Toxicological Methods* 7, 289-309 (1997).

Sefarin JA. Avian species differences in the intestinal absorption of xenobiotics (PCB, dieldrin, Hg2+). *Comparative Biochemistry and Physiology* 78, 491-496 (1984).

Selgrade MK, Cooper KD, Devlin RB, van Loveren H, Biagini RE and Luster MI. Immunotoxicity--bridging the gap between animal research and human health effects. *Fundamental and Applied Toxicology* 24, 13-21 (1995).

Seward JR, Sinks GD and Schultz TW. Reproducibility of toxicity across mode of toxic action in the *Tetrahymena* population growth impairment assay. *Aquatic Toxicology* 53, 33-47 (2001).

Shea TB, Clarke JF, Wheelock TR, Paskevich PA and Nixon RA. Aluminum salts induce the accumulation of neurofilaments in perikarya of NB2a/dl neuroblastoma. *Brain Research* 492, 53-64 (1989).

Sheets L. Experience Conducting Developmental Neurotoxicity Studies. Transcript from a presentation delivered to The Toxicology Forum – Summer Meeting 2003. Aspen, CO: Toxicology Forum (2003).

Shepard TH. *Catalog of Teratogenic Agents*, 9th Ed, 593 pp. Baltimore, MD: Johns Hopkins University Press (1998).

Sinks GD and Schultz TW. Correlation of *Tetrahymena* and *Pimephales* toxicity: evaluation of 100 additional compounds. *Environmental Toxicology and Chemistry* 20, 917-921 (2001).

Smialowicz RJ, Riddle MM, Williams WC and Diliberto JJ. Effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) on humoral immunity and lymphocyte subpopulations: differences between mice and rats. *Toxicology and Applied Pharmacology* 124, 248-256 (1994a).

Smialowicz RJ, Riddle MM and Williams WC. Species and strain comparisons of immunosuppression by 2-methoxyethanol and 2-methoxyacetic acid. *International Journal of Immunopharmacology* 16, 695-702 (1994b).

Smialowicz RJ, Riddle MM, Williams WC, Copeland CB, Luebke RW and Andrews DL. Differences between rats and mice in the immunosuppressive activity of 2-methoxyethanol and 2-methoxyacetic acid. *Toxicology* 74, 57-67 (1992).

Spielmann H and Liebsch M. *INVITOX Protocol No. 46: Balb/c 3T3 NRU Cytotoxicity Test.* European Commission GD-JRC, ECVAM, SIS Database. Website http://ecvam.jrc.it. Ispra, Italy: ECVAM (1992).

Spielmann H, Genschow E, Liebsch M and Halle W. Determination of the starting dose for acute oral toxicity (LD50) testing in the Up and Down Procedure (UDP) from cytotoxicity data. *Alternatives to Laboratory Animals* 27, 957-966 (1999).

Spielmann H, Pohl I, Döring B, Liebsch M and Moldenhauer F. The embryonic stem cell test (EST): an *in vitro* embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitro Toxicology* 10, 119-127 (1997).

Spielmann H and Gerbracht U. The use of dogs as second species in regulatory testing of pesticides. Part II: subacute, subchronic and chronic studies in the dog. *Archives of Toxicology* 75, 1-21 (2001).

Stanton ME and Spear LP. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity – Workgroup 1 report. *Neurotoxicology and Teratology* 12, 261-267 (1990).

Taylor D, Jortner BS, Walton A and Ehrich M. An immunohistochemical study of the cytoskeleton of SH-SY5Y human neuroblastoma cells, and the effects of acrylamide. *In Vitro Toxicology* 8, 323-333 (1995).

Tilson HA, Jacobson JL and Rogan WJ. Polychlorinated biphenyls and the developing nervous system: Cross-species comparisons. *Neurotoxicology and Teratology* 12, 239-248 (1990).

Tilson HA. Study design considerations in developmental neurotoxicology. *Neurotoxicology and Teratology* 14, 199-203 (1992).

Tilson HA. The concern for developmental neurotoxicology: Is it justified and what is being done about it? *Environmental Health Perspectives* 103 (Suppl 6), 147-151 (1995).

Tilson HA. Neurotoxicology risk assessment guidelines: Developmental neurotoxicology. *NeuroToxicology* 21, 189-194 (2000).

Trevan JW. The error of determination of toxicity. *Proceedings of the Royal Society (London), Series B* 101, 483-514 (1927).

Trinchieri G and Perussia B. Human natural killer cells: biologic and pathologic aspects. *Lab Invest.* 50, 489-513 (1984).

Trosko JE and BL. The emperor wears no clothes in the field of carcinogen risk assessment: ignored concepts in cancer risk assessment. *Mutagenesis* 20, 81-92 (2005),

Tsuchiya S, Yamabe M, Yamaguchi Y, Kobayashi Y, Konno T and Tada K. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *International Journal of Cancer.* 26, 171-176 (1980).

Tyl RW and Sette WF. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity, work group III report: Weight of evidence and quantitative evaluation of developmental neurotoxicity data. *Neurotoxicology and Teratology* 12, 275-280 (1990).

Ulleras E, Trzaska D, Arkusz J, Ringerike T, Adamczewska V, Olszewski M, Wyczolkowska J, Walczak-Drzewiecka A, Al-Nedawi K, Nilsson G, Bialek-Wyrzykowska U, Stepnik M, Loveren HV, Vandebriel RJ, Lovik M, Rydzynski K and Dastych J. Development of the "Cell Chip": a new *in vitro* alternative technique for immunotoxicity testing. *Toxicology* 206, 245-256 (2005).

Van Loveren H, De Jong WH, Vandebriel RJ, Vos JG and Garssen J. Risk assessment and immunotoxicology. *Toxicol Letters* 102-103, 261-265 (1998).

Veronesi B. *In vitro* screening batteries for neurotoxicants. *Neurotoxicology* 13, 185-196 (1992a).

Veronesi B. The use of cell culture for evaluating neurotoxicity. In: Tilson, H. and Mitchell, C. (Eds.) *Target Organ Toxicology Series: Neurotoxicology*, pp. 21-49, New York, NY: Raven Press (1992b).

Wakuri S, Izumi J, Sasaki K, Tanaka N and Ono H. Cytotoxicity study of 32 MEIC chemicals by colony formation and ATP assays. *Toxicology In Vitro* 7, 517-521 (1993).

Weil CS and Scala RA. Study of intra- and inter-laboratory variability in the results of rabbit eye and skin irritation tests. *Toxicology and Applied Pharmacology* 19, 276-360 (1971).

Werner KL. Companies Question Agency's Changes for Developmental Neurotoxicity Study. 24 BNA 535 (March 20 2000).

Worth A and Balls M (Eds.). Alternative (non-animal) methods for chemicals testing: Current status and future prospects. *Alternatives to Laboratory Animals* 30 (Suppl. 1), 1-125 (2002).

Zbinden G and Flury-Roversi M. Significance of the LD50 test for the toxicological evaluation of chemical substances. *Archives of Toxicology* 47, 77-99 (1981).

Zeiger E. Carcinogenicity of mutagens: Predictive capability of the Salmonella mutagenesis assay for rodent carcinogenicity. *Cancer Research* 47, 1287-1296 (1987).