# From Alternative Methods to a New Regulatory Toxicology

*Thomas Hartung*

Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Department of Environmental Health Sciences, Doerenkamp-Zbinden Chair for Evidence-based Toxicology, Baltimore, MD, USA

**Summary**

*The 3Rs concept to replace, reduce, and refine animal experiments celebrated its 50th anniversary in 2009. Meanwhile, a mechanistic toxicology has evolved that effectively relies to a large extent on methodologies that substitute or complement traditional animal tests. The biotechnology and informatics revolution of the last decades has made such technologies broadly available and useful.*

*Regulatory toxicology has only slowly begun to embrace these new approaches. Major validation efforts, however, have delivered the evidence that new approaches do not lower safety standards and can be integrated into regulatory safety assessments.*

*Political pressures such as the REACH legislation and the 7th amendment to the cosmetics legislation further prompt the need for new approaches, especially in the EU. In the US, the NAS vision report for a toxicology in the 21st century and its most recent adaptation by the EPA have initiated a debate regarding their toxicity testing strategy and how to create a novel approach based on human cell cultures, lower species, high-throughput testing, and modeling. The human toxome, a systematic mapping of the entirety of pathways of toxicity, is now underway.*

*The lessons learned from the development, validation, and acceptance of alternative methods for the creation of a new approach to regulatory toxicology are reviewed herein. Besides the technical development of new approaches, a case is made that we need both conceptual steering and an objective assessment of current practices by evidence-based toxicology. The application of an approach modeled on Evidence-based Medicine (EBM) has been suggested, as for the last two decades EBM has demonstrated that rigorous systematic reviews of current practices and meta-analyses of studies offer powerful tools to provide health care professionals and patients with the current best scientific evidence for diagnostic and treatment options. Similarly, a portal for high-quality reviews of toxicological approaches and tools for the quantitative meta-analyses of data promise to serve as a door opener for the new regulatory toxicology. The Evidence-based Toxicology Collaboration was created in the US in 2011, and a European equivalent in 2012.*

*Keywords: evidence-based toxicology, regulatory toxicology, human toxome*

## 1 Introduction

In the last 150 years, chemists have synthesized about 70 million substances. More than 100,000 of these are found in consumer products of daily use, in drugs, in cosmetics, in detergents, in our food, our clothes and – last but not least – as contaminants of our natural environment. When Bayer brought Aspirin™ – our eldest synthetic drug – to the market 111 years ago, no regulatory safety assessments on animals were stipulated by legislation. The producer was held liable for any problems with his products. This did not always turn out as positive as in case of aspirin; only a few days after aspirin, the same chemist, Felix Hoffmann (1868-1946), synthesized a sedative for coughs: heroin… (Fig.1). Not every product is as harmless as aspirin. It is easy to understand that, with each safety scandal, the desire for safety assessments grows. In the 1920s, scientists started to use mice and rats broadly for laboratory research. Until then, it was considered absurd that these animals could mirror humans. It was clearly convincing, however, just how fast experiments could be performed with them: the animals did not cost a lot, they reproduced quickly, and a large number of them could be kept in a small cage. This created a real research rush, similar to today's introduction of stem cells.

With every scandal the toolbox of toxicology grew as chemists sought to prevent a similar occurrence. In the early 1930s in the US LashLure created a scandal: The cosmetic product was used to dye lashes permanently; unfortunately, the anilin dye it contained sometimes led to strong inflammation. More than 3,000 reports of collateral effects were collected: five women were blinded, and one woman died. This prompted the first regulation of cosmetics, which have since been controlled in the US by the FDA (Food and Drug Administration). Their employee, John H. Draize (1900-1992) in 1944 developed the Draize rabbit eye test, where a chemical is applied into the eye of a rabbit (Hartung et al., 2010). Today, many perceive this procedure as cruel, but in fact, for 65 years this test prevented the recurrence of a case like LashLure. In this manner, toxicology grew with every scandal, pieced together like a patchwork quilt.

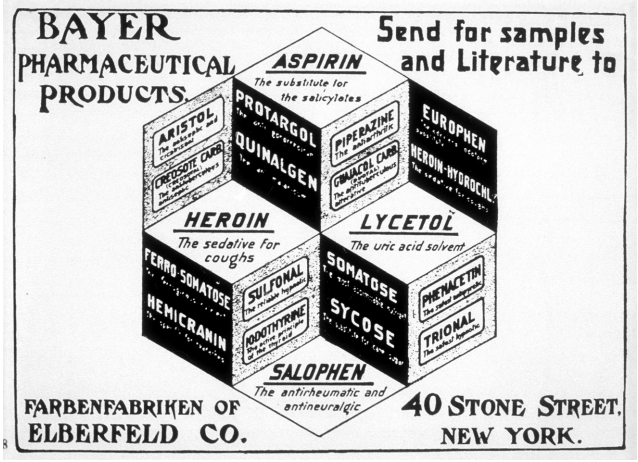


**Fig. 1: Advertisement for aspirin and heroin at the beginning of the 20th century**
(Archive of the author)


**Fig. 2: Title page from December 1962 of the German weekly journal "Der Spiegel" on the Contergan/thalidomide scandal**
(Archive of the author)

The thalidomide (Contergan™ ) scandal of the 1960s (Fig. 2) led toxicologists to test for malformation of embryos, for example.

## 2 The birth of doubt in animal experiments

Concerns about animal experimentation and the killing of animals have a long history. Even the ancient Greeks discussed whether we should kill animals. In Germany in the 1920s there were 700 animal welfare associations. However, it wasn't until 1959 that Bill Russell and Rex Burch in England developed what they called "the principles of humane experimental technique" (Russell and Burch, 1959). They referred to these principles as "the 3Rs," i.e., Replacement, Reduction, and Refinement. One must substitute animals with non-sentient test systems (Replacement), when an alternative exists. One must reduce the number of animals used wherever possible, if the same result can be obtained with fewer animals (Reduction). One must avoid unnecessary suffering and distress by using, for example, analgesics or working under narcosis (Refinement). Any suggestion of Replacement was considered utopian 50 years ago. At that time, cell culture and computer programs were in their infancy, and few scientists could imagine that such methods might lead to success. Over the last few decades, however, industry, science, and politics have demonstrated a commitment to the 3Rs that has led to compromise with those who would prefer to see animal experiments end today rather than tomorrow. This represented the basis for a credible investment in overcoming animal testing. In fact, animal experiments decreased until the turn of the century by an estimated two-thirds since their peak in the mid-Seventies. Since then, however, numbers have been increasing again, due largely to the new techniques for manipulating individual genes in mice, which have become very popular scientific models.

In the meantime, we have a number of examples that 3Rs approaches have indeed come to fruition. For example, the $LD_{50}$ test has been used since the 1920s. This test determines the lethal dose of a chemical that kills 50% of treated rats. Until 1989, 150 animals per substance were used for this purpose (10 female and 10 male at 7 dosages each, plus one untreated control group of 10 animals). This resulted in an enormous number of animals being used, especially since almost any substance going to the market was tested. Apparently, the lethality of this test led to labeling with the famous skull and crossbones as an indicator of poison. Both the protection of workers and safety measures for the transport of substances also were determined on this basis. In 1989, after an analysis of test data, a revision of guidance took place on the OECD (Organization for Economic

Co-operation and Development) level. The OECD, now 34 industrialized countries, achieved agreement to drastically reduce animal numbers. Since then, groups of 5 animals of one gender have been used, thus reducing the number of rats from 150 to 45 per substance. In the 1990s, a further step was taken. The idea was simple: Why should all animals be treated simultaneously? When starting with just one dose, a higher dose can be tested next if animals survive. If the animals die, the dose has to be lowered. At the same time, it was shown that groups of three rats suffice. Consequently, three methods were accepted internationally in 2001. On average, these tests use only 8 to 12 animals. From 150 to 45 to just 8 to 12 animals – an enormous reduction indeed. In addition, one of these methods introduced the notion that the animal does not have to await death but rather can be euthanized humanely when it shows signs it will not survive or will be severely damaged. This is an example of "Refinement" – the second R for the amelioration of pain and distress in animal experiments. Another classical example is testing for skin allergy. Traditionally, this has been done with guinea pigs. The Local Lymph Node Assay (LLNA) represents both a reduction and a refinement alternative in mice, as it uses fewer animals, involves a shorter treatment period, and ends the experiment at the stage of lymph node swelling instead of waiting for the skin lesions to occur.

## 3 We can do it differently – the replacement of animal experiments

Increasingly, animal tests in toxicology can be fully substituted – the third R for "Replacement." As an example, human skin obtained from surgical procedures can be grown further in the laboratory. A small tissue sample can produce several square meters of skin. These technologies were developed originally for skin transplantation after burn injury, for example. Quickly, the idea arose that this tissue also could be used for testing chemicals. In fact, it was possible to demonstrate that artificial human skin is as suitable as rabbit skin to test skin corrosion or irritation by chemicals. The respective international test guidelines have been agreed upon. This was not only a milestone for the cosmetic industry (Hartung, 2008b), but also a proof-of-principle that international consensus can be achieved regarding the replacement of an animal test with an animal-free method (Hartung and Daston, 2009).

## 4 Validation of alternative methods – animal welfare must not impair safety

The prerequisite for the acceptance of 3Rs approaches, however, is that these approaches must not lower safety standards for consumers. For this reason, the concept of formal validation was introduced. In 1991, the European Center for the Validation of Alternative Methods (ECVAM) was created for this purpose in Ispra, Italy. About 50 alternative methods have been validated there so far, and a number are currently undergoing ring trials and peer-reviews. An American (1995), a Japanese (2005), and a Korean (2011) and Brazilian (2011) equivalent followed, and the creation of similar centers currently are being discussed in India and China.

When validating an alternative method today, several things have to be shown (Bottini and Hartung, 2009): (1) Has the method been clearly defined – especially, is it clear when to use it and when not? (2) Does the method have a scientific basis that reflects our understanding of pathophysiology in humans and in animals? (3) Is the method reproducible, i.e., do we get the same results when repeating the method in other laboratories? (4) Are the results of the method relevant, i.e., in general: Can the method predict the outcome of the traditional (animal) test (Hartung et al., 2004)?

The last aspect is certainly the most critical (Hartung, 2007a). Most animal tests themselves have never been tested as to their relevance. Data from poison centers or clinical trials are only rarely available to compare with humans. However, we can carry out the same animal experiment with different species and ask, for example, how well do rats predict mice or hamsters predict guinea pigs? Obviously, there is no reason why any of these species should predict humans better than they predict each other. Many rodent species clearly are closer to each other than they are to humans. Even non-human primates have a certain evolutionary distance. The result is worrisome – the correlation between laboratory animal species usually only ranges between 60-70% (Hartung and Daston, 2009). What can we do? Traditionally, two paths are followed: One tests in two animal species, or one renders the tests precautionary, e.g., by testing extremely high dosages. Not "more is better" but "more kills better."

## 5 How reliable are animal experiments?

Animal tests have made the world safer, but they also have created quite a few problems. We sort out more and more substances because of possible problems. The example of Aspirin is most interesting (Hartung, 2009a): Today aspirin would fail almost all safety tests. Aspirin kills half of the rats ($LD_{50}$) at doses we use as maximal allowed daily dose in humans. Today we typically request safety margins of a factor of 100, which means, in general, that we use doses that are at least 100 times smaller than those that harmed animals. Aspirin is an irritant to eye, skin, and lung. Aspirin has had ambiguous results in genotoxicity assays and, while not actually carcinogenic in the respective animal test, it augmented the carcinogenic effect of other substances when co-applied. Furthermore, aspirin led to embryonic malformations in practically every species tested (rats, mice, rabbits, cats, dogs, and monkeys). Note that these are all tests as they are used today for drugs, pesticides, and industrial chemicals. We know a lot about aspirin – 23,000 scientific publications are available, and a trillion (one thousand billion) tablets have been swallowed. None of the animal findings are really relevant for humans. But this shows that it would be impossible to bring aspirin to the market today. These, too,

are the costs of our desire for safety. This attitude is killing people as well – a new drug that is not allowed to go into the clinics because of alerts in precautionary animal tests is a drug that cannot cure patients.

At the same time, testing new substances in animals cannot prevent all dangers. When, after the respective animal tests, drugs are tested on volunteers and patients, 10-30% show toxic effects that will not allow developing them further (Kola and Landis, 2004). We simply are not 70 kg rats... There remains uncertainty on both sides – the false positive and false negative results. Animals represent only a model of humans, and all models are wrong, though some are still useful (Hartung, 2008a). It is most important that we are clear that we are using models that reflect only part of reality. Cell cultures (Hartung, 2007b) and computer models (Hartung and Hoffmann, 2009) have their own limitations. It is of utmost importance that we start analyzing the strengths and weaknesses of all our tools.

## 6 "Toxic Ignorance" – the REACH project and toxicity testing in the 21st century

Another problem is that testing in animals is far too expensive and laborious: To determine whether a substance is carcinogenic, for example, takes four years and costs about one million dollars. It is no surprise, then, that in the last 30 years in Europe only 14 of 5,000 new industrial chemicals were tested for their carcinogenic potential; out of more than 100,000 chemicals on the market, these represent only about 3,000. This has been termed "toxic ignorance" (Roe et al., 1997). The European REACH legislation aims to tackle this problem (Hartung, 2010a), but with traditional animal tests we will not achieve the throughput necessary (Hartung and Rovida, 2009; Rovida and Hartung, 2009). We simply do not have enough laboratories to test that many substances within a reasonable time frame. For this reason, REACH asks for new methods, but the implementation of the regulation is already foreseen for the next decade. This leaves little room to develop and validate new approaches.

In addition to the ethical criticisms of animal tests, we must increasingly add a practical one: We cannot assess the safety of new substances coming to the market with sufficient certainty and speed (Hartung, 2009b). The renowned US National Academy of Sciences suggested in 2007 (http://www.nap.edu/catalog.php?record_id=11970) that toxicity testing in the 21st century has to move away from animal testing and establish a new safety testing paradigm. This has created an enormous atmosphere of departure. Currently, discussions are taking place at many venues regarding how to implement this (Collins et al., 2008; Hartung, 2009c; Firestone et al., 2010). Experts discuss a "Human Toxicology Project" (Seidle and Stephens, 2009), similar to the human genome project. We will see whether this can be financed. It promises to move the safety testing of products onto a new level, at least, but a lot of steering will be necessary (Hartung, 2009b). Most remarkably, the EPA already has made this their novel toxicity testing paradigm (Firestone et al., 2010). So we see very different approaches in the US and Europe: While the EU much earlier took up the challenge of old chemicals and only later aimed to reduce animal testing for animal welfare reasons, the US systematically developed a new approach based on new technologies although a testing program did not come about until now (Hartung, 2010b).

## 7 The technologies of the 21st century for the toxicology of the 21st century

What are the prospects and what are the new technologies? It has been claimed that knowledge in the life sciences doubles every seven years. In this case, we now have about 1,000 times more knowledge than was available at the time when most animal tests were devised. The revolutions in biotechnology and informatics we have seen occur not only on the stock market. Today, we have cell cultures for practically all tissues and organs of the human body. We know many pathways, we know how cells work, and we know how synthetic substances disturb these. Precise analytical methods, robotized testing, and complex measurements now allow enormous quantities of information to be obtained, and modern computers enable the analysis. The buzzword "systems toxicology" (Hartung et al., 2012) was coined to describe the systematic combination of existing knowledge via computer models with large datasets – from gene chips, for example – which can include all of the roughly 30,000 human genes (Hartung and Leist, 2008). This determines which genes in contact with a given poison are switched on or off. Similar poisons lead to similar responses ("signatures"). This also can be studied on the level of proteins produced or the changes in metabolite concentrations. Increasingly, we can deduce from this the pathways of toxicity that caused these molecular changes. The mapping of the entire pathways of toxicity, the human toxome, has been proposed (Hartung and McBride, 2011). Automated image analysis frequently plays a role, too. Still, a safety assessment that relies only on such methods and uses no animals remains a utopian vision. Twenty years ago, however, this held true for the mobile telephones we now take for granted, as well as today's internet, which also was only emerging. In the developing laboratories we can already find the new toxicology techniques – they only have to be optimized to find their market.

And a market is there for sure: Each year industry spends about $ 3 billion on safety assessments worldwide (Bottini and Hartung, 2009). The European REACH program for old chemicals alone, which has just started will produce data costing $ 13 billion over the next ten years, and this is only the beginning: Nanoparticles, genetically modified food, cell therapies... new products lead to new challenges to control their risks (Hartung, 2010c; Hartung and Koeter, 2008). The job market for toxicologists is huge, and so it is good that some universities have again started to invest in their education.

Whether such a novel approach will improve how predictive toxicology is has yet to be seen. Everything starts, however, with no longer pretending how safe things are when they

enter the market. The pressure to develop a novel toxicology results only when the need for new technologies is clear. The use of animals is, in the end, a technology, the systematic use of a problem-solving approach. Blind faith in the meaningfulness of results from animal tests makes them an animal sacrifice for the invocation of a bright future for our products. A realistic judgment of their strengths and weaknesses, on the contrary, allows them to be used in a targeted manner to provide consumers with safe products at a more acceptable expense to animals. It also helps producers understand their safety gaps and toxicologists to develop new approaches allowing animal use to decline automatically. Because patients and consumers are of primary concern, the animals need not be secondary.

## References


Bottini, A. A. and Hartung, T. (2009). Food for thought … on economics of animal testing. *ALTEX 26*, 3-16.

Collins, F. S., Gray G. M., and Bucher, J. R. (2008). Toxicology: Transforming environmental health protection. *Science 319*, 906-907.

Firestone, M., Kavlock, R., Zenick, H., et. al. (2010). The U.S. environmental protection agency strategic plan for evaluating the toxicity of chemicals. *J Toxicol Environ Health B Crit Rev 13*, 139-162.

Hartung, T., Bremer, S., Casati, S., et al. (2004). A modular approach to the ECVAM principles on test validity. *Altern Lab Anim 32*, 467-472.

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

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

Hartung, T. (2008a). Food for thought … on animal tests. *ALTEX 25*, 3-9.

Hartung, T. (2008b). Food for thought … on alternative methods for cosmetics safety testing. *ALTEX 25*, 147-162.

Hartung, T. and Leist, M. (2008). Food for thought … on the evolution of toxicology and phasing out of animal testing. *ALTEX 25*, 91-96.

Hartung, T. and Koeter, H. (2008). Food for thought ... On food safety testing. *ALTEX 25*, 259-264.

Hartung, T. (2009a). Per aspirin ad astra... *Altern Lab Anim 2*, 45-47.

Hartung, T. (2009b). A toxicology for the 21st century: Mapping the road ahead. *Toxicol Sci 109*, 18-23.

Hartung, T. (2009c). Toxicology for the twenty-first century. *Nature 460*, 208-212.

Hartung, T., van Vliet, E., Jaworska, J., et al. (2012). Food for though ... systems toxicology. *ALTEX 29* , 119-128.

Hartung, T. and Daston, G. (2009). Are in vitro tests suitable for regulatory use? *Tox Sci 111*, 233-237.

Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature 460*, 1080-1081.

Hartung, T. and Hoffmann, S. (2009). Food for thought … on in silico methods in toxicology. *ALTEX 26*, 155-166.

Hartung, T. (2010a). Food for thought … on alternative methods for chemical safety testing. *ALTEX 27*, 3-14.

Hartung, T. (2010b). Lessons learned from alternative methods and their validation for a new toxicology in the 21st century. *J Toxicol Environ Health B Crit Rev 13*, 277-290.

Hartung, T. (2010c). Food for thought ... on alternative methods for nanoparticle safety testing. *ALTEX 27*, 87-95.

Hartung, T., Bruner, L., Curren, R., et al. (2010). First alternative method validated by a retrospective weight-of-evidence approach to replace the Draize eye test for the identification of non-irritant substances for a defined applicability domain. *ALTEX 27*, 43-51.

Kola, I. and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature Rev Drug Discov 3*, 711-715.

National Research Council – NRC (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington DC, USA: The National Academies Press. http://www.nap.edu/catalog.php?record_id=11970

Roe, D., Peace, W., Florini, K., et al. (1997) *Toxic Ignorance*. New York: Environmental Defense Fund.

Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements. *ALTEX 26*, 187-208.

Russel, W. M. S. and Burch, R. L. (1959). *The principles of humane experimental technique*. London, UK: Methuen. http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc

Seidle, T. and Stephens, M. L. (2009). Bringing toxicology into the 21st century: A global call to action. *Toxicol In Vitro 23*, 1576-1579.


## Correspondence to


Thomas Hartung, MD PhD
Center for Alternatives to Animal Testing
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
W7032, Baltimore, MD 21205, USA
e-mail: thartung@jhsph.edu