Food for Thought ... on Animal Tests

Thomas Hartung
EC Joint Research Centre, IHCP/ECVAM, Ispra, Italy

The fourth article in this series addresses the target of all work on alternative methods, the animal test. It is not easy to say something new here, since anti-vivisectionists have collected many arguments over a long time. Surprisingly, there are few relevant reviews retrievable from the scientific literature (MedLine). I hope this does not indicate that the scientific community is less interested to discuss the inherent problems of this fundamental tool in the life sciences. Animal models have advantages and disadvantages compared to other approaches. In few instances they represent the only reasonable approach. Increasingly modern methods allow reducing, refining and replacing (3R principle) animal experiments, as required wherever possible by European legislation. This article summarises typical limitations of animal models. However, this does not mean that each and every limitation holds true for all animal models or that their alternatives bear fewer limitations, for further explanation see (Hartung, 2007b).

A detailed assessment of the models that are in use is necessary to draw conclusions and make decisions in an evidence-based manner. Only the thorough assessment of the performance characteristics of any tool used in research allows interpreting results and properly complementing the method employed. There is growing public concern regarding the welfare of animals in general (e.g. Eurobarometer on "Social values, Science and Technology" published in June 2005, where more than four in five EU citizens stated that the Community has a duty to protect animal rights regardless of the cost). The EU acknowledged the importance of the protection and welfare of animals used for scientific purposes on an international level, and in 1999 became party to the Council of Europe Convention on the protection of vertebrate species used for experimental and other scientific purposes, ETS 123 (Council Decision 1999/575/EC).

Hypothesis 1: In the absence of a common philosophy on killing, we have to live with case-by-case decisions, ambiguity and dissent

There are questions of life and death also in science; life and death of patients and experimental animals for example. The border between what is allowed or at least questionable, endangering or taking life, depends on purpose and circumstances. Table 1 tries to capture this; it illustrates that we accept killing the same beings as pests or health hazards that we protect in case of research (with minor geographical differences). Such protection tends to expand (e.g. in the ongoing revision of European laboratory animal welfare legislation which will include embryonic stages in the last third of gestation, some molluscs and crustaceans). This can be seen to parallel the development and spread of humanity.

It is evident that nature is full of killing. It is evident that we might kill small animals literally with every step we take. It is evident that we consume animals, that we protect ourselves and our food, that we sacrifice animals to find new cures for disease. Thus, there is no "no-kill" option, and we need to define a threshold. In a continuum of evolutionary complexity (from amoebocytes to humans) thresholds defining what species to protect to which extent need to be set. The most sensible approach seems to be to limit ourselves to what is unavoidable and justifiable – certainly making any decision utilitarian and dependent on the circumstances. Thus, we can promote societal and technical developments that make animal tests avoidable and less justifiable. This is the spirit of the 3R approach, a continuous call to replace, reduce and refine further. A larger ethical debate is beyond the scope of this article. However, the animals' ability to suffer forms the broadly accepted basis (Jeremy Bentham, An Introduction to the Principles

Tab. 1: The license to kill depends on the purpose

	Insects	Fish / birds	Rodents	Other mammals	Non-human primates	Great apes	Humans
Plant protection products/ biocides							
Food & products							
Research				EU			Forbidden
				US			
Self defence							

White = generally accepted; grey = not applicable; black = restricted in EU and US, respectively.
The figure tries to capture that for different purposes the killing of animals and even humans is generally accepted – major prohibitions exist only in the field of research, with notable differences between especially EU and US legislation.

of Morals & Legislation, 1789) for the recognition of animals as sentient beings and so represents a solid point of origin. The European Science Foundation has phrased it thus recently, "Laboratory animals not only have an instrumental value, but also an intrinsic value in themselves, which must be respected. Animals must always be treated as sentient creatures." (European Science Foundation, Policy Briefing, September 2000)

Hypothesis 2: No model is perfect

The term "model" implies a deviation from reality, usually simplifying, reducing variables. The only perfect experiment in human life sciences would be on a large and diverse human group under natural exposure conditions – certainly not often possible in a prospective setup. The deviations from this theoretically ideal case are unavoidable if any progress is to be made, but we tend to forget the compromises we have made to end up with something feasible. The limitations of *in vitro* models have been summarized recently (Hartung, 2007b). Limitations also exist for animal models of human health effects, e.g. in toxicology, from which key examples are drawn below. The situation is not different in drug discovery. The famous NIH pharmacologist Bernard Brodie coined it already in 1964: "…it is often a matter of pure luck that animal experiments lead to clinically useful drugs".

Species-differences – rarely can the same experiment be performed both on humans and on experimental animals; where this is possible (e.g. skin irritation (York et al., 1996; Basketter et al., 2004)) a correlation of about 60% is found. However, different laboratory animal species can be subjected to the same experiment more easily. Even among rodent species, typical correlations of only 70% are found (for some references see Hoffmann and Hartung, 2006), while notably most laboratory animal species are much more similar to each other than to humans. There is no reason to assume that any species would predict health effects on humans better than on any other species. Already bioavailability, i.e. how much of an orally given substance is taken up and can possibly exert an effect, differs enormously between species as impressively shown by Grass and Sinko (2002, Fig. 1, page 437). Even monkeys can differ considerably, as experienced tragically in 2006 when the TeGenero anti-CD28 antibody, after testing safe at 500-times higher concentrations in monkeys, led to multiple organ failure within hours in six human volunteers (Bhogal and Combes, 2006).

Pharmaco- and toxicokinetics of substances differ between animals and humans – humans are not 70 kg rats: there are tremendous differences to all laboratory animal species, e.g. in metabolic rates, food and water intake, pH in the gut, body temperature, metabolism of xenobiotics with regard to capacity, velocity and metabolites formed, body fat distribution, elimination routes, carrier proteins / protein binding, defence mechanisms, life span. Some browsing for example in Michael Derelanko's "Toxicologist's Pocket Handbook" is eye-opening. In regulatory toxicology an assessment factor of 100 is typically used, i.e. it is assumed that humans are up to 100-fold more sensitive than the laboratory animal to account for inter-species and inter-individual differences. However, it has recently been shown that even a factor of 100 is not always sufficient, and that especially sensitive individuals are not covered by this (Wood, 1998; Falk-Filipsson et al., 2007; Bokkers and Slob, 2007). It appears to be more appropriate to speak of uncertainty factors (IGHRC, 2003), because that is what they are. However, this addresses only the quantitative error of the assessment, since it is well argued that the hazard matters and the quantitative assessment follows more complex considerations. Very little is known, however, with regard to the qualitative mistake, i.e. assigning a toxic effect or not. In general, the inter-individual differences have received much more attention in pharmacology (for review, see for example Inaba et al., 1995; Lin, 2007) than toxicology (Lovell, 1993; Lipscomb, 2002), although there is no reason to assume that there is less impact.

In-bred strains – the broad use of in-bred strains has helped to reduce the variability of experimental results. However, working with identical twins implies that the variability of the (animal) population is no longer reflected (see Festing, 1993; Lovell, 1993; Kacew and Festing, 1996).

Young animals – typically, for cost and logistic reasons, young animals are used for experiments; however most health effects are found in the elderly, who may also have other underlying diseases. Recently it has been proposed that one should go even further and consider the impact of the epigenome, comprised of chromatin and a covalent modification of DNA by methylation, as the interface between the dynamic environment and the inherited static genome (Szyf, 2007).

Restriction to one sex – this is not meant as a call for default testing in both sexes, but again the spectrum of sensitivities is reduced and there should be at least some reasoning on the choice of sex. The contribution of sex differences is better established for drug effects (for review see Gandhi et al., 2004) than for toxic effects, but there is no reason to assume a lesser impact.

Group sizes – usually the smallest possible group size is selected for animal tests, often without performing a power analysis which would indicate whether a significant result can be expected at all. The problem is aggravated because often maximum information is sought by studying multiple endpoints or time points without adjusting the statistics for multiple testing, e.g. standard test guidelines for repeat dose toxicity include up to 40 evaluated endpoints. the easiest, but conservative Bonferroni correction for example would request to lower the significance level from typical p=0.05 to 0.00125 (i.e. 0.05 divided by 40) for each endpoint in this case. Assuming that an effect is just significant at p=0.05 with the typical group size of 10 of the repeated dose test (OECD test guideline 407), we would need 22 animals per

group to make it significant at 0.00125 in the unpaired two-tailed t-test following this example. There is good reason why clinical trials with thousands of enrolled patients have only one principal endpoint.

Cost – the maintenance of adequate animal numbers and the employment of qualified staff is extremely costly. This can amount to tens to a hundred thousand € for example for higher apes. In consequence, the number of replicates and repetitions are typically limited. But also standard animal tests can be quite costly: A survey of testing costs in Europe (Fleischer, 2007) showed average costs of e.g. 1,200 € for skin irritation, 1,350 € for eye irritation, 50,000 € for 28d-repeated-dose oral studies (twice this for the inhalation route), 330,000 € for a two-generation reproductive toxicity study or 780,000 € for a rat carcinogenicity bioassay.

Unrealistic dosages and exposure durations – high dose treatments (in toxicology often at the maximum tolerated dose, i.e. up to 10% of the animals dying) are applied, which hardly reflect human exposure (Sumner and Stevens, 1994; Mehendale, 1995). While there is not very much evidence for additional effects occurring at lower doses (for summary see Kamrin, 2007), many additional irrelevant effects occur when extreme tissue concentrations are reached and defence systems are overwhelmed.

Many health effects are multi-factorial – but for example the disease models used are mostly based on one (artificial) cause only. This might be more obvious for human disease models, where almost always the respective disease has genetic, psychological, social, dietary and/or environmental risk factors. Human intoxication also has many cofactors, including health state, genetic predisposition, and behaviour to coexposure.

Animals are stressed – there are many causes of stress in the animals: transport, insufficient time for them to accustom, cages without enrichment (e.g. opportunities for species-specific behaviour) (Wolfer et al., 2004), handling, procedures during daytime on night-active species, changes in groups for randomisation, too many male animals caged together, to name some prominent ones.

Adaptive responses of animals are underestimated – this is especially relevant for the fashionable knock-out mouse experiments; we often forget that these animals have had a lifetime to compensate for their genetic defect, but we simplify our interpretation as if only the one gene were lacking. Also, any exposure to xenobiotics or trauma can induce mechanisms of defence, inflammation or tissue repair, influencing the response to subsequent (toxic) challenge.

Proprietary data – especially in the regulatory context, an enormous scientific limitation is the fact that data on animal tests are not shared; this is often done to protect commercial interests (transparency toward competitors, barriers to make use of the data for their developments), but sometimes it is just the hassle of and lack of incentive for publication.

Acute models of chronic phenomena – there are enormous pressures to produce results quickly, leading us to develop animal models, even of chronic health effects, which establish symptoms quickly.

Post-hoc definition of the experimental hypothesis – especially in basic research, it is quite common to change a hypothesis according to the results obtained to "tell a story" in a publication; few are aware that this impacts on the statistics to be used (*a posteriori* testing).

Tests are either sensitive or specific – these characteristics exclude each other; thus, we can either rely on positive or on negative results of a test; however, typically we accept both outcomes of animal tests with the same confidence (Hoffmann and Hartung, 2005).

Prevalences of toxicological effects are neglected when interpreting animal test results – even an almost perfect animal test would not help us as a stand alone. The reason is known in diagnostic medicine as "prevalence". In simple words, if an event is unlikely, the frequency of false-positive results of a test becomes more decisive for the reliability – if I have few real positives among the positive test results, I cannot rely on the positive outcome of the test. We are typically looking for rare properties of substances, both in agent discovery (a hit in a hundred or thousand) and toxicology (2-10% of the substances). The proportion of either toxic or effective substances in a panel of tested substances is thus normally in the low percentage range, and even tests with good specificity will produce some percent false-positives, i.e. in a similar or higher proportion than the real positives. We have alluded to this previously (Hoffmann and Hartung, 2005). The typical combination of several tests, for example in different laboratory animal species, to increase sensitivity, e.g. in order not to miss any toxic compound, further reduces the specificity and thus increases the number of false positives.

Hypothesis 3: Lethality in acute toxicity might sometimes be just a side-effect

Mortality of animals is often the result of lack of food and water only, not the primary effect of the substance. If small rodents are not capable of feeding, they die within hours – most probably many substances would not be toxic if a simple sugar solution were injected. Similarly, the loss of body heat most probably contributes to the lethality of many treatments. Body temperature is affected by a lack of movement (resulting in >80% heat production), unfavourable surface/body weight ratio of small animals, and hypothermia instead of a fever reaction in rodents. Most probably, it is often the translocation of bacteria from the gut which kills the animal, especially if orally administered substances damage the GIT barriers. I would like to make use of a model calculation (based on Ernst Rietschel, Borstel, personal communication), which at first sight seems to have little to do with toxicology: the gut of one human being contains 1 kg of Gram-negative bacteria, with

about 50 g of endotoxin. When injected, this amount of endotoxin would suffice to kill one million humans or induce fever in one billion. Rats and mice are 100-fold less sensitive to endotoxin, and I do not know the mass and composition of their faeces, but certainly there is enough to kill a lot of them "endogenously". This does not even account for the translocation of live bacteria and ensuing infection, a major contributing factor to mortality in similarly traumatized and intoxicated human patients (for review see Deitch et al., 1996; Stechmiller et al., 1997; Gatt et al., 2007).

I would like to turn an argument around to substantiate the hypothesis that, especially in acute oral intoxications, the animals do not die from the toxin but from secondary effects to the GIT: It has been documented in at least three major attempts, that cytotoxicity correlates pretty well with acute oral toxicity (see Halle register, MEIC study and the more recent ICCVAM/NICEATM/ECVAM validation study). Actually, this makes little sense if we assume that the substances are taken up, distributed and metabolised with complex kinetics and can affect more than 400 different tissues with various sensitivities. Might it be that the animal experiment simply measures cytotoxicity to the GIT epithelium, which results in translocation of bacteria? Ironically, this would mean that we can pretty well predict this animal test *in vitro*, because the animal test measures a phenomenon (cytotoxicity to the intestine) that is irrelevant for humans (we would vomit – which rodents cannot do – or remove the intoxication before it reaches the intestine, supply intensive care treatment, etc.). Instead of our 9 million € effort of A-Cute-Tox (http://www.acutetox.org/), a well designed series of animal experiments (blasphemy from the head of ECVAM…) might demonstrate that the reference method is meaningless. Efforts by the U.K. centre NC3R to abolish the requirement for acute toxicity studies for pharmaceuticals (Chapman and Robinson, 2007) – because they are useless for pharmaceuticals – deserve our full support and should be extended to chemicals, plant protection products and cosmetics soon.

Hypothesis 4: Can animal tests predict chronic, systemic toxicity for humans at all?

Certainly, we get predictions from animal tests, but are they sufficiently good? A prediction is acceptable if its uncertainty is acceptable. However, in order to establish uncertainty, we would need to know what is true. How should this be possible with the tremendous differences between human intoxication and the experimental situation described above? Besides, the major limitation for extrapolation to humans is typically the exposure part: Humans are exposed to thousands of substances at the same time and only few in a controlled/controllable and continuous manner – how should we single out individual substances that cause a problem? The case of tobacco with its limited number of health effects has shown how difficult it is to trace back a toxic effect to its cause. How should this be possible in less evident, less traceable cases? We will always only get what we know ("wykiwyg = what you know is what you get"), i.e. at best we will be able to confirm a positive finding in an animal test. Given the high number of false-positive results of precautionary testing and the enormous effort of epidemiological studies, it is unlikely that many causal relationships can be established. So, it is no risk for a risk assessor to claim a substance may represent a human health hazard – if there are no additional supportive data, nobody will take the effort to verify. Given unclear exposure scenarios, nobody will be able to verify.

There is one way to estimate the reliability of toxicological assessments: compare experimental animal species. This will not tell you the truth (human hazard) but give an idea about the reliability of results in general terms. It will be biased, since different rodent species that are typically compared are more closely related to each other than to humans, but this is at least an indication. With the notable exception of the cancer bioassay, few such comparisons are available. This is especially true for repeated-dose toxicity. The database established by the German Fraunhofer Institute in Hannover (Bitsch et al., 2006) represents an exemption and a valuable resource for such comparisons.

If we expand the concept of uncertainty factors (Tab. 2), we can start to think about the rank order of our uncertainty: The figures might be challenged (or better substituted by more correct estimates), but the rough calculation would suggest that 50% of our (qualitative) hazard assignments are wrong and that quantitatively we can err by more than 1250fold with regard to the potency of a toxicant. Does this represent an acceptable uncertainty?

Tab. 2: The toxicological uncertainty chain

	Animal test reproducibility	Species extrapolation	Inter-individual differences	Resulting hazard in men	Exposure uncertainty	Resulting risk men
Quantitative = dose response (uncertainty factor)	**1.25** C.V. typically 20-30%	**10**	**>10**	**>125**	**>10** (personal estimate)	**>1250**
Qualitative = Hazard or not (% correctly identified)	**90%**	**60-70%** (sens. 90%, spec. 10-20%)	**90%** (personal estimate; more quantitative)	**53%**		

Hypothesis 5: Economic factors impact on animal use and its substitution

Animal testing is a business: No official, aggregated data are available about many key data of the sector, but the study performed by Prognos (http://ec.europa.eu/environment/chemicals/lab_animals/ia_en.htm) on behalf of the European Commission in the context of the revision of the Directive for the protection of laboratory animals has estimated a turnover of about 3 billion € per year for animal testing in the EU; about 1,330 establishments (industry, contract research laboratories and universities) in the EU perform animal tests. The number of breeders and suppliers of animals for experimental purposes, however, is most likely in the range of several dozen only. About 9,300 new projects involving animal tests are started annually in the EU. An average project was estimated to cost about 300,000 Euros over three years. This is by itself an enormous market with business interests, lobbyists, etc. It is an industry with the natural mechanisms of resistance under pressure to change, but this extends also to the people working in the field: their skills, their basis for empowerment and reputation, are under pressure to change. Somebody running an animal test laboratory will tend to find some way to continue using it. An expert in designing, carrying out or interpreting animal tests will tend to apply these tools again. In the field of safety, it is especially difficult to object: The safety considerations of the toxicologist are difficult to challenge – who would take the liability of objecting to a safety concern? Decimating substances (by precautionary test design and interpretation) only increases this psychological power base.

However, safety assessments are also in the interest of companies: First of all, certainly there is a strong interest to put safe products on the market, but secondly this also serves to formally reduce liability ("everything possible done to assure safety" – sarcastically, the blessing given after the ritualistic sacrifice of animals, or the 3rd millennium variant of the medieval letter of indulgence). A third factor might be that the increased development costs of raising safety assessment requirements favour large over small enterprises; a similar process occurred in the pharmaceutical area, though here the clinical trials drove costs to a situation where finally only a dozen companies can still afford to develop new products up to market launch. Given the much smaller profit margins of chemical industry, similar effects might be caused by the animal testing requirements for chemicals, which can amount to several million Euros for high-production volume chemicals.

Another economical factor is represented by the non-tariff trade barriers created by different safety testing requirements in different economic areas. Fortunately, the European unification, OECD and WTO have tackled this problem, but global markets will lead to the continued use of the traditional methods until the last important market has converted to the alternatives (Bottini et al., 2007).

Hypothesis 6: It is belief not science that drives our expectations as to the validity of many animal tests

Let's be provocative: A field that is based on belief resembles religion more than science. A field that is satisfied with its state of the art is at risk of becoming system of belief. As Richard Dawkins phrased it, "one of the truly bad effects of religion is that it teaches us that it is a virtue to be satisfied with not understanding" (Dawkins, 2006, p.126). The reliance on traditional approaches and the resistance to move to an up-to-date mechanistic and technological approach in toxicology bears characteristics of a religious approach. It is tempting to adapt (replacing "religion") the saying by Oscar Wilde, "Truth, in matter of (toxicology), is simply the opinion that has survived".

Face-validity versus validation – we tend to believe in the relevance of animal tests if some of their features resemble our expectations; in contrast, validation means to systematically challenge the relevance of a model. An interesting example might be the gastric ulcer (Lee, 2000). As long as we believed that overproduction of acids was the key pathomechanism, we used animal models reflecting this and resulting in the expected ulcer. When *Helicobacter pylori* was found to be the leading cause of ulcers, new models were established. This means, that the face-validity, i.e. the obvious reflection of our understanding of pathogenesis, was driving the choice and belief in the respective animal model.

Standardisation versus tradition – while standardisation is a prerequisite for mutual acceptance of data, especially in the regulatory context (Bottini et al., 2007), and also for the accumulation of sufficient knowledge of the model itself, it needs to be distinguished from a "we have always done it like this" approach (tradition). This notion points to the dilemma that fixing a procedure in time bears the risk of missing innovation, while lack of standardisation impairs mutual acceptance. This means that models should be modified if there is strong reason for this, but this needs to be documented and spread at the same time to create a new standard. As much as standardisation is the prerequisite of validation, it is the confidence created by validation that forms the basis for change of standard procedures.

Most animal test protocols are not validated – the systematic assessment of performance characteristics (reproducibility, limit of detection, relevance of results, mechanistic basis) of animal models is typically lacking. Full stop. While this is now largely accepted for the introduction of new animal methods for which validation is requested (bravo!), this is not at all accepted for traditional approaches. There is apparently a lot of fear of reopening these cases and questioning earlier assessments. We might lose our "state of innocence", in which we can state that everything was done to the best of our knowledge.

Right target species does not equate with relevance - In ecology, ecotoxicology or veterinary medicine, the claim of relevance is often made, since testing is carried out on the target species. Another "intellectual shortcut" or convenient belief: with about 400,000 fish species, there is most probably no lesser problem in modelling each species' sensitivities and specificities than with mammals. Most limitations accounted for above hold true also when testing in the target species.

Organisms are incredibly complex – but we use them to answer simple questions; any interpretation is an oversimplification of what has really happened in the experiment, usually overwhelming our capabilities to interpret. Thus, we tend to follow "wykiwyg" (what you know is what you get, see above, but perhaps we should say, what you believe is what you get), i.e. to interpret the experimental results according to our current understanding of pathophysiology.

Prospects

So, we live in an imperfect world, working with imperfect tools. What are the prospects of still producing relevant science, protecting the consumer and identifying new agents with high probability? First, we need to understand the limitations of the tools, which requires, where possible, optimisation, standardisation and validation (Hartung, 2007a), which can lead us to an evidence-based approach (Hoffmann and Hartung, 2006). Bertrand Russel's famous quote, how to explain to God why he did not believe "Not enough evidence, God, not enough evidence" might be adapted to question some areas of toxicological tradition. Second, we must not rely on one tool only: the combination of tools with complementary strengths, such as a sensitive and a specific test, can overcome weaknesses. Integrated testing, especially using different technologies, is most promising. However, the tools to compose testing strategies systematically as well as to validate them are only emerging. Third, stop believing and start doubting. There is a tendency to believe *a priori* in the relevance of animal models ("*in vivo veritas*"). This is further augmented by the desire to come to a definitive conclusion (following economic or publication pressure). Believing is most probably the most non-scientific approach.

The prospects? 3R! However, most probably the low-hanging fruits to be harvested by current *in vitro* and amended *in vivo* techniques are already largely exploited. What *in silico* technologies, which are coming closer and closer to validation, can add will need to be explored (some thoughts on this in one of the next "Food for thought…" articles). However, a really new quality might emerge from combining approaches using (high-throughput and high-content) *in vitro* models with bioinformatics and mechanistic information in systems biology approaches. This avenue has been recently illustrated by an important vision document on behalf of the US National Academy of Sciences (Committee on Toxicity and Assessment of Environmental Agents, 2007). But, again, vision is a rather religious approach…

References

Basketter, D. A., York, M., McFadden, J. P. and Robinson, M. K. (2004). Determination of skin irritation potential in the human 4-h patch test. *Contact Dermatitis 51,* 1-4.

Bhogal, N. and Combes, R. (2006). TGN1412: time to change the paradigm for the testing of new pharmaceuticals. *ATLA 34*, 225-239.

Bitsch, A., Jacobia, S., Melbera, C. et al. (2006). REPDOSE: A database on repeated dose toxicity studies of commercial chemicals – a multifunctional tool. *Reg. Tox. Pharmacol. 46*, 202-210.

Bokkers, B. G. and Slob, W. (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. *Crit. Rev. Toxicol. 37*, 355-373.

Bottini, A. M., Amcoff, P. and Hartung, T. (2007). Food for thought … on globalization of alternative methods. *ALTEX 24*, 255-261.

Brodie, B. (1964). Of mice, microsomes and man. *Pharmacologist 6*, 12-26.

Chapman, K. and Robinson, S. (2007). Challenging the regulatory requirement for acute toxicity studies in the development of new medicines – a workshop report. http://www.nc3rs.org.uk/downloaddoc.asp?id=559&page=22&skin=0

Committee on Toxicity and Assessment of Environmental Agents, National Research Council. (2007). Toxicity Testing in the Twenty-first Century: A Vision and a Strategy, http://www.nap.edu/catalog/11970.html

Dawkins, R. (2006). *The god delusion*. London: Bantam Press.

Deitch, E. A., Rutan, R. and Waymack, J. P. (1996). Trauma, shock, and gut translocation. *New Horiz. 4,* 289-99.

Falk-Filipsson, A., Hanberg, A., Victorin, K. et al. (2007). Assessment factors--applications in health risk assessment of chemicals. *Environ. Res. 104*, 108-127.

Festing, M. F. (1993). Genetic variation in outbred rats and mice and its implications for toxicological screening. *J. Exp. Anim. Sci. 35*, 210-220.

Fleischer, M. (2007). Testing cost and testing capacity according to REACH requirements – results of a survey of independent and corporate GLP laboratories in the EU and Switzerland. *J. Business Chem. 4*, 96-114.

Gandhi, M., Aweeka, F., Greenblatt, R. M. and Blaschke, T. F. (2004). Sex differences in pharmacokinetics and pharmacodynamics. *Annu. Rev. Pharmacol. Toxicol. 44*, 499-523.

Gatt, M., Reddy, B. S. and MacFie, J. (2007). Review article: bacterial translocation in the critically ill - evidence and methods of prevention. Aliment. *Pharmacol. Ther. 25*, 741-757.

Grass, G. M. and Sinko, P. J. (2002). Physiology-based pharmacokinetic simulation modelling. *Adv. Drug Delivery Rev. 43*, 433-451.

Hartung, T. (2007a). Food for thought … on validation. *ALTEX 24(2)*, 67-72.

Hartung, T. (2007b). Food for thought … on cell culture. *ALTEX 24(3)*, 143-147.

Hoffmann, S. and Hartung, T. (2005). Diagnosis: Toxic! – Trying to apply

approaches of clinical diagnostics and prevalence in toxicology considerations. *Tox. Sci. 85*, 422-428.

Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Human Exp. Toxicol. 25*, 497-513.

IGHRC (The Interdepartmental Group of Health Risks from Chemicals UK). (2003). Uncertainty factor: their use in human health risk assessment by UK government. MRC Institute for Environment and Health, 1-68, http://www.le.ac.uk/ieh/

Inaba, T., Nebert, D. W., Burchell, B. et al. (1995). Pharmacogenetics in clinical pharmacology and toxicology. *Can. J. Physiol. Pharmacol. 73*, 331-338.

Kacew, S. and Festing, M. F. (1996). Role of rat strain in the differential sensitivity to pharmaceutical agents and naturally occurring substances. *J. Toxicol. Environ. Health 47*, 1-30.

Kamrin, M. A. (2007). The "low dose" hypothesis: validity and implications for human risk. *Int. J. Toxicol. 26*, 13-23.

Lee, A. (2000). Animal models of gastroduodenal ulcer disease. *Baillieres Best Pract. Res. Clin. Gastroenterol. 14*, 75-96.

Lin, J. H. (2007). Pharmacokinetic and pharmacodynamic variability: a daunting challenge in drug therapy. *Curr. Drug Metab. 8(2)*, 109-136

Lipscomb, J. C. and Kedderis, G. L. (2002). Incorporating human interindividual biotransformation variance in health risk assessment. *Sci. Total Environ. 288*, 13-21.

Lovell, D. P. (1993). Impact of pharmacogenetics on toxicological studies. Statistical implications. *J. Exp. Anim. Sci. 35*, 259-281.

Mehendale, H. M. (1995). Toxicodynamics of low level toxicant interactions of biological significance: inhibition of tissue repair. *Toxicology 105*, 251-266.

Stechmiller, J. K., Treloar, D. and Allen, N. (1997). Gut dysfunction in critically ill patients: a review of the literature. *Am. J. Crit. Care 6*, 204-209.

Sumner, D. D. and Stevens, J. T. (1994). Pharmacokinetic factors influencing risk assessment: saturation of biochemical processes and cofactor depletion. *Environ. Health Perspect. 102 Suppl 11*, 13-22.

Szyf, M. (2007). The dynamic epigenome and its implications in toxicology. *Toxicol. Sci. 100*, 7-23.

Wolfer, D. P., Litvin, O., Morf, S. et al. (2004). Laboratory animal welfare: Cage enrichment and mouse behaviour. *Nature 432*, 821-822.

Wood, A. J. (1998). Ethnic differences in drug disposition and response. T*her. Drug Monit. 20*, 525-526.

York, M., Griffiths, H. A., Whittle, E. and Basketter, D. A. (1996). Evaluation of a human patch test for the identification and classification of skin irritation potential. *Contact Dermatitis 34*, 204-212.

Acknowledgement

The continuous discussions with colleagues and friends, especially the ECVAM staff, which shape these ideas is gratefully appreciated. Limitations of animal models have recently been summarized also for the proceedings of the 2007 LIMPE seminar Experimental Models in Parkinson's disease, Alghero, Sardinia, Italy, September 23 - 25, 2007 (Hartung T. Thoughts on limitations of animal models. Parkinsonism & Related Disorders 2008, in press).

Correspondence to

Prof. Dr. Dr. Thomas Hartung
EU Joint Research Centre
Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)
21020 Ispra
Italy
e-mail: thomas.hartung@ec.europa.eu