



Comment

REPRINT: Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice¹

Summary

Modern toxicology has embraced in vitro methods, and major hopes are based on the omics technologies and systems biology approaches they bring along (Hartung and McBride, 2011; Hartung et al., 2012). A culture of stringent validation has been developed for such approaches (Leist et al., 2010, 2012a,b), while the quality and usefulness of animal experiments have been little scrutinized. A new study (Seok et al., 2013) now shows the low predictivity of animal responses in the field of inflammation. These findings corroborate earlier findings from comparisons in the fields of neurodegeneration, stroke and sepsis. The low predictivity of animal experiments in research areas allowing direct comparisons of mouse versus human data puts strong doubt on the usefulness of animal data as key technology to predict human safety.

Regulatory toxicology is involved with the prediction of human risk and with regulatory approaches to limit such 'assumed/predicted' risks to humans. This is a very particular form of science, in that it deals mostly not with facts (concerning human hazard), but with assumptions and predictions derived from models. For most compounds, the human hazard is (fortunately!) not known. The art of predictive toxicology lies in its construction of an intricate web of cross-relationships, to anchor the assumed human hazard to sets of real data. Animal experiments are to date the most important source of such data. A key question of the discipline is: how can we get information on the appropriateness of these data as anchor point for the toxicological 'spider net' of cross-references and extrapolations? 'The proof of the pudding is the eating', that is, human data are needed to control the validity of the prediction network.

The evaluation of whether animal experiments provide a solid starting point for the prediction of human hazard can follow two major lines. The first collects evidence from cases of human poisoning. At least for some compounds, this allows a direct comparison of effects on animals and on man. Prominent examples of case studies that suggest poor predictivity are the experience with thalidomide, or with the TG1412 drug candidate, which caused terrible effects in man that had not been predicted from the available animal data (Stebbins et al., 2007). Lack of correlation is also seen the other way around, that is, when rodent data predict cancer for compounds that are safe in man

(Gold et al., 2005; Basketter et al., 2012). For some compound classes, there are also positive examples of animal data quantitatively predicting toxicity. However, in many areas of toxicology (for instance in the field of pesticides) such comparative data are hardly available. Moreover, this inductive approach (using individual case studies) does not allow conclusive general statements on the usefulness of animal experiments. Therefore, as a second line, deductive strategies to approach the question have been devised. Such approaches require answers to two types of question. For instance: (a) is there at least one field in which high-quality comparative data can be obtained? (b) can one show, or reasonably assume, that the predictivity of animals for man does not differ fundamentally in different fields of biomedical research? If answers to these questions can be obtained, a third step would be the combination of the answers for deduction of a generalized conclusion.

We will deal here with the second question only briefly. The answer from screening the scientific literature must be clearly 'yes'. Tens of thousands of publications, all peer-reviewed, often in high-impact journals, are based on the assumption that animals are predictive of man in all the different research areas of animal use. Comparative claims that one area is particularly well or particularly badly predicted cannot be substantiated by the available scientific literature. Huge amounts of public money are spent on the assumption that animals are useful for all biomedical areas. No granting agency has ever declared a par-

¹ Reprinted with permission from Archives of Toxicology: Leist, M. and Hartung, T. (2013). *Arch Toxicol* 87, 563-567.



ticular field of medical research to be pointless for animal-based research. Animals are applied uniformly as model in all areas of pharmacology, toxicology and general research in disease biology. This use is endorsed by committees of scientific experts, by ethical review boards, by the funding agencies and by political decision makers that channel the huge sums for funding of research and development into the different areas. The increasing use of animals for research in the last years has been accelerated by the widespread generation of transgenic mice. The increase in animal experimentation in most biomedical areas has overcompensated all successful efforts to substitute animals in some research fields (Hartung and Leist, 2008; Blaauboer et al., 2012; Leist et al., 2012a; Hasiwa et al., 2011). The hard evidence for the belief in the usefulness of animal experimentation across fields (in terms of hundreds of millions of dollars and euros invested on basis of this assumption) is overwhelming.

This also relates to the field of toxicology, which cannot be separated from other biomedical research areas, as far as biological mechanisms and their correlation in man and animals are concerned (Leist et al., 2008a; Hartung, 2009). Toxicology has profited a lot from findings and methods of other fields, and it is generally assumed that biochemical and physiological regulations, as well as their pathological counterparts discovered by different medical disciplines, do also apply to the field of safety sciences (Leist et al., 2008b; Rossini and Hartung, 2012). We can thus safely assume that the predictivity of animal models is judged to be equally high in pharmacology and toxicology, and the following part will concentrate on where to find good comparative data of animals versus man.

An answer has been provided by a recent noteworthy study of Seok et al. (2013) from the 'large scale collaborative research program on inflammation and the host response to injury'. They chose inflammation as a field of medical research, in which human data are available and in which the mouse models seem to have a very good mechanistic resemblance to the human disease situation. The biological response to injury was analyzed on a molecular level, by looking into the regulation of about 5000 human genes relevant to inflammation and by comparing them to the murine counterpart responses. The result was surprising, almost shocking: the correlation was not only poor, it was virtually absent for the main study areas: burns, trauma, endotoxemia. When the study was expanded to other areas, such as sepsis and infection, poor correlations of human and mouse data were confirmed. Thus, responses in mice cannot predict human responses; at least in these fields. Based on the above considerations (question (b)), there is no reason to believe that the correlation would be better in any other field.

It might be argued, that this is only one study, and only one very particular and small field. In this context, it is important to look at the reasons, why these experiments were performed. The paper by Seok is not a stand-alone study, but it was triggered by worrying findings of 20 years of research, which suggested that non-predictive animal models might be the reason for the many clinical failures of new drugs in the field of sepsis. Sepsis is a systemic inflammatory response and still one of the leading causes of death on intensive stations worldwide. For this reason,

enormous resources have been devoted to basic research into its mechanisms and to the discovery of drugs. Countless papers appeared in top-impact journals already in the 90s, but translation of any animal finding into clinics failed. Opal and Cross (1999) summarized already then 'It has become painfully evident that animal models provide misleading and overly optimistic estimates of the survival benefit of specific antisepsis drugs when compared to clinical efficacy in actual human sepsis'. This situation did not become better with more time for trials and optimization of animal studies (Buras et al., 2005). When the only treatment discovered by this approach, activated C-reactive protein, had to be withdrawn from the market in 2011, more than 100 additional clinical trials had been performed, and it became evident that every single approach that had been successful in animals had failed (Rittirsch et al., 2007; Christaki et al., 2011). Nevertheless, animal-based studies in this field still continue to be financed. This somehow rings memories of how prince Hamlet's behavior was described by Polonius: 'Though this be madness, yet there is method in 't'.

Mice continue to be used as models, as their failure in the past has been claimed to be not due to a general inaptness of animal models, but rather to the poor quality, standardization and adaptation to clinical questions of such studies. It is in fact true that there is strong evidence for deficits in the quality and reporting of animal studies (Hartung, 2008; Macleod and van der Worp, 2010; Kilkenny et al., 2010; van der Worp et al., 2010; van der Worp and Macleod, 2011). On this basis, one may ask whether the translational value (Hackam and Redelmeier, 2006; Rice, 2012) is high enough to justify further use.

Possibly, the poor correlation, and its connection with the quality of animal experimentation are particular features of research on inflammation and infection. To examine this, it is worth taking a look at an entirely different research field: ischemic stroke. It shares one important feature with inflammation research: the animal models are thought to be conceptually very close to the human situation. In human ischemic stroke, the blood circulation is occluded and exactly the same is modeled in animals. In endotoxemia, infection or burn injury, the stimuli in humans and mice are exactly the same. This is a favorable situation, compared to the fields of age-related neurodegeneration, cardiovascular disease, type II diabetes, asthma or cancer, which require generation of quite artificial animal models. Back to stroke: how well do the animal models work? They work similar as in inflammation: not at all. Apart from thrombolysis, every single neuroprotective treatment for cerebral ischemia that has worked successfully in animals (over 500 have been reported (van der Worp et al., 2010)) has failed in man. This negative statement is based on large numbers of trials, comprising dozens of large studies and hundreds of smaller ones (De Keyser et al., 1999; Gladstone et al., 2002; O'Collins et al., 2006; Savitz and Fisher, 2007). Also, in this field, the failure has been attributed to the poor research standards, and quality criteria have been defined to amend this situation. Failure of drugs, despite adherence to such criteria, then triggered the design of new criteria (Dirnagl and Fisher, 2012; Savitz and Fisher, 2007). Also, in this respect, stroke research resembles inflammation research. The alterna-

tive conclusion, that animal studies are inherently not suited to predict the human situation, is considered more rarely (Musch et al., 2006; Matthews, 2008).

Before a quick generalization of the conclusions, it is certainly prudent to take a closer look at adjacent research areas. A field related to inflammation and infection is the research that deals with ‘countermeasures to biological and chemical terrorism and warfare’. This example is highlighted here, as the department of defence of the US sponsored a National Academy of Science of the USA report on ‘animal models for assessing countermeasures to bioterrorism agents’, published in December 2011 (NRC, 2011). The usefulness of animal models was evaluated by renowned scientists, and the conclusion of the report was that animal models would not be useful. Instead, a recommendation was issued that human cell-based 3D *in vitro* systems should be developed. This decision was taken so serious that altogether 200 Mio \$ have been made available since for research in this field (Hartung and Zurlo, 2012). Scientific areas related to the one of ischemic stroke are neurodegenerative diseases such as Alzheimer’s or Parkinson’s disease. This field has seen vast investments of the private and public sector. Dozens of neuroprotective drugs and principles have been discovered in animals, none of them ever worked in man.

Nevertheless, the belief that mouse research can result in information about human disease and its treatment is still held firmly by most scientific funding agencies. Findings on species differences tend to be neglected, and the skewing of the available data by a publication bias towards positive animal findings has only recently been unraveled (Sena, 2010). Just to name few examples, it has been clear before the Seok study that TLR4 signaling, a pivotal process in the inflammatory response, is different in man and mouse (Schmidt et al., 2010), and it is generally known that many inflammatory mediators take very different roles in different species. Even fundamental regulations ranging from neural control of airways (Schleppütz et al., 2012) to the biology of stem cells (Schnerch et al., 2010) are very different between species. All this evidence suggests that animals are not particularly good predictors of humans, in the areas where we have comparative data on different species. Is toxicology an exception? At least some comparative data are available from drugs that have been evaluated first in animals, then in man. The largest comparative study in this area (Olson et al., 2000) finds a poor (i.e., 43%) predictivity of rodents for man. It is stated explicitly that this is not necessarily due to different metabolism, but possibly due to a different biology. Some examples for such molecular differences in toxicodynamics are well-known. For instance, man is about 1000-fold more sensitive to inhibition of the Na/K ATPase by the cardiac glycoside ouabain than mice (Kent et al., 1987), and the difference in sensitivity to bacterial endotoxin may even be in the million-fold range (Seok et al., 2013; Hasiwa et al., 2013). Thus, there are many individual examples suggesting that humans are not simply 70-kg mice, neither in pharmacology, nor in toxicology. The recent study of Seok et al. (2013) has corroborated this notion, based on a broad systematic approach. The statements of this paper have been endorsed by renowned scientists that have

themselves relied on animal studies in the past. Their statement, on the failure of mice to predict for man in an important area of pharmacology, should be taken seriously – and also serve as food for thought in toxicology.

References

- Basketter, D. A., Clewell, H., Kimber, I., et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing - t⁴ report. *ALTEX* 29, 3-91.
- Blaauboer, B. J., Boekelheide, K., Clewell, H. J., et al. (2012). The use of biomarkers of toxicity for integrating *in vitro* hazard estimates into risk assessment for humans. *ALTEX* 29, 411-25.
- Buras, J. A., Holzmann, B., and Sitkovsky, M. (2005). Animal models of sepsis: setting the stage. *Nat Rev Drug Discov* 4, 854-65.
- Christaki, E., Anyfanti, P., and Opal, S. M. (2011). Immunomodulatory therapy for sepsis: an update. *Expert Rev Anti Infect Ther* 9, 1013-33.
- De Keyser, J., Sulter, G., and Luiten, P. G. (1999). Clinical trials with neuroprotective drugs in acute ischaemic stroke: are we doing the right thing? *Trends Neurosci* 22, 535-40.
- Dirnagl, U. and Fisher, M. (2012) REPRINT: International, multicenter randomized preclinical trials in translational stroke research: it is time to act. *Stroke* 43, 1453-4.
- Gladstone, D. J., Black, S. E., Hakim, A. M., and Heart and Stroke Foundation of Ontario Centre of Excellence in Stroke Recovery (2002). Toward wisdom from failure: lessons from neuroprotective stroke trials and new therapeutic directions. *Stroke* 33, 2123-36.
- Gold, L. S., Manley, N. B., Slone, T. H., et al. (2005). Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature through 1997 and by the National Toxicology Program in 1997-1998. *Toxicol Sci* 85, 747-808.
- Hackam, D. G. and Redelmeier, D. A. (2006). Translation of research evidence from animals to humans. *JAMA* 296, 1731-2.
- Hartung, T. and Leist, M. (2008). Food for thought ... on the evolution of toxicology and the phasing out of animal testing. *ALTEX* 25, 91-102.
- Hartung, T. and McBride, M. (2011) Food for thought ... on mapping the human toxome. *ALTEX* 28, 83-93.
- Hartung, T., van Vliet, E., Jaworska, J., et al. (2012). Systems toxicology. *ALTEX* 29, 119-28.
- Hartung, T. and Zurlo, J. (2012). Alternative approaches for medical countermeasures to biological and chemical terrorism and warfare. *ALTEX* 29, 251-60.
- Hartung, T. (2009). Toxicology for the twenty-first century. *Nature* 460, 208-212.
- Hartung, T. (2008). Food for thought ... on animal tests. *ALTEX* 25, 3-9.
- Hasiwa, N., Daneshian, M., Bruegger, P., et al. (2013). Evidence for the detection of non-endotoxin pyrogens (NEPs) by the whole blood monocyte activation test. *ALTEX* 30, 169-208.
- Hasiwa, N., Bailey, J., Clausing, P., et al. (2011). Critical evalu-



- ation of the use of dogs in biomedical research and testing in Europe. *ALTEX* 28, 326-40.
- Kent, R. B., Emanuel, J. R., Ben Neriah, Y., et al. (1987). Ouabain resistance conferred by expression of the cDNA for a murine Na⁺, K⁺-ATPase alpha subunit. *Science* 237, 901-3.
- Kilkenny, C., Browne, W. J., Cuthill, I. C., et al. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* 29, e1000412.
- Leist, M., Hartung, T., and Nicotera, P. (2008a). The dawning of a new age of toxicology. *ALTEX* 25, 103-14.
- Leist, M., Kadereit, S., and Schildknecht, S. (2008b). Food for thought... on the real success of 3R approaches. *ALTEX* 25, 17-32.
- Leist, M., Efremova, L., and Karreman, C. (2010). Food for thought ... considerations and guidelines for basic test method descriptions in toxicology. *ALTEX* 27, 309-17.
- Leist, M., Lidbury, B. A., Yang, C., et al. (2012a) Novel technologies and an overall strategy to allow hazard assessment and risk prediction of chemicals, cosmetics, and drugs with animal-free methods. *ALTEX* 29, 373-88.
- Leist, M., Hasiwa, N., Daneshian, M., and Hartung, T. (2012b). Validation and quality control of replacement alternatives – current status and future challenges. *Toxicol Res* 1, 8-22.
- Macleod, M. and van der Worp, H. B. (2010). Animal models of neurological disease: are there any babies in the bathwater? *Pract Neurol* 10, 312-4.
- Matthews, R. A. (2008). Medical progress depends on animal models - doesn't it? *J R Soc Med* 101, 95-98.
- Musch, T. I., Carroll, R.G., Just, A., et al. (2006). A broader view of animal research. *Br Med J* 334, 274.
- NRC – National Research Council, Committee on Animal Models for Assessing Countermeasures to Bioterrorism Agents (2011). *Animal Models for Assessing Countermeasures to Bioterrorism Agents*. Washington, DC, USA: The National Academies Press. <http://dels.nationalacademies.org/Report/Animal-Models-Assessing-Countermeasures/13233>
- O'Collins, V. E., Macleod, M. R., Donnan, G. A., et al. (2006). 1,026 experimental treatments in acute stroke. *Ann Neurol* 59, 467-77.
- Olson, H., Betton, G., Robinson, D., et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56-67.
- Opal, S. M. and Cross, A. S. (1999). Clinical trials for severe sepsis. Past failures, and future hopes. *Infect Dis Clin North Am* 13, 285-97.
- Rice, J. (2012). Animal models: Not close enough. *Nature* 484, S9.
- Rittirsch, D., Hoesel, L.M. and Ward, P.A. (2007). The disconnect between animal models of sepsis and human sepsis. *J Leukoc Biol* 81, 137-143.
- Rossini, G. P. and Hartung, T. (2012). Food for thought ... towards tailored assays for cell-based approaches to toxicity testing. *ALTEX* 29, 359-72.
- Savitz, S. I. and Fisher, M. (2007). Future of neuroprotection for acute stroke: in the aftermath of the SAINT trials. *Ann Neurol* 61, 396-402.
- Schleppütz, M., Rieg, A. D., Seehase, S., et al. (2012). Neurally mediated airway constriction in human and other species: a comparative study using precision-cut lung slices (PCLS). *PLoS One* 7, e47344.
- Schmidt, M., Raghavan, B., Müller, V., et al. (2010). Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. *Nat Immunol* 11, 814-9.
- Schnerch, A., Cerdan, C., and Bhatia, M. (2010). Distinguishing between mouse and human pluripotent stem cell regulation: the best laid plans of mice and men. *Stem Cells* 28, 419-30.
- Sena, E. S., van der Worp, H. B., Bath, P. M., et al. (2010). Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* 8, e1000344.
- Seok, J., Warren, H. S., Cuenca, A. G., et al. (2013). Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A.* 110, 3507-12.
- Stebbins, R., Findlay, L., Edwards, C., et al. (2007). "Cytokine storm" in the phase I trial of monoclonal antibody TGN1412: better understanding the causes to improve preclinical testing of immunotherapeutics. *J Immunol* 179, 3325-31.
- van der Worp, H. B., Howells, D. W., Sena, E. S., et al. (2010). Can animal models of disease reliably inform human studies? *PLoS Med* 7, e1000245.
- van der Worp, H. B. and Macleod, M. R. (2011). Preclinical studies of human disease: time to take methodological quality seriously. *J Mol Cell Cardiol* 51, 449-50.

Marcel Leist, PhD
Doerenkamp-Zbinden Chair for in vitro toxicology and biomedicine
University of Konstanz
Box M657
78457 Konstanz
Germany
e-mail: marcel.leist@uni-konstanz.de

Thomas Hartung, MD PhD
Doerenkamp-Zbinden Chair for Evidence-based Medicine
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
W7032, Baltimore, MD 21205, USA
e-mail: thartung@jhsph.edu