Biogenic Amines, Vol. 19, No. 2, pp. 97–145 (2005)

Also available online - www.vsppub.com

The future of teratology research is *in vitro*

JARROD BAILEY [1,*], ANDREW KNIGHT [2] and JONATHAN BALCOMBE [3]

[1] *School of Surgical and Reproductive Sciences, 3rd Floor, Leech Building, Medical School, University of Newcastle upon Tyne, NE2 4HH, UK*
[2] *Independent consultant*
[3] *Physicians Committee for Responsible Medicine, 5100 Wisconsin Ave., Suite 400, Washington, DC 20016, USA*

Received 1 November 2004; accepted 28 December 2004

Abstract—Birth defects induced by maternal exposure to exogenous agents during pregnancy are preventable, if the agents themselves can be identified and avoided. Billions of dollars and man-hours have been dedicated to animal-based discovery and characterisation methods over decades. We show here, via a comprehensive systematic review and analysis of this data, that these methods constitute questionable science and pose a hazard to humans. 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. Despite strict validation criteria that animal-based teratology studies would fail to meet, three *in vitro* alternatives have done so. The embryonic stem-cell test (EST) is the best of these. We argue that the poor performance of animal-based teratology alone warrants its cessation; it ought to be replaced by the easier, cheaper and more repeatable EST, and resources made available to improve this and other tests even further.

Keywords: Teratogen; teratology; developmental toxicology; birth defects; thalidomide; animal model.

INTRODUCTION

Teratology, the study of abnormal prenatal development and congenital malformations induced by exogenous chemical or physical agents, continues to be a burgeoning area of medical research in the quest for the eradication of preventable birth defects. Identification of agents with teratogenic potential from the plethora of drugs and chemicals that human beings come into contact with in their everyday environment is crucial. Although only some 10% of congenital anomalies are thought to be caused by teratogens (Brent, 1995) representing roughly one in every

*To whom correspondence should be addressed. E-mail: jarrod.bailey@ncl.ac.uk

thousand live births, they compromise the quality of life for millions of individuals worldwide and cost billions of dollars in health care every year. Knowledge of the most hazardous substances would enable medical professionals and would-be mothers to minimize foetal exposure to them, helping to achieve the laudable goal of abolishing teratogen-induced malformations.

The burden of this goal currently rests heavily upon animal-based testing. In this review, we examine the dogma of animal-oriented teratology, consider its positive and negative aspects, cite specific examples with regard to inter-species concordance or discordance and extrapolation to humans, and assess alternatives to the use of animals from a scientific standpoint.

HISTORY AND PRINCIPLES OF TERATOGENESIS

Teratology as a science was born in the 1920s and 1930s, when the birth of malformed piglets from mothers fed an experimental diet high in fat or deficient in vitamin A elicited the shocking realization that the conceptus was not, as had been believed, in a privileged and highly protected position when within the mother's 'impervious womb', but was susceptible to environmental conditions with potentially serious effects (Hale, 1933; Schardein, 1993). All of these piglets suffered a variety of malformations, predominantly a lack of eyes. Subsequent evidence came to light over the next two decades; correlation of particular birth defects of children with maternal Rubella infection in 1941 (Gregg, 1941) and with environmental mercury contamination in 1956 (Igata, 1993); malformed rats born following the inclusion of a growth-inhibiting amino-acid mimic in their mothers' diet (Murphy and Karnovsky, 1956), and malformed children born following failed aminopterin-induced abortions (a substance that shows no teratogenic effect in other primates, mice or cats; Thiersch, 1956). Then, infamously, the thalidomide catastrophe in 1961 (McBride 1961; Lenz 1966) represented the first case of a substance producing minimal toxicity in adults but considerable toxicity in the developing human embryo. The characteristic malformations of thalidomide embryopathy subsequently manifested themselves in 20–30% of children born to mothers who had taken the drug at any point during the four years it had been available; in some cases, even one single low dose was sufficient (Toms, 1962). Thalidomide is widely believed to be the catalyst that prompted regulatory agencies such as the United States Food and Drug Administration (FDA) to instigate requirements for new drugs to be thoroughly tested on animals prior to approval for marketing. Since this time, a number of basic principles and concepts applicable to teratology research have been formulated. The confounding nature of many of the results from these animal-based studies has required these principles to be elaborated and revised (Wilson, 1977; Finnell, 1999). Many variables have been found to interfere with interspecies and animal-human comparisons, and these must be considered when designing developmental and reproductive toxicology studies (Nielsen *et al*., 2001; Palmer, 1986). They can be summarized as follows.

Principles of teratogenesis

Susceptibility to teratogenesis depends on the genotype of the conceptus and how it interacts with the environment. Inter- and intra-species variability (in terms of susceptibility to a teratogen and the resultant phenotype) is clearly evident in all teratology studies, both qualitatively and quantitatively. These differences may be due to the genetic constitution of species and individuals, environmental factors, differences in metabolic pathways and products and placental properties (Schardein, 1993).

Susceptibility to a teratogenic agent varies with the developmental stage at which the exposure occurs. This was first evidenced by the fact that it was the time of treatment, rather than the dose, that was important in thalidomide teratogenesis (Wilson, 1972). The period of maximum sensitivity to a teratogen corresponds to the 'critical period of organogenesis' (from days 22–55 of gestation in humans): earlier exposure during neurulation often damages the central nervous system (CNS), whereas later exposure may cause urogenital and growth problems. Prior to the critical period of organogenesis, the embryo will either continue to develop normally or spontaneously abort: after organogenesis, the foetus becomes less susceptible to teratogenic effects, though the CNS continues to mature and form synapses, which renders it vulnerable to teratogens throughout gestation; some teratogenic agents cause malformations when administered before organogenesis because they require time to become active (such as actinomycin D and cyclophosphamide) (Wilson, 1972; Wilson and Fouts, 1966).

Teratogenic agents act in specific ways (mechanisms) on developing cells and tissues to initiate abnormal embryogenesis (pathogenesis). Teratogenic pathogenesis often involves changes in apoptosis, biosynthesis and morphogenesis (i.e. breakdown of an otherwise normal development process, such as through vasoconstriction; Wilson, 1973a, b), via mechanisms including mitotic interference, alterations in RNA or protein synthesis, deficiencies in substrates, precursors or energy sources, alterations in membrane transport processes, modifications of the cell surface or matrix that lead to altered cell migration, or osmolar imbalances. Phenotypic changes are not necessarily specific to each teratogen, nor do particular teratogens always induce the same malformations.

The final manifestations of abnormal development are death, malformation, growth retardation and functional disorder. For many teratogenic agents, any of the above possible manifestations can be realized depending on the time of embryonic or foetal exposure and the amount or dose encountered. For example, insult during pre-implantation usually results in embryonic death; during early organogenesis, malformations of the CNS, eyes and limbs occur; later in organogenesis, the ears, external genitalia, teeth and palate are affected; during

subsequent histogenesis and functional maturation, structural tissue defects and/or functional loss (growth/CNS abnormalities and behavioural effects), plus growth retardation as a consequence of general cell necrosis are common. However, in accordance with Principle 1 (see above), the end-points of many teratogens frequently encompass a range of all the possible manifestations in any sample of individuals.

One major problem with the detection of final manifestations in animal teratology studies is that more subtle signs are often missed such as learning or behavioural difficulties. These characteristics are often seen in epidemiological and case studies involving human teratogens and developmental toxicants.

Access of an adverse environmental agent to developing tissues depends on the nature of the agent (influences). Several factors affect the ability of a teratogen to contact a developing conceptus, such as the nature of the agent itself, route and degree of maternal exposure, rate of placental transfer and systemic absorption, and composition of the maternal and embryonic and foetal genotypes The latter elicits differences in cell sensitivity, receptor binding and distribution, and the make-up of the mother's metabolism in terms of its ability to deal with and clear the xenobiotic substance (Polifka and Friedman, 1999). Cigarette smoke and alcohol are salient examples of human teratogens whose destructive potential is highly variable and a function of genetic variability (Finnell *et al*., 2002). The significance of the placental transfer factor, a major determinant of interspecies variability in teratogen presentation to the foetus, depends on the nature of the agent: teratogens with a molecular weight less than 600 and low ionic charge cross the placenta by simple diffusion, whereas those of greater molecular weight and ionic charge must do so via facilitated diffusion or active transport.

The manifestations of deviant development increase in degree as dosage increases from the no-effect to the lethal level. Exposures to teratogenic agents only result in teratogenic effects when a certain threshold dose has been exceeded (Brent, 1995). For every teratogen there is a level of exposure below which no adverse effect to the embryo occurs, known as the 'no observable effect level' (NOEL); above this threshold, teratogenic effects steadily increase in a dose-response relationship. This dose-response curve is generally quite steep followed by a plateau, and is an essential criterion for the identification of true teratogens (Wilson, 1973). Short-term dosing schedules elicit greater teratogenic responses for some agents than extended ones, for example in the response of the rat to dactinomycin treatment (Wilson and Fouts, 1966). Sometimes, however, chronic exposures harm the well-being of the foetus more than acute exposures at similar doses, as evidenced by foetal alcohol syndrome. These aspects of exposure are especially important when considering teratogenicity in cases such as suicide attempts (acute exposure) *versus* long-term drug abuse and occupational hazards (chronic exposure; Polifka and Friedman, 1999).

PREDICTING HUMAN TERATOGENS: PROBLEMS WITH ANIMAL-BASED TESTING AND EVALUATION METHODS

Animal-based studies of developmental toxicology provide the initial guidelines on whether a drug or chemical may present a teratogenic risk during pregnancy. Typically, a range of doses administered via the most appropriate route (usually oral, but occasionally dermal or via inhalation) is given to pregnant animals during the period of embryonic organogenesis, and the outcomes compared to control untreated animals. Safety testing regulations generally require testing on two species, one of which must be a non-rodent. The most prevalent species used are mice, rats and rabbits, although if a particular agent is highly likely to be encountered by pregnant women, non-human primates may be included. The usual sample size is 20 pregnant females per dose, and the dose range is selected so that the highest dose causes some signs of maternal toxicity, the lowest causes no discernable effect in the mother or foetus, and at least one intermediate dose.

When one considers the many physiological and biochemical differences between animal species used in teratology research and humans, it is of no great surprise that no one species can be shown to be the experimental 'animal of choice'. Desired characteristics for the ideal teratology animal have been proposed in a 'wish list', but none comes close to fulfilling the criteria. No one species absorbs, metabolises and eliminates test substances just like a human nor possesses the same placental transfer properties; no one species has the same pre-term developmental and metabolic patterns as do humans; and, even if it were possible to meet all of these standards, the animal would be unlikely to meet other 'ideal' criteria such as producing large litters after a short gestation, inexpensive maintenance, and an inability and unwillingness to 'bite, scratch, kick, howl or squeal' (*sic*; Wilson, 1975).

Despite these liabilities, rodents have become the most commonly used species for evaluating potential human teratogens. Proponents of animal use, while admitting that it can only ever give an approximation of effects in humans, praise the rat model because for many years, all human teratogens identified exhibited teratogenesis in rats (Tuchmann-Duplessis *et al.*, 1972). There are, however, important exceptions, such as with the prostaglandin E1 analogue misoprostol: treatment of humans with this drug for peptic ulcer disease or to initiate labour has a strong association with foetal malformations known as Moebius syndrome (Pastuszak *et al.*, 1998), but is not teratogenic in the rat even up to 10 times the human dose (Klasco and Heitland, 2003). But it is the extrapolation of results from animals to humans that is of most concern; there is little point in demonstrating that a known human teratogen is also teratogenic to some species of animal. As will be subsequently shown in this review, there can be little confidence in the extrapolation to humans of negative results from the common species used in teratogenicity testing, and there are many examples of positive results in these species that have little or no effect in humans ('false positives'), especially at normal exposures and therapeutic dose levels. Notable examples include glucocorticoids and benzodiazepines, which induce oral clefts in

rats, mice and rabbits but not in humans (Baxter and Fraser, 1950; Buresh and Urban, 1970; Czeizel 1987; Fainstat, 1954; Fraser and Sajoo, 1995; Miller and Becker, 1975; Pinsky and DiGeorge, 1965; Rosenberg *et al.*, 1983; Shepard 1994; Shiono and Mills, 1984; Walker 1971; Wilson *et al.*, 1970), and aspirin, which causes cardiac malformations in several species of animals (such as the rat and the rhesus monkey: Beall and Klein, 1977; Klein *et al.*, 1981; Slone *et al.*, 1976; Werler *et al.*,1989; Wilson *et al.*, 1977) but is harmless in man.

Because of the inherent problems and inadequacies of teratology testing and research with the five groups of animals most commonly used (mouse, rat, rabbit, hamster, and monkey), scientists have tried to incorporate ever more species into their experiments in an attempt to find something approaching that elusive 'ideal animal'. Dogs (more specifically beagles) were tested with an array of known teratogenic compounds, but deemed unsuitable due to poor sensitivity; in any case, it is known that many drugs are metabolised differently in the dog, and there are particular problems with extrapolation from dogs to humans with reference to steroids (Schardein, 1993). With cats there was some promising concordance with several compounds, though in common with dogs they are known to metabolise a significant number of drugs differently, some uniquely, and there were discordant results compared to other species with compounds such as hydantoins, thalidomide, and especially the anti-leukaemia drug and abortifacient aminopterin (which is highly teratogenic in humans but not at all in cats) (Schardein, 1993). Pigs were found to be as insensitive as dogs, ferrets did not live up to early expectations, and non-human primates, despite their close phylogenetic relationship to humans, have been particularly disappointing as a predictive model (Schardein, 1993). Over 100 teratogenic agents classified as 'possible' or 'probable' have been tested in non-human primates, and the vast majority showed a high level of discordance; of the known human teratogens tested, only about half were found also to be teratogenic in one or more primate species.

This futile search, despite all the evidence, for a species of animal that responds as a human does to potential teratogens is not surprising when one considers all the biological variables. Numerous inter-species differences must be accounted for when designing animal developmental toxicology studies and extrapolating to humans (Nielsen, 2001; Pratt, 1980).

In summary:

1. Anatomical differences: for example, there is one placenta in humans, but two in rodents and rabbits.
2. Metabolic differences in both adults and foetuses of different species; these affect the absorption, distribution, metabolism and excretion (i.e. pharmacokinetics) of agents tested. Metabolic differences particularly affect the teratogenesis of drugs that require metabolic activation (for example the anti-cancer agent cyclophosphamide is activated to phosphoramide mustard) (Fantel *et al.*, 1979), and also determine the half-life of drugs in particular species due to varied inactivation rates. Occasionally metabolic activation via the embryo or foetal tissues

is of greater importance than that of the maternal system, for example with the activation of some nucleotide antimetabolites (Brown and Fabro, 1983). The half-life of some teratogenic metabolites is so short that they must be formed actually within or in close proximity to their target tissues. This is one reason why rodents can display poor correlation to the human situation; some xenobiotic-metabolising P450 isoenzymes are not produced by rodent foetuses or pups until just prior to or just after birth respectively (Brown and Fabro, 1983).

3. Variations in response to potential teratogens. These are discussed in detail below.
4. Sensitivity of animals to environmental factors. The metabolism of test animals can be affected by pesticide residues in bedding (Pratt, 1980); congenital malformations are induced in some species by alterations in temperature, barometric pressure, audiovisual stimuli and diet. Results can also be significantly affected by poor treatment, excess or shortage of companions, maternal age and the season of the year.
5. Route of administration and dose of the test substance. and the vehicle used. The route of administration of drugs or chemicals in animal-based teratology experiments can produce markedly different results between animals. Often, these differences are attributable to variable absorption rates and metabolic fates. In some cases, agents affect the palatability of food and drink to such an extent that nutritional imbalances may occur. However, many instances exist of confounding results that cannot be explained in this way. For example, teratogens may be active only when administered orally and not intraperitoneally (Nora *et al.*, 1965), whereas for others the opposite is the case, even when higher doses are used to compensate (Cahen, 1966). Some induce foetal malformations when administered by gavage but not in the mother's diet (Kavlock *et al.*, 1982), whereas for others the reverse is true (Staples *et al.*, 1976). Certain non-teratogenic substances and physical states are able to potentiate the effect of a known teratogen when administered together; immobilization of rats potentiates vitamin A teratogenicity. Furthermore, two teratogens can act synergistically to produce a combined effect that is much more than additive (Datta and Singh, 1999; Sisodia, 1972).

   Dose levels in animal experiments usually extend far beyond those ever likely to be encountered by humans. Justifications for this include overcoming the limited sensitivity of these studies due to small sample sizes, mimicking long-term chronic human exposure where the exposure period is necessarily much shorter in the laboratory, and that humans are simply more sensitive than commonly used laboratory animals to doses expressed as mg/kg (Nielsen *et al.*, 2001). Intuitively, this seems to be 'bad science': increasing dose to account for a limited sample size of animals may push it over the threshold level so fundamental to teratology studies; altering the time and duration of treatment by mimicking long-term exposure via the instigation of much higher dose short-term regimens is well known to alter the teratogenicity of many substances;

and if humans are more sensitive to various drugs and chemicals than other species, perhaps this should be deemed an observation rather than a problem to overcome.

Nevertheless, huge numbers of teratology research programs continue to investigate the responses of many animal species to physiologically large doses of drugs and chemicals, far in excess of anything likely to be encountered by humans. It is difficult to conceive any advantage in this regime, not only due to the myriad of interspecies variables and inconsistencies discussed here, but also in light of Karnofsky's Law, which states 'Anything can be teratogenic if given in the right dose, to the right species, at the right time'. If every single drug, chemical and indeed substance can be teratogenic in some particular animal at some specific dose, then to produce a positive result one need only find a suitably sensitive species and administer a suitably high dose (Scialli, 1992). This may well explain the inclusion of various everyday substances, some of which are intrinsic components of the mother's body and/or the developing conceptus and which are essential for life itself, in the list of teratogens in 'Dangerous Properties of Industrial Materials@ (Lewis, 1989). These include drinking water (Turbow *et al.*, 1971) and table salt (Nishimura and Miyamoto, 1969); oxygen; sugars in the form of sucrose and lactose; palm oil, corn oil and nutmeg oil; other naturally occurring food constituents such as cholesterol and papain (prevalent in pineapples); vitamins such as A, D2, K, B7 and B12 (the B vitamins are frequently found in pregnancy supplements); naturally occurring and essential hormones such as estradiol, progesterone and various prostaglandins; the amino acid methionine, and the DNA constituent adenine.

6. Other factors. Other confounding variables include the standardization of litter size, postnatal effects, and the manner in which animals are housed and handled. Litter sizes are sometimes 'standardized' to 8 pups by culling for generation studies to produce more uniform pup weights. However, this is known to have no effect on variation of pup weight, and may introduce bias by the random elimination of runts (Palmer, 1986). Postnatal effects may be of concern when the mother suffers physiological or behavioural changes from the administration of the potential teratogen; maternal care and lactation may be compromised, adversely affecting the offspring. Cross-fostering techniques are used as a solution to this problem, although the requirement of more resources and the resultant greater expense often prohibit this. Finally, it is acknowledged that the handling of animals can induce physiological stress responses and cause alterations in behaviour; both of these factors can affect teratogenicity results. Standardized conditions are often employed in an effort to reduce these stresses with regard to temperature, humidity, noise, light, living quarters and handling, although there is evidence that laboratory conditions and participation in toxicology research are so inherently stressful that this aspect can never be excluded (Balcombe *et al.*, 2004).

RELIABILITY AND CONCORDANCE OF ANIMAL DATA

Despite the problems cited above, animal-based experiments persist as the first port of call in teratology research because they have the advantage of being relatively inexpensive, avoid unnecessary human exposure, and provide the ability to test one variable at a time (Schardein, 1993). Taking into account the time, money and resources dedicated to them over the past forty years, it would seem desirable if not imperative that the assertion 'They are highly predictive of human teratogenicity' be added to the list. Some scrutiny of the substantial amount of existing animal data goes some way to explaining why this statement cannot be included: we achieved this using the REPRORISK system (TERIS and REPROTOX databases, MICROMEDEX; Klasco and Heitland 2003), by consulting reference books such as Schardein (1993), and via PubMed literature searches using the following terms (Teratogen, Teratology, Teratogenicity, Concordance, Discordance, Results, Species, Extrapolation, Differences, Human, Animal).

Although the ultimate goal of teratology testing is to predict with confidence both the positive and negative teratogenic potential of any substance in humans, ostensibly by extrapolation of data from animal experiments, examination of inter-species correlation and concordance between the various species of test animals underscores the problematic nature of this approach. Even a cursory glance at the information presented in Tables 1a and 1b (derived from data in Schardein, 1993) depicting the responses of up to 12 animal species to 11 groups of known human teratogens, reveals the disarray in the data. While positive predictability is 75% for the hamster, it is just 40% for the rabbit, which also exhibits a false negative rate of 40%. The mean positive predictability rate in those species tested for at least 9 of the 21 agents (namely the first six categories, i.e. mouse, rat, rabbit, hamster, primate, dog) was under 55%, and the number of equivocal results remained high across these six species at just under 25%. Table 2, derived from information published in Schardein (1993) and Klasco and Heitland (2001), summarizes similar data for 35 individual substances linked with human teratogenicity (Koren *et al.*, 1998; Polifka and Friedmann, 1999, 2002; State of California EPA Prop 65, 2003), rather than groups of substances. Of the 139 individual classifications across the species tested, a total of 78 (56%) were positive; the remaining 44% of results were almost entirely negative. The only encouraging aspect to come from these statistics appears to be the high positive predictability score for the hamster; however, the USFDA published a report detailing the responses of the mice, rats, rabbits, hamsters and monkeys to 38 known human teratogens in which the high scoring hamster produced only a 45% rate of correct positives (USA FDA Federal Register 'Caffeine', 1980). Furthermore, the mean percentage of correct positives from any one of these species was only 60%.

The USFDA report also analysed the rate of concordance between these species and humans for 165 compounds known to be non-teratogenic in the latter; the 'order of merit' for each species and its negative predictive value were completely different to that for the positive predictive values, ranging from 80% in monkeys to 35% in

Table 1a.
Teratological classification in animal species of groups of substances universally recognized as human teratogens: summary of classifications by substance group

Substance group	Species												Result summary	
	Mouse	Rat	Rabbit	Hamster	Primate	Dog	Cat	Pig	Ferret	Guinea Pig	Sheep	Cow	+/ ± /−	%+
Androgenic hormones		+	+	+	+	+		+		+	−	+	8/0/1	89
Anticancer antimetabolites	±	+	±	+	±	±	±	+		+	+		5/5/0	50
Anticancer alkylating agents	+	+	+		+				+				5/0/0	100
Anticonvulsant hydantoins	+	+	+	−	+	−	−						4/0/3	57
Anticonvulsant diones	+	±			±								1/2/0	33
Coumarin anticoagulants	−	+	−										1/0/2	33
Antithyroid drugs	+	±	±							+	+	−	3/2/1	50
Aminoglycoside antibiotics	±	+	−		−					±			1/2/2	20
Vitamin A, analogues	+	+	+	+	+	+		+	+	+			9/0/0	100
PCBs	−	−	−		±	+		±		−		−	1/2/5	12.5
Tetracyclines	±	±	−			−				+			1/2/2	20
Totals													39/15/16	Mean = 51%

Eleven groups of substances known to be teratogenic in humans are listed above, along with their teratogenic classifications in 12 species of animal (Schardein, 1993), where + = teratogenic; ± = variably teratogenic; − = not teratogenic. No entry = no classification available. The numbers of positive, variable and negative results for each group of substances in the 12 named species is summarized, along with the combined percentage of positive results in all species for each group of substances. Concordance varies greatly from 100% for the anticancer alkylating agents and vitamin A analogues, down to 12.5% for PCBs. Notably, the total proportion of positive classifications and therefore correlation to the human situation amounts to only 51%; no better than by pure chance.

Table 1b.
Teratological classification in animal species of groups of substances universally recognized as human teratogens: summary of classifications by animal species

Species	Substance group											Results	
	Androgenic hormones	Anticancer anti-metabolites	Anticancer alkylating agents	Anti-convulsant hydantoins	Anti-convulsant diones	Coumarin anti-coagulants	Antithyroid drugs	Amino-glycoside antibiotics	Vitamin A, analogues	PCBs	Tetra-cyclines	+/±/−	%+
Mouse		±	+	+	+	−	+	±	+	−	±	5/3/2	50
Rat	+	+	+	+	±	+	±	+	+	−	±	7/3/1	64
Rabbit	+	±	+	+		−	±	−	+	−	−	4/2/4	40
Hamster	+	+		−					+			3/0/1	75
Primate	+	±	+	+	±			−	+	±		4/3/1	50
Dog	+	±		−					+	+	−	3/1/2	50
Cat		±		−								0/1/1	0
Pig	+	+							+	±		3/1/0	75
Ferret			+						+			2/0/0	100
Guinea Pig	+	+					+	±	+	−	+	5/1/1	71
Sheep	−	+					+					2/0/1	67
Cow	+						−			−		1/0/2	33
Totals												39/15/16	Mean = 56%

Eleven groups of substances known to be teratogenic in humans are listed above, along with their teratogenic classifications in 12 species of animal (Schardein, 1993). + = teratogenic; ± = variably teratogenic; − = not teratogenic. No entry = no classification available. The numbers of positive, variable and negative results in each species for all groups of substances is summarized, along with the combined percentage of positive results for all groups of substances in each species. The maximum and minimum concordance of results with respect to human classifications are 100% and 0% in the ferret and the cat respectively, although only 2 results are listed for both of these species. While positive predictability is 75% for the hamster, it is just 40% for the rabbit, which also exhibits a false negative rate of 40%. The mean positive predictability rate in those species tested for at least 9 of the 21 agents (namely the first six categories i.e. mouse, rat, rabbit, hamster, primate, dog) was under 55%, and the number of equivocal results remained high across these six species at just under 25%. Taking all 70 results into account leads to an overall positive rate of only 56%; only slightly better than by chance.

Table 2.
Teratological classification in animal species of individual substances universally recognized as human teratogens

Substance/group	Species												Summary of results	
	Mouse	Rat	Rabbit	Hamster	Primate	Dog	Cat	Pig	Ferret	Guinea Pig	Sheep	Cow	+/±/−	%+
Aminopterin	−	+			−	+	−	+			+		4/0/3	57
Amiodarone	−												0/0/1	0
Androgens		+	+	+	+	+		+		+	−	+	8/0/1	89
Captopril		−	−	−							−		0/0/4	0
Carbamazepine	+	+											2/0/0	100
Clonazepam	+	−	−										1/0/2	33
Cocaine	±	±	±										0/3/0	0
Coumarin	−		−										2/0/0	100
Cyclophosphamide	+	+	+		+								4/0/0	100
Diazepam (Valium)	+	−		+									2/0/1	67
Diethylstilbestrol (DES)	+	−	−	−	−				−				1/0/5	17
Dilantin (Diphenylhydantoin)	+	+	+		−	−	−						3/0/3	50
Enalapril		−	−										0/0/2	0
Ethyl alcohol	+	+	+		+	+		+	+	+	+		9/0/0	100
Fluconazole		+	−										1/0/1	50
Indomethacin	+	+	−										2/0/1	67
Isotretinoin	+	+	+	+	+								5/0/0	100
Lithium	±	±	−		−			±					0/3/2	0
Methimazole			−										0/0/1	—
Methotrexate	+	+	+		±	−	+						4/1/1	67
Methyl mercury	+	+	+	+	−	+	+		+				7/0/1	87
Misoprostol		−	−										0/0/2	0
PCBs	−	−	−		±	+		±		−		−	1/2/5	12.5
Penicillamine	+	+		+									3/0/0	100
Phenobarbital	+	+	+										3/0/0	100
Primidone	+												1/0/0	—

Table 2.
(Continued)

Substance/group	Species												Summary of results	
	Mouse	Rat	Rabbit	Hamster	Primate	Dog	Cat	Pig	Ferret	Guinea Pig	Sheep	Cow	+/±/−	%+
Quinine	−	−	+		−	−				+			2/0/4	33
Radioiodine	−	−								+			1/0/2	33
Streptomycin	−	+	−							−			1/0/3	25
Tetracycline	−	+	−			−				+			2/0/3	40
Thalidomide	+*	+*	+	+*	+#	+	+	±	+	−			—	—
Trimethadione	+	±			+								2/1/0	67
Trimethoprim		+	−										1/0/1	50
Valproic acid	+	+	+	+	+								5/0/0	100
Warfarin	−	+	−										1/0/2	33
													78/10/51	Mean = 56%

Thirty-five substances known to be teratogenic in humans are listed above, along with their teratogenic classifications in 12 species of animal (Schardein, 1993). + = teratogenic (*only particular strains are reactive; #only particular species are reactive); ± = variably teratogenic; − = not teratogenic. No entry = no classification available. The numbers of positive, variable and negative results for each substance across all species tested is summarized, along with the combined percentage of positive results for each substance. Though the numbers involved are small (there is a maximum of twelve possible results for each substance), the percentage of positive results varies greatly, from 0% to 100%. The mean number (weighted) of positive results in all animal species for the 35 human teratogens listed is 78 from a possible 139, i.e. 56%. In other words, 44% of results in various animal species tested with known human teratogens were negative.

mice and hamsters. The mean negative predictive value for any of these species was 54%. Taken together, these predictive values of 60% and 54% for human teratogens and human non-teratogens, respectively, represent a poor return on the investment of animals, time, labour and money. The 57% mean value is little better than the 50% that would have been obtained by pure chance.

A broader view can be had by examining data not only for those substances linked with teratogenicity in humans, but by taking into account all substances for which some teratological data has been determined in animals. Table 3 lists categories of drugs, chemicals and other substances as classified in (Schardein, 1993), along with a measure of the discordance in the results across the various species of animals in which they were tested for teratogenicity. Any substance for which only one species was tested is omitted, because no evidence of discordance could be determined. The total number of remaining substances in each group that showed inconsistent results across species is shown out of the total number of possible substances, and this discordance is shown as a percentage. In the 70 groups of substances examined, a total of 1396 individual substances had been tested in more than one animal species. Of these, 401 (just under 30%) showed discordance, i.e. a mixture of positive, equivocal and negative results. A number of these are discussed in more detail below.

Advocates of animal teratology investigations may raise the point that the USFDA report declared a 97% correlation between known human teratogens and the chance of a positive result in any one of the five animal species tested. Not only is this not surprising (it seems logical that if enough animal species are tested with a compound already known to be teratogenic in humans, at least one will prove positive), but it is surely of dubious worth. It is imprudent to place any confidence in the ability of the animal model as being positively predictive for human teratogens simply because every agent that elicits teratogenesis in humans does likewise in one animal species; if only the reverse were true, then some of the unfortunate consequences of misplaced animal-human extrapolation of experiments involving teratogenic substances would have been avoided (Heilla *et al.*, 1994; McElhatton, 2002; Peters and Schaefer, 2001). Often, positive results with known human teratogens occur only in certain animal species when given a massive dose at an exact time of gestation. Also, many drugs and chemicals that are teratogenic in at least one (and frequently several) animal species are not so in humans: the value of extrapolation in this direction, from animal to human, is the salient point and must be scientifically examined, not the dubious worth of the retrospective support of an animal model after the event in humans has already been documented.

As already discussed, many substances following animal-based tests will exhibit teratogenic effects at some particular dose (often artificially high) at a particular time of gestation via a particular dose route. Of the thousands of agents that have been tested to some degree, the number of known teratogens is low; of 3301 tested prior to 1993, just 7% were teratogenic, 21% were classed as probably teratogenic, and 9% only possibly so; the remaining 63% were deemed non-

Table 3.
Degree of discordance in teratogenicity results among animal species for seventy groups of substances

Substance group	Discordance	
	Individual substances showing discordance/total substances in group	% substances in group showing discordant results across species
Antihyperlipoproteinemic drugs	2/18	11
Diuretics	3/18	17
Antihypertensive agents	7/31	23
Cardioactive drugs	1/25	4
Vasodilator drugs	2/24	8
Agents affecting blood	2/13	15
Narcotic analgesics/antigout agents	6/11	54
Analgesics, Anti-Inflammatory and Antipyretic agents	11/67	16
Anesthetics	6/19	32
Anticonvulsants	5/9	56
Tranquillisers	8/31	26
Antipsychotic drugs	10/21	48
Antidepressants	4/29	14
CNS stimulant and Anorectic agents	2/10	20
Sedatives — Hypnotics	4/18	22
Thalidomide and related chemicals	7/21	33
Androgens	1/12	8
Oestrogens	4/10	40
Progestogens/progestogen-containing contraceptives	8/17	47
Various contraceptive agents	1/8	12
Hormone antagonists	4/8	50
Adenohypophyseal hormones	5/10	50
Corticosteroids	8/17	47
Respiratory drugs	13/49	27
Sulphonamides	4/14	29
Antibiotics	15/71	21
Tuberculostatic and leprostatic agents	4/9	44
Antifungal agents	3/18	17
Antiviral drugs	4/8	50
Misc. antibacterial agents	10/38	26
Antimalarials	2/3	67
Antihelmintics	6/17	35
Misc. antiparasitic drugs	2/11	18
Hypoglycaemic drugs	4/13	31
Thyroid-acting agents	4/5	80
Oxytocic agents	0/4	0
Prostaglandins	3/6	50
Neuromuscular drugs	2/12	17
Gastrointestinal drugs	7/45	16
Anticancer alkylating agents	6/12	50

Table 3.
(Continued)

Substance group	Discordance	
	Individual substances showing discordance/total substances in group	% substances in group showing discordant results across species
Anticancer antibiotics	4/8	50
Anticancer antimetabolites	9/22	41
Misc. anticancer agents	7/28	25
Immunological agents	7/16	44
Chemical antagonists	9/29	31
Foods and nutrients	2/11	18
Hypovitaminoses	3/9	33
Hypervitaminoses	1/11	9
Misc. vitamin analogues and congeners	2/7	29
Dietary deficiencies	2/7	29
Food additives	14/79	18
Cosmetic agents	0/2	0
Misc. drugs	10/47	21
Fungicides	15/23	65
Herbicides	10/31	32
Insecticides	19/42	45
Misc. pesticides	3/12	25
Lead compounds	3/4	75
Mercury compounds	3/9	33
Misc. metals	5/17	29
Industrial solvents	15/46	33
Radioactive chemicals	2/6	33
Diagnostic chemicals	3/11	27
Dyes	3/8	38
Plastic and rubber chemicals	10/26	38
Toxins	2/5	40
Plants and phytotoxins	18/38	47
Halogenated biphenyls	2/7	29
Misc. pollutants	2/7	29
Misc. chemicals	19/46	41
Totals:	401/1396	Mean = 28.7%

A large number of substance groups were examined to identify the proportion of individual substances therein showing discordant results across animal species. Specific substances were only included in the analysis if they had been tested in more than one species of animal, i.e. were able to show discordance. The number of substances in each grouping that produced irregular results is shown, out of the total number of substances that had been tested in more than one species of animal; this proportion is also represented as a percentage. The degree of discordance revealed ranges from 0% in the case of cosmetic and oxytocic agents, to 80% in the case of thyroid-acting agents. However, it must be noted that the sample sizes for these groups were small, at only 2, 4 and 5 substances respectively. Out of a total of 1396 substances that had undergone teratological testing in 2 or more species of animal, the results for 401 showed some discordance, i.e. just under 30%.

teratogenic (Schardein, 1993). Of the 37% of agents associated with teratogenesis, only a very small proportion has been linked to teratogenesis in humans. Numbers vary: for example, Schardein (1993) cited '19 or so drugs or groups of drugs plus 2 chemicals', and the Mountain States Genetics Network "17 medications or medication categories, plus 7 types of maternal infectious agents, 3 chemicals, 5 classes of maternal disorders and 4 'reproductive toxins' " (Genetic Drift, Teratogen Update, 1995). This means that of 1223 definite, probable and possible animal teratogens, fewer than 2.3% were linked to human birth defects.

SPECIES DIFFERENCES IN INDIVIDUAL TESTING OF POTENTIAL AND KNOWN HUMAN TERATOGENS

In spite of the positive evidence in support of epidemiology and case studies, animal-based teratology persists regardless of the *in vitro* alternatives that scientific progress has offered. Before summarizing and briefly evaluating these alternatives, it is interesting to examine in more detail some results of animal-based experiments involving specific drugs and chemicals. Information for the following précis was obtained from results reviewed in Schardein (1993) and Klasco and Heitland (2003); inclusion was based on the National Toxicology Program's list of agents with defined teratogenicity (National Toxicology Program, 1983), or as items of special interest from a wide variety of publications. Substances were only considered, and are only mentioned here, if they had been tested in more than one animal species. Those described as generating discordant results across species showed at least one different result in one of the species tested when compared to the others. Where known, human classifications are included in brackets after the drug name; for example (D) denotes category D, and C^1 represents a category C risk only in the first trimester. These categories are in agreement with FDA regulations as described in Millstein (1980):

Category A — No demonstrated risk; possibility of foetal harm is remote.

Category B — Either no evidence of foetal risk in extrapolation from animal data plus no information from human studies, or evidence of adverse effects in animal studies that could not be confirmed in human studies.

Category C — Either evidence of adverse effects in animal studies combined with no studies in humans, or studies in animals and humans are not available. Drugs only given to pregnant women when the possible benefit outweighs the risks.

Category D — Positive evidence of foetal risk. Drugs may be given to pregnant women when the benefits are acceptable despite the risk.

Category E — Known not to affect animal reproduction but no human data.

Category X — Animal or human studies or human experience have revealed foetal abnormalities. Risk of use in pregnant women outweighs any possible benefit.

Angiotensin Converting Enzyme (ACE) inhibitors

ACE inhibitors are anti-hypertensive drugs prescribed to control high blood pressure. From 31 anti-hypertensives of various types that had been tested in more than one species, 7 (23%) showed discordance. The ACE inhibitor *captopril* presents a slight teratogenic risk in the 2nd and 3rd trimesters of human pregnancy; it causes a high incidence of foetal death in rabbits and sheep. *Enalapril* (C^1, D^{2+3}) also offers a slight teratogenic risk in the 2nd and 3rd trimesters in humans; it is not teratogenic in baboons or rabbits although there is evidence of increased foetal loss in the latter with 4-40x the human dose; evidence for rats is equivocal.

Anticonvulsants

5 of 9 (55%) anticonvulsants used in the treatment of epilepsy showed discordant results. All epidemiological studies of drugs in this class are confounded by the fact that many are prescribed and administered concurrently. *Phenytoin* (C), one of the hydantoins, is surrounded by controversy. Although a 'foetal hydantoin syndrome' is evident in 10% of children born to epileptic mothers treated with phenytoin during pregnancy, underlying familial factors and the effects of epilepsy itself have not been discounted as causal factors for this. Phenytoin is teratogenic in rabbits, structurally and behaviourally so in rats, and also in mice although this is strain-dependent. In rhesus monkeys it is not teratogenic despite high maternal toxicity, although there is some evidence of behavioural problems in offspring.

Primidone (D), a barbiturate, is almost certainly a human teratogen, although much of this evidence comes from studies not confounded by the inclusion of other anticonvulsants in any maternal treatment. It is teratogenic in the mouse although not in a dose-dependent manner, and causes behavioural problems and increased foetal death in rats.

Clonazepam (D), a benzodiazepine, is not teratogenic in humans, rats or rabbits, but there is some evidence of minor developmental defects in the mouse.

Trimethadione is associated with 'Foetal trimethadione syndrome' in the form of multiple structural abnormalities in humans, but is teratogenic in mice only at 8–22x the human dose. Rats and rhesus monkeys display signs of teratogenesis at up to 10x the human dose.

Valproic acid (D) is well associated with 'Foetal valproate syndrome' in humans. It is also teratogenic in the rhesus monkey. In the rat, hamster and mouse it is teratogenic at high doses but not at doses comparable to those in humans, and in mice this is strain dependent.

Antifungal agents and fungicides

3 of 18 anti-fungal agents (17%) and 15 of 23 fungicides (65%) show discordant results between animal species.

Fluconazole (C), used in the treatment of mycotic infections, is not a human teratogen at normal therapeutic doses, although there are some possible examples of teratogenicity when used at high chronic doses. It is teratogenic in rats but only at 5–20 times the typical human dose, and it causes a greater rate of foetal death in rabbits at these same high doses.

Ethylenethiourea is teratogenic to rats and rabbits, but no human data are available.

Antimicrobial agents

15 of the 71 listed antibiotics (21%) show discordant results in animals; most results are negative.

Streptomycin (D) is classed as a non-teratogen in humans, mice, rabbits and guinea pigs, but was classed as being teratogenic in rats where it caused inner ear damage. However, more recent experiments have confounded these views; rats treated with 10 x the human dose in one experiment had no damage to their hearing, whereas injections of streptomycin sulphate are now believed to cause ototoxic harm to the human foetus.

Tetracycline is known to cause dental staining in 2nd and 3rd trimester human foetuses, and in one case study has been linked to transposition of the great arteries. Animal results are extremely conflicting. It is also classified as a teratogen in rats and guinea pigs, though results for rats are equivocal. It is not teratogenic in mice or rabbits, nor is the related compound oxytetracycline although this is teratogenic in dogs.

Trimethoprim (C), an antimicrobial agent often used in conjunction with sulphonamides, shows equivocal results in humans. It is teratogenic in rats only at doses 16–117 times that of the usual human dose. Teratogenesis was evident in mice only when administered in combination with sulphamethoxazole at high doses; similar treatment of rabbits produced no foetal malformation, but increased foetal loss.

Isoniazid, used in the treatment of tuberculosis, is not a human, mouse or rabbit teratogen, but is teratogenic in rats in which it induces skeletal abnormalities.

Antihistamines

The results for 13 of 49 ‘Drugs for respiratory and allergic disorders’ were discordant across animal species (27%).

The antihistamine *diphenhydramine* (*Benadryl*) was found to be non-teratogenic in humans in all but one study. It shows no signs of teratogenicity in the mouse or the rabbit even up to 33x and 17x the human clinical dose respectively, but is teratogenic in rats at 1.5–2.5x the usual human dose.

Anti-cancer chemotherapeutic agents

In total, discordance was evident in 26 of 70 anti-cancer drugs (37%). This comprised 6/12 (50%) of alkylating agents, 4 of 8 anticancer antibiotics (50%), 9 of 22 anticancer antimetabolites (41%), and 7 of 28 miscellaneous anticancer agents (25%).

The alkylating agents *chlorambucil* and *cyclophosphamide* have demonstrated teratogenesis in every species so far tested, and unlike many other anticancer agents in their class show good concordance across animal species. There is a paucity of human data for chlorambucil other than that it elicits chromosome damage in human cells after therapy, but it is teratogenic in mice and rats at doses of approximately twice that used in humans where it can cause CNS and skeletal defects; doses equivalent to those for humans were foetotoxic. Cyclophosphamide is teratogenic to humans during the 1st trimester and may be slightly so in the 2nd and 3rd, causing a range of skeletal, limb, palate and eye defects. Other affected species include rhesus monkeys, rats, mice and rabbits, though the deformities induced can differ among species and the doses used were significantly higher than human therapeutic doses.

The cytotoxic antibiotic *Dactinomycin* (*actinomycin D*) (D) is also teratogenic in all species tested to date (albeit only at doses 2 to 13 times greater than the human therapeutic dose), inducing predominantly CNS defects in rats, rabbits, mice, hamsters, and chick embryos. However, human studies have revealed no cases of teratogenesis.

The case is similar for the antimetabolite *fluorouracil*. There is a great deal of evidence for its teratogenicity in many species, including the mouse, rat, hamster, rabbit, and guinea pig at doses of 1–3.3 times the human therapeutic dose. The case for the rhesus monkey is equivocal, and human data, other than a small number of case reports, are classed as either negative or as producing 'not sufficient evidence' to establish a teratogenic link.

Aminopterin, also used as a rodenticide and an abortifacient, is classed as being a moderate to high-risk teratogen during the first trimester of human pregnancy. It is also teratogenic in the rat, sheep, dog, pig and rabbit, and causes increased foetal loss in cats and mice.

Hydroxyurea, an inhibitor of DNA synthesis, is classed as a 'universal teratogen'. A huge range of malformations have been induced in experimental animals, though at doses 5–67 times the human dose in rats, and 2–25 times the human dose in mice. Adverse effects have not been reported in any human data so far examined.

Insulin and hypoglycaemic agents

4 of 13 (31%) hypoglycaemic drugs yielded discordant results across animal species.

Insulin (B) is teratogenic in rats, mice and rabbits, although the nature of malformations differs among these species, and there have been negative reports in rats even for doses much greater than those used for human therapy. However, it

is believed that insulin is not teratogenic in humans. Indeed, insulin administration may actually decrease the incidence of adverse foetal outcomes via better control of the metabolic imbalances that can be induced by the diabetic state itself.

Carbutamide is teratogenic in rats at 7–60 times the human dose, mice at 20–40 times the human dose, and rabbits at 40 times the human dose but not 10–30 times. There are no human teratology data for this drug.

*Oestrogens (*X*)*

Although very few oestrogens appear to have teratogenic potential according to animal experiments, 4 out of 10 still show some discordance, with mice being particularly sensitive.

Diethylstilbestrol (*DES*) (X), a non-steroidal synthetic oestrogen used to treat ovarian insufficiency, to prevent miscarriages and as a post-coital contraceptive, is a known human teratogen that predominantly causes genital and pregnancy problems in the female offspring of exposed mothers. Although initially classed as a non-teratogen in rats, rabbits, hamsters, monkeys and ferrets, more recent experiments have detected malformations in the offspring of these species, as well as in pigs and guinea pigs.

Estradiol-17β (E2) is naturally produced by the ovary, and believed to be the most active oestrogen in non-pregnant women. Although there is no evidence of E2 teratogenicity in humans, it can cause foetal abnormalities in mice, rats and opossums, though not in monkeys.

Ethinyl estradiol is a synthetic oestrogen used in the treatment of menopausal symptoms and menstrual disorders, and is also an ingredient in some oral contraceptives. Results of teratology studies are mostly negative; there is no evidence of foetal malformations in rats, rabbits or three species of non-human primates although it is classed as a mouse teratogen when used at a dose 20–2000 times the human contraceptive dose. Despite one confounding human study that associated foetal limb defects with maternal contraceptive use, the data overwhelmingly reject its classification as a human teratogen.

Steroids

8 of 17 (47%) corticosteroids showed discordant results across various animal species. *Dexamethasone*, though classified as a teratogen in rats, mice, rabbits, non-human primates, hamsters, pigs and dogs, was not in sheep or horses. In spite of teratogenesis in 7 species, there is no evidence whatsoever to associate dexamethasone use to foetal abnormalities in humans.

Cortisone (D) is a naturally occurring glucocorticoid excreted by the adrenal cortex, and is used to treat allergic and inflammatory diseases. It is teratogenic in all animals tested, including beagle dogs (albeit at 25 times the human dose), mice, hamsters, rabbits, and rats. However, human studies suggest only non-teratogenicity.

Danazol (X), an anabolic steroid used to treat gynaecologic and menstrual disorders, has a strong association with human teratogenesis. However, there are currently no animal data.

Tranquillisers

8 of 31 (26%) tranquillisers showed discordance between species. *Diazepam* (*valium*), a benzodiazepine sedative often prescribed for short-term anxiety problems, yields wildly discordant results despite an abundance of available data. Epidemiology and case studies in humans have been utterly equivocal; there are as many reports associating 1st trimester exposure to diazepam with foetal malformations such as oral clefts as there are reports refuting the link; both types are numerous. Animal research fares no better: rats exposed to doses 100 times that in humans have shown no teratogenic effects, while other studies describe persistent depression of immune response and/or neurobehavioral defects in the offspring of exposed mothers. Mice and hamsters produce more malformed progeny after exposure to diazepam, but only at doses dozens to hundreds of times that used in humans; also, the defects produced are different. Rabbits show no teratogenic response even at doses up to 100 times the human dose.

Meprobamate, a carbamide tranquilliser also used to treat anxiety, has also produced inconsistent results. As with diazepam, human studies have been equivocal; there are studies where an association appears very strong, particularly, for example, with cardiac defects, although no two studies indicating a positive link describe the same malformations. In contrast, several studies have found no association (Belafsky *et al.*, 1969; Heinonen *et al.*, 1977). Again, as for diazepam, the animal research is confusing. Birth defects have been linked with exposure to meprobamate in mice, rats and rabbits, though the majority of studies, even using doses of up to 16 times the human dose, have found no increase in foetal malformations in these species.

*Acitretin (*X*)*

Acitretin is a retinoid, given to treat skin conditions such as psoriasis. 29% of related substances (vitamin analogues and congeners) appear to show discordant results in animal tests. Although teratogenic in all species tested so far, minimum teratogenic doses vary; while mice and rabbits require doses only 1x or 2x the human therapeutic dose, hamsters are only affected at a minimum of 3x the human dose, and rats at 4x to 10x the minimum human dose.

Acetazolamide

Acetazolamide is a diuretic, of which 3 of 18 (17%) show discordant animal teratogenicity results. Also used therapeutically to treat glaucoma and epilepsy, there is no significant association between its use and human birth defects. It is

teratogenic in rats, mice and hamsters at high doses, may be teratogenic in rabbits, but is not so in the rhesus monkey.

*Amiodarone (*C/D*)*

Amiodarone is a vasodilatory drug, used to treat cardiac arrhythmias. The majority of these drugs are not teratogenic in any species, and results exhibit only an 8% discordance (2 of 27). Amiodarone has produced conflicting teratogenicity data in humans; there are links to foetal harm of the thyroid and heart, and studies demonstrate an absence of long-term defects. Rats, rabbits and mice suffer increased foetal loss at 1x to 3x the human dose.

*Aspirin (acetylsalicylic acid) (*C/D*)*

Analgesic, anti-inflammatory and antipyretic agents show teratogenic discordance in 11 of 67 cases (16%) including aspirin, one of the more common oral analgesics. In spite of a great many studies and a vast amount of data, there has never been any conclusive evidence that salicylates are teratogenic in humans; indeed, they have been shown to be beneficial in certain pregnancies. Conversely, there is no doubt that the drug is extremely foetotoxic and teratogenic in other animals. Teratogenic effects at doses similar to or greater than those used to treat rheumatic disease in humans have been observed in mice (where the drug potentiates the teratogenic and foetotoxic effects of ethanol), rats (particularly when food is restricted), dogs, cats, rabbits and monkeys.

Caffeine

The overwhelming majority of human studies demonstrate that caffeine does not induce teratogenic effects; those that have suggested a teratogenic property generally propose only a weak association, and suffer from methodological limitations, poor study design, and the influence of confounding factors. Caffeine-induced malformations in the offspring of mice, rats and monkeys are reported frequently, though at daily equivalent doses of >5–10 cups of coffee, i.e. at least 3 times normal human exposure.

*Carbon dioxide (*B+*)*

The only reported effect of carbon dioxide exposure in pregnant women is increased foetal breathing rate; there is no evidence of any adverse effect to foetuses exposed to increased levels of the gas. Conversely, it is known to be teratogenic to rats, rabbits, and also some strains of mice. The malformations observed differ between these species, from ectrodactyly (absence of toes) in susceptible mice, to vertebral and lung problems in rabbits exposed to 10% carbon dioxide, and cardiac and dental problems in rats exposed to only 3% and 10% carbon dioxide respectively.

Cocaine (C)

Although cocaine is regarded as a human teratogen, it is accepted that much of the literature regarding its effects must be interpreted with great caution. There are problems with design and reporting in many studies, and detailed documentation of maternal cocaine use, especially in combination with other illicit drugs and alcohol, is often inadequate or even poor. There is also a well-documented publication bias favouring adverse effects, although significant numbers of papers denying any association between maternal cocaine use and foetal anomalies do exist. The weight of evidence, however, in spite of these caveats, does suggest a pattern of cocaine-induced foetal abnormalities, though it seems that no specific defect will ever be conclusively linked solely to cocaine. This confusion is also common to animal studies. Foetal rats are apparently susceptible to cocaine-induced genitourinary defects at doses similar to those commonly encountered by humans, with accompanying limb and tail defects at slightly higher doses. Mice appear particularly sensitive to the teratogenic effects of cocaine; other reports have suggested that no adverse foetal effects are seen upon exposing pregnant rats, mice and rabbits to similar doses of the drug; some reports link exposure of these species to cocaine with brain defects, although these are also inconsistent. The case for the rhesus monkey is also equivocal, with both positive and negative associations reported. Subcutaneous administration of cocaine to rats causes ocular defects in pups; this has also been reported for mice along with genitourinary and skeletal defects. Intraperitoneal injection, however, is believed not to increase the rate of malformations in the offspring of these species.

Cytochalasin D

Plants and phytotoxins, which include the experimental agent cytochalasin D (12), show discordance in 18 of 38 cases (47%). Cytochalasin D is used frequently in research to disrupt microfibril organization, and has been found to exert teratogenic effects in hamsters, in certain strains of mice, but not rats.

Dimethyl mercury

Metals and metal compounds have produced a confusing array of teratology results; 3 of 4 lead compounds (75%) showed discordance for example, while only 3 of 9 (33%) mercurial compounds did so. For metals in general, discordant results occurred in 5 of 17 cases (27%).

Although experimental animal models of teratogenicity and embryotoxicity with regard to mercury and mercury compounds have demonstrated some common neurological effects, a wide range of birth defects have been noted, many of which have not been identified in humans.

*Dopamine (*C*)*

Dopamine, a naturally occurring catecholamine, precursor of norepinephrine, and an important neurotransmitter in many animals, has been found to be a teratogen in chick embryos, rats and rabbits. However, despite its therapeutic use during human pregnancy to maintain blood pressure and renal perfusion, no adverse effects have been reported.

Ethylnitrosourea (ENU)

ENU is used experimentally as a mutagen, and is classified by Schardein (1993) as a 'miscellaneous chemical'. Of 46 miscellaneous chemicals tested in more than one animal species, 19 showed discordant results (41%). However, ENU is not one of them; a teratogenic effect has been demonstrated in mice, rats, hamsters, rabbits and pigs, although there is little human data with which to compare these effects.

Fever

Exposure to hyperthermic conditions during pregnancy appears to induce teratogenesis in all species tested, including humans, although malformations differ among species. These effects are thought to be directly mediated via altered expression of heat-shock proteins. Most prevalent are CNS defects, although cardiovascular anomalies are most common in rats, and skeletal malformations in mice along with significant maternal toxicity.

Folic acid

Hypervitaminoses, or dietary vitamin deficiencies, are generally teratogenic in animals. Of 9 hypervitaminoses tested in more than one species, 3 showed discordant results (33%). Perhaps the most widely known nutrient deficiency linked to teratogenesis is folic acid (or pteroylglutamic acid, PGA), whose protective effect during early gestation with respect to neural tube defects in humans has been well documented and supported. However, in other animals results have been equivocal, with protection from neural tube defects evident in some mammalian models but not others (Harris and Juriloff, 1999; Juriloff and Harris, 2000; Moffa and White, 1983; Shin and Shiota, 1999). For instance, folic acid deficiency is non-teratogenic in mice and guinea pigs, and is only teratogenic in rats when associated with concurrent vitamin C deficiency or with supplementation of a folic acid antagonist.

Formaldehyde

Antibacterial agents such as formaldehyde exhibit teratogenic discordance in 10 of 38 cases (26%). However, the case for formaldehyde, an antibacterial agent, preservative, and component of everyday items such as clothing and cosmetics, appears to be relatively straightforward. The only associations reported have been

weak, and in poorly conducted and/or confounding investigations. Consequently, formaldehyde is not considered a teratogen in any species so far tested, including mice, rats, hamsters, dogs and humans.

*Indomethacin (*B/D in 3rd trimester*)*

This drug has analgesic, anti-inflammatory and anti-pyretic properties, and, as a prostaglandin synthase inhibitor, is commonly used to arrest premature labour and to treat polyhydramnios. It is classified as a teratogen in mice and rats but not in rabbits, though results have been inconsistent. Some studies have demonstrated teratogenic effects with doses of 1–4x the human dose of indomethacin, whereas others have shown no adverse effects with doses up to 100x the human dose and/or up to the maternal lethal dose. Human studies on the whole have been negative, with the exception of a small number of case studies.

Isotretinoin (Accutane/Roaccutane)

Teratogenic effects have been ascribed to almost all retinoids and retinoid analogues in at least one species, though discordant results were present in 2 of 7 cases (29%). Isotretinoin (13-*cis*-retinoic acid) has been extensively studied, and is teratogenic in all species tested, although teratogenesis in non-human species such as rabbits, monkeys, hamsters, rats and mice requires doses several to many times greater than those used in humans.

Lactose

Lactose is one of a surprising variety of foodstuffs that have been shown in at least one species to exert a teratogenic effect: in this case, a variety of defects in rats and mice. Of 79 food additives tested in more than one animal species, the results of 14 (18%) were inconsistent across species.

Lead

Organic salts of lead have generally not been associated with teratogenic effects. However, inorganic lead salts have a strong correlation with teratogenesis, although there is a high degree of discordance in the results (3 of 4 salts tested in more than one animal species gave inconsistent results). For example, lead acetate is teratogenic in rats, hamsters and primates, but not in sheep, cows, mice, rabbits or guinea pigs. Lead carbonate is teratogenic in mice but not rabbits, and lead nitrate in rats, mice and hamsters but not guinea pigs. Behavioural studies in rats have been highly contradictory. In humans, there is no clear evidence to link maternal exposure to lead and its salts with structural defects in the newborn, though there is compelling evidence to associate it with intrauterine growth retardation, spontaneous abortion, problems with neural development and mental retardation, depending upon the level and timing of exposure.

*Lithium (*D*)*

Of the miscellaneous metals and their salts, 5 of 17 gave inconsistent results across species. Lithium salts are used to treat psychiatric illnesses, and extensive case studies and epidemiology have shown no convincing link to teratogenesis. Only one study has reported an association between maternal use of lithium salts during early pregnancy with cardiovascular malformations in offspring, but no satisfactory control data were available. In animals, serum levels of lithium salts within the human therapeutic range have elicited teratogenesis in mice and rats, whereas other studies in rats and rabbits have produced negative results. Other investigations have given equivocal results in pigs, and mice, negative results in monkeys, and strain-dependent results in rats.

*Methimazole (*D*)*

Methimazole is a thioamide given to treat hyperthyroidism. Of 5 thyroid-acting agents tested in more than one species, 4 showed discordant results. Human case reports are inconclusive; although a number of reports have been suggestive of a 'Methimazole embryopathy', no causal relationship has been established. Methimazole was not teratogenic to rabbits, but CNS abnormalities and behavioural anomalies have been reported in rats.

*Methotrexate (*X*)*

Methotrexate is an anticancer antimetabolite, which is also used at lower doses in the treatment of rheumatic disease. Of 22 such agents tested in more than one species, 9 show discordance (41%). It is a known human teratogen when administered in the first trimester of pregnancy, although the frequency of the characteristic craniofacial anomalies associated with its use in the offspring of mothers treated with the drug is not considered to be particularly high. In animal studies, it has been shown to be teratogenic at doses similar to the human therapeutic dose in chicks (Zamenhof, 1985), rabbits (DeSesso and Goeringer, 1992) and cats (Khera, 1976), and in mice only at doses four times greater (Skalko and Gold, 1974). In rats, methotrexate was teratogenic in embryonic culture only; administration to pregnant animals had no effect (Schmid, 1984). Experiments with primates have produced largely negative results, though a non dose-dependent teratogenic effect has been reported in the rhesus monkey at very high doses (Wilson *et al*., 1979). No teratogenic effect was found in experiments with dogs (Schardein, 1993).

Mirex

Insecticides such as Mirex show a considerable degree of discordance in their results; 19 of 42, almost half, gave inconsistent results across species. Mirex itself has been shown to be teratogenic in rats at moderate doses, where it affects predominantly the foetal cardiovascular system. In mice, however, no teratogenicity has been reported.

*Misoprostol (*X*)*

Misoprostol is an analogue of prostaglandin E1, used therapeutically as an abortifacient, in the treatment of peptic ulcer disease, and in the induction of labour. Of 6 prostaglandins tested for teratogenic potential in animals, 3 showed a high degree of discordance across species (50%). Consequently, it is no great surprise to find substantial discordance in results with misoprostol; a great deal of evidence supports its classification as a human teratogen from case studies that associate its use with a wide range of birth defects, although this is confounded by a significant number of reports that show no increase in anomalies. Animal experiments have also been equivocal, with similar numbers of positive and negative associations involving rats and rabbits.

Hyperbaric oxygen

Hyperbaric oxygen has been associated with teratogenesis in experiments involving rats, rabbits and hamsters, but only with prolonged exposure. Lower levels of exposure, however, do not show any teratogenic effect in these species. There is currently no data assessing the reproductive hazard in humans.

*Penicillamine (*D*)*

Penicillamine is classified as a 'Chemical antagonist', of which 9 from a total of 29 (31%) showed discordant results in animal testing. It is used as a chelating agent, and in the treatment of rheumatoid arthritis and cystinuria. Although classified as teratogenic in all 3 tested species (mice, rats and hamsters), deeper analysis of the literature revealed inconsistency; just as many studies showed high doses of the drug to be teratogenic in rats as did not, and some mouse studies associated penicillamine use with connective tissue abnormalities in offspring but not others. Human case reports are also contradictory.

*Quinine (*X*)*

Quinine is used at small doses to treat leg cramps, in higher doses as an anti-malarial, and at very large doses as a post-coital contraceptive and abortifacient. It is also a component of cold remedies and some foods. Of 3 anti-malarial drugs tested in more than one species, 2 showed discordant results. Results for quinine in particular varied considerably; although there is a teratogenic effect in the rat, guinea pig and chinchilla, there is no effect in rats, mice, dogs or primates. In humans, there is a strong link between high doses of quinine and teratogenesis when administered in the first trimester of pregnancy only.

Thalidomide

This sedative, prescribed to pregnant women in the 1950s to control nervousness and nausea, has become the most notorious teratogen known to man. Various

characteristic malformations, particularly limb-reduction defects, were associated with its use during the fifth and sixth weeks of pregnancy in 20% of cases, even with doses as little as one 50 mg capsule. Later investigations in pregnant animals were performed in the early 1960s using pregnant mice, rats and guinea pigs, but revealed no malformations in their offspring. Eventually, one particular strain of rabbit (New Zealand) was found to be teratologically sensitive to thalidomide during a specific 2-day window of pregnancy, and subsequent extensive investigations have demonstrated extreme variability in species susceptibility. The embryonic development of cats, hamsters, rats and mice has been adversely affected, though all four species required large doses of the drug and only a very few strains of the latter three species were reactive. Other strains of rabbit than the New Zealand are now known to be reactive, though again only at very large doses. Several species of primate have shown sensitivity to thalidomide including baboons, macaques and green monkeys, though bushbabies have not. Dogs are classified as positive due the induction of malformations in mongrels, and in beagles at high doses; results for pigs have been equivocal, and for guinea pigs entirely negative.

Other drugs in the sedatives–hypnotics class showed discordant results across species in 4 of 18 instances (22%); thalidomide, its analogues and related chemicals did so in 7 of 21 cases (33%).

*Warfarin (*D*)*

Of 13 drugs classified as 'affecting blood', only 2 showed irregular results among different animal species; one of these is warfarin, which is used in the treatment of a variety of thromboembolic disorders, such as deep vein thrombosis (DVT). It has gained notoriety for teratogenesis in humans, being associated with characteristic defects known as foetal warfarin syndrome. These defects are predominantly skeletal when exposure is in the first trimester of pregnancy, and of the CNS in the second or third trimesters. Teratology studies have produced discordant results in animals; studies in mice showed no teratogenic effects and produced only an increase in stillbirths at high doses, which may have been a result of warfarin's vitamin K antagonism. No teratogenicity was observed in rabbits, but birth defects have been induced in rats irrespective of vitamin K status.

Alcohol

Alcohol (specifically ethanol), widely consumed as a recreational drug, has long been strongly associated with human birth defects. A characteristic set of abnormalities known as Foetal Alcohol Syndrome (FAS) is present in up to 1 in 3 children of alcoholic mothers; manifestations include growth deficiency, CNS problems including mental retardation and behavioural difficulties, and characteristic facial features and organ malformations. These abnormalities are generally similar to those in the many animal species that have been the subject of teratology testing with human-equivalent doses of ethanol, including rats, mice, rabbits, ferrets, guinea pigs, sheep,

pigs, dogs, and non-human primates. Indeed, of all the teratology results for all of the substances considered here, the cross-species results for ethanol arguably show the highest level of concordance.

ROLE OF EPIDEMIOLOGY AND CASE STUDIES

The significant fact is that virtually every substance or dietary deficiency currently recognized as being teratogenic in humans was initially identified as a result of case reports and clinical series (Polifka and Friedmann, 1999). This is not to say that it is desirable to rely upon teratogenic effects in humans to derive an 'after the event' classification; simply that these forms of data continue to be the most powerful and reliable way to classify substances with respect to their teratogenic potential, despite the huge amount of animal-based information generally available. Epidemiology, i.e. 'studies to observe the effect of exposure of a single factor upon the incidence of disease in human populations' (Black and Lilienfield, 1984), is therefore arguably the critical factor in the identification of human teratogens, and has been the major contributor to the characterization of almost all (if not all) human teratogenic agents. Infamous examples include thalidomide, foetal alcohol syndrome, foetal hydantoin syndrome, spina bifida, foetal rubella syndrome, and the effects of diethylstilbestrol (DES), mercury, and folic acid deficiency (Greek R, date unknown). The discovery of associations such as these was and remains possible because of characteristic patterns of anomalies exhibited in the offspring of mothers who have been exposed to particular agents at similar times during pregnancy. These observations are widely regarded as the most reliable and indeed only certain means of human teratogen identification through quantitative estimation of the statistical significance of any suspected associations, and of ensuring the subsequent public protection that follows from them (Polifka and Friedmann, 1999; Schardein, 1993). Critics of the importance of epidemiology and case studies point towards its 'after the fact' nature and reliance upon the existence of people already afflicted; this aspect should not detract from the contribution these methods have made to preventing many thousands of congenital abnormalities that could not have been predicted by relying upon animal data. Indeed, central birth defect registries have been set up in many countries (for example Canada and Australia) to expedite the discovery of potential teratogens through retrospective epidemiology methods (Schardein, 1993).

IN VITRO ALTERNATIVES TO ANIMAL-BASED TERATOLOGY

Animal teratology experimentation has not been significantly effective in the prevention of teratogen-induced birth defects, which affect some 2–3% of the 4% of all children born in the United States with developmental anomalies. Despite the continued predominance of animal-based teratogen testing and the ostensible intransigence of the scientific community involved in it with regard to alternative

methods, these do exist, with the potential to revolutionize teratology research by streamlining it in terms of time, cost, and improved efficacy. Surprisingly though, considering the difficulty in interpreting animal experiments and extrapolation to humans, the development of these methods could barely have progressed more slowly, and their adoption into common practice has arguably been resisted in many quarters. Only relatively recently have scientists begun to turn to *in vitro* methods to compensate for the inherent problems and discordance in animal teratology. These methods are now well established and invaluable for conducting mechanistic studies, are extremely useful for 'secondary testing', i.e. the screening of chemicals structurally related to a known teratogen (Brown *et al.*, 1995), and are less expensive, faster, and much more reproducible. In any case, there is now an absolute necessity for alternatives to conventional animal-based methods due to the immense number of substances not only already in the marketplace but being released into the environment every year. There are an estimated 50 000 to 70 000 chemicals to which humans are exposed on a regular basis, and up to 400 new drugs introduced to the market each year (Bournais-Vardiabasis, 1994). It was generally accepted long ago that *in vivo* teratology testing could not possibly be applied to all new chemical substances due to the scale of the demand upon time and resources (Kotwani *et al.*, 1995). Crucially, a European Union (EU) white paper published in 2001 proposed the harmonization of testing requirements for chemicals marketed before September 1981 (Registration, Evaluation and Authorization of CHemicals, 'REACH'), which amounts to approximately 30 000 substances (Hartung *et al.*, 2003). If current animal-based methodology were adhered to, this endeavour would require the use and sacrifice of an estimated 12.8 million animals at a cost of 8.68 billion Euros (11 billion US dollars) over a period of 45 years (Hartung *et al.*, 2003). Clearly, there can no longer be any justified resistance to the development and adoption of alternative teratology methods.

A wide variety of alternatives have been researched, all of which have sought to follow the principle of the 'Three Rs' (Reduction, Refinement and Replacement), which provide a rational strategy to minimize and eventually replace animal models with non-animal alternatives without compromising high quality science (Russell and Burch, 1959). These 'Replacement' alternatives can be divided into the following categories (Nielsen *et al.*, 2001).

Information

Some data from teratology studies are not freely available in the public domain, whether from proprietary considerations due to commercial significance, or simply as a result of access problems and/or difficulties with the construction and maintenance of appropriate databases. Unrestricted and easier access to this information would avoid repetition of completed animal-based work, and allow data comparison in the search for alternative methods without new animal work being done.

Computer based systems

Computer-based research is already invaluable in the field of drug discovery and development, and has been responsible for the elimination of many animal tests in the pre-screening of candidate drug compounds. Methods employing physiologically based pharmacokinetic modelling (PBPK) and structure activity relationship (SAR) analyses have already been widely and successfully used: PBPK acts to predict their effects through the integration of cross-species physiological parameters coupled with specific information about chemicals and their metabolites. SAR uses information concerning the molecular structure and properties of chemicals to predict biological responses to them, and has shown very high accuracy in the prediction of dermal sensitisation, carcinogenicity and teratogenicity.

Physico-chemical techniques

Often used in conjunction with SAR, knowledge of the physico-chemical aspects of chemicals can allow the prediction of their biological effects.

Human studies

Almost all exposures currently recognized as being teratogenic in humans were initially identified in case reports and clinical series (Polifka and Friedman, 1999). Indeed, links between maternal exposure to particular agents during pregnancy and characteristic patterns of congenital anomalies provide the only means of quantitating the statistical significance of clinical associations between chemical agents and teratogenesis. Case reports are typically the first accounts of foetal abnormalities following maternal exposure to, for example, a newly marketed drug, and can be extremely useful in establishing strong associations even from small numbers of cases, when a drug is rarely used or malformations rare. Warfarin (Saxen, 1975), DES (Herbst and Scully, 1970) and isotretinoin (Rosa, 1983) are important examples. Epidemiological studies may take two forms: Cohort studies use a group of mothers who were exposed to a particular substance during pregnancy as a basis, and examine their children for any increased occurrence of abnormalities compared to children of unexposed mothers. Prospective cohort studies have been useful in establishing the safety of Bendectin, isotretinoin, fluoxetine and acyclovir (Goldstein *et al.*, 1997; Lammer *et al.*, 1985). Case-control studies, on the other hand, use children with specific malformations as a basis, and compare maternal exposure with mothers of non-affected children. Additional measures have been taken to augment the information provided by these studies. Teratology information services worldwide now record data prospectively and track cases on a long-term basis. This not only reduces the chances of recall bias in the mothers involved, but collaboration between such services permits the more effective study of rare events, while the longer-term attention to cases allows neurobehavioral effects to be detected; examples include carbamazepine,

isotretinoin, phenytoin, valproic acid and warfarin (Koren *et al.*, 1998). Also, meta-analyses are becoming more and more frequent, in which several studies are combined yielding statistically more robust results (Einarson *et al.*, 1988).

Lower organisms and embryo stages/cell, tissue and organ cultures

These latter two groups constitute the main basis of the *in vitro* systems currently being used, developed and researched in the field of teratogen testing, and may themselves be grouped into four categories (Brown, 1995): (a) Established cell lines (cell lines and embryonic stem cells), (b) Primary cell cultures, (c) Non-eutherian embryos (both vertebrate and non-vertebrate), (d) Cultured mammalian embryos/primordia.

Initially, cell lines such as human embryonic palate mesenchymal (HEPM) cells (Pratt *et al.*, 1982), mouse ovarian tumour (MOT) cells (Braun *et al.*, 1982), and neuroblastoma cells (Mummery *et al.*, 1984) were used in screening experiments, though the over-simplified nature of these systems resulted in an unacceptably high number (>50%) of false positives. Subsequent work with super-confluent 'micromass' cultures (Umansky, 1966) showed them to be superior to these cell lines (discussed in more detail below), though they require the primary culture of cells from embryos. Totipotent embryonic stem cell (ES) lines derived from mammalian blastocysts are arguably better still, allowing the development of various desirable and appropriate phenotypes *in vitro*, and all from permanent cell-lines. These also are discussed in more detail below.

In addition to cell lines, numerous tests involving sub-mammalian and invertebrate embryos have been described to assay for teratogenic potential, such as fish (Guiney *et al.*, 1980; Melancon *et al.*, 1977), frogs (Dawson *et al.*, 1989; Sabourin and Faulk, 1987), crickets (Waltan, 1983), *Drosophila* (Bournais-Vardiabasis and Teplitz, 1982), *Hydra* (Johnson and Gabel, 1983) and chicks (Gebhardt, 1972). Initial results have been promising with many systems: for example, the *Hydra* test has shown a good degree of accuracy, and is predictive, for example, for thalidomide teratogenesis (Mayura *et al.*, 1991; Wilby *et al.*, 1986). The *Drosophila* system, based upon the monitoring of cell death and proliferation events, has shown a high degree of concordance not only with *in vivo* animal studies but also with human epidemiological data; furthermore, the vast and detailed knowledge of *Drosophila* molecular biology could potentially be of use in the derivation of teratogenic mechanisms (Bournais-Vardiabasis, 1994). Use of the chick embryo has been extensively investigated, and its suitability questioned in some quarters (Karnofsky, 1965) while, with some revisions, favoured in others (Jelinek, 1977; Wilson, 1978). Initially, *in ovo* testing was thought to be too sensitive and variable, though the refined chick embryotoxicity screening test (CHEST) allows easy monitoring of growth retardation, malformation and death, and the determination of dose-response relationships (Jelinek, 1977). Although not permitting a comprehensive study of embryogenesis, the alternative *in vitro* chick embryo culture system is easier, cheaper, more ethical (not requiring animal sacrifice) and appears to give comparable results (Kucera

Table 4.
Performance characteristics of FETAX

Performance characteristics	FETAX, with and without metabolic activation, compared to combined laboratory mammal (using TI > 1.5)	FETAX, with and without metabolic activation, compared to human (using MCIG/LC_{50} < 0.30)	Combined laboratory mammal compared to human
Accuracy	61% (55/90)	70% (19/27)	63% (19/30)
Sensitivity	82% (41/50)	67% (8/12)	71% (10/14)
Specificity	35% (14/40)	73% (11/15)	56% (9/16)
Positive predictivity	61% (41/67)	67% (8/12)	59% (10/17)
Negative predictivity	61% (14/23)	73% (11/15)	69% (9/13)
False positive rate	65% (26/40)	27% (4/15)	44% (7/16)
False negative rate	18% (9/50)	33% (4/12)	29% (4/14)

Taken from NICEATM FETAX Background Review Document: Executive Summary, 10th March 2000, http://iccvam.niehs.nih.gov/about/overview.htm. Performance characteristics of FETAX were compared to three laboratory animal species combined (rat, mouse and rabbit), based on a Teratogenic Index (TI) value greater than 5, where TI = LC_{50} (concentration inducing lethality in 50% of exposed embryos) divided by EC_{50} (concentration inducing malformations in 50% of exposed embryos). Performance was also compared against human teratogenic data, based on the minimum concentration to inhibit growth (MCIG)/LC_{50} ratio of less than 0.30. Numbers in parenthesis indicate the number of accurate results/the total number of substances compared.

et al., 1993). While a report from the twelfth workshop of the European Centre for the Validation of Alternative Methods (ECVAM) recommended further consideration and analysis of the CHEST method, it also concluded that the only non-avian vertebrate system worth pursuing was that of the South African clawed frog (*Xenopus laevis*), known as the Frog Embryo Teratogenesis Assay (FETAX). This is a 96-hour whole embryo *in vitro* test based on the fact that the first 96 hours of *Xenopus* development parallels many of the major processes of human organogenesis (NICEATM FETAX exec. summary document, 2000). The specificity and accuracy of the FETAX system is enhanced by the addition of an exogenous metabolic activation system (MAS) to account for an inherent deficiency of *Xenopus* in this respect, and the primary endpoints include mortality, malformations and growth inhibition. Using these endpoints to assess the teratogenic effects of up to 90 substances, FETAX demonstrated more accurate, more specific, and more predictive results when compared to human teratology data than when compared to animal teratology data (Table 4), and although the false negative rate was higher, the false positive rate was considerably lower. While FETAX has been proposed as a screen

for hazard identification as opposed to an outright replacement for animal-based studies, one major advantage of its use would be a highly significant reduction and refinement in the numbers of animals used: fewer organisms are used per dose level, these animals are both non-mammalian and embryonic, and it is estimated that each successful FETAX assay would spare the use of 190 rats and 112 rabbits (ASTM, 1998). Animal use would also be reduced if FETAX was used (a) in the preliminary stages of product development to aid in the selection of compounds least likely to show teratogenic effects, (b) to compare the teratogenic activity of a chemical highly structurally related to one that has already undergone *in vivo* testing, and (c) to evaluate compounds where the anticipated exposure is very low (Spielmann *et al.*, 1998).

The culture of mammalian embryos has also been extensively researched in a number of species, and is discussed at length below. Organ culture has also been investigated, which has a number of advantages and disadvantages when compared to the use of whole live animals. It allows precise control of teratogen dose and exposure time to heterogeneous tissue, permits close observation of development and detection of assayable endpoints such as cell death, differential growth through regional proliferation, and cell differentiation and movement, and enables these studies to be performed free of confounding factors. Disadvantages include its delicate, labour intensive and short-lived nature. Also to be considered is the ethically charged subject of human organ culture from embryos resulting from therapeutic abortions; the advantages are obvious for investigating human teratogens. Its value can be seen in an example involving maternal rubella infection and congenital cataracts in children (Pratt, 1980). Having found no good animal model for the infection of embryonic eyes with rubella virus, the investigators discovered that human embryos gave outstanding results: eyes infected with rubella developed lens cataracts, while non-infected controls remained cataract-free.

At the 17th meeting of the ECVAM Scientific Advisory Committee in 2001, three of the aforementioned methods were endorsed as 'scientifically validated' and ready for consideration for regulatory acceptance and application, and therefore deserve a more detailed discussion. Detailed reports on the performance of these validated tests have recently been published (Genschow *et al.*, 2004; Piersma *et al.*, 2004; Spielmann *et al.*, 2004) These are; the Embryonic Stem-Cell Test (EST), the Micromass test (MM), and the Whole Embryo Culture test (WEC):

The Embryonic Stem-cell Test (EST)

In the EST, two permanent murine cell lines are used to assess teratogenic potential (Genschow *et al.*, 1999, 2000; Scholz *et al.*, 1999a, b; Spielmann *et al.*, 1997): embryonic stem (ES) cells known as D3 cells represent embryonic tissue, and 3T3 fibroblast cells represent adult tissue. The D3s are maintained in an undifferentiated stage in the presence of leukaemia inhibiting factor (LIF), then released from LIF and allowed to form embryo bodies that differentiate into cardiomyocytes. The cultured D3 and 3T3 cells are then exposed to a range of concentrations of the

Table 5.
Overall statistics for the EST, MM and WEC tests following validation studies

	EST %	MM %	WEC PM1 %	WEC PM2 %
Predictivity for non-embryotoxic	73	57	56	70
Predictivity for weakly embryotoxic	69	71	75	76
Predictivity for strongly embryotoxic	100	100	79	100
Mean	81	76	70	82
Precision for non-embryotoxic	68	80	70	80
Precision for weakly embryotoxic	84	60	45	65
Precision for strongly embryotoxic	83	69	94	100
Mean	78	70	70	82
Accuracy	78	70	68	80

Predictivity is an estimate of the likelihood that a positive or negative prediction correctly identifies an embryotoxic/non-embryotoxic substance respectively, under the proposed condition of use. Precision is defined as the proportion of correctly classified chemicals from *in vitro* tests relative to *in vivo* tests. Accuracy takes into account all of these results. Performance is classed as being 'sufficient' if $\geqslant$ 65%; 'good' if $\geqslant$ 75%; 'excellent' if $\geqslant$ 85%. Predictivity, precision and accuracy figures shown were calculated without outliers; including outliers, they were higher still. (Adapted from information from a report of the 17th meeting of the ECVAM scientific advisory committee, and *Alternatives to Laboratory Animals* **32** (3) 2004).

potentially embryotoxic substance, and after 10 days of culture three endpoints are assayed: the inhibition of D3 cell differentiation, the inhibition of D3 cell growth, and the inhibition of 3T3 cell growth. Three values are derived: (i) the concentration of substance at which there is 50% inhibition of D3 differentiation (ID_{50}), (ii) the concentration at which there is 50% inhibition of D3 growth ($IC_{50\,D3}$), and (iii) the concentration at which there is 50% inhibition of 3T3 growth ($IC_{50\,3T3}$). These values are then channelled through a number of equations in the 'Improved Prediction Model', the results of which allow the classification of the substance under examination into one of three classes; *not embryotoxic, weak embryotoxic*, and *strong embryotoxic*. When applied to a range of test compounds in the official validation study and compared to the equivalent *in vivo* results, the predictability, precision and accuracy of the EST scores highly (Table 5). The predictability for non-and weakly embryotoxic substances is classed as sufficient, approaching good, and for strongly embryotoxic substances it is 100% accurate. Precision for non-embryotoxic substances was sufficient, but good and approaching excellent for both weakly and strongly embryotoxic chemicals. Overall accuracy was scored at 78%, or good. These results are impressive, but it is particularly encouraging to examine efforts being made to optimise the protocol and further improve its performance. These involve the use of alternative endpoints of a molecular nature, as opposed to the interpretive aspects of the current microscopic methods. They include, for example, the detection of tissue-specific protein expression via immunofluorescent antibodies (such as Fluorescence Activated Cell Sorting or FACS: Buesen *et al.*,

2003); using reporter-genes in stably-transfected cells activated by cardiac muscle-cell development (Kolossov *et al.*, 1998); and assaying development-specific gene expression by RT-PCR methods (Wiles, 1993). The outlook for the EST appears promising; even if it is accepted as a pre-screening method rather than a form of replacement for animal-based techniques, it will have achieved a '3Rs' success by greatly reducing the numbers of animals involved and sidestepping a need for the use of pregnant animals to provide tissue for *in vitro* culture, and will still provide useful and pertinent information for human teratology assessments.

THE MICROMASS (MM) TEST

The MM test is based upon rat embryonic limb mesenchyme cells which, when cultured in small volumes at high density, form foci of differentiating chondrocytes within a background of non-differentiated cells (Flint and Orton, 1983). Cultured embryonic rat midbrain cells forming foci of neurons were also examined, though subsequent experiments produced more variable results. The foci of differentiating chondrocytes exhibit cell proliferation, differentiation, movement, cell-to-cell communication and interactions with the extracellular matrix, essentially reproducing cartilage histogenesis, a fundamental step in skeletal morphogenesis.

In practice, the end-point used in the MM test is a 50% inhibition of differentiation (ID_{50}) as assessed by the staining of cartilage cells with an alcian blue dye. In the validation study managed by ECVAM, a prediction model similar to that for the EST was then used to classify test substances as non-, weakly, or strongly embryotoxic. When applied to a range of test compounds and compared to the equivalent *in vivo* results, the MM test was found to be highly reproducible, with correlation to *in vivo* results designated as 'sufficient' with an accuracy of 70% (Table 5). Predictability for strongly embryotoxic chemicals was 'excellent' at 100%, with sufficient precision at 69%. The test did, however, show a slightly lower overall predictability and precision for teratogens than the EST. Due to the 'insufficient' classifications of non-embryotoxic predictability and weakly embryotoxic precision (57% and 60%, respectively), the MM test was proposed for consideration for the screening and identification of strong embryotoxic chemicals only.

WHOLE EMBRYO CULTURE (WEC)

Teratogen screening systems using whole mouse (Sadler *et al.*, 1982), rat (Schmid, 1985) and rabbit (Ninomiya *et al.*, 1993) embryos cultured for short periods during the phase from fertilization to the end of organogenesis have been described. These methods involve the dissection of embryos at the head-fold or early somite stage away from maternal tissue, parietal yolk sac and Reichert's membrane, leaving the visceral yolk sac and ectoplacental cone intact. The conceptus is then cultured in an appropriate high-serum medium for 24–48 hours, during which time the

test substance is added. Metabolic activation systems may also be included, such as S9 or microsomal fractions from liver, or co-culture with hepatocytes (Fantel *et al.*, 1979; Kitchin *et al.*, 1981; Zhao *et al.*, 1993). In terms of appropriate end-points and to correctly classify the embryotoxic potentials of test chemicals, adverse effects on yolk sac development, embryonic growth and differentiation are assayed in addition to dysmorphogenesis. An elaborate morphological scoring system was initially suggested to derive some measure of embryo malformation and/or retardation (Brown and Fabro, 1981). This became difficult to apply reproducibly in practice, so two prediction models were developed (PM1 and PM2; Genschow *et al.*, 2000). PM1 involves two end-points; IC_{50} for malformation (the concentration of test substance at which 50% of embryos are malformed), and IC_{NOEC} for TMS (the maximum concentration that has no observable effect on the total morphological score, TMS). However, this predictive model considers only differentiation and developmental parameters and not cytotoxicity data, and so PM2 was developed by incorporating IC_{50} 3T3 data from the accompanying EST results. The results for both predictive models are shown in Table 5, and indicate that the incorporation of EST information, and thus the results from the PM2 model, increases predictability, precision and accuracy considerably. For PM2, both predictability and precision for strongly embryotoxic chemicals were 100%, and overall accuracy was classed as 'excellent' at 80%. On the basis of these results, the ECVAM scientific advisory committee advised that the WEC test had been scientifically validated, and was ready for regulatory review. The technique does, however, have certain negative aspects: it is complex, each test can cover only one part of organogenesis, it requires high levels of skill, can be costly, and still requires animal use to provide tissue and serum.

Considering the poor performance of animal models with regard to human predictability, it may seem strange that their *in vitro* alternatives should be compared directly to them, instead of against human case studies, epidemiological data and meta-analyses. Selection of an appropriate panel of test substances for validation purposes has proved to be a significant challenge, which comes as no surprise considering the complexity and disarray in the *in vivo* data. Thirty reports of *in vitro* validation studies were published prior to 1997, in which over 300 chemicals were tested; all proposed lists of 'ideal' substances for validation were subsequently found to be deficient (Brown, 2002) in terms of over-simplification, inconsistency and bias. Table 6 summarizes animal data and human classifications (according to the TERIS database, MICROMEDEX) for the eventual list of twenty chemicals finally selected for use in the EST, MM test and WEC test validation studies by ECVAM. Even this list seems far from ideal when the available animal and human information for the constituent chemicals is listed, a criticism accepted in Brown's report. For example, two of the 'non-embryotoxic' chemicals show teratogenic effects in the rat; of the weakly embryotoxic chemicals, insufficient information was available to assess the teratogenic risk in humans for three of the seven, and the risk of another was defined as 'unknown'. Of the six 'strongly embryotoxic'

Table 6.
Summary of animal data and human classifications for the eventual list of twenty chemicals selected for use in the EST, MM test and WEC test validation studies by ECVAM

In vivo embryotoxicity	Substance	Species											Human classification
		Mouse	Rat	Rabbit	Hamster	Primate	Dog	Cat	Pig	Ferret	Guinea Pig	Sheep	Cow	
None	Acrylamide	−	−											NI
None	Diethyl phthalate	−	+											NI
None	Isobutyl-ethyl-valproic acid													NI
None	D-(+) Camphor		−	−										Unlikely
None	Diphenhydramine	−	−	−										Unlikely
None	Penicillin G	−		−										—
None	Saccharin (sodium)	−	±	−										Unlikely
Weak	Boric acid	+	+											Unknown
Weak	Salicylic acid	+	+											Minimal risk
Weak	Valproic acid (VPA)	+	+	+	+	+								Moderate risk
Weak	Pentyl-4-yn-VPA													NI
Weak	Lithium chloride	+	+											Small risk
Weak	Dimethadione	+	±			+								NI
Weak	Methoxyacetic acid		+											NI

Table 6.
(Continued)

In vivo embryotoxicity	Substance	Species												Human classification
		Mouse	Rat	Rabbit	Hamster	Primate	Dog	Cat	Pig	Ferret	Guinea Pig	Sheep	Cow	
Strong	Hydroxyurea	+	+	+	+	+	+	+	+					Undetermined
Strong	Aminonicotinamide	+	+	+	+	±			+					Unlikely
Strong	Bromodeoxyuridine	+	+		+									NI
Strong	Methylmercury chloride	+	+	+	+	−	+	+		+				Small to moderate
Strong	Methotrexate	+	+	+		±	−	+						Moderate to high
Strong	*All trans* Retinoic acid	+	+	−	+	+				+				Undetermined

Substances are grouped by classifications into non-, weak, and strong embryotoxicants. Teratological classifications in various animal species are shown where data are available: + = teratogenic, ± = equivocal results, − = non-teratogenic. The teratogenic risk in humans, as defined by the REPROTOX database (Micromedex) is indicated, where 'NI' = no information available, and '-' = non-teratogenic. This list seems far from ideal when the available animal and human information for the constituent chemicals is listed, a criticism accepted in Brown's report. For example, two of the 'non-embryotoxic' chemicals show teratogenic effects in the rat; of the weakly embryotoxic chemicals, insufficient information was available to assess the teratogenic risk in humans for three of the seven, and the risk of another was defined as 'unknown.' Of the six 'strongly embryotoxic' substances, the primate gave equivocal results for two and a negative result for one: negative classifications were also obtained for dogs and rabbits. Only one of the six substances was classed as a moderate to high risk for humans, the other five being undetermined, unlikely, or lower risk. Of these twenty chemicals, just seven were listed as having human data: two non-embryotoxicants were classed as non-teratogens in humans, one of which had supporting and concurrent animal data. Two chemicals in the 'weak embryotoxicant' class were defined as positive and suspected positive human teratogens respectively. Three 'strong embryotoxicants' were all defined as positive human teratogens; one was without supporting animal data, one had concurrent animal data, and the other, methylmercury, had contrasting animal data (the results of three of four animal experiments did not show teratogenicity).

substances, the primate gave equivocal results for two and a negative result for one: negative classifications were also obtained for dogs and rabbits. Only one of the six substances was classed as a moderate to high risk for humans, the other five being undetermined, unlikely, or lower risk. Of these twenty chemicals, just seven were listed as having human data: two non-embryotoxicants were classed as non-teratogens in humans, one of which had supporting and concurrent animal data. Two chemicals in the 'weak embryotoxicant' class were defined as positive and suspected positive human teratogens respectively. Three 'strong embryotoxicants' were all defined as positive human teratogens; one was without supporting animal data, one had concurrent animal data, and the other, methyl mercury, had contrasting animal data (the results of three of four animal experiments did not show teratogenicity). The committee was unclear as to how much weight should be given to human effects in the face of contrasting animal effects and vice versa, and to animal effects in the absence of human data. The relationship between maternal and developmental toxicity ('A/D') was used as a discriminator, though this is known not to be constant across species (Daston *et al.*, 1991). The committee also relied primarily on animal data due to the paucity of systematic human data for negative chemicals, and many known human teratogens were not considered for reasons ranging from known mechanisms, 'complex mixtures', and the fact that certain substances were known to act via well-characterized receptor interactions for which assays could easily be assembled.

DISCUSSION

Identification of environmental agents that cause damage to unborn children is absolutely imperative. Armed with the knowledge of those drugs, chemicals and foodstuffs that human beings are routinely exposed to that are truly teratogenic, exposures can be avoided, reduced or limited, and the incidence of tragic teratogen-induced birth defects drastically lowered. In the four decades since the thalidomide tragedy, the burden of identifying of human teratogens has rested squarely on the shoulders of animal-based research. The fruits of these forty years of endeavour, as summarized and presented in this review from a purely scientific viewpoint, amount to little more than an almost futile effort to achieve the stated goal, and a huge waste of resources.

Our examination of the data for 11 groups of known human teratogens across 12 animal species (Tables 1a and 1b) reveals huge variability in positive predictability, the mean of which is only 61%. This is little better than chance. Some species also show a high rate of false negative results, and almost a quarter of all outcomes in the six main species used (mouse, rat, rabbit, hamster, primate and dog) are equivocal. When the focus of attention was switched to individual substances linked with human teratogenicity rather than groups of substances (Table 2), only 56% of classifications were positive, the remaining 44% were almost entirely negative, with only a small number equivocal. Mean positive

predictability was 60%, and negative predictability just 54%. Of all substances for which there is teratology data in more than one species, discordance is apparent in just under 30% of 1396 results (Table 3); inconsistency was particularly prevalent in the results for narcotic analgesics, anticonvulsants, anti-psychotic drugs, corticosteroids, fungicides, insecticides, and plants and phytotoxins.

Just less than 40% of substances that undergo teratogenicity testing in animals show teratogenic potential, which may be defined as positive, probable or possible. Teratogenic findings for such benign substances as water, table salt, oxygen, sugars, food oils, and various vitamins and hormones raise serious questions as to the validity of established animal-based methods. Furthermore, fewer than 1 in 40 of the substances designated as potential teratogens from animal studies have been conclusively linked to human birth defects, and of all these known human teratogens only roughly half are teratogenic in more than one primate species. Are we really so resistant to the threat of teratogenesis compared to other animals, or should the value of the animal data be called into question? It is startling and disconcerting to realize that virtually all human teratogens so far identified have been discovered as a result of case reports and clinical series, i.e. despite animal studies, rather than because of them. Those that defend the animal model point to the fact that all human teratogens are also teratogenic in at least one species of animal: the worth of this statement is difficult to appreciate. There is little value in showing that one of the relatively few substances known to be teratogenic in humans can also be teratogenic in at least one of, say, twelve animal species at a particular dose by a particular route at a particular time and in a particular strain. Conversely, it means everything to show that a substance that exhibits teratogenesis in animals poses a genuine threat to unborn human beings, or that a substance which shows no teratogenic effects in animals will also be non-teratogenic in humans. Unfortunately, this, as we have shown, is where the animal-human extrapolation system has failed.

The degree of disarray between animal species and the problematic nature of extrapolation to humans comes as no surprise when one considers the number and variety of factors involved. In summary: susceptibility to teratogenesis varies between different species, different strains and among individuals; affected individuals frequently show different phenotypes, and all these aspects are influenced by genetic make-up, environmental factors, and metabolic and placental differences. Results are further affected by anatomical differences, differences in routes of administration, dose levels and strategies, differences in absorption, distribution, metabolic activation, sensitivity and excretion, and by typically stressful laboratory handling and housing conditions which can impair health.

We propose that the demonstrably poor performance of animal models as a basis for teratology research is sufficient argument against its continued practice. When coupled with the ethical problems of large-scale animal use and a commitment to the '3Rs', in addition to the question of the quality of the science and ultimate impact upon human beings, the *status quo* is untenable. This is underscored by the availability and advancement of the *in vitro* alternatives, some of which were

first mooted over twenty years ago. The perceived intransigence of the teratology community with regard to these alternative methods may have been evidenced by the demanding set of validation criteria attached to them, criteria that have never been applied to the animal tests against which they are being measured. Nevertheless, three *in vitro* alternatives were finally endorsed as scientifically validated by ECVAM in 2001; the embryonic stem-cell test (EST), the micromass test (MM) and whole embryo culture (WEC). The best results (as determined by ECVAM, comparing the *in vitro* results to data largely derived from animal tests) were obtained by a combination of WEC and EST tests known as 'WEC test, predictive model 2'. This was only marginally better than the EST alone however, and the EST has the distinct advantage of not requiring the use of animals at all. These validations may be timely: we are exposed regularly to an estimated 50 000–70 000 chemicals, only a fraction of which have been tested for teratogenicity; approximately 400 new drugs are introduced to the marketplace every year, and new European harmonization of chemical testing requirements is about to demand the generation of toxicology data for 30 000 substances. Left to animal-based methods alone, this task would take half a century, and fail to keep pace with the constant introduction of new chemicals. Alternative methods are therefore an absolute necessity. Not only will the technology develop and the tests become more reliable, but the *in vitro* methods are already more reproducible, provide easier end-points, present no problems with respect to 'route of exposure', placental transfer and metabolic differences, and are devoid of the plethora of confounding factors associated with the animal tests. Perhaps most importantly, the alternatives provide a means to establish vital mechanistic models of teratogenic action and reduce the human impact of the false positive and false negative results generated by unreliable and confusing animal models. As human cell culture technology improves, particularly regarding stem cells, new methods will undoubtedly evolve that will enable a closer *in vitro* approximation of *in vivo* human teratogenesis. We believe the information presented here provides a scientific basis for teratology research to move away from the old, inconsistent, futile animal models, through the adoption of now validated *in vitro* methods, towards the development of new, cheaper, easier, more reliable, more predictive *in vitro* screening techniques.

Acknowledgements

This work was funded by the Physicians Committee for Responsible Medicine, 5100 Wisconsin Ave., Suite 400, Washington, DC 20016, USA.

REFERENCES

ASTM (1998). Standard guide for conducting the frog embryo teratogenesis assay-Xenopus. Designation E 1439-91, in: *Annual Book of ASTM Standards*, Vol. 11.5, pp. 825–836. American Society for Testing and Materials, Philadelphia.

Balcombe, J., Barnard, N. and Sandusky, C. (2004). Laboratory routines cause animal stress. *Contemporary Topics in Laboratory Animal Science* (in press).

Baxter, H. and Fraser, F. C. (1950). The production of congenital defects in the offspring of female mice treated with cortisone, *McGill Med. J.* **19**, 245–249.

Beall, J. R. and Klein, M. F. (1977). Enhancement of aspirin-induced teratogenicity by food restriction in rats, *Toxicol. Appl. Pharmacol.* **39**, 489–495.

Belafsky, H. A., Breslow, S., Hirsch, L. M., *et al.* (1969). Meprobamate during pregnancy, *Obstet. Gynecol.* **34**, 378–386.

Black and Lilienfield. Epidemiologic proof in toxic tort litigation, *52 Fordham Law Rev.* **52**, 732–755.

Bournais-Vardiabasis, N. (1994). An alternative *in vitro* method to detect teratogens utilizing *Drosophila melanogaster* embryos, *Humane Innovations and Alternatives* **8**, 630–634.

Bournais-Vardiabasis, N. and Teplitz, R. L. (1982). Use of Drosophila embryo cell cultures as an *in vitro* teratogen assay, *Teratogenesis Carcinogen Mutagen* **2**, 335–341.

Braun, A. G., Nichinson, B. B. and Horowicz, P. B. (1982). Inhibition of tumour cell attachment to concanavalin A-coated surfaces as an assay for teratogenic agents: approaches to validation, *Teratogenesis, Carcinogenesis, Mutagenesis* **2**, 343–354.

Brent, R. L. (1995). The application and principles of toxicology and teratology in evaluating the risks of new drugs for treatment of drug addiction in women of reproductive age, *NIDA Res. Monogr.* **149**, 130–184.

Brown, N. A. (2002). Selection of test chemicals for the ECVAM international validation study on *in vitro* embryotoxicity tests, *Alt. Lab. Anim.* **30**, 177–198.

Brown, N. A. and Fabro, S. (1981). Quantitation of rat embryo development *in vitro*: a morphological scoring system, *Teratology* **24**, 65–78.

Brown, N. A. and Fabro, S. (1983). The value of animal teratogenicity testing for predicting human risk, *Clin. Obstet. Gynecol.* **26** (2), 467–477.

Brown, N. A., Spielmann, H., Bechter, R., *et al.* (1995). Screening chemicals for reproductive toxicity: the current alternatives, *Alt. Lab. Anim.* **23**, 868–882.

Buesen, R., Seiler, A., Visan, A., Slawik, B., Genschow, E. and Spielmann, H. (2003). Improving the embryonic stem cell test (EST) by establishing molecular endpoints of tissue-specific development, *In Vitro Cell. Dev. Biol.* **39**, Abstr. *2003 Congress on In Vitro Biology*, Portland, Oregon, USA.

Buresh, J. J. and Urban, T. J. (1970). The teratogenic effect of the steroid nucleus in the rat, *J. Dent. Res.* **43**, 548–554.

Cahen, R. L. (1966). Experimental and clinical chemoteratogenesis, *Adv. Pharmacol.* **4**, 263–349.

Czeizel, A. (1987). Lack of evidence of teratogenicity of benzodiazepine drugs in Hungary, *Reprod. Toxicol.* **1**, 183–188.

Daston, G. P., Rogers, J. M., Versteeg, D. J., Sabourin, T. D., Baines, D. and Marsh, S. S. (1991). Interspecies comparisons of A/D ratios: A/D ratios are not constant across species, *Fundamental Appl. Toxicol.* **17**, 696–722.

Datta, N. and Singh, G. (1999). Potentiation of teratogenicity by acute hypoxia in animal model and its hypothetical clinical implications, *J. Anatomical Soc. India* **48** (2), 83–89.

Dawson, D. A., Fort, D. J., Newell, D. L. and Bantle, J. A. (1989). Developmental toxicity testing with FETAX:evaluation of five compounds, *Drug Chem. Toxicol.* **12**, 67–75.

DeSesso, J. M. and Goeringer, G. C. (1992). Methotrexate-induced developmental toxicity in rabbits is ameliorated by 1-(*p*−tosyl)-3,4,4-trimethylimidazolidine, a functional analogue for tetrahydrofolate-mediated one-carbon transfer, *Teratology* **45**, 271–283.

Einarson, T. R., Leeder, J. S. and Koren, G. (1988). A method for meta-analysis of epidemiological studies, *Drug Intell. Clin. Pharm.* **22**, 813–824.

Fainstat, T. (1954). Cortisone-induced congenital cleft palate in rabbits, *Endocrinology* **55**, 502–508.

Fantel, A. G., Greenaway, J. C., Juchau, M. R. and Shepard, T. H. (1979). Teratogenic bioactivation of cyclophosphamide *in vitro*, *Life Sci.* **25** (1), 67–72.

Finnell, R. H. (1999). Teratology: General considerations and principles, *J. Allergy Clin. Immunol.* **103** (2, Pt. 2), S337–S342 (review).

Finnell, R. H., Waes, J. G., Eudy, J. D. and Rosenquist, T. H. (2002). Molecular basis of environmentally induced birth defects, *Annu. Rev. Pharmacol. Toxicol.* **42**, 181–208.

Flint, O. P. and Orton, T. C. (1983). An *in vitro* assay for teratogens with cultures of rat embryonic midbrain and limb cells, *Toxicol. Appl. Pharmacol.* **76**, 383–395.

Fraser, F. C. and Sajoo, A. (1995). Teratogenic potential of corticosteroids in humans, *Teratology* **51**, 45–46.

Gebhardt, D. O. E. (1972). The use of the chick embryo in applied teratology, in: *Advances in Teratology*, Wollam, D. H. M. (Ed.), Vol. 5, pp. 97–111. Academic Press, London.

Genetic Drift Teratogen Update (1995). Vol. 12. (http://www.mostgene.org). Mountain States Genetics Network, Albuquerque NM.

Genschow, E., Scholz, G., Brown, N. A., Piersma, A. H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S., Becker, K. and Spielmann, H. (1999). Development of prediction models for three *in vitro* embryotoxicity tests which are evaluated in an ECVAM validation study, *ALTEX* **16** (2), 73–83.

Genschow, E., Scholz, G., Brown, N. A., Piersma, A. H., Brady, M., Clemann, M., Huuskonen, H., Paillard, F., Bremer, S. and Spielmann, H. (2000). Development of prediction models for three *in vitro* embryotoxicity tests in an ECVAM validation study, *In vitro Molecular Toxicol.* **13**, 51–65.

Genschow, E., Spielmann, H., Scholz, G., Pohl, I., Seiler, A., Clemann, N., Bremer, S. and Becker, K. (2004). Validation of the embryonic stem cell test in the international ECVAM validation study on three *in vitro* embryotoxicity tests, *Alt. Lab. Anim.* **32** (3), 209–244.

Goldstein, D. J., Corbin, L. A. and Sundell, K. L. (1997). Effects of first trimester fluoxetine exposure on the newborn, *Obstet. Gynecol.* **89**, 713–718.

Greek R. (undated) The history of medical progress. In defense of animal reports, at http://www.idausa.org

Gregg, N. McA. (1941). Congenital cataract following German measles in the mother, *Trans. Ophthalmol. Soc. Aust.* **3**, 35–46.

Guiney, P. D., Lech, J. J. and Peterson, R. E. (1980). Distribution and elimination of polychlorinated orpphenyl during early life stages of rainbow trout (*Salmo girdneri*), *Toxicol. Appl. Pharmacol.* **53**, 521–529.

Hale, F. (1933). Pigs born without eyeballs, *J. Hered* **24**, 105–110.

Harris, M. J. and Juriloff, D. M. (1999). Mini-review: Toward understanding mechanisms of genetic neural tube defects in mice, *Teratology* **60**, 292–305.

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Roi, A. J., Prieto, P., Sabbioni, E., Worth, A. and Zuang, V. (2003). ECVAM's response to the changing political environment for alternatives: consequences of the European union chemicals and cosmetics policies, *Alt. Lab. Anim.* **31**, 473–481.

Heilla, A. M., Erkkola, R. U. and Nummi, S. E. (1994). Use of medication during pregnancy — a cohort study on use and policy of prescribing, *Ann. Chirurg. Gynaecol.* **83** (Suppl. 208), 80–83.

Heinonen, O. P., Slone, D. and Shapiro, S. (1977). *Birth Defects and Drugs in Pregnancy*, pp. 336–337. John Wright-PSG, Littleton, MA, USA.

Herbst, A. L. and Scully, R. E. (1970). Adenocarcinoma of the vagina in adolescence: a report of 7 cases including 6 clear-cell carcinomas, *Cancer* **25**, 745–757.

Hogan, M. D. and Hoel, D. G. (1982). Extrapolation to man, in: *Methods in Toxicology*, Hayes, A. W. (Ed.), p. 711. Raven Press, New York.

Igata, A. (1993). Epidemiological and clinical features of Minamata disease, *Environ. Res.* **63** (1), 157–169.

Jelinek, R. (1977). The chick embryotoxicity screening test (CHEST), in: *Methods in Prenatal Toxicology*, Neubert, D., Merker, H. J. and Kwasigroch, T. E. (Eds), pp. 381–386. Thieme, Stuttgart.

Johnson, E. M. and Gabel, B. E. G. (1983). An artificial embryo for detection of abnormal developmental biology, *Fundamental Appl. Toxicol.* **3**, 243–249.

Juriloff, D. M. and Harris, M. J. (2000). Mouse models for neural tube closure defects, *Human Mol. Genet.* **9** (6), 993–1000.

Karnofsky, D. A. (1965). Mechanisms of actions of certain growth inhibiting drugs, in: *Teratology, Principles and Techniques*, Wilson, J. G. (Ed.), pp. 185–213. University of Chicago Press, Chicago, IL, USA

Kavlock, R. J., Chernoff, N., Gray, L. E., Jr., Gray, J. A. and Whitehouse, D. (1982). Teratogenic effects of benomyl in the Wistar rat and CD-1 mouse, with emphasis on the route of administration, *Toxicol. Appl. Pharmacol.* **62** (1), 44–54.

Khera, K. S. (1976). Teratogenicity studies with methotrexate, aminopterin, and acetylsalicylic acid in domestic cats, *Teratology* **14**, 21–28.

Kitchin, K. T., Schmid, B. P. and Sanyai, M. K. (1981). Teratogenicity of cyclophosphamide in a coupled microsomal activating/embryo culture system, *Biochem. Pharmacol.* **30**, 59–64.

Klasco, R. K. and Heitland, G. (Eds) (2003). REPRORISK® System. MICROMEDEX, Greenwood Village, Colorado (Edition expires 12/2003).

Klein, K. L., Scott, W. J. and Wilson, J. G. (1981). Aspirin-induced teratogenesis: a unique pattern of cell death and subsequent polydactyly in the rat, *J. Exper. Zool.* **216**, 107–112.

Kolossov, E., Fleischmann, B. K., Liu, Q., Bloch, W., Viatchenko-Karpinski, S., Manzke, O., Ji, G. J., Bohlen, H., Adicks, K. and Hescheler, J. (1998). Functional characteristics of ES cell-derived cardiac precursor cells identified by tissue-specific expression of the green fluorescent protein, *J. Cell Biol.* **143**, 2045–2056.

Koren, G., Pastuszak, A. and Ito, S. (1998). Drug therapy: drugs in pregnancy, *New Engl. J. Med.* **338**(16), 1128–1137.

Kotwani, A., Mehta, V. L., Gupta, U., Prabhu, S. and Bapna, J. S. (1995). Methods for teratogenicity testing — existing and future models, *Indian J. Pharmacol.* **27**, 204–213.

Kucera, P., Cano, E., Honegger, P., Schilter, B., Zulstra, A. and Schmid, B. (1993). Validation of whole chick embryo cultures, whole rat embryo cultures and aggregating embryonic brain cell cultures using six pairs of coded compounds, *Toxicology in vitro* **7**, 785–798.

Laboratory Animal Science teaching materials; Case 7 (alternatives). The Royal Veterinary and Agricultural University, Denmark. Division of laboratory animal science and welfare.

Lammer, E. J., Chen, D. T., Hoar, R. M., *et al.* (1985). Retinoic acid embryopathy, *New Engl. J. Med.* **313**, 837–841.

Lenz, W. (1966). Malformations caused by drugs in pregnancy, *Am. J. Dis. Child.* **112**, 99–106.

Lewis, R. J., Sr. (1989). *Sax's Dangerous Properties of Industrial Materials*. 7th edn. John Wiley, New York.

Mayura, K., *et al.* (1991). Evaluation of the developmental toxicity of chlorinated phenols utilizing Hydra attenuata and post-implantation rat embryos in culture, *Toxicol. Appl. Pharmacol.* **108**, 253–266.

McBride, W. G. (1961). Thalidomide and congenital abnormalities, *Lancet* **2**, 1358.

McElhatton, P. (2002) Teratogenic drugs — part 1, *Adv. Drug React. Bull.* **213**, 815–818.

Melancon, M. R., Jr., Saybold, J. and Lech, J. J. (1977). Effect of piperonyl butoxide on disposition of di-2′ethylhexylphthalate by rainbow trout, *Xenobiotica* **7**, 633–640.

Miller, R. P. and Becker, B. A. (1975). Teratogenicity of oral diazepam and diphenylhydantoin in mice, *Toxicol. Appl. Pharmacol.* **32**, 53–61.

Millstein, L. G. (1980). FDA's 'Pregnancy Categories'. *New Engl. J. Med.* **303** (12), 706.

Moffa, A. M. and White, J. A. (1983). The effect of periconceptional supplementation of folic acid on the incidence of open neural tube defects in Golden Hamster embryos, *Teratology* **27**, 64A.

Mummery, C. L., van der Brink, C. E., van der Saag, P. T. and De Laat, S. W. (1984). A short term screening test for teratogens using differentiating neuroblastoma cells *in vitro*, *Teratology* **29**, 271–279.

Murphy, M. L. and Karnofsky, D. A. (1956). Effect of azaserine and other growth-inhibiting agents on foetal development of the rat, *Cancer* **9**, 955–962.

National Toxicology Program (1983). List of agents with defined teratogenicity, *Teratogenesis, Carcinogenesis and Mutagenesis* **3**, 461–480.

NICEATM FETAX (2000). Background Review Document: Executive Summary, 10th March 2000. http://iccvam.niehs.nih.gov/about/overview.htm

Nielsen, E., Thorup, I., Schnipper, A., Hass, U., Meyer, O., Ladefoged, O., Larsen, J.C. and Østergaard, G. (2001). Environmental project number 589: Children and the unborn child; Exposure and susceptibility to chemical substances — an evaluation, Danish Environmental Protection Agency Website.

Ninomiya, H., Kishida, K., Ohno, Y., Tsurumi, K. and Eto, K. (1993) Effects of trypan blue on rat and rabbit embryos cultured *in vitro*, *Toxicology in vitro* **7**, 707–717.

Nishimura, H. and Miyamoto, S. (1969). Teratogenic effects of sodium chloride in mice, *Acta Anat. (Basel)* **74** (1), 121–124.

Nora, J. J., Trasler, D. G. and Fraser, F. C. (1965). Malformations in mice induced by dexamphetamine sulphate, *Lancet* **13** 2(7420), 1021–1022.

Palmer, A. K. (1986). A simpler multigeneration study, in: *International Congress of Pesticide Chemistry* (abstract).

Pastuszak, A. L., Schuler, L., Speck-Martins, C. E., Coelho, K. E., Cordello, S. M., Vargas, F., Brunoni, D., Schwarz, I. V., Larrandaburu, M., Safattle, H., Meloni, V. F. and Koren, G. (1998). Use of misoprostol during pregnancy and Mobius' syndrome in infants, *New Engl. J. Med.* **25**, 1881–1885.

Peters, P. and Schaefer, C. H. (2001). General commentary to drug therapy and drug risks in pregnancy, in: *Drugs During Pregnancy and Lactation*, Schaefer, C. H. (Ed.), pp. 1–13. Elsevier; Amsterdam, The Netherlands.

Piersma, A. H., Genschow, E., Verhoef, A., Spanjersberg, M. Q. I., Brown, N. A., Brady, M., Burns, A., Clemann, N., Seiler, A. and Spielmann, H. (2004). Validation of the post-implantation rat whole-embryo culture test in the international ECVAM validation study on three *in vitro* embryotoxicity tests, *Alt. Lab. Anim.* **32** (3), 275–308.

Pinsky, L. and DiGeorge, A. M. (1965) Cleft palate in the mouse: a teratogenic index of glucocorticoid potency, *Science* **147**, 402–403.

Polifka, J. E. and Friedman, J. M. (1999). Clinical teratology: identifying teratogenic risks in humans, *Clin. Genet.* **56**, 409–420.

Polifka, J. E. and Friedman, J. M. (2002). Medical genetics: 1. Clinical teratology in the age of genomics, *Can. Med. Assoc. J.* **167**, 265–273.

Pratt, D. (1980). *Alternatives to Pain in Experiments on Animals*. Argus Archives, New York.

Pratt, R. M., Grove, R. I. and Willis, W. D. (1982). Pre-screening for environmental teratogens using cultured mesenchymal cells from the human embryonic palate, *Teratogenesis, Carcinogenesis, Mutagenesis* **2**, 313–318.

Rosa, F. W. (1983). Teratogenicity of isotretinoin, *Lancet* **2**, 513.

Rosenberg, L., Mitchell, A. A., Parsells, J. L., Pashayan, H., Luvik, C. and Shapiro, S. (1983). Lack of relation of oral clefts to diazepam use during pregnancy, *New Engl. J. Med.* **309**, 1282–1285.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. Methuen, London.

Sabourin, T. D. and Faulk, R. T. (1987). Comparative evaluation of a short-term test for developmental effects using frog embryos, in: *Developmental Toxicology: Mechanisms and Risk*, Banbury Report, Vol. 26, McLachlan, J. A., Pratt, R. M. and Markert, C. L. (Eds), pp. 203–223. Cold Spring Harbor Laboratory, New York.

Sadler, T. W., Horton, W. E. and Warner, C. W. (1982). Whole embryo culture: a screening technique for teratogens? *Teratogenesis, Carcinogenesis, Mutagenesis* **2**, 243–253.

Saxen, I. (1975). Associations between oral clefts and drugs taken during pregnancy, *Int. J. Epidemiol.* **4**, 37–44.

Schardein, J. L. (1993). *Chemically Induced Birth Defects*, 2nd edn rev. Marcel Dekker, New York.

Schmid, B. P. (1984). Monitoring of organ formation in rat embryos after in vitro exposure to azathioprine, mercaptopurine, methotrexate or cyclosporin A, *Toxicology* **31**, 9–21.

Schmid, B. P. (1985). Teratogenicity testing of new drugs with the postimplantation embryo culture system, in: *Concepts in Toxicology*, Homburger, F. (Ed.), Vol. 3, pp. 46–57. Karger, Basel.

Scholz, G., Genschow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Moldenhauer, F., Southee, J. and Spielmann, H. (1999a). Prevalidation of the embryonic stem cell test (EST) — A new *in vitro* embryotoxicity test, *Toxicology in vitro* **13**, 675–681.

Scholz, G., Pohl, I., Glenschow, E., Klemm, M. and Spielmann, H. (1999b). Embryotoxicity screening using ES cells *in vitro*: correlation to *in vivo* teratogenicity, *Cells Tissues Organs* **165**, 203–211.

Scialli, A. R. (1992). *A Clinical Guide to Reproductive and Developmental Toxicology*. CRC Press, Ann Arbor, Mich.

Shepard, T. H. (1994). *Catalog of Teratogenic Agents*, 7th edn. Johns Hopkins University Press, Baltimore, MD, USA.

Shin, J.-H. and Shiota, K. (1999). Folic acid supplementation of pregnant mice suppresses heat-induced neural tube defects in the offspring, *J. Nutr.* **129**, 2070–2073.

Shiono, P. H. and Mills, J. L. (1984). Oral clefts and diazepam use during pregnancy, *New Engl. J. Med.* **311**, 919–920.

Sisodiam, P. (1972). Teratogenic effects of drugs, *Ind. J. Pharmac.* **4** (2), 51–56.

Skalko, R. G. and Gold, M. P. (1974). Teratogenicity of methotrexate in mice, *Teratology* **9**, 159–164.

Slone, D., Siskind, V., Heinonen, O. P., Monson, R. R., Kaufman, D. W. and Shapiro, S. (1976). Aspirin and congenital malformations, *Lancet* **1**, 1373–1375.

Spielmann, H., Pohl, I., Döring, B., Liebsch, M. and Moldenhauer, F. (1997). 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.

Spielmann, H., Scholz, G., Bremer, S. and Holzhuetter, H. G. (1998). In vitro testing of embryotoxicity using the embryonic stem cell test (EST): an interlaboratory comparison, *In Vitro Cell Dev. Biol.-Animal.* **34** (3 pt. 2), 11A.

Spielmann, H., Genschow, E., Brown, N. A., Piersma, A. H., Verhoef, A., Spanjersberg, M. Q. I., Huuskonen, H., Paillard, F. and Seiler, A. (2004). Validation of the rat limb bud micromass test in the international ECVAM validation study on three *in vitro* embryotoxicity tests, *Alt. Lab. Anim.* **32** (3), 245–274.

Staples, R. E., Kellam, R. G. and Haseman, J. K. (1976). Developmental toxicity in the rat after ingestion or gavage of organophosphate pesticides (Dipterex, Imidan) during pregnancy, *Environ. Health Perspect.* **13**, 133–140.

State of California Environmental Protection Agency (2003). Office of Environmental Health Hazard Assessment; Safe Drinking Water and Toxic Enforcement Act of 1986: Chemicals known to the state to cause cancer or reproductive toxicity, June 13, 2003. Proposition 65 list.

Thiersch, J. B. (1956). Therapeutic abortions with folic acid antagonist 4-amino pteroylglutamic acid (4-amino PGA) administered by oral route, *Am. J. Obstet. Gynecol.* **63**, 1298–1304.

Toms, D. A. (1962). Thalidomide and congenital abnormalities, *Lancet* **2**, 400.

Tuchmann-Duplessis, H., David, G. and Hagel, P. (1972). *Illustrated Human Embryology*, Translated by Hurley, L. S. Masson, Paris.

Turbow, M. M., Clark, W. H. and Dipaolo, J. A. (1971). Embryonic abnormalities in hamsters following intrauterine injection of 6-aminonicotinamide, *Teratology* **4** (4), 427–431.

Umansky, R. (1966). The effect of cell population density on the developmental fate of reaggregating mouse limb bud mesenchyme, *Developmental Biology* **13**, 31–56.

United States Food and Drug Administration. (1980). Caffeine: deletion of GRAS status, proposed declaration that no prior sanction exists, and use on an intern basis pending additional study. Federal Register; 45: 69817.

Walker, B. E. (1971). Induction of cleft palate in rats with anti-inflammatory drugs, *Teratology* **4**, 39–42.

Waltan, B. T. (1983). Use of the cricket embryo (Acheta domesticus) as an invertebrate teratology model, *Fund. Appl. Toxicol.* **3**, 233–236.

Werler, M. M., Mitchell, A. A. and Shapiro, S. (1989). The relation of aspirin use during the first trimester of pregnancy to congenital cardiac defects, *New Engl. J. Med.* **321**, 1639–1642.

Wilby, O., *et al.* (1986). A Hydra assay as a pre-screen for teratogenic potential, *Food Chem. Toxicol.* **24**, 651–652.

Wiles, M. V. (1993). Embryonic stem cell differentiation *in vitro*, *Methods in Enzymol.* **225**, 900–917.

Wilson, J. G. (1972). Abnormalities of intrauterine development in non-human primates, *Acta Endocrinol. Suppl (Copenhagen)* **166**, 261–292.

Wilson, J. G. (1973a). Present status of drugs as teratogens in man, *Teratology* **7** (1), 3–15.

Wilson, J. G. (1973b). Mechanisms of teratogenesis, *Am. J. Anat.* **136** (2), 129–131.

Wilson, J. G. (1975). Reproduction and teratogenesis: current methods and suggested improvements, *J. Assoc. Off. Anal. Chem.* **58** (4), 657–667.

Wilson, J. G. (1977). Current status of teratology. General principles and mechanisms derived from animal studies, in: *Handbook of Teratology*, pp. 1–47. Plenum Press, New York.

Wilson, J. G. (1978). Survey of *in vitro* systems. their potential use in teratogenicity screening, in: *Handbook of Teratology*, Wilson, J. G. and Fraser, F. C. (Eds), pp. 135–154. Plenum Press, New York.

Wilson, J. G., Fradkin, R. and Schumacher, H. J. (1970). Influence of drug pretreatment on the effectiveness of known teratogenic agents, *Teratology* **3**, 210–211 (abstract).

Wilson, J. G., Ritter, E. J., Scott, W. J. and Fradkin, R. (1977). Comparative distribution and embryotoxicity of acetylsalicylic acid in pregnant rats and rhesus monkeys, *Toxicol. Appl. Pharmacol.* **41**, 67–78.

Wilson, J. G., Scott, W. J., Ritter, E. J. and Fradkin, R. (1979). Comparative distribution and embryotoxicity of methotrexate in pregnant rats and rhesus monkeys, *Teratology* **19**, 71–80.

Wilson, J. T. and Fouts, J. R. (1966). The effect of actinomycin D on post-phenobarbital activity of hepatic microsomal drug-metabolising enzymes in the rat, *J. Biol. Chem.* **241**, 4810–4815.

Zamenhof, S. (1985). Differential effects of antifolate on the development of brain parts in chick embryos, *Growth* **49**, 28–33.

Zhao, K., Krafft, N., Terlouw, G. D. C. and Bechter, R. (1993). A model combining the whole embryo culture with human liver S9 fraction for human teratogenic prediction, *Toxicology in vitro* **7**, 827–831.