



Future of Regulatory Safety Assessments

Gerd Bode and Petra Starck-Lantova

Contents

Introduction	2
Review of History, Methods, Regulatory, and Industrial Environment	2
Animal Models	5
The Controversy of the Animal Use in Pharmaceutical R&D	8
Suggestions for Improvement	9
Consider Entering the Clinic Without Animal Studies	9
Reduce the Need of Animal Studies by Gaining Information in Exploratory Clinical Trials	10
Use Alternative Approaches	10
Validation of Alternative Methods	13
Minimalize Bias in Experimental Data and Mind Good Research Practices	13
Conclusions and Outlooks	18
References and Further Reading	19

Abstract

Drug development contributes to improve health, duration, and quality of life. Lethal diseases have turned into chronic tolerable conditions, but medical need for many pathological processes continues. Concerns appear that in spite of extensive workload, success of pharmaceutical activity, and included

facilitated access to novel drugs, may slow down. The preclinical testing via in vitro and animal experimentation reveals limitations to select the right promising candidates, most likely to be effective in humans and predict undesirable side effects early on.

Therefore, constant efforts are necessary to improve the strategies. Courage needs to be stimulated to leave traditional paths and find new and better ways. This “rethinking” process needs directions to focus on additional options: use of more in silico data, deeper insight via cell cultures or receptor studies, new methods to explore more intensively relevant mechanisms of diseases and pharmacodynamics, more comparative data from different animal

G. Bode (✉)
Institute of Pharmacology and Toxicology, University
Medical Center, Goettingen, Germany
e-mail: gerd-bode@t-online.de
P. Starck-Lantova
The University of Bonn/German Society for Regulatory
Affairs, Bonn, Germany

models, which species really deliver signals relevant for patients; for this objective, disease models or implementation of human conditions into transgenic animals may be supportive. More rigorous randomized designs of preclinical studies and their blinded assessment may improve reproducible and therefore validated results.

In times of “big data” regulatory agencies, academic and industry researchers (possibly under political pressure) should feel obliged to stop selective publications (only positive effects) but create access also to options to learn from failures. The use of available knowledge (literature, experience, scientific advice) may limit the risks of reducing attrition rates and help to shorten timelines. Discussions with agencies have already facilitated a number of strategies. Examples are ICH guidelines M3 (allowing early access to new compounds for women of childbearing potential) or S 9 (reducing the preclinical development package for patients suffering from tumors).

The purpose of this chapter is to prompt openness and imagination to use new methods, more science, experience, and communication among researchers to the benefit of patients.

Introduction

The dualism of desirable effects and undesired reactions by chemical and pharmaceutical compounds on biological systems continues to be a fascinating and difficult phenomenon. Research, preclinical and clinical developments carry on to identify the main characteristics. Pharmacology detects new mechanisms for therapeutic purposes and thereby improves quality of life and survival, while toxicology and safety pharmacology submit their results to rigorous evaluations when extrapolation to humans takes place. The success is based on refinement of analytical methods, on paradigm change from morphology to inclusion of physiological functions, on better understanding of diseases and options for correcting such dysfunctions. *In silico*, *in vitro*, and *in vivo* studies support this process, but high-throughput

screening does not seem to lead to higher success rates, just the opposite, the attrition rates slow down.

This chapter tries to recommend some options how regulatory safety assessments could try to counterbalance this growing weakness.

Review of History, Methods, Regulatory, and Industrial Environment

Patent expirations, fast-rising competition of developing countries, and an increasingly complicated regulatory environment has reduced the attrition rates. Further, growing cost-constraining healthcare systems reduce the potential of many new drug discoveries to generate revenues sufficient to cover the costs of development, which raised for an individual substance up to US \$ 2.6 billion (Kaitin and DiMasi 2011).

The continuous pressure, fueled by public interest for effective and safe medicines, tightens regulatory requirements on the safety assurance before launch. On the other hand, payers enforce drug pricing limits and proofs of therapeutic or economic advantages over existing products. Biotechnological products can serve as an example for this trend. Due to their novelty and their substantial impact on disease progression, biotech products were originally well received in health care systems. Their acceptance, however, quickly became unsustainable when during the last few years, more than ten new products treating cancer were priced at over \$100,000 per year per patient, along with a continuing stream of many new orphan drugs costing from \$150,000 to \$500,000 per year, hepatitis drugs at \$80,000 per year, and others. This resulted in the public pressure to advocate for heavy price discounts and formal restrictions (Evens 2016).

As a counterbalance, pharmaceutical companies try to reduce their R&D costs by building strategic alliances with academic institutions, Contract Research Organizations (CROs), patient groups and other developers. Promising drug candidates are often acquired through mergers and acquisitions. This increases complexity,

fragmentation, obstructed communication and coordination among globally active partners and can result in insular solutions and addressing only part of a larger systemic problem. Mergers keep many pharmaceutical companies in a continuous state of reorganization; this does not favor stable research departments allowing to pursuing long-term scientific goals, like drug development, needing mostly more than 10 years of perseverance (Schueler and Buckley 2014; DiMasi et al. 2014).

No doubt, the tragedies of the past (Drug, Food and Cosmetics Act – USA, 1938 – due to mortalities of more than 100 patients after using sulfanilamide elixir with diethylene glycol as a solvent or thalidomide (Contergan) intake leading to phocomelia and other deformities) caused rigid legislations and requirements for animal studies to confirm safety before human treatment with new drugs.

Accordingly, the Kefauver Harris Amendment to the US Drug, Food and Cosmetics Act in 1962 defined the need to prove safety and efficacy of new pharmaceuticals (mostly through animal testing) before their exposure to humans and later approval; the predecessor of the EU, the European Economic Community, reacted similarly three years later by introducing the Directive 65/65/EEC. Since then, public attention and expectations on medicines safety is rising and pharmaceutical manufacturers are required to perform safety tests on their new drugs and submit the data to supervisory organs before being allowed to market their products.

The characterization of new products starts today often with high-throughput screening (leading to a ten-fold reduction of the cost of testing compound libraries), with combinatorial chemistry (increasing by 800-fold the number of new molecular entities to be potentially synthesized), or new generations of DNA sequencing (Scannel et al. 2012). However, although the turnover of tested substances had significantly increased, this enormous effort did not improve the final goal. Out of every 10,000 new molecular entities discovered, only one receives regulatory approval to be marketed, and the percentage of drugs entering clinical trials resulting in an approved medicine

is estimated to less than 11.3% compared to 16.4% success rate ten years ago (tufts CSDD 2015; PhRMA 2015).

This failure rate can partially be explained by the fact that many companies investigate new molecules before they clarify the pathological mechanisms of the diseases they are trying to treat. A more successful strategy would, therefore, attempt to understand the pathophysiology and epidemiology of the disease as early as possible before embarking expensive development programs (pwc 2012).

Some discrepancies can also be stated when looking at the relationship of preclinical and clinical results. The US-American FDA published already 2004 a white paper that critically evaluated preclinical animal models, recognizing them as one of the reasons for the disconnection between increased expenditure in R&D and attrition rate in drug discovery (FDA 2004). It was suggested that the high attrition rates of clinical trials indicate a discrepancy between the promising studies in animal models documenting the efficacy of a drug, and the real effects of the drug in human trial subjects. The quality and translational value of nonclinical research, particularly animal studies, has been therefore questioned.

Two directives referring to animal studies stress different sides of this issue.

In the European Union (EU), the rationale and requirements for animal testing in the development of medicinal products for human use are defined in Directive 2001/83/EC **Annex I**, which states that: “*An integrated and critical assessment of the non-clinical evaluation of the medicinal product in animals/in vitro shall be required. . .*” (Annex I, Part I, Art. 2.4), “*Clinical trials must always be preceded by adequate pharmacological and toxicological tests, carried out on animals . . .*” (Annex I, Part I, Art. 5.2b) with the option that “*Studies in animals can be substituted by validated in vitro tests provided that the test results are of comparable quality and usefulness for the purpose of safety evaluation*” (Annex I, Part I, Art. 4.2.3f). Animal studies have therefore become a standard component of pharmaceutical R&D.

This fact is in contrast to Directive 2010/63/EU on the protection of animals used for scientific purposes. This Directive has taken full effect on 1 January 2013 and refers directly to the principles of **Three Rs** (Refinement, Reduction, and Replacement). According to this Directive *“the use of animals for scientific or educational purposes should only be considered where a non-animal alternative is unavailable”* (preamble 12) and *“Member States shall ensure that, wherever possible, a scientifically satisfactory method or testing strategy, not entailing the use of live animals, shall be used instead”* (Art. 4.1).

Over the past three decades, the preclinical safety evaluation paradigm has developed in two parallel branches. For new chemical entities, the general approach has provided common ground for evaluation across different product classes; for new biological entities, where classical toxicology was acknowledged to be less relevant, a more product specific approach evolved. This resulted in two guidelines, ICH M3(R2) and ICH S6(R1), both published first 1997 in the Proceedings of the International Congresses on Harmonization (ICH) and then later revised, which is discernible by the letter R in the title for the recent guidelines.

The guideline ICH M3(R2) delivers practical recommendations for timing (when to conduct which safety studies) and conditions to include different patient population from male adults, over women without or with child bearing potential to pregnant women and children.

The guideline ICH S6(R1), on the other hand, stimulates to reconsider the traditional strategies in preclinical development. As one example, in case of recombinant proteins no studies on carcinogenicity and metabolism are required (because proteins are not metabolized to reactive species) and off-target effects are not expected (because of a high specificity of the recombinant proteins to the target). Another example is the assessment of potential effects on the cardiovascular, respiratory and nervous system (safety pharmacology): instead of stand-alone studies, these functions are recommended to incorporate into the pivotal chronic toxicity studies.

ICH S6(R1) emphasizes especially the selection of a relevant animal species, which are able to predict human reactions. The selection should be usually accomplished by an in-vitro comparison of binding affinity or functional activity of the product in human and animal cells, followed by in-vivo confirmation of the pharmacological activity or cross reactivity in that test species. In case of monoclonal antibodies, relevant species for testing are those that express the desired epitope and demonstrate a similar tissue cross-reactivity profile as for human tissues.

A number of flexible options should be considered: e.g. one relevant species may suffice, e.g., when only one relevant species can be identified, or where the biological activity of the biopharmaceutical is well understood. Or when no relevant species exists, the use of relevant transgenic or gene knock-out animals expressing the human receptor or homologous proteins could be chosen, no doubt, this option will prolong the evaluation process. In humanized mice, the comparability of pharmacodynamics in the animal model and humans is an important conclusion to consider the mouse as a suitable relevant model (van Meer et al. 2015).

The S 6(R1) guideline recognizes also animal models of disease as a relevant option. These models were originally used mainly to better understand the pharmacological action of the product, the pharmacokinetics and dosimetry. In all cases, the use of animal models of disease to support safety should be scientifically justified (ICH S6(R1)).

Nevertheless, the guideline caused criticism by some researchers: e.g. the guideline missed the chance to catch up with scientific progress (critical review by Kooijman et al. 2012). Indeed, for example the option to use in-vitro alternative approaches is mentioned only briefly, even in the guideline's revised version S6(R1) (approved 2011): *“Although not discussed in this guidance, consideration should be given to the use of appropriate in-vitro alternative methods for safety evaluation. These methods, if accepted by all ICH regulatory authorities, can be used to replace current standard methods.”* (Part II, Chap. 1.1).

It was therefore recommended (e.g., by Kooijman et al. (2012)) to make safety evaluations on a case-by-case basis, driven by product specific aspects such as the cause, mechanisms and reversibility of adverse effects. However, not much of experience and scientific expertise with these products could be gained in the meantime. In addition, the flexible case-by-case approach may lead to diverging interpretations and inconsistency of opinions between regulatory agencies.

The Tegenero case (monoclonal antibody) from 2006, leading unexpectedly to a cytokine storm inducing severe shock symptoms in volunteers, stimulated Agencies to reconsider their recommendations for the First in Man use of new compounds. The EMA guidance for first-in-man studies, “*Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products*” (EMA 2017a) came in February 2018 into effect. Requested are a better integration of pharmacokinetic, pharmacodynamic data and toxicological findings into the overall risk assessment; non-clinical data should help to define the estimated therapeutic dose, the maximum dose, and dose steps and intervals. The plead includes stronger stress on using alternative methods and encourages to *use more in vitro studies whenever scientifically relevant and sufficiently validated.*” The “weight of evidence” should include a comparison with humans in regard to target expression, distribution and primary structure, pharmacodynamics, metabolism and other PK aspects, and on- and off-target binding affinities and receptor/ligand occupancy and kinetics. Nevertheless, the guideline warns that even a high degree of homology between the selected animal model and human, or a similar response in human and animal cells in vitro, does not necessarily imply comparable effects in vivo. “*For example, there might be differences in affinity of the new candidate for molecular targets, physiology differences in tissue distribution of the molecular target, cellular consequences of target binding, cellular regulatory mechanisms, metabolic pathways, or compensatory responses to an initial physiological perturbation.*” In such situation understanding of the relevance of the animal models and their

translational differences may be improved by using in vitro human cell systems or human-derived materials.

Animal Models

Animal models certainly play an important role for the overall assessments. The request is for new pharmaceuticals to include relevant models, relevant for the prediction of reactions in humans.

No animal model can fully reproduce all features of human diseases. And no human models (volunteers or Phase II patients) can predict all reactions possibly seen later under broad exposure of thousands of patients under differing life styles and co-morbidities. But animal models allow to gain early on important signals for any severe effects. But selecting the optimal model is not a trivial task. Despite the S7A ICH Guideline recommends that “*consideration should be given to the selection of relevant animal models or other test systems so that scientifically valid information can be derived*”, the selection of animal species follows rather long-established practices and less scientifically justified deliberations.

A short reflexion about animal models may be helpful. Traditionally, over 90% of animals used in drug discovery are mice and rats. In drug development, the two species-testing is the rule. Rodent experiments should be completed with non-rodent species like dogs, non-human primates or minipigs. Non-human primates (NHPs) should only be used when the purpose of the study cannot be achieved by any other species (Article 8.1(b) Directive 2010/63/EU).

Mouse: Easily available; low cost; ease of handling; fast reproduction rate, important for reproductive toxicity studies. Many transgenic models.

Many well-established disease models created in mice, allowing both pharmacology and toxicology investigations. The limited blood volume can be overcome today by new sampling techniques and refined analytical methods, allowing micro-dosing studies.

Rats: Enormous historical background; important animal model for research in psychology and

biomedical science, especially cancer research. Their advantage: good availability; larger body size; easy handling. Gene knockout and embryonic stem cell techniques are relatively more difficult in rats. The role of genes is easy to study: many inbred strains, all members are closely genetically identical.

Syrian hamsters: Less used today; few times for assessing the potential for carcinogenesis, metabolic diseases, non-cancer respiratory diseases, cardiovascular or infectious diseases; less easy to handle because of fighting of males.

Rabbits: Particularly useful for assessing ocular and dermal irritation; primary non-rodent species for embryo-fetal developmental toxicity studies ever since the tragedy of thalidomide.

Dogs: Most frequently used non-rodent species in preclinical drug development; genetically diverse; a convenient model for many human diseases, e.g. the bone cancer osteosarcoma (Rowell et al. 2012). Dogs naturally develop beta-amyloid plaques (the protein's amino acid sequence in dogs is identical to humans), they show cognitive decline when growing older (Davis and Head 2014). Good model for complex neurocognitive disorders such as Alzheimer's disease. Dogs are expensive; need more sophisticated housing conditions and higher doses of an experimental drug. In addition, there is elevated public scrutiny and reluctance for the public and many evaluators.

Nonhuman primates (NHP): Use allowed when scientifically justified for safety testing. Most frequently used model to study potential adverse effects of monoclonal antibodies (mAbs). High public scrutiny. Target expression and function comparable to humans. Relatively large body size (allows repeated blood sampling), good availability of reagents, assays and methods (often adapted from humans), and generally good availability of animals are advantages of NHPs.

Potential limitations: high costs, limited group size, and often heterogeneous population with occasional background infections (like humans?). Testing often requires adult animals (4–5 years). NHPs inadequate for carcinogenicity testing and inconvenient for reproductive toxicity studies with their low fertility rate, high spontaneous

abortion rate and long pre- and postnatal development times (Baumann et al. 2014). When testing mAbs, results may be confounded by anti-drug antibodies formation, leading to a neutralization of pharmacologic effect or a clearance of the mAbs from the circulation (Bussiere et al. 2009).

Chapman et al. (2012) demonstrated that in some therapeutic areas rodents can support biologic development and provide relevant data and could therefore reduce the use of NHP. In controversial discussion is the duration of repeat-dose studies: six months provide sufficient data and nine months or longer did not bring any additional benefit (Clarke et al. 2008). There is also a plea to use only two dose groups instead of standard three, leading equally to relevant data (Chapman et al. 2010).

Parallel to public demands, criticism of unnecessary or even uninformative use of NHPs can be heard from professional circles too. Van Meer et al. (2013) evaluated safety studies in NHPs for mAbs registered in the EU and concluded that NHPs have been used even when there were other pharmacologically-responsive species available and the testing was in some cases not informative. The authors could also show that pharmacology-mediated adverse effects of mAbs are highly predictive from *in vitro* studies.

Minipigs: Increasingly used for toxicity testing of pharmaceuticals, experience in Europe (Ganderup et al. 2012), also in USA and Japan. Increasing acceptability by regulatory agencies, e.g. FDA. Up today: mostly testing on small molecule-based therapeutics and dermal administration (Ganderup et al. 2012), but increasingly also for repeat-dose administration of biologics (reviewed in Baumann et al. 2014). Zheng et al. (2012) demonstrated on several human mAbs that they show low clearance, long half-life and low volume of distribution in minipigs and therefore good translation to humans. Also, according to Baumann et al. (2014), studies on tissue cross reactivity of biopharmaceuticals as well as safety pharmacology and fertility endpoints in repeat-dose studies can be carried out in minipigs.

Advantage of the minipigs: immune system with structures and functions largely analogous to the human immune system (Bode et al. 2010).

Additionally, minipigs show less undue effects than dogs (Weaver et al. 2016). Also the complete genome (of a size as well as the number of genes comparable to humans) is known due to intensive data information based on research of house swines. Minipigs are not genetically transformed but the result of chronic breeding selection. Housing and handling easy with training.

Disadvantage: lack of placental transfer of macromolecules (Bode et al. 2010), which may limit their role in developmental toxicity testing of mAbs. In addition, rapid body weight gain, requiring more flexible testing strategies (shorter studies and use of younger lighter animals) and lack of published experience may be considered disadvantageous (Baumann et al. 2014).

Animal models of human disease: Primarily utilized to gain insight into the potential efficacy and mode of action of novel pharmaceuticals. Their value in understanding safety risks of compounds begins to be recognized (Morgan et al. 2013). Their use as part of a preclinical safety submission/dossier has been driven by the need to test a specific hypothesis and combine efficacy and safety evaluations; these models are even recommended by regulatory authorities (Cavagnaro and Lima 2015). Examples of the use of disease models include: infected animals to test the efficacy of a vaccine, mice inoculated with xenogenic (human) tumors expressing the target antigen, or genetically modified animals that develop spontaneous disease (Bussiere et al. 2009; Te Koppele and Witkamp 2008).

In general, animal models of human disease reflect rather simple mechanistic pathways while human diseases are mostly complicated by multifactorial pathological processes, often poorly understood. From a pathology perspective, the evaluation of animal disease models is challenging as the induced disease results in effects confounding safety assessment (Morgan et al. 2013). Historical data on spontaneous background finding are usually missing. Therefore, discerning whether the clinical and anatomic pathology findings are attributable to incidental age-related or background changes, tested agent, or disease manifestations require additional experience and background data need to be

accumulated. Multigenerational studies or increased numbers of control animals may be necessary (Cavagnaro and Lima 2015; Morgan et al. 2013; Bussiere et al. 2009). First observations signal that lifespan of disease models may be limited, therefore, adequacy of such animals in regard to chronic experiments may be an issue (Cavagnaro and Lima 2015).

Further, animal models of disease have intrinsic variability and immutable genetic and species differences. These factors can complicate the interpretation of the data. Investigators should therefore carefully evaluate the results and keep in mind that over- or under-estimating of adverse side effects may be possible. Also, analyze the target behavior in the animals, for low molecular weight chemicals the metabolite profile, for recombinant proteins the pharmacological effect (e.g., activity, clearance, target expression, immune phenotype, and immunogenicity) (van Meer et al. 2015).

Transgenic animals: The most common models are gene-targeted or knock-out (KO) animals; they lack an endogenous gene and therefore fail to express the related protein(s). This property offers the chance to assess drug specificity, investigate mechanisms of toxicity, screen for mutagenic and carcinogenic activities of therapeutic candidates, or study target blockade by novel therapeutic candidates (Bussiere et al. 2009). KO-mice have been valuable to study obesity, heart disease, diabetes, arthritis, substance abuse, anxiety, aging and Parkinson disease (NIH 2009).

The transfer of new genetic information can overexpress a target protein. The “humanized” knock-in animals with a human gene can evaluate the efficacy and toxicity of human biopharmaceuticals that are not pharmacologically active in normal rodents. Transgenic mice generated to carry cloned oncogenes and knock-out mice lacking tumor suppressing genes have provided good models for studying risk of human cancer; but in spite of their recommendation by ICH S1A, their use is still limited for the assessing the carcinogenic potential of new drugs (Friedrich and Olejniczak 2011).

For testing recombinant proteins and cell therapy products compensatory mechanisms may take

on the function of the absent protein(s) or target (e. g., induction of other calcium transporter genes in calbindin-D9k gene KO-mice described by Lee et al. 2007). Additionally, physiological effects of genetic mutations underlying diseases may differ in humans and in mice (Hirano et al. 2007).

The Controversy of the Animal Use in Pharmaceutical R&D

The selection of animal models should be based on their relevance for humans. In practice, the selection of species beyond those cited above is limited. There is a striking paucity of quantitative comparative data for animal models (Schein et al. 1970; Heywood 1981; Greaves et al. 2004; Matthews 2008). This makes any request using validated models and methods difficult. Moreover, literature offers only informative data, where animal models were positively contributing for Market Authorization. Data on failures and lack of success are not selected by researchers and editors for publication: reports about unacceptable adverse effects in animals are unattractive for journals; such data do not raise public interest; there could be also reasons for commercial confidentiality (Matthews 2008). Such data are thus stored in internal databases of pharmaceutical companies or research organizations and forlorn for public research.

This dilemma was addressed by Olson et al. in 2000: they compiled a survey of 150 compounds which revealed to be toxic under clinical test conditions in humans. There was a true positive human toxicity concordance rate of 71% for rodent and non-rodent species, with non-rodents alone being predictive for 63% of human toxicity and rodents alone for 43%. The authors appraised safety testing on (healthy) animals as significantly beneficial. But Matthews (2008) criticizes Olson's analysis for being inconclusive or even misleading because the authors did not attempt to estimate the corresponding specificity (true negative rate) and sensitivity (true positive rate) without which it is impossible to assess the evidential weight provided by the animal models. Van Meer (2013) contributes to this discouraging interpretation

and predicts that this poor outcome increases when looking at highly complex and species specific protein drugs, which are usually immunogenic in animals.

Perrin (2014) fortifies this issue by reporting that more than 80% of animal studies on safety and efficiency of potential therapeutics fail to predict the desired success rate in patients. Similar data by Hay et al. (2014) on success rates of 835 drug developers show that the proportion of therapies advancing from Phase 1 to regulatory approval is only around 10%. Bailey et al. (2014) analyzed datasets of 2366 drugs with both animal (rat, mouse and rabbit as preclinical species) and human data. The authors concluded that the absence of toxicity in the animals provided little or virtually no evidential signals for the lack of adverse drug reactions in humans.; a (re-)analysis of data specific for dogs from the same original dataset reinforced recent criticism that dogs are used mainly for historical instead of scientific reasons (Bailey et al. 2013): No evidence appeared, that canine data would predict efficacy and toxicology of medical compounds in clinical trials; they suggested that alternative methods are urgently required (Bailey et al. 2013).

Finally, van Meer et al. (2012), when focusing on post-marketing data, confirmed that animal data were not predictive for detecting serious adverse drug reactions in patients. Because 63% of adverse drug reactions had no counterparts in animals and less than 20% of serious reactions had an actual positive corollary in animal studies the authors conclude that animal safety studies in their current form should not be included in prospective pharmacovigilance studies.

Is there any chance to overcome this dilemma?

Kooijman (2013) explains the persistent use of animal studies in drug development by inertia of the system with animal studies embedded in a set of institutions (i.e., regulations, norms and values) that are taken for granted, normatively endorsed, and backed up by regulatory authorities. This is the motivation why the industry stays reluctant to move away from established conservative models. Although the standard animal models are as the result of guidelines by definition not binding, and it is possible to provide preclinical

data from own scientifically justified experimental programs adapted to a specific product, but this option is in practice not frequently used. This reluctance is the result of fears about possible delays or even failures of the marketing authorization process. Despite of some official support for progressive trend-setting new trends, skepticism prevails towards new strategies and conservatism dominates in industry.

Admittedly the situation of regulators is challenging. On one hand they try to promote innovations and recognize rapid growth in knowledge and technologies (Cavagnaro and Lima 2015), on the other hand they have to protect patients from risks.

Van Meer (2013) attacks both the industry and the regulators by saying:

The adoption of the precautionary principle by the regulatory authorities and the relative ease with which this burden of proof is accepted by the pharmaceutical industry – without attempts to improve the current paradigm – has created a stalemate in which animal studies, predictive or not, continue to exist with little room for innovation.

Therefore, all stakeholders should critically rethink their developmental strategies and should be encouraged to implement new technologies that predict the safety and efficacy of therapeutics better than current animal studies do.

Suggestions for Improvement

To break successfully with long-term traditions is only possible when academic and industrial researchers and developers cooperate openly with regulators and transmit their innovative thoughts into new guidelines and/or good practices. Some options will be illustrated in the following sections.

Consider Entering the Clinic Without Animal Studies

Biosimilars are taken as an example. The option to circumvent in-vivo preclinical studies has been

recently recommended in revised guidelines on biosimilar medicinal products (EMA 2012, 2014).

In contrast to generic medicinal products, non-clinical data, including animal studies, have been traditionally requested for marketing authorization of biosimilars. Nonetheless, this praxis has been recently abandoned by European regulators, because animal models turned out not to be sensitive enough to provide sufficient information on pharmaceutical similarity of these products (EMA 2012, 2014). The guidelines on biosimilars acknowledge that “*in-vitro assays may often be more specific and sensitive to detect differences between the biosimilar and the reference product than studies in animal*” and, therefore, “*these assays can be considered as paramount for the non-clinical biosimilar comparability exercise.*” Therefore, in-vivo testing should no longer be performed by default and its necessity should be considered on a case-by-case basis in a stepwise approach where the extent and nature of the development program depends on the level of evidence obtained in the previous step(s). This regulatory decision is regarded as revolutionary, it opens new ways for pharmaceutical developments with no new animal testing at all, and it implies that regulators may even discourage developers from performing such studies (van Aerts et al. 2014).

This applies especially to highly specific mAbs, where only NHPs are pharmacologically responsive. Therefore, as toxicological studies in NHPs have notably small group sizes, their conduct has been explicitly not recommended for biosimilars. In situation, when no relevant in-vivo animal model is available, the guidance leaves the option to proceed directly to human studies while applying principles to mitigate any potential risk (EMA 2012, 2014).

The recommended step-wise approach should proceed as follows: after physicochemical and biological characterization of the product, pharmacodynamic comparability should be evaluated in in-vitro assays. Assays using human cells or human receptors can be used to assess binding to the target and the subsequent functional effects. Pharmacokinetic comparability can then be best evaluated directly in clinical studies. When close similarity of the biological and its reference

product can be demonstrated, it is highly unlikely that new safety issues different from the reference product, with the exception of immunogenicity issues, would arise. For immunogenicity, animal studies have no predictive value anyway. Only after performing this biosimilarity exercise, it should be determined whether additional in-vivo non-clinical work is deemed necessary (EMA 2012, 2014; van Aerts et al. 2014).

Reduce the Need of Animal Studies by Gaining Information in Exploratory Clinical Trials

Exploratory clinical trials are an approach described in ICH M3(R) guideline, which recognizes that in some cases early access to human data can provide valuable information on human physiology/pharmacology, on drug candidate behavior, and on therapeutic target relevance to disease. Such data can reduce the need of information gained in animal studies. Central to this approach is the concept that “the best model for man is man.” Exploratory clinical trials are conducted in early Phase I (sometimes called Phase 0), have no therapeutic intent, are not intended to explore clinical tolerability, and can be conducted on patients or healthy individuals (ICH M3R). Their advantage is that they may give information on exposure and allow early comparison of kinetic/metabolic data between animal models and humans. They certainly help very early on to prioritize compounds when several candidates are available; these aspects again help to reduce animal usage compared to traditional development.

ICH M3(R) recommends also several approaches based on applying micro- or sub-therapeutic doses. **Microdosing** (most often: single microdose of 100 µg) is a method assessing the basic behavior of drugs by applying small doses directly to human volunteers. The doses are well below those expected to produce whole-body effects but high enough to allow the cellular response to be studied. A candidate drug is labeled by radiocarbon isotopes and extremely sensitive analytical methods (mostly positron emission

tomography (PET) and accelerator mass spectrometry (AMS)) are used for its biochemical quantitation. AMS is used for determining PK data by taking blood samples over time, processing the samples in the laboratory, and then analyzing their drug content. PET provides primarily PD data through real-time imaging and some limited PK data. The method provides important information about pharmacokinetics and pharmacodynamics, but it does not reveal information about toxicity or toxicology. Those endpoints will be addressed by supportive rigidly reduced conventional study designs.

Eliminating less promising molecules saves costs, resources, animals, and time. It avoids unnecessary exposure of the participants in clinical trials. Because the trials mostly involve a single dose administration (usually $1 \times 100 \mu\text{g}$, the alternative is $5 \times 100 \mu\text{g}$), the method poses very little risk of human toxic side effects (low dose and short duration of exposure). Very limited number of subjects is usually involved. Further, preclinical safety package required by authorities can be smaller as compared to the traditional Phase I studies, less animals are needed, and also only small quantity of the test drug is required. Other valuable advantages of microdosing studies are that they help to establish a likely pharmacological dose and select the first dose for the subsequent Phase I studies. A limitation of the method is shortage of data that exemplify whether the body’s reaction to a particular compound is similar when applied as microdose or in its pharmacologically active dose (Tiwari 2014).

Use Alternative Approaches

Alternative models should be more efficient and provide additional information to supplement the results from traditional animal models. Although animal models are still often considered to be a “gold standard,” they have never undergone validation to the same extent as non-animal technologies.

The need for improvement is recognized by Agencies. There is a new *Regulatory guidance on alternative/3Rs testing* approaches in

discussion: “Guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches” (EMA 2016a) and related reflection papers (still in the form of drafts – EMA 2016b, c). The guideline provides information on the scientific and technical criteria for regulatory acceptance of alternative/3Rs testing approaches and encourages stakeholders and authorities to initiate, support, and accept development and use of such approaches. The reflection papers summarize the main animal tests required for the regulatory testing of medicinal products and presents opportunities for limiting the use of animals.

The guideline recommends the following criteria: availability of test methodology, test protocols with clearly defined and scientifically sound endpoints; relevance of the test for a particular purpose and accuracy/extent to which the test correctly measures the biological effect of interest; robustness of the test (i.e., reproducibility of the test results); a comparison with existing methods; and a description of circumstances under which the 3Rs testing approach is/is not applicable.

The reflection paper on opportunities for implementation of the 3Rs during regulatory testing of medicinal products for human use provides an overview of options to limit or completely skip the use of animal studies in nonclinical evaluation of drug substances. The paper also clearly indicates that the 3R approach is in the state of dynamic development and there will be more options coming. It is, however, already clear that, for example, toxicity evaluation will change. Traditionally, repeat dose toxicology studies follow a standard design and in rodents and nonrodents yield information on general characteristics of the toxicity, the target organs of toxicity, the dose–response (curve) for each toxicity endpoint, responses to toxic metabolites formed in the organism, delayed responses, cumulative effects, the margin between toxic and nontoxic dose, information on reversibility/irreversibility of the effect, and NOAEL (no observed adverse effect level), NOEL (no observed effect level) for toxicity (EMA 2008, 2010; ICH M(R2)). In contrast to this standard approach, the reflection paper

(EMA 2016b) concedes the option to perform the tests on one species only (“*on a case by case approach, and if clearly justified*”) and to possibly omit a study on reversibility of compound-related effects.

Also changes for safety evaluation paradigms are recommended; for instance, in vivo genotoxicity can be assessed by integrating this endpoint into repeated dose toxicity studies, usually of 4 weeks duration. The reflection paper recommends a standard test battery (in-vitro tests plus in vivo genotoxicity integrated in repeated dose toxicity study) without the isolated single in-vivo study. Likewise, carcinogenicity and reproductive toxicity test requirements are currently under revision with the aim to induce new testing paradigms based on a more comprehensive weight-of-evidence approach and potential to replace in-vivo studies or not doing them at all (Bode and Van der Laan 2016). “Core battery” tests for safety pharmacology could also be integrated in repeated dose toxicity studies. And a variety of tests aiming at manufacture, characterization, and control of the drug substance should be primarily performed in-vitro unless thoroughly justified. Other more specific examples of the recommended 3Rs approaches involve avoiding physiological distribution test of radiopharmaceutical preparations as required by the Ph.Eur., using duck cells rather than live animals when testing plasma derived hepatitis B vaccine, or discouraging from using animals for potency testing of investigational, or biological medicinal products.

Alternatives to animal testing (called 3Rs testing approaches in EMA’s guidelines) are being developed for – besides ethical reasons – their time efficiency, less man power required, and cost effectiveness. Two most important approaches involve in-vitro cell culture techniques and in-silico computer simulations. These two approaches are then combined in another method known under the name “organs on a chip.” Also microdosing described above can be considered as an alternative method. All these approaches do not replace animals completely; however, they help to significantly reduce animal numbers needed.

There is knowledge that alternative methods also have their specific advantages and drawbacks. For example, cell cultures are criticized for not providing enough information about the complex interactions of living systems, computer simulations for using data from prior animal experiments, and microdosing for not revealing information about toxicity or toxicology. Thorough knowledge of the strengths and limitations of one's model is therefore crucial for its appropriate use and interpretation of results.

For cells and tissue cultures, in vitro tests are recommended in **OECD guidelines**. Examples are:

TG 428	Skin absorption: in vitro method
TG 430	In vitro skin corrosion: transcutaneous electrical resistance (TER)
TG 431	In vitro skin corrosion: human skin model test
TG 432	In vitro 3 T3 NRU phototoxicity test
TG 437	Bovine corneal opacity and permeability test method for identifying ocular corrosives and severe irritants
TG 438	Isolated chicken eye test method for identifying ocular corrosives and severe irritants
TG 439	In vitro skin irritation: reconstructed human epidermis (RhE) test method

Various types of cultures like cell culture, callus culture, tissue culture, organ culture, or separated cellular components are used for various purposes. For instance, for safety testing, bovine corneal organ culture can replace rabbits eye irritancy test, or models of human skin derived from cultured human skin (Corrositex[®], EPISKIN[™], EpiDerm[™]) can replace animal-based skin irritative and corrosive studies. Test systems based on the activation of human monocytes or monocytoid cell lines have been developed that take advantage of the role of these cells in the fever response and can replace rabbit pyrogen test. Similarly, mouse fibroblast (3 T3) and normal human keratinocyte (NHK) cells can be used in basal cytotoxicity test (e.g., phototoxicity

and support to determine the starting dose for the acute oral systemic toxicity test method and thereby reducing overall animal use requirements (NTP 2017). Cell cultures are further used to measure the rate of chemical absorption by the skin or phototoxic reactions and cultured cells have been developed to create monoclonal antibodies (Hester et al. 2006; Doke and Dhawale 2015).

Another example is represented by **tissue models**. For example, in-vitro metabolism studies have traditionally involved cells cultured into monolayers. However, because the interactions of cells with their surrounding environment can greatly affect shape, cell function and gene expression, two-dimensional or three-dimensional models have been developed. These models are supposed to better mimic mechanisms such as cell-to-cell adhesion and resistance to drug-induced apoptosis. Among the 3D-tissue reconstruction models are models of epidermis, full-thickness skin models, respiratory epithelia, keratinocyte eye cornea, vaginal epithelia, oral epithelia, and even models of the blood–brain barrier or three-dimensional models such as placenta, lymph node, and liver (Liebsch et al. 2011).

Organs on a Chip

Organ on a chip is a multichannel 3D microfluidic cell culture chip that simulates to some extent the activities, mechanics, and physiological response of entire organs. The chip is formed by small chambers containing a sample of tissue from a particular organ. When nutrients, air, blood and test compounds, such as experimental drugs, are pumped through the chambers, the cells replicate some of the key functions of that organ, just as they do in the body. By recapitulating the multicellular architectures, tissue-tissue interfaces, physicochemical microenvironments, and vascular perfusion of the body, these devices produce levels of tissue and organ functionality not possible with conventional 2D or 3D-culture systems. Biochemical, genetic, and metabolic activities of the cells are then measured by sensors and transferred for computer analysis. In the context of drug discovery and development, this technology is valuable for the study of molecular

mechanisms of action, prioritization of lead candidates, toxicity testing, and biomarker identification (Bhatia and Ingberg 2014; Prot and Leclerc 2012).

Validation of Alternative Methods

Alternative approaches may produce relevant and reliable results, but all new methods must be confirmed as suitable for its scientific **and** regulatory purpose. These methods are used routinely and repeatedly; they should be acceptable across countries; formal validation is a necessity. Therefore, it is recommended to involve regulators already in the process of definition of performance standards. Such cooperation will facilitate regulatory acceptance and help to implement new test methods (Liebsch et al. 2011).

The need of new alternative and validated methods is expressed in the EU legislation. Directive 2010/63/EU describes the coordination of formal validation studies at EU level to facilitate rapid uptake of new methods and approaches to replace reliance on animal testing as one of its key tasks. For this purpose, the European Union Reference Laboratory for alternatives to animal testing, **EURL ECVAM**, was established by this Directive. Through its network of laboratories (EU-NETVAL, European Union Network of Laboratories for the Validation of Alternative Methods), EURL ECVAM focuses on the validation of 3Rs methods for safety testing and efficacy/potency testing of chemicals, biologicals, and vaccines. It offers to research laboratories to scientifically validate alternative methods to animal testing. Through a dialogue with the stakeholder community and provision of information systems (DataBase service on Alternative Methods, DB-ALM, QSAR Model database and TSAR tracking system on alternative methods), EURL ECVAM further promotes the use and acceptance of new alternative methods in industry, academia, and by regulators. Examples are non-animal approaches for skin sensitization (allergy) testing, or co-developing two new (VICH) guidelines for the reduction of animal

tests for the quality control of veterinary vaccines (EURL ECVAM 2017).

Method validation relates mainly to:

1. The repeatability and reproducibility of results obtained
2. The test's relevance for measuring or predicting relevant biological effects

Validity assessment can include general knowledge of the method, the scientific principles on which it is based, historical data from using the method, and the use of pilot studies (when using in vivo methods) with smaller numbers of animals before embarking on a full scale study (European Commission 2016). The formal validation process involves multiple phases including preparatory method refinement, small-scale transfer studies, and finally large-scale international collaborative studies with manufacturers and national control laboratories (EMA 2016a). Alternatively, testing approaches that have sufficiently demonstrated their scientific validity according to the criteria described but have not been assessed in a formal validation process can also be evaluated on a case-by-case basis by the competent authorities (EMA 2016a).

Minimalize Bias in Experimental Data and Mind Good Research Practices

Animal studies can elucidate normal biology and improve the understanding for the pathogenesis of a disease, a deficiency often appearing when developing therapeutic interventions. However, animal studies produce insights only if tests are carefully designed, critically interpreted, and thoroughly reported. These quality features amplify good laboratory practices (GLP), compliant to which many animal studies (e.g., safety studies) should be performed. GLP ensures traceability and uniform, reproducible quality, but it does not guarantee the quality of the animal model or scientific valuable interpretation of the outcome for human purposes.

The lack of methodological rigor in preclinical studies acts as a barrier to translation of research

findings and represents a major source of reduced attrition rates in drug development (Glasziou 2014; Green 2015). The Lancet published 2009 a review on the production and reporting of biomedical research in which it was calculated that 85% of basic and clinical research is wasted because of inadequate or inappropriate design, nonpublication, and poor reporting (Chalmers and Glasziou 2009). This represents an estimated annual loss of over \$100 billion research funding. Clinical trials erroneously based on poorly conducted preclinical safety studies may lead to unnecessary exposure of trial participants to potentially harmful agents or to prevent them from participating in other trials with possibly effective products (Landis et al. 2012).

Particularly widespread are deficiencies in reporting key methodological parameters and poor experimental designs, both correlating with overstated findings (Landis et al. 2012; Gulin et al. 2015). Scientists from hematology and oncology department at the biotechnology firm Amgen (Begley and Ellis 2012) tried to confirm published findings related to their work and despite efforts to avoid technical differences they could confirm scientific findings in only 11% of cases. Reproducible studies were mainly those, in which authors had paid close attention to controls, reagents, investigator bias and describing complete data set. In the other cases, results could not be reproduced, the data were not routinely analyzed by investigators blinded to the experimental versus control groups and/or only selected experimental results supporting an underlying hypothesis were presented (Begley and Ellis 2012). Corresponding results were reported by Bayer HealthCare who could validate only about 25% of published preclinical studies (Prinz et al. 2011).

The recognition grows, that the use of techniques that assess the impact of publication and study-quality biases on estimates of efficacy in animal experiments is necessary (Sena et al. 2007). An adoption of newly and better defined quality standards would lead to improved effectiveness and efficiency in the selection of promising candidate drugs.

There are a number of sources of experimental bias which reduce the quality of the research.

Bias from Poor Reporting

Reporting details of a study including methods of statistical analyses used, sample sizes, inclusion/exclusion criteria, methods of randomization, blinding, gender, strain, species selection, and age of animals is essential to avoid publication bias, assist replication, and justify the research. Meanwhile, several guidelines have been issued to improve poor reporting, among them the ARRIVE guidelines (“Animal Research: Reporting of In-Vivo Experiments,” 2010), the GSPC (“Gold Standard Publication Checklist,” 2011), or the checklist of the Nature Journal (2013). Although the guidelines list suggestions for improved reporting, lack of pressure to apply these suggestions and report comprehensively and uniformly leads to noticeably inconsistency, obstructing correct assessment of reported results (Green 2015). Here, journal editors and regulators/assessors of clinical trial applications can support improvements considerably.

Bias from Nonpublication

Selective reporting is another reason for the lack of translation from basic research to the clinical situation. An increasing number of studies demonstrate publication bias that only about 50% of animal research results are published. The main motivation seems to be the lack of statistical significance as there is relatively little incentive for journals to publish negative, non-novel, or repeated findings (Korevaar et al. 2011; Sena et al. 2010; Ter Riet et al. 2012; Tsilidis et al. 2013). Nonpublication causes unnecessary duplication of research and poses a serious problem for performing valid literature syntheses. There should be a plea to publish **all** results regardless of whether the outcomes are positive or negative; all studies (equivalent to existing registers of clinical trials) should be registered in professional circles (Kimmelman and Anderson 2012). Registration of animal trials would impede retrospective changes of endpoints and study protocols and not publishing negative or unfavorable results. And today such publications of failures should

be easy since safety studies are always done under GLP conditions and collecting data and archiving is performed on local computers.

Bias from Using Inappropriate Animal Model

Prestige, economy, convenience, and poor awareness of the translation of basic research into medical practice influences decisions on animal studies more than scientific rigor and patient need (Green 2015). For instance, laboratory mice are disproportionately more often used than any other animal species (JAXmice[®] alone stocks tens of thousands of types of strains of mouse models to choose from), and the common practice of using inbred rodent strains completely ignores genetic variation of target populations. Bennani (2012) points out that for some conditions (e.g., influenza, bacterial, and fungal infections, measuring CVD and LDL and simple blood chemistry) animal models are more reliable predictors, whereas for other diseases (oncology, immunology, psychiatry, HIV, etc.), animal models are to large extent nonpredictive of clinical outcome. The importance of selecting the best possible animal model should be therefore not underestimated.

Bias from the Regulation of Animal Research

Regulatory agencies require sometimes preclinical investigations that use animal models known to have no predictive value. Among such problematic disease areas are oncology, immunology, or diseases of the central nervous system (Bennani 2012). In addition, compliance with the 3Rs and animal welfare are in many countries controlled by veterinary inspectors and ethic committees. The assessment process is however neither open nor transparent and relies on individual opinions of the experts (Green 2015). Pressure to apply the 3Rs principles may be overstretched; it may reduce the statistical power of experiments under meaningful values. Scott et al. (2008) demonstrated that the failure of murine amyotrophic lateral sclerosis treatments to translate to the clinic was due to small group size numbers and underpowered experiments.

Many good research principles are actually long known, it seems that they sometimes get forgotten in the complex process as R&D of new medicines. Awareness of quality guidelines for biomedical research should therefore be reanimated, examples are, for example, good research practice system of the World Health Organization (2006), guidelines published by the Research Quality Association in the UK (2008), or the Quality Assurance Toolkit developed at the University of Minnesota, USA (Michelson Prize and Grants 2014). Few details should be stressed.

Planning an Experimental Protocol

The methodological quality of an animal study starts with preparing a detailed experimental protocol. Checklist of factors listed, for example, in the ARRIVE guidelines, can be meaningful. Variations in the experiments must be considered and outlined in the protocol. Results from control animals need to be known and interpretation should benefit from these historical data. Study directors should seek consultancy from interdisciplinary interactions of the primary investigative team with experts in ancillary disciplines (statistics, laboratory animal science, pathology, etc.) and include the data generation and collection process (Everitt 2015). The experimental hypothesis to be tested must be well explained and defined as well as the experimental aims, design, and endpoints.

Recognizing Sources of Variation

Sources of variation can include inherent factors of the animal (e.g., stock/strain/substrain, source, sex, age, weight, source, pathogen status, etc.) as well as the animal facility environment (diet, bedding, housing, water delivery, lighting, noise, vibration, temperature, humidity, etc.). For this reason, harmonization of international standards for animal care would already reduce one important source of internal variation. Other factors are the methods used, dose form and timing of dose administration, types and preparation of excipients and vehicles, blood and tissue sampling sites and methods, handling of subjects, etc. (Everitt 2015). Example is the significant difference in the serum hepatic enzyme, alanine transaminase, which can occur if mice are handled by the body

instead of the tail (Swaim et al. 1985). Similarly, significant differences have been reported in research endpoints, such as cytokine concentrations, depending on method/site of blood removal (Mella et al. 2014).

Collecting Data

Although usually inbred rodent strains with minimal genetic differences are used, data show how important randomization is. For instance, in a systematic review of hypothermia in experimental stroke, nonrandomized studies overstated the reduction in infarct volume by 27% and studies without blinded outcome assessment overstated efficacy by 19% when they were compared to randomized and blinded studies, respectively (van der Worp et al. 2005). To minimize bias resulting from internal variation in the data, following steps should be always taken:

- **Randomization:** Animals should be assigned randomly to the various experimental groups and the method of randomization reported. Information on the allocation, treatment and handling of animals across study groups, the selection and source of control animals, including whether they are true littermates of the test groups should be provided. Data should be collected and processed randomly or appropriately blocked.
- **Blinding:** The investigator should be unaware of the group to which the next animal taken from a cage will be allocated (allocation concealment). Animal caretakers and investigators conducting the experiments should be blinded to the allocation sequence (blinded conduct of the experiment). Investigators assessing, measuring, or quantifying experimental outcomes should be blinded to the intervention (blinded assessment of outcome). This may hold true for all instances of the experiment, including also post-mortal investigations like macroscopical and pathohistological inspections and assessments.
- **Sample-size estimation:** Underpowered experiments with low predictive value may either falsely conclude that interventions are without efficacy or provide falsely positive

results leading to needless subsequent studies building upon the incorrect results. Too large studies will be unnecessarily costly. Both cases mean wasted resources in terms of time, money, and animals. An appropriate sample size should be therefore computed when the study is being designed and the statistical method of computation reported, which would also provide some assurance that sample size has not been increased incrementally in the light of ongoing analyses. Statistical methods that take into account multiple evaluations of the data should be used when an interim evaluation is carried out (Sena et al. 2007; Landis et al. 2012).

- **Data handling:** Rules for stopping data collection should be defined in advance. Also criteria for inclusion and exclusion of data should be established prospectively. How outliers will be defined and handled should be decided when the experiment is being designed, and any data removed before analysis should be reported. The primary endpoint should be prospectively selected. If multiple endpoints are to be assessed, then appropriate statistical corrections should be applied. Pseudoreplicate issues need to be considered before determining study design and analysis. For example, when analyzing effects of pollutants on reproductive health, multiple sampling from a litter, regardless of how many littermates are quantified, provides data from only a single biologic replicate. Investigators should also report how often a particular experiment was performed and whether results were substantiated by repetition under a range of conditions. Additionally, it should not be forgotten that a significant result does not provide information on the magnitude of the effect and thus does not necessarily mean that the effect is robust and highly reproducible (Landis et al. 2012).
- **Fighting experimental noise sound:** Irrelevant animals like those that die for reasons unrelated to disease (such as mishandling) should not be counted in results. Reasons for exclusion should be well documented. Whenever possible, numbers of males and females should be balanced because they can show sex-

dependent differences in symptoms that obscure modest drug effects. Littermates should be splitted among experimental groups (Perrin 2014).

- **Retrospective primary end-point selection:** Selection of a primary end-point only after data have been analyzed inflates the type-I error (false-positive results). This can be avoided by specifying a primary end point before the study is undertaken, the time(s) at which the end point will be assessed, and the method(s) of analysis. Significant findings for secondary end-points can and should be reported but should be delineated as exploratory in nature.
- **Reporting of individual data:** Nonrodent data are usually reported and interpreted on the basis of individual observations, reactions, and results. With rodent data using considerable higher numbers of animals, the statistical results often prevail and rare individual reactions get lost. The rule should be that preclinical investigations are handled like clinical results: Individual by individual, and not as a group mean. Hereby, possibly human relevant, but rare reactions do not get lost.
- **Avoid publication bias. Register all experiments. Use systematic reviews**

Systematic review (SR) (Sanderscock and Roberts 2002) is a simple technique developed to provide summary information by combining results from different sources and to make judgments on possible translation into clinical trials. In contrast to a narrative review which has no standardized methodology, the SR is a type of review that is structured, thorough, and transparent. Performing such appraisal can save resources and improve safety for participants in clinical trials achieved (van Lujik et al. 2014; Ritskes-Hoitinga et al. 2014; Vesterinen et al. 2014). Examples of the use of SR include, for instance, the study of Horn et al. (2001), who found no evidence to justify the start of clinical trials of nimodipine for focal cerebral ischemia in humans. The study emerged, however, only after 7665 patients participated in clinical trials. Comparably, Pound et al. (2004) demonstrated that drug side effects (in this case, excess risk of intracranial hemorrhage

after thrombolysis treatment for acute stroke) found during a clinical trial could have been identified beforehand if a SR of preclinical animal studies had been performed.

When performing a SR, it is important to evaluate the quality of data collected by other researchers. Its relevance can be illustrated on Alzheimer's disease, a condition which is despite of decades of experimental research known for a lack of effective disease modifying interventions. Egan et al. (2016) performed a SR and a meta-analysis of interventions tested in transgenic mouse model of the disease and after analyzing 427 publications describing 357 interventions in 55 transgenic models, involving 11,118 animals in 838 experiments, the authors found that the quality of these experiments was relatively poor – less than one in four publications reported blinded assessment of outcome or random allocation to group and no study reported a sample size calculation. Additionally, “trim and fill” analyses suggested that one in seven pathological and neurobehavioral experiments remained unpublished.

Likewise, Tsilidis et al. (2013) evaluated 4445 animal studies or 160 candidate treatments of neurological disorders and observed that 1719 of them had a “positive” result, whereas only 919 studies would a priori be expected to have such a result. From these 160 treatments, only 8 should have been subsequently tested in humans. These examples illustrate not only historical methodological weaknesses in preclinical animal testing but also insufficient critical appraisal of existing animal data before starting clinical research. Considering ethical issues and enormous financial costs related to clinical trials this is a rather alarming finding.

The bias resulting from not publishing could be significantly reduced by registering all experiments in a system similar to the one established for clinical trials. In this way, negative data would be published, unnecessary duplication of experiments would be prevented, investigators would receive credit for their work done and those seeking to summarize what is known would have

access to all relevant data. The registration to the system could be flexible, with information embargoed for a time to protect intellectual property (Macleod 2011).

To facilitate assessment of data collected, and to point out critical factors several study-quality checklists have been proposed. Among these are the CAMARADES checklist (Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Stroke), Macleod et al. 2004), Stroke Therapy Academic Industry Roundtable (STAIR 1999), Amsterdam criteria (Horn et al. 2001), Utrecht criteria (van der Worp et al. 2005), ARRIVE Guidelines (Animal Research: Reporting of In Vivo Experiments), Kilkenny et al. 2010), and the “Guidance for the Description of Animal Research in Scientific Publications” (National Research Council [US] Institute for Laboratory Animal Research 2011). Factors itemized on the checklists are, e.g., publication in peer-reviewed journal, assessment of functional and histological outcome, replication in two laboratories, testing both males and females, behavioral outcome measured for at least 1 month, assessment made in acute and chronic phase, randomization of treatment or control, blinded assessment of outcome, sample-size calculation before start of an experiment, and others.

Conclusions and Outlooks

Preclinical development and especially here animal studies have been identified as possible factors, responsible for the insufficient efficiency of nonclinical pharmaceutical R&D. Quantitative analyses of publicly available animal toxicity studies revealed that their results were inconsistent predictors of undesirable or toxic responses in humans. There is a lack of powerful data and sometimes only a poor basis available for deciding whether a compound should proceed to clinical testing (e.g., Bailey et al. 2014). The selection and justification of the studies is frequently based on regulatory principles 50 years old. The way how they are designed and performed is in many cases decided rather on habit and tradition than on

modern, scientifically sound justifications. Yet, although new approaches and technologies are being developed in a fast pace, their integration into drug development is rather slow.

But progress and optimization is required and can be achieved by opening and accepting new pathways: better animal models are needed, and better predictive non-animal models required. Provided that fit-for-purpose animal models are used and the design and execution of the testing is implemented according to stringent quality criteria *in vitro* and *in vivo* experiments can provide valuable information for the clinical performance of the drug. Unfortunately, besides a few exceptions, like development of humanized experimental animals, investment in development of more predictive animal models has been during the last decades considerably lower than in development of new technologies in areas such as molecular biology or clinical trial biomarkers (Denayer et al. 2014).

Better predictive state-of-the-art *in vitro* assays and *in silico* data, applied during early stages of drug discovery, can facilitate the long-term process of drug development. Replacing current acute and selected chronic *in vivo* regulatory toxicology studies by validated *in vitro* replacements would result in reduced animal use in pharmaceutical development of individual compounds. Following such strategy can be already observed in, e.g., some OECD test guidelines. The guideline no. 404 (Acute Dermal Irritation/Corrosion) recommends the conduct of *in vitro* assays (EOECD TG 430, 431) to limit the severity of toxicity for compounds that progress to *in vivo* evaluation.

The concept of the 3Rs exists since almost 60 years but its value has been mostly perceived only as a European regulatory issue (Chapman et al. 2013). Recognition is growing during recent years that there are benefits for improving the quality of research and reducing costs. The quality, reliability, and predictive value of many well established methods have not sufficiently been validated, but to abandon them is associated with insecurity and reduced trust in the view of remaining data. Lack of confidence in novel approaches is rooted in limited experience among researchers and Agencies and additionally in the

lack of historical data. There is fear to change long established pathways, which have been successful in the past. Joined effort of all parties involved is therefore needed to achieve progress and better acceptance of new approaches. Sharing knowledge (positive as well as negative experience) among stake-holders would facilitate the selection of the most promising methods. Such cooperation would fasten the transition towards novel approaches and reveal gaps for future research. To achieve the objective of such a paradigm change is only successful with full governmental support: the common goal should be to develop, optimize, and validate new translational tools, to revize some of the older guidelines and harmonize their acceptance on a global level. On the regulatory side, first attempt to catch up with progress is already happening: example the publication of the new guidelines on the evaluation of biosimilars on the European level. The International Conferences on Harmonization continue their awareness of scientific advances and their Expert Working Groups modify and improve the important recommendations of their global guidelines.

The objective of this chapter is to propose several approaches that can contribute to improving efficiency and translational value of non-clinical testing. The suggestions target at:

1. Enlarge in silico data bases and improve their accessibility.
2. For researchers and editors: Publish **all** data, knowledge and experience. Include **all** data, positive and negative results.
3. Expand options for in vitro methods and elucidate their advantages and limitations.
4. Improve the selection of validated approaches and document the real values of animal models and applicability of methods.
5. Improve the design and execution of experiments and use fully randomization, blinding, optimal statistical interpretation etc.
6. Reflect clinical conditions in preclinical studies: Standard programs, biomarkers, specified endpoints. Understand the mechanisms of disease.

7. Identify weaknesses of methods: Which are the most relevant and predictive models and methods for human conditions?.
8. Introduce state-of-the-art methods into daily practice: GLP, statistics, combine functions with morphology, use non-invasive methods, provide support by kinetic data, etc.
9. Improve quality of reporting: analyze and assess all results, from **all** studies, focus on mean and individual effects.
10. Conscientious review of literature: built up “weight of evidence” approach, use information from Quality, Safety and Efficacy.
11. Encourage an open dialogue among researchers from all disciplines in industry and agencies.
12. Use scientific advice offered by Agencies to facilitate the decisions for best strategy during all phases of development.
13. Gain meaningful information of animal usage for human conditions at every step of development, get human data early as possible, use expedited explorations.
14. Better prediction of drug reactions in humans based on modern intelligent complex approaches will fasten access to efficient and safe drugs. For all these objectives, courage should be stimulated to swing from preconceived concepts to new methods and pathways. Only frank imaginative discussions will open the doors to an optimum way forward.

References and Further Reading

- Akhtar A (2015) The flaws and human harms of animal experimentation. *Camb Q Healthc Ethics* 24(4):407–419
- Arlington S (2012) From vision to decision: pharma 2020. Price Waterhouse Coopers (PwC), London. Available at <https://www.pwc.com/gx/en/pharma-life-sciences/pharma2020/assets/pwc-pharma-success-strategies.pdf>. Accessed on 3 July 2017
- ARRIVE – Animal Research: Reporting of In Vivo Experiments. Available at <https://www.nc3rs.org.uk/arrive-animal-research-reporting-vivo-experiments>. Accessed on 3 July 2017
- Arrowsmith J (2011) Trial watch: phase III and submission failures: 2007–2010. *Nat Rev Drug Discov* 10:87

- Bailey J, Thew M, Balls M (2013) An analysis of the use of dogs in predicting human toxicology and drug safety. *Altern Lab Anim* 41(5):335–350
- Bailey J, Thew M, Balls M (2014) An analysis of the use of animal models in predicting human toxicity and drug safety. *Altern Lab Anim* 42:181–199
- Baumann A, Flagella K, Forster R, de Haan L, Kronenberg S, Locher M et al (2014) New challenges and opportunities in nonclinical safety testing of biologics. *Regul Toxicol Pharmacol* 69(2):226–233
- Begley C, Ellis L (2012) Drug development: raise standards for preclinical cancer research. *Nature* 483:531–533
- Bennani YL (2012) Drug discovery in the next decade: innovation needed ASAP. *Drug Discov Today* 16(17–18):779–792
- Bhatia SN, Ingberg DE (2014) Microfluidic organs-on-chips. *Nat Biotechnol* 32:760–772
- Bluemel J (2012) Considerations for the use of nonhuman primates in nonclinical safety assessment. In: Weinbauer GF, Vogel F (eds) *Challenges in nonhuman primate research in the 21st century*, Waxman (2012). Charles River Publication Series, pp 59–70. ISBN 978-3-8309-2839-3
- Bode G, Van der Laan JW (2016) Paradigm change in cancerogenicity, Presentation: German Pharm-Tox Summit, Berlin, 29-02 to 03-03-2016, Berlin, ACS Publications, pubs.acs.org/crt, Symposium 19
- Bode G, Clausing P, Gervais F, Loegsted J, Luft J, Noguez V, Sims J et al (2010) The utility of the minipig as an animal model in regulatory toxicology. *J Pharmacol Toxicol Methods* 62(3):196–220
- Bussiere JL, Martin P, Horner M, Couch J, Flaherty M, Andrews L, Beyer J, Horvath C (2009) Alternative strategies for toxicity testing of species-specific biopharmaceuticals. *Int J Toxicol* 28(3):230–253
- CAMARADES: Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies. Available at <http://www.dcn.ed.ac.uk/camarades/>. Accessed 3 July 2017
- Casty F, Wieman M (2013) Drug development in the 21st century. The synergy of public, private, and international collaboration. *Ther Innov Regul Sci* 47(3):375–383
- Cavagnaro J, Lima BS (2015) Regulatory acceptance of animal models of disease to support clinical trials of medicines and advanced therapy medicinal products. *Eur J Pharmacol* 759:51–62
- Chalmers I, Glasziou P (2009) Avoidable waste in the production and reporting of research evidence. *Lancet* 374:86–89
- Chapman KL, Pullen N, Andrews L, Ragan I (2010) The future of non-human primate use in mAb development. *Drug Discov Today* 15:235–242
- Chapman K, Andrews L, Bajramovic JJ, Baldrick P et al (2012) The design of chronic toxicology studies of monoclonal antibodies: implications for the reduction in use of non-human primate. *Regul Toxicol Pharmacol* 62:347–354
- Chapman KL, Holzgrefe H, Black LE, Brown M, Chellman G, Copeman C, Couch J et al (2013) Pharmaceutical toxicology: designing studies to reduce animal use, while maximizing human translation. *Regul Toxicol Pharmacol* 66:88–103
- Cherkasov A, Muratov E, Fourches D, Varnek A, Baskin I, Cronin M et al (2014) QSAR modeling: where have you been? Where are you going to? *J Med Chem* 57(12):4977–5010
- Clarke J, Hurst C, Martin P, Vahle J, Ponce R, Mounho B, Heidel S, Andrews L, Reynolds T, Cavagnaro J (2008) Duration of chronic toxicity studies for biotechnology-derived pharmaceuticals: is 6 months still appropriate? *Regul Toxicol Pharmacol* 50:2–22
- Tufts center for study of drug development (CSDD) (2015) Outlook 2015. Available at: <http://csdd.tufts.edu/files/uploads/Outlook-2015.pdf>. Accessed 5 Feb 2017
- Davis PR, Head E (2014) Prevention approaches in a preclinical canine model of Alzheimer's disease: benefits and challenges. *Front Pharmacol* 5:47
- Denayer T, Stöhr T, van Roy M (2014) Animal models in translational medicine: validation and prediction. *New Horiz Transl Med* 2:5–11
- DiMasi J, Kim J, Getz K (2014) The impact of collaborative and risk-sharing innovation approaches on clinical and regulatory cycle times. *Ther Innov Regul Sci* 48(4):482–487
- DiMasi J, Grabowski H, Hansen R (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 47:20–33
- Directive 65/65/EEC of 26 January 1965 on the approximation of provisions laid down by Law, Regulation or Administrative Action relating to proprietary medicinal products. *Off J Eur Union* 22, 09/02/1965, pp 369–373
- Directive 2001/82/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to veterinary medicinal products. *Off J Eur Union L* 311, 28/11/2001, pp 1–66
- Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Off J Eur Union L* 276, 20/10/2010, pp 33–79
- Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. *Off J Eur Union L* 311, 28/11/2001, pp 37–128
- Doke SK, Dhawale SC (2015) Alternatives to animal testing: a review. *Saudi Pharm J* 23(3):223–229
- Egan KJ, Vesterinen HM, Beglopoulos V, Sena ES, Macleod MR (2016) From a mouse: systematic analysis reveals limitations of experiments testing interventions in Alzheimer's disease mouse models. *Evid-Based Preclin Med* 3(1):12–23
- EMA (2008) Note for guidance on non-clinical safety studies for the conduct of human clinical trials and marketing authorisation for pharmaceuticals. CPMP/ICH/286/95
- EMA (2010) Guideline on repeated dose toxicity. CPMP/SWP/1042/99 Rev 1 Corr

- EMA (2012) Guideline on similar biological medicinal products containing monoclonal antibodies – non-clinical and clinical issues. EMA/CHMP/BMWP/403543/2010
- EMA (2014) Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues. EMEA/CHMP/BMWP/42832/2005 Rev1
- EMA (2016a) Guideline on the principles of regulatory acceptance of 3R (replacement, reduction, refinement) testing approaches. EMA/CHMP/CVMP/JEG-3Rs/450091/2012
- EMA (2016b) Reflection paper providing an overview of the current regulatory testing requirements for medicinal products for human use and opportunities for implementation of the 3R (EMA/CHMP/CVMP/JEG3Rs-3Rs/742466/2015) – published for consultation
- EMA (2016c) Reflection paper providing an overview of the current regulatory testing requirements for veterinary medicinal products and opportunities for implementation of the 3Rs (EMA/CHMP/CVMP/JEG-3Rs/164002/2016) – published for consultation
- EMA (2017a) Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products (EMA/CHMP/SWP/28367/07 Rev. 1
- EMA (2017b) Reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development (EMA/CHMP/138502/2017) – published for consultation
- EURL ECVAM (2017) Homepage of the European Commission, The European Union Reference Laboratory for alternatives to animal testing (EURL-ECVAM). Available at <https://eurl-ecvam.jrc.ec.europa.eu/>. Accessed 1 Sept 2017
- European Commission 2013: Report from the Commission to the Council and the European Parliament: seventh report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union, SEC(2010) 1107, 15pp. European Commission, Brussels. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52013DC0859&from=EN>. Accessed 13 Jan 2017
- European Commission (2016) Animals used for scientific purposes. Available at http://ec.europa.eu/environment/chemicals/lab_animals/3r/acceptance_en.htm. Accessed 4 July 2017
- Evens R (2016) Pharma success in product development – does biotechnology change the paradigm in product development and attrition. *AAPS J* 18:281
- Everitt JI (2015) The future of preclinical animal models in pharmaceutical discovery and development: a need to bring in cerebro to the in vivo discussions. *Toxicol Pathol* 43:70–77
- FDA (2004) Innovation or stagnation: challenge and opportunity on the critical path to new medical products. In: *Challenges and Opportunities Report* US Department of Health and Human Services. Available at: <https://www.fda.gov/downloads/scienceresearch/specialtopics/criticalpathinitiative/criticalpathopportunitiesreports/ucm113411.pdf>. Accessed 13 Sept 2017
- Friedrich A, Olejniczak K (2011) Evaluation of carcinogenicity studies of medicinal product for human use authorized via the European centralized procedure (1995–2009). *Regul Toxicol Pharmacol* 60:225–248
- Ganderup NC, Harvey W, Mortensen JT, Harrouk W (2012) The minipig as nonrodent species in toxicology – where are we now? *Int J Toxicol* 31(6):507–528
- Getz K (2011) Transforming legacy R&D through open innovation. *Monit* 4(3):16–21
- Getz KA, Kaitin KI (2015) Why is the pharmaceutical and biotechnology industry struggling? In: Schüller P, Buckley B (eds) 2015: re-engineering clinical trials. Best practices for streamlining the development process. Academic Press/Elsevier. ISBN: 9780124202467
- Glasziou P (2014) The role of open access in reducing waste in medical research. *PLoS Med* 11:e1001651
- Greaves P, Williams A, Eve M (2004) First dose of potential new medicines to humans: how animals help. *Nat Rev Drug Discov* 3:226–236
- Greek R, Hansen LA (2013) The strengths and limits of animal models as illustrated by the discovery and development of antibacterials. *Biol Syst* 2:109
- Green SB (2015) Can animal data translate to innovations necessary for a new era of patient-centred and individualised healthcare? Bias in preclinical animal research. *BMC Med Ethics* 16:53
- Groenink L, Folkerts G, Schuurman HJ (2015) Editorial, special issue on translational value of animal models. *Eur J Pharmacol* 759:1–2
- Gulin JE, Rocco DM, García-Bournissen F (2015) Quality of reporting and adherence to ARRIVE guidelines in animal studies for Chagas disease preclinical drug research: a systematic review. *PLoS Negl Trop Dis* 9(11):e0004194
- Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. *Nat Biotechnol* 32:40–51
- Hester RE, Harrison M, Illing P (2006) General overview of the safety evaluation of chemicals. In: Hester RE, Harrison M (eds) *Alternatives to animal testing*. RSC Publishing. ISBN 978-0-85404-211-1
- Heywood R (1981) Target organ toxicity. *Toxicol Lett* 8:349–358
- Hirano E, Knutsen RH, Sugitani H, Ciliberto CH, Mecham RP (2007) Functional rescue of elastin insufficiency in mice by the human elastin gene: implications for mouse models of human disease. *Circ Res* 101(5):523–531
- Horn J, de Haan RJ, Vermeulen M, Luiten PGM, Limburg M (2001) Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. *Stroke* 32:2433–2438
- ICH guideline S3A: Note for guidance on toxicokinetics: the assessment of systemic exposure in toxicity studies. Oct 1994
- ICH guideline S1A: Need for carcinogenicity studies of pharmaceuticals. Nov 1995

- ICH guideline S1B: Testing for carcinogenicity of pharmaceuticals. July 1997
- ICH guideline S 7 B: The nonclinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals. May 2005
- ICH guideline S1C(R2): Dose selection for carcinogenicity studies of pharmaceuticals. March 2008
- ICH guideline M3(R2): Non-clinical safety studies for the conduct of human clinical trials and marketing authorisation for pharmaceuticals. June 2009
- ICH Guideline S2(R1): Guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use. Nov 2011
- ICH guideline S3B Pharmacokinetics: guidance for repeated dose tissue distribution studies. Oct 1994
- ICH Guideline S5(R2): Detection of toxicity to reproduction for medicinal products & toxicity to male fertility. June 1993
- ICH guideline S6 (R1): Preclinical safety evaluation of biotechnology-derived pharmaceuticals. June 2011
- ICH guideline S7A: Safety pharmacology studies for human pharmaceuticals. Nov 2000
- ICH guideline S9: Nonclinical evaluation for anticancer pharmaceuticals. Oct 2009
- Kaitin K, DiMasi J (2011) Pharmaceutical innovation in the 21st century: new drug approvals in the first decade 2000–2009. *Clin Pharm Therap* 89(2):183–188
- Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010) Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 8:e1000412
- Kimmelman J, Anderson JA (2012) Should preclinical studies be registered? *Nat Biotechnol* 30(6):488–489
- Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3:711–715
- Kooijman M (2013) Why animal studies are still being used in drug development. *Altern Lab Anim* 41(6):79–81
- Kooijman M, van Meer P, Moors E, Schellekens H (2012) Thirty years of preclinical safety evaluation of biopharmaceuticals: did scientific progress lead to appropriate regulatory guidance? *Expert Opin Drug Saf* 11(5):797–801
- Korevaar DA, Hooft L, Ter Riet G (2011) Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments. *Lab Anim* 45:225–230
- Lamberti M, Mathias A, Myles J, Howe D, Getz K (2012) Evaluating the impact of patient recruitment and retention practices. *Drug Inf J* 46(5):573–580
- Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG et al (2012) A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 490:187–191
- Langley G, Austin CP, Balapure AK, Birnbaum LS, Bucher JR, Fentem J, Fitzpatrick SC et al (2015) Lessons from toxicology: developing a 21st-century paradigm for medical research. *Environ Health Perspect* 123:A268–A272
- Lee GS, Lee KY, Choi KC, Ryu YH, Paik SG, Oh GT, Jeung EB (2007) Phenotype of a calbindin-D9k gene knockout is compensated for by the induction of other calcium transporter genes in a mouse model. *J Bone Miner Res* 22(12):1968–1978
- Liebsch M, Grune B, Seiler A, Butzke D, Oelgeschläger M et al (2011) Alternatives to animal testing: current status and future perspectives. *Arch Toxicol* 85(8):841–858
- Macleod M (2011) Why animal research need to improve. *Nature* 477:511
- Macleod MR, O'Collins T, Howells DW, Donnan GA (2004) Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* 35:1203–1208
- Matthews R (2008) Medical progress depends on animal models – doesn't it? *J R Soc Med* 101(2):95–98
- Mella JR, Chiswick EL, King E, Remick DG (2014) Location, location, location: cytokine concentrations are dependent on blood sampling site. *Shock* 42:337–342
- Michelson Prize & Grants (2014) The quality assurance toolkit. Available at <http://www.michelsonprizeandgrants.org/resources/qa-toolkit>. Accessed 29 Aug 2017
- Morgan SJ, Elangbam CS, Berens S, Janovitz E, Vitsky A, Zabka T, Conour L (2013) Use of animal models of human disease for nonclinical safety assessment of novel pharmaceuticals. *Toxicol Pathol* 41(3):508–518
- National Research Council [US] Institute for Laboratory Animal Research 2011: Guidance for the Description of Animal Research in Scientific Publications
- NIH, National Human Genome Research Institute (2009) Knockout mice. Available at: <https://www.genome.gov/12514551/>. Accessed 2 Aug 2017
- NTP (2017) <https://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/acute-systemic-tox/in-vitro-validation/index.html>
- Nuffield Council on Bioethics (2005) The ethics of research involving animals. Available at: <https://nuffieldbioethics.org/wp-content/uploads/The-ethics-of-research-involving-animals-full-report.pdf>. Accessed 12 Sept 2017
- OECD (2015) Guidance document on revisions to OECD genetic toxicology test guidelines. Available at <https://www.oecd.org/chemicalsafety/testing/Genetic%20Toxicology%20Guidance%20Document%20Aug%2031%202015.pdf>
- Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J et al (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32(1):56–67
- Paul S, Mytelka D, Dunwiddie C, Persinger C, Munos B, Lindborg S, Schacht A (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 9:203–214
- Perrin S (2014) Preclinical research: make mouse studies work. *Nature* 507:7493

- PhRMA (Pharmaceutical Research and Manufacturers of America) 2015: 2015 biopharmaceutical research industry profile. Available at: http://www.phrma.org/sites/default/files/pdf/2015_phrma_profile.pdf. Accessed 4 Feb 2017
- Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I (2004) Where is the evidence that animal research benefits humans? *BMJ* 328(7438):514–517
- Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 10(9):712
- Prot J, Leclerc E (2012) The current status of alternatives to animal testing and predictive toxicology methods using liver microfluidic biochips. *Ann Biomed Eng* 40(6):1228–1243
- Pwc (PricewaterhouseCoopers) 2012: Pharma 2020: The vision. Which path will you take? Available at: <http://www.pwc.com/gx/en/industries/pharmaceuticals-life-sciences/pharma-2020/pharma-2020-vision-path.html>. Accessed 4 Feb 2017
- Research Quality Association (2008) Guidelines for quality in non-regulated scientific research. Available at: <http://www.therqa.com/publications/booklets/guidelines-quality-non-regulated-scientific-research/>. Accessed 29 Aug 2017
- Ritskes-Hoitinga M, Leenaars M, Avey M, Rovers M, Scholten R (2014) Systematic reviews of preclinical animal studies can make significant contributions to health care and more transparent translational medicine[editorial]. *Cochrane Database Syst Rev* (3). <https://doi.org/10.1002/14651858.ED000078>
- Rowell JL, McCarthy DO, Alvarez CE (2012) Dog models of naturally occurring cancer. *Trends Mol Med* 17(7):380–388
- Russell WM, Burch RL (1959) The principles of humane experimental technique. Methuen, London. ISBN 9780900767784
- Sandercock P, Roberts I (2002) Systematic reviews of animal experiments. *Lancet* 360(9333):586
- Scannel J, Blanckley A, Boldon H, Warrington B (2012) Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* 11:191–200
- Schein PS, Davis RD, Carter S, Newman J, Schein DR, Rall DP (1970) The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther* 11:3–40
- SCHER (The Scientific Committee on Health and Environmental, The European Commission) (2009) Opinion on “The need for non-human primates in biomedical research, production and testing of products and devices”. Available at: http://ec.europa.eu/health/scientific_committees/opinions_layman/en/non-human-primates/1-3/2-research-safety-testing.htm#0p0. Accessed 7 Aug 2017
- Schueler P, Buckley B (eds) (2014) Re-engineering clinical trials: best practices for streamlining the development process. Academic Press. ISBN-10: 0124202462
- Scott S, Kranz JE, Cole J, Lincecum JM, Thompson K, Kelly N et al (2008) Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph Lateral Scler* 9:4–15
- Sena E, van der Worp B, Howells D, Macleod M (2007) How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci* 30(9):433–439
- Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol* 8:e1000344
- STAIR (Stroke Therapy Academic Industry Roundtable) (1999) Recommendations for standards regarding pre-clinical neuroprotective and restorative drug development. *Stroke* 30:2752–2758
- Swaim LD, Taylor HW, Jersey GC (1985) The effects of handling techniques on serum ALT in mice. *J Appl Toxicol* 5:160–162
- Taylor K, Gordon N, Langley G, Higgins W (2008) Estimates for worldwide laboratory animal use in 2005. *Altern Lab Anim* 36:327–342
- Te Koppele J, Witkamp R (2008) Use of animal models of disease in the preclinical safety evaluation of biopharmaceuticals. In: Cavagnaro JA (ed) *Preclinical safety evaluation of biopharmaceuticals: a science-based approach to facilitating clinical trials*. Wiley, pp 293–308. ISBN: 978-0-470-10884-0
- Teilmann K, Hohbach C, Lehmann H (1986) Preclinical safety testing of species specific proteins produced with recombinant DNA-techniques. An attempt to transfer current experience into future testing strategies. *Arch Toxicol* 59:195–200
- Ter Riet G, Korevaar DA, Leenaars M, Sterk PJ, VanNoorden CJF, Bouter LM, Lutter R, Elferink RPO, Hooft L (2012) Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. *PLoS One* 7:e43404
- Tiwari A (2014) Microdosing: a new approach to clinical drug development. Available at: <https://de.slideshare.net/drashutoshtiwari/microdosing-phase-0-studies>. Accessed on 22 Sept 2017
- Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, Salman RA, Macleod MR, Ioannidis JPA (2013) Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS One* 11:e1001609
- Van Aerts LA, De Smet K, Reichmann G, van der Laan JW, Schneider CK (2014) Biosimilars entering the clinic without animal studies: a paradigm shift in the European Union. *MAbs* 6(5):1155–1162
- van der Worp HB, de Haan P, Morrema E, Kalkman CJ (2005) Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. *J Neurol* 252:1108–1114
- Van Luijk J, Bakker B, Rovers MM, Ritskes-Hoitinga M, de Vries RB, Leenaars M (2014) Systematic reviews of animal studies; missing link in translational research? *PLoS One* 9:e89981

- Van Meer P (2013) The scientific value of non-clinical animal studies in drug development. PhD thesis, Utrecht University
- Van Meer P, Gispens-de Wied C, Moors E, Schellekens H (2012) The ability of animal studies to detect serious post marketing adverse events is limited. *Regul Toxicol Pharmacol* 64(3):345–349
- Van Meer PJ, Kooijman M, van der Laan JW, Moors EH, Schellekens H (2013) The value of non-human primates in the development of monoclonal antibodies. *Nat Biotechnol* 31(10):882–883
- Van Meer PJK, Graham ML, Schuurman HJ (2015) The safety, efficacy and regulatory triangle in drug development: impact for animal models and the use of animals. *Eur J Pharmacol* 759:3–13
- Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, Antonic A, Howells DW, Macleod MR (2014) Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods* 221:92–102
- Weaver ML, Blak Grossi A, Schützack J, Parish J et al (2016) Vehicle systems and excipients used in minipig drug development studies. *Toxicol Pathol* 44(3):367–372
- WHO (World Health Organization) (2006) Handbook: quality practices in basic biomedical research. Available at: <http://www.who.int/tdr/publications/training-guideline-publications/handbook-quality-practices-bio-medical-research/en/>. Accessed 29 Oct 2017
- Zbinden G (1987) Biotechnology products intended for human use, toxicological targets and research strategies. *Prog Clin Biol Res* 235:143–159
- Zheng Y, Tesar DB, Benincosa L, Birnböck H (2012) Minipig as a potential translatable model for monoclonal antibody pharmacokinetics after intravenous and subcutaneous administration. *MAbs* 4(2):243–255