3-2012

Alternative Approaches for Medical Countermeasures to Biological and Chemical Terrorism and Warfare

Thomas Hartung
Johns Hopkins University

Joanne Zurlo
Johns Hopkins University

Recommended Citation

Hartung, T., & Zurlo, J. (2012). Alternative approaches for medical countermeasures to biological and chemical terrorism and warfare. ALTEX-Alternatives to animal experimentation, 29(3), 251-260. DOI https://doi.org/10.14573/altex.2012.3.251

Food for Thought ...
Alternative Approaches for Medical Countermeasures to Biological and Chemical Terrorism and Warfare

Thomas Hartung [1,2] *and Joanne Zurlo* [1]

[1]Johns Hopkins University, Bloomberg School of Public Health, CAAT, Baltimore, USA; [2]University of Konstanz, CAAT-Europe, Germany

Summary

The desire to develop and evaluate drugs as potential countermeasures for biological and chemical threats requires test systems that can also substitute for the clinical trials normally crucial for drug development. Animal models have limited predictivity for drug efficacy, as is well known from many disappointments in clinical trials. Traditional in vitro and in silico approaches are not really game changers here, but the substantial investment into novel tools now underway might bring about a second generation of alternative approaches. The avenue pursued focuses primarily on the development of a Human on a Chip, i.e., the combination of different three-dimensional (stem) cell-based organ equivalents combined with microfluidics. The prospects of such approaches, their impact on the field of alternative approaches, and necessary complementary activities are discussed. The need to adapt quality assurance measures and experiences from validation is stressed.

Keywords: human on a chip, biological warfare, chemical warfare, countermeasures, alternative methods

Introduction

Countermeasures evaluation is the new kid on the block of alternative approaches. And it is a big kid that appeared suddenly: Over the last six months three funding opportunities in the US totaled more than $ 200 million, all with the goal of advancing the *Human on a Chip* concept. An alliance of US agencies is tackling the problem of evaluating drugs for which there are no patients and, hopefully, never will be patients. Three different calls from NIH, FDA, and the Department of Defense (DoD) agencies, DTRA (Defense Threat Reduction Agency) and DARPA (Defense Advanced Research Agency), all aim for producing a number of three-dimensional human organ equivalents based on stem cell technology and combining them with microfluidics on a chip. The design criteria for the two consortia sponsored by DARPA/FDA[1], with about $ 35 million each, are very demanding: 10 organs to be maintained for 4 weeks. The projects supported by NIH/FDA[2] will develop individual 3D organ equivalents to opt into the platforms established in the former call. It is the first time that DARPA and NIH have teamed up, with each agency committing $ 70 million. The NIH contribution will come from its director's discretionary "Common Fund" but will be administered through the new National Center for Advancing Translational Sciences (NCATS)[3]. The FDA will advise the agencies on how to meet its requirements for safety and effectiveness as part of their medical countermeasure initiative[4]. DTRA seeks independently[5] "*a platform comprised of in vitro human organ constructs in communication with each other [(i.e., liver, lung, heart, kidney, vasculature, and Blood Brain Barrier (BBB) with neuronal component] that will accurately assess efficacy, toxicity, and pharmacokinetics of drugs in a way that is relevant to humans*."

The activities are prompted by the perceived need to have medical countermeasures (MCM) at hand in case something happens. (Note, however, that the EU does not follow the same path.) The obvious problem is the lack of patients for clinical development, which make a traditional product registration with FDA impossible. The original response was the Animal Rule, i.e., the suggestion to use appropriate animal models instead. In May 2002, FDA issued the final rule *New Drug and*

1 http://www.darpa.mil/NewsEvents/Releases/2011/2011/09/16_DARPA_TO_DEVELOP_PLATFORM_FOR_MORE_EFFECTIVE_TESTING_OF_DRUGS_AND_VACCINES.aspx

2 http://www.nih.gov/news/health/sep2011/od-16.htm

3 http://www.ncats.nih.gov/

4 http://www.fda.gov/EmergencyPreparedness/MedicalCountermeasures/default.htm

5 http://www.prweb.com/releases/DTRA/XCEL/prweb9048655.htm

Biological Drug Products; Evidence Needed to Demonstrate Effectiveness of New Drugs When Human Efficacy Studies Are Not Ethical or Feasible[6] (Kwik et al. 2007). FDA's own summary of the animal rule is reproduced in Box 1. In a nutshell, FDA allows substituting for evidence of efficacy (not safety!) in humans with animal studies if a "*reasonably well understood pathophysiological mechanism for the toxicity … and its amelioration or prevention by the product*" is given, "*effect is demonstrated in more than one animal species*" or "*a single animal species … predicting the response in humans,*" "*study endpoint is … generally the enhancement of survival or prevention of major morbidity*," and "*pharmacokinetics and pharmacodynamics … in animals and humans is sufficiently well understood*."

It appears that the obvious non-fit for purpose of animal models paired with the need to regulate these new products opens doors for new approaches. This is very reminiscent of the introduction of the Limulus assay (LAL) in 1986 by FDA, when the duration of the rabbit assay did simply not allow testing short-lived radiopharmaceuticals. Once introduced, the LAL became broadly applied, making it probably the most successful alternative method to date. Similarly, FDA's new interest in predictive *in vitro* tools for MCM might be a door-opener for the evaluation of drugs in general.

Box 1

FDA Animal Rule Summary[7]

In assessing the sufficiency of animal data, the agency may take into account other data, including human data, available to the agency. Under this rule, FDA can rely on the evidence from animal studies to provide substantial evidence of the effectiveness of these products when:

1. There is a reasonably well understood pathophysiological mechanism for the toxicity of the chemical, biological, radiological, or nuclear substance and its amelioration or prevention by the product;
2. The effect is demonstrated in more than one animal species expected to react with a response predictive for humans, unless the effect is demonstrated in a single animal species that represents a sufficiently well characterized animal model (meaning the model has been adequately evaluated for its responsiveness) for predicting the response in humans;
3. The animal study endpoint is clearly related to the desired benefit in humans, which is generally the enhancement of survival or prevention of major morbidity; and
4. The data or information on the pharmacokinetics and pharmacodynamics of the product or other relevant data or information in animals and humans is sufficiently well understood to allow selection of an effective dose in humans, and it is therefore reasonable to expect the effectiveness of the product in animals to be a reliable indicator of its effectiveness in humans.

All studies subject to this rule must be conducted in accordance with preexisting requirements under the good laboratory practices (21 CFR part 58) regulations and the Animal Welfare Act (7 U.S.C. 2131 et seq.). Safety evaluation of products is not addressed in this rule. Products evaluated for effectiveness under subpart I of part 314 and subpart H of part 601 will be evaluated for safety under preexisting requirements for establishing the safety of new drug and biological products.

The agency believes that the safety of most of these products can be studied in human volunteers similar to the people who would be exposed to the product. FDA recognizes that some safety data, such as data on possible adverse interactions between the toxic substance itself and the new product, may not be available. This is not expected to keep the agency from making an adequate safety evaluation. FDA's procedures and standards for evaluating the safety of new drug and biological products are sufficiently flexible to provide for the safety evaluation of products evaluated for efficacy under 21 CFR subpart I of part 314 and subpart H of part 601. This rule will not apply if product approval can be based on standards described elsewhere in our regulations (e.g., accelerated approval based on human surrogate markers or clinical endpoints other than survival or irreversible morbidity).

Consideration 1: There is no such thing as a sufficiently predictive animal model for countermeasures

The US Department of Defense sponsored a National Academy of Sciences report, *Animal Models for Assessing Countermeasures to Bioterrorism Agents*, published in December 2011[8]. One author (TH) had the privilege of being part of the committee. The key findings of the report are reproduced in Box 2. In a nutshell, neither animal nor alternative methods are available for this purpose, but the committee discouraged the development of further animal models while proposing the exploitation of new alternative approaches.

[6] http://www.gpo.gov/fdsys/pkg/FR-2002-05-31/pdf/02-13583.pdf

[7] Taken from the introduction in the final regulation, New Drug and Biological Drug Products; Evidence Needed to Demonstrate Effectiveness of New Drugs When Human Efficacy Studies Are Not Ethical or Feasible, Federal Register, Vol. 67, No. 105, Friday, May 31, 2002, https://federalregister.gov/a/02-13583. Last accessed, June 21, 2012.

[8] http://dels.nationalacademies.org/Report/Animal-Models-Assessing-Countermeasures/13233

Box 2

Key Findings NAS report *Animal Models for Assessing Countermeasures to Bioterrorism Agents* (NRC, 2011)

- Currently available animal models are imperfect representations of the human-pathogen interaction with several important limitations, such as methodological differences and a lack of sufficient human data and knowledge of the natural history of diseases of interest. However, at this time animal models remain central to the development of countermeasures against biothreats when testing the efficacy of therapeutics or vaccines would otherwise involve exposing human volunteers or warfighters to a potentially lethal or permanently disabling toxic substance or microorganism.
- Because animal models may be imperfect for a specific need and are expensive to employ (they require large numbers of animals and must be used in secure biocontainment facilities), current models should be reevaluated for their limitations as well as their presumed advantages. For example, methodological differences among similar animal models may result in differences in how well those models accurately mirror the human response to infection or treatment. Consequently, expanded collection and analysis of human clinical data from natural infections could help verify and augment the strengths of available models.
- Developing new animal models for biodefense research cannot adequately resolve in a reasonable time frame the limitations of the currently available ones. It would be more useful for Transformational Medical Technologies to support a more thorough qualification of currently available animal models to advance the predictive capacity of animal-derived data than to create new models.
- *In vitro* and *in silico* methods are not yet advanced enough (in part due to the absence of human data) to reliably replace animals in biodefense research on a large scale.
- The Committee suggests that Transformational Medical Technologies undertake an analysis of the discovery, development, and approval process for medical countermeasures to identify:
  - Scientific gaps in terms of utilizing alternative methods to animal models and how to address them
  - Specific areas in which use of in vitro and in silico methods could be sufficient, or an adjunct, to the use of animals
  - Criteria for choosing and utilizing the most suitable technologies to replace animal use in biodefense research in the near future
- Changing the standard practice of animal experimentation where feasible to approximate the clinical course of treatment for humans could provide a more reasonable prediction of the usefulness of countermeasures during the development process.
- Potential advances in knowledge regarding biothreats and medical countermeasures should be weighed against the duration and severity of animal pain and distress.
- A comprehensive strategy to improve the gathering and sharing of data from animal models (and their alternatives) would significantly increase the efficiency and productivity of research into bioterrorism countermeasures as well as improve laboratory animal welfare, if it includes:
  - Compartmentalization – experiments designed to yield information from components of the animals (organs, cells, and systems) rather than data derived from the whole organism;
  - The use of systems biology and in vitro or in silico methods;
  - Systemic collection of, and access to, experimental data;
  - Publication of negative results;
  - Enhanced collection and analysis of human data;
  - Added clinical veterinary care.
- Where possible, Transformational Medical Technologies should encourage efforts to replace nonhuman primates as the animal of choice in biodefense research. Such efforts, coupled with unhindered access to data and publishing of all results – even negative ones – help ensure that these data are beneficial, animals are used judiciously, and unnecessary duplication of work is avoided.

This judgment is remarkable, coming unanimously from such an esteemed institution as the National Academies. It extends the earlier report, *Toxicity Testing for the 21st Century: A Vision and a Strategy*[9], to drug development, in the sense that, after taking stock of the state of the art of animal-based evaluations, a call is made for novel approaches based on today's biotechnology. Though the committee had no mandate to discuss the animal rule, it effectively called for the development of a new approach. Similarly, we might apply this thinking to challenge current animal-based drug development more generally. This is quite in line with the ongoing devaluation of preclinical drug evaluation, giving rise to explorative human trials (Robinson, 2007; Coleman, 2011), orphan drug development (Bashaw and Fang, 2012), microdosing (Garner, 2005; Boyd and Lalonde, 2007; Lappin and Garner, 2008), etc.

A critical point of comparison for the evaluation of MCM is the regular drug development process to understand the probabilities of arriving, finally, at a safe and efficient drug. The complete process – from drug discovery to FDA approval – takes an average of 10 to 15 years and costs more than $ 1 billion (Mundae and Ostör, 2010; Tamimi and Ellis, 2009; Gilbert et al., 2003). In some estimates, when the costs of failed prospective drugs are factored in, the cost of a single drug development

[9] free PDF available at: http://www.nap.edu/catalog.php?record_id=11970

has soared from $ 1.1 billion in 1995 to $ 1.7 billion in 2002. The figures are not very different for biopharmaceuticals compared to small molecules (DiMasi and Grabowski, 2007).

Typically, MCM can be developed until phase I clinical trials, so the point of comparison needs to be the clinical development phase and its success/attrition rate. Approximately 8% of drugs that now enter phase I studies eventually become FDA approved products, compared to 14% in the eighties[10]. The success rates from the first study in humans to launch are now <10% (Peck, 2007), according to the C.M.R. International 2006/7 Pharmaceutical R&D Factbook cited. The attrition rate in phase II is now more than 70% and rising, and even in phase III one-third to half the molecules fail (Kola and Landis, 2004; DiMasi, 2001a)[11]. Obviously, recent biomedical research breakthroughs have not improved our ability to identify successful candidates.

The main causes of failure in the clinic include safety problems (about 20%) and lack of effectiveness (about 40%), both predicted by a series of animal models before entering the most costly part of drug development. The inability to predict these failures before human testing or early in clinical trials dramatically escalates costs. In the infectious disease area, data from the ten biggest drug companies during 1991-2000 showed a success rate of about 15%, while the average of all indications was 11% (DiMasi and Grabowski, 2007). Similarly, DiMasi et al. (2010) showed a success rate for systemic infectious disease of 15.6% during 1994 and 2003. Noteworthy, from 1981 to 1992 the success rate of anti-infective drugs was 28.1% (DiMasi, 2001b). Overall, biopharmaceuticals appear to have the higher success rate (all indications) of 30.2% (Gilbert et al., 2003).

A key question, then, is whether countermeasures to bioterrorism have a higher likelihood of success in a (theoretical) human trial? A number of aspects actually argue against this, as summarized by one author (TH) for the NAS report:

– The type of diseases are peracute systemic infections, most closely related to sepsis patients, a disease entity most notorious in failing clinical trials (see below).
– The pathogenesis of these rare or even unknown diseases is little known to guide the development process.
– The pathogens are likely to be optimized to stand interventions, e.g., by introduced antibiotic resistance.
– The clinical setting is likely one of mass infection, possibly combined with other threats, hardly comparable to the randomized clinical trials of hospitalized patients.
– The biosafety levels and the strong reliance on non-human primates limit the number of animal studies.
– Most development took place with less than average development expenditure by entities not experienced with full clinical development of drugs.

It must be concluded that a success rate of normal drug development can only be used as a best-case scenario. Note that this predictivity is not that of a single animal model but rather of the combination of all efforts of preclinical development.

Sepsis, as an uncontrolled systemic infection with high mortality, is a clinical condition most closely reflecting the clinical features of a bioterroristic agent's clinical picture. Buras et al. summarized their value as follows: "*Animal models have been developed in an effort to create reproducible systems for studying sepsis pathogenesis and preliminary testing of potential therapeutic agents. However, demonstrated benefit from a therapeutic agent in animal models has rarely been translated into success in human clinical trials*." Buras et al. (2005) also give reasons for the difficulties of the bacterial infection models used.

"*Caveats of sepsis animal models…*
Bacterial infection
- *Requires growth and quantification of bacteria prior to administration*
- *Significant inter-laboratory variability*
- *Large quantity of bacteria used may elicit confounding toxicosis response*
- *Host response is dependent on infecting bacterial strain*
- *Different host responses with infection of different compartments*
- *Variable host response dependent on bacterial load and infusion time*
- *Genetic background affects host responses to specific pathogens*
- *Human therapy potentially withheld could detract from validity of therapeutic agent*

Variability factors…
Bacterial infection models
- *Bacterial load*
- *Route of administration*
- *Timing of infusion*
- *Bacterial strain*
- *Host strain*
- *Antibiotics/fluid resuscitation*"

Similarly, Opal and Cross (1999) summarize: "*It has become painfully evident that animal models provide misleading and overly optimistic estimates of the survival benefit of specific antisepsis drugs when compared to clinical efficacy in actual human sepsis*." The clinical studies in sepsis following the successful completion of preclinical animal models was summarized by Opal and Cross (1999): "*Since 1982, the field of clinical sepsis research has suffered a long series of failed clinical trials that studied new treatment strategies for septic shock. Over 13,000 patients have been enrolled in at least 23 multicenter, double-blind, placebo-controlled, clinical trials in an attempt to develop new treatments for septic shock. Unfortunately, the results have been generally disappointing and interspersed with some*

[10] Food and Drug Administration, Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products (March 2004), p. 8. http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm113411.pdf

[11] http://www.fda.gov/NewsEvents/Speeches/ucm053444.htm?utm_campaign=Google2&utm_source=fdaSearch&utm_medium=website&utm_term=Lester%20M.%20Crawford,%20D.V.M.,%20Ph.D.%20-%20Healthcare%20Institutional%20Investor%20Conference&utm_content=2

spectacular failures and unexpected toxicities." Opal updated these data for 2009 in his presentation to the committee reporting that 42 clinical studies led to 39 cases of lack of effect, one small effect, and two cases where the situation of the patients worsened (Christaki et al., 2011).

In conclusion, it must be assumed that the likelihood of success in humans of countermeasures to bioterrorism would be considerably lower than the average 11% success rate of drugs at a similar stage of development. There is no evidence that any additional animal model would improve such a success rate. This is especially apt if this animal model represents a repetition of a model used during the development process: In fact, it would be rather unlikely that a most promising animal model would not have been used during the development process and left for the final phase of the clinical trial equivalent.

Consideration 2: Complexity versus simplicity in modeling complex biology

The *Human on a Chip* concept aims for the combination of human stem cell-based 3D organ systems to be combined in a microfluidic platform. This combines a number of desirable features:

1) Organotypic coculture of cells: The concept of mirroring the complexity of the organism in the test models is appealing. The coculture of cells allows for cellular interactions and mutual influences for cell development and differentiation. However, achieving natural proportions of cell types and tissue architecture is a major challenge.
2) 3D systems: The third dimension adds another physiologic component. While traditional cell cultures have only 1% of the cell density of tissue and less than 10% of normal cell-cell-contacts (Hartung, 2007a), 3D models reflect the tissue situation but pose a problem of nutrient and oxygen supply in the absence of a circulatory system and hemoglobin.
3) Perfusion: While traditional cell culture does not achieve homeostasis but is characterized by repeated exchange of culture media with an interim decrease in nutrients and accumulation of waste products, cell perfusion culture can maintain stable conditions. However, stable conditions are not yet physiological conditions and, typically, large culture media volumes are needed. If recirculating the medium, volumes can be reduced but stability of media composition is reduced if no regulated nutrient supply and excretion systems are built in. One author (TH) had early experiences in two EU-funded consortia for perfused cultures of kidney and liver (Koppelstaetter et al., 2004; Jennings et al., 2004). These taught us the advantages of such systems but also the difficulties in standardization, long-term maintenance, throughput, etc.
4) Human stem-cell based systems: It is a common assumption that human cells reflect human reactivity better than animal cells. Though experimental evidence is rather limited, it is certainly wise to work with cells from one species, preferably humans. Studies in human *versus* mouse bone marrow (Pessina et al., 2003) suggest that human reactivity is indeed reflected on an *in vitro* level to an extent that species differences *in vivo* can be estimated. In the same vein, the MEIC study, which tested 50 reference compounds in 61 *in vitro* assays, showed the best predictivity of human lethal plasma concentrations with human cells (Ekwall, 1999). Stem cells promise to be a resource of human quasi-primary cells, but we should keep in mind that we do not currently have protocols for final differentiation of most cell types (with the exception of cardiomyocytes), and the generation of pure cell types is still beyond reach.
5) Chip-based systems: They combine both miniaturization (few cells and little test agent) and opportunities for continuous measurements (Khetani and Bhaita, 2006; Ni et al., 2009; Zhang et al., 2009; Gupta et al., 2010; Shintu et al., 2012). Promising examples exist (van Vliet et al., 2007; Huh et al., 2010; Robinette et al., 2011), some of which use functional endpoints to predict hazards. At this juncture, however, it is mainly electrophysiology that allows continuous measurement but limits us to neurons and cardiomyocytes. Unspecific measurements such as impedance are on the way, but continuous functional markers are rare.

As appealing as a combination of so many good things is ("*Too much of a good thing is wonderful*," Mae West), the question emerges whether the enormous efforts to create such systems are necessary to predict human reactions. Challenges include:

– Final differentiation of stem cells to the different cell types of the organ, including drug metabolism capacity.
– Finding compromise cell culture conditions to maintain all the different cells and organ equivalents.
– Balancing organ equivalent sizes and the perfusion liquid compartment to allow close to physiological kinetics.
– Continuous supply with nutrients and oxygen as well as extraction of waste products.
– Biocompatible materials for all cell types supporting but not interfering with differentiation.
– Sufficiently noninvasive measurements for the small amounts of cells on a chip and the kinetics of the test substance.

This is not just an engineering challenge but also a standardization and reproducibility challenge. A key lesson from the validation of alternative methods (Hartung, 2010a) is the pivotal role of this challenge, and usually this requires simplicity. The more variables in a system, the more standardization is required and the more opportunities for introducing variability.

Consideration 3: Three things count for the new tools for MCM: quality, quality, and quality

Cell cultures are prone to artifacts (Hartung, 2007a): Far too many artificially chosen and difficult to control conditions influence our experiments. Quality assurance is the gift from alternative methods to the life sciences. This blunt statement might be challenged by those involved with Good Laboratory Practice (GLP) or ISO quality assurance. However, while GLP (at least

originally) addressed only regulatory *in vivo* studies, and ISO guidance is not really specific for life science tools, neither addresses the key issue, i.e., the relevance of a test. This is the truly unique contribution of validation, which is far too rarely applied in other settings.

The limited applicability of GLP to *in vitro* studies was first addressed in an ECVAM workshop in 1998 (Cooper-Hannan et al., 1999). Parallel initiatives involving one author (TH) (1996 in Germany and 1999 in Bologna at the Third World Congress on Alternatives and Animal Use in the Life Sciences) led to a declaration toward Good Cell Culture Practice – GCCP (Gstraunthaler and Hartung, 1999):

> "*The participants … call on the scientific community to develop guidelines defining minimum standards in cell and tissue culture, to be called Good Cell Culture Practice … should facilitate the interlaboratory comparability of in vitro results … encourage journals in the life sciences to adopt these guidelines*..."

A GCCP task force was then established, which produced two reports (Hartung et al., 2002; Coecke et al., 2005). The maintenance of high standards is fundamental to all good scientific practice, and it is essential for ensuring the reproducibility, reliability, credibility, acceptance, and proper application of any results produced. The aim of GCCP is to reduce uncertainty in the development and application of *in vitro* procedures by encouraging the establishment of principles for the greater international harmonization, standardization, and rational implementation of laboratory practices, nomenclature, quality control systems, safety procedures, and reporting, linked, where appropriate, to the application of the principles of Good Laboratory Practice (GLP). GCCP addresses issues related to:

– Characterization & maintenance of essential characteristics
– Quality assurance
– Recording
– Reporting
– Safety
– Education and training
– Ethics

The GCCP documents formed a major basis for a GLP advisory document by OECD for *in vitro* studies (OECD, 2004), which addresses:

– Test Facility Organization and Personnel
– Quality Assurance Program
– Facilities
– Apparatus, Materials, and Reagents
– Test Systems
– Test and Reference Items
– Standard Operating Procedures
– Performance of the Study
– Reporting of Study Results
– Storage and Retention of Records and Materials

Therefore, both guidance documents have a lot in common: Inherent variation of *in vitro* test systems calls for standardization, and both the GLP advisory document and the GCCP guidance are intended to support best practice in all aspects of the use of *in vitro* systems, including the use of cells and tissues. When comparing GLP and GCCP, there also are some major differences: GLP still gives only limited guidance for *in vitro*. GLP cannot normally be implemented in academia on the grounds of costs and lack of flexibility. GCCP, on the other hand, also aims to give guidance to journals and funding bodies. Note that guidance also has been developed for the publication of *in vitro* journal articles (Leist et al., 2010). A CAAT workshop was held in March 2012 in San Francisco, and a taskforce was formed to further this work.

All quality assurance of an *in vitro* system starts with its definition and standardization, which include:

– A definition of the scientific purpose of the method
– A description of its mechanistic basis
– The case for its relevance
– The availability of an optimized protocol, including:
  - standard operation procedures
  - specification of endpoints and endpoint measurements
  - derivation, expression, and interpretation of results (preliminary prediction model)
  - the inclusion of adequate controls
– An indication of limitations (preliminary applicability domain)
– Quality assurance measures

The novel types of *Human on a Chip* test will represent additional challenges as to standardization of design and generation of cultures and devices. The systems are considerably more complex than traditional *in vitro* approaches, involving various cell types and engineering.

This standardization forms the basis for formal validation, as developed by ECVAM, adapted and expanded by ICCVAM and other validation bodies, and, finally, internationally harmonized by OECD (OECD, 2005). Validation is the independent assessment of the scientific basis, the reproducibility, and the predictive capacity of a test. It was redefined in 2004 in the Modular Approach (Hartung et al., 2004) but needs to be seen as a continuous adaptation of the process to practical needs and a case-by-case assessment of what is feasible (Hartung, 2007b; Leist et al., 2012). The most important changes to the Modular Approach were: the introduction of an applicability domain (borrowing the concept from QSAR), the use of existing data (retrospective validation), the independence of reproducibility and relevance assessment, allowing leaner study designs and performance standards for similar tests to be considered equivalent to a validated one.

Applicability domain describes the range of test materials to which the test can be applied and reliable predictions obtained (e.g., which chemical class(es), types of products). A later change in, or extension of, the applicability domain might require additional validation work and a new peer-review. While earlier validation studies had not taken into consideration any existing data on a test, the introduction of retrospective validation allowed for their use as sole source or in combination with prospective data generation. Traditional validation studies need a certain number of test items in at least three labs to assess both reproducibility and relevance. Often, however, fewer test items are required to establish reproducibility, allowing the testing of further items in one laboratory only once reproducibility has been confirmed, which can eliminate half the testing with-

out major decrease in statistical power (Hoffmann and Hartung, 2006). Last, the introduction of performance standards, i.e., minimum criteria to be fulfilled for any later test development to prove equivalence to the validated test, represents a key tool to open the market for competing developments while also accommodating changes to the validated protocol without embarking on a full validation study.

It became evident, however, that it is difficult to adapt these schemes to such complex technologies as toxicogenomics (Corvi et al., 2006). Therefore, the concept of applying some tools of evidence-based medicine for validation purposes was put forward (Hartung, 2010b). A key problem for the new technologies, and also for new products like MCM and endpoints, is the absence of a point of reference, i.e., a "traditional test" or "gold standard." For MCM we have neither human data, other drugs for the same purpose, nor established animal tests fit for purpose. A 2008 workshop (Hoffmann et al., 2008) discussed similar issues for toxicology, identifying three types of reference points: reference method/results where available; expert consensus to establish a putative point of reference where data are ambiguous and incomplete; and cases of no point of reference methods, such as latent class analysis. As laid out earlier (Hartung, 2010b), in the absence of reference data, the scientific validation needs to be stressed. Figure 1 shows the classical validation scheme and its adaptation to such situations. The process of defining the point of reference for MCM evaluation will be very challenging. It should be started early enough, as it will guide test development, but it

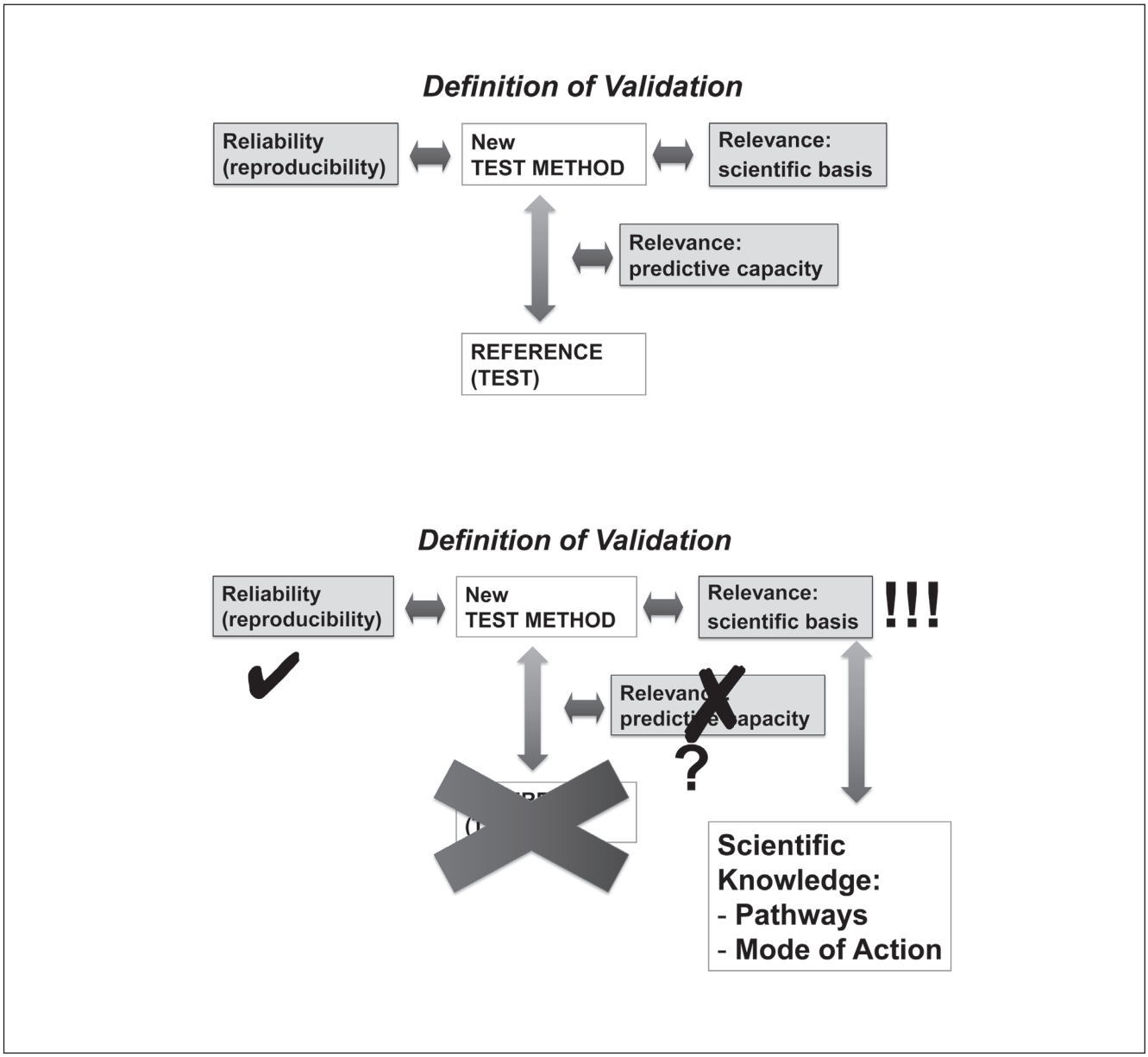

Fig. 1: The traditional validation scheme and its adaptation to situations without reference test

is important not to use up all reference information during the design phase of an assay.

The framework of evidence-based medicine is increasingly being translated to toxicology (Hoffmann and Hartung, 2006), and it recently led to the creation of the Evidence-based Toxicology Collaboration (EBTC)[12] (Zurlo, 2011).

Consideration 4: The investment in superior *in vitro* models will promote toxicology and pharmacology for the 21st century, even if it does not result in a routine predictive tool

The convergence of new technologies, the needs of the pharmaceutical industry, and the goals of regulatory agencies further inform the dialogue that began with the publication of *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC, 2007). The need to understand pathways of toxicity in order to better predict the toxicity of environmental chemicals has now spilled over into drug development and safety assessment of pharmaceuticals and biologicals, as well as the assessment of the efficacy of countermeasures to biological and chemical terrorism agents. The need to develop approved pharmaceutical and biological interventions without clinical trials poses a unique challenge. The added benefit in pursing this challenge is that the *in vitro*, human-based, mechanistically-oriented organ-on-a-chip or combined to *Human on a Chip* and other test systems developed for that purpose will also serve to meet the needs of the other disciplines that currently rely on animal tests to predict human responses.

Our discussions for the last few years have centered very much on toxicology for the 21st century (Tox-21c). However, neither is toxicology the key problem of MCM (safety assessments including phase I trials can actually be done in a very conventional way, with the possible caveat that biologicals such as vaccines and immune response modifiers are not really testable for side effects or excess-pharmacology in animals), nor is pharmacology short of novel tools. Whatever gives a cutting edge to drug development is typically applied. However, so far the new tools have been used to convince the management to make certain business decisions, rather than to make the regulators admit a product to the market. For this reason, the very same tools require some independent endorsement so that regulators can feel comfortable in basing decisions on them. This will hold true whether we talk about animal models or alternative approaches. For this reason, MCM open up opportunities for Pharma-21c, i.e., embracing novel technologies for regulatory decisions on the likely efficacy of drug candidates.

Creating organotypic cell models is not necessarily the most convincing approach to produce credible data given the artificial nature of the constructs and the degree of innovation required to design them. We could as well imagine approaches of integrated testing strategies (ITS) or pathway-based approaches, as suggested in the roadmap for systemic toxicity testing (Hartung and McBride, 2011; Basketter et al., 2012) to produce more predictive systems. The last article in this Food for thought … series (Hartung et al., 2012) laid out the vision of a systems toxicology. Similarly, a systems pharmacology (van der Greef and McBurney, 2005; Berger and Iyengar, 2011; Hansen et al., 2012) can be envisioned and is currently emerging to model drug interventions in the organism. In both cases it will be necessary to convince the regulatory community to base decisions on this novel type of data, which will be best achieved by scientific rigor and the continuous exposure to new evidence from opinion leaders and market forces.

A very interesting opportunity lies in making use of the new organotypic models for pathway identification. If the improved culture conditions boost relevance of the models, the pathway identification in these models should be even more relevant.

Conclusions

The new interest in predictive *in vitro* systems in the US will revamp the field of alternative approaches. The substantial funding opportunities will bring researchers and engineers into the field. It will be important to acquaint them with the lessons learned over the last two decades of developing predictive models and their quality assurance. Toxicology has served as a pilot, but all areas of life sciences have similar needs, and drug development can benefit in similar ways.

While the *Human on a Chip* approach is not the only way to construct novel predictive test strategies, it complements approaches based on integrated testing strategies or pathway-based approaches as they are mainly pursued in toxicology (Tox-21c).

The validation process as defined originally by ECVAM has been proven to work. The ECVAM principles on validation were taken up by ICCVAM (USA) and internationally by the OECD in GD 34 on the validation and international acceptance of new or updated test methods for hazard assessment. They can be reasonably translated to drug development purposes. The validation process is in constant evolution, and MCM will need such adaptations. Validation of tests is an essential quality assurance process, similar to EBM. The Evidence-based Toxicology Collaboration (EBTC) promises to be a tool for validation of 21st century methods.

References

Bashaw, E. D. and Fang, L. (2012). Clinical pharmacology and orphan drugs: an informational inventory 2006-2010. *Clin Pharmacol Ther 91*, 932-936.

Basketter, D. A., Clewell, H., Kimber, I., et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX 29*, 3-89.

[12] http://ebtox.com/

Berger, S. I. and Iyengar, R. (2011). Role of systems pharmacology in understanding drug adverse events. *Wiley Interdiscip Rev Syst Biol Med 3*, 129-135.

Boyd, R. A. and Lalonde, R. L. (2007). Nontraditional approaches to first-in-human studies to increase efficiency of drug development: will microdose studies make a significant impact? *Clin Pharmacol Ther 81*, 24-26.

Buras, J. A., Holzmann, B., and Sitkovsky, M. (2005). Animal models of sepsis: setting the stage. *Nat Rev Drug Discov 4*, 854-865.

Christaki, E., Anyfanti, P., and Opal, S. M. (2011). Immunomodulatory therapy for sepsis: an update. *Expert Rev Anti Infect Ther 9*, 1013-1033.

Coecke, S., Balls, M., Bowe, G., et al. (2005). Guidance on Good Cell Culture Practice. *Altern Lab Anim 33*, 261-287.

Coleman, R. A. (2011). Efficacy and safety of new medicines: a human focus. *Cell Tissue Bank 12*, 3-5.

Cooper-Hannan, R., Harbell, J., Coecke, S., et al. (1999). The principles of Good Laboratory Practice: application to in vitro toxicology studies. *Altern Lab Anim 27*, 539-577.

Corvi, R., Ahr, H. J., Albertini, S., et al. (2006). Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. *Environ Health Perspect 114*, 420-429.

DiMasi, J. A. (2001a). Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther 69*, 297-307.

DiMasi, J. A. (2001b). Risks in new drug development: Approval success rates for investigational drugs. *Clin Pharmacol Ther 69*, 297-307.

DiMasi, J. A. and Grabowski, H. G. (2007). The cost of biopharmaceutical R&D: Is biotech different? *Manage Decis Econ 28*, 469-479.

DiMasi, J. A., Feldman, L., Seckler, A., et al. (2010). Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther 87*, 272-277.

Ekwall, B. (1999). Overview of the final MEIC results: II. The in vitro-in vivo evaluation, including the selection of a practical battery of cell tests for prediction of acute lethal blood concentrations in humans. *Toxicol In Vitro 13*, 665-673.

Garner, R. (2005). Less is more: the human microdosing concept. *Drug Discov Today 10*, 449-451.

Gilbert, J., Henske, P., and Singh, A. (2003). Rebuilding big pharma's business model, in vivo. *The Business & Medicine Report, Windhover Information 21*, No. 10.

Gstraunthaler, G. and Hartung, T. (1999). Bologna declaration toward Good Cell Culture Practice. *Altern Lab Anim 27*, 206.

Gupta, K., Kim, D-H., Ellison D., et al. (2010). Lab-on-a-chip devices as an emerging platform for stem cell biology. *Lab Chip 10*, 2019-2031.

Hansen, J., Zhao, S., and Iyengar, R. (2012). Systems pharmacology of complex diseases. *Ann N Y Acad Sci 1245*, E1-E5.

Hartung, T., Balls, M., Bardouille, C., et al. (2002). Report of ECVAM task force on good cell culture practice (GCCP). *Altern Lab Anim 30*, 407-414.

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

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

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

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

Hartung, T. (2010b). Evidence based-toxicology – the toolbox of validation for the 21st century? *ALTEX 27*, 241-251.

Hartung, T. and McBride, M. (2011). Food for thought … on mapping the human toxome. ALTEX 28, 83-93.

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

Hoffmann, S., Edler, L., Gardner, I. et al. (2008). Points of reference in validation – the report and recommendations of ECVAM Workshop. *Altern Lab Anim 36*, 343-352.

Hoffmann, S. and Hartung, T. (2006). Designing validation studies more efficiently according to the modular approach: retrospective analysis of the EPISKIN test for skin corrosion. *Altern Lab Anim 34*, 177-191.

Huh, D., Matthews, B. D., Mammoto, M., et al. (2010). Reconstituting organ-level lung functions on a Chip. *Science 328*, 1662-1668.

Jennings, P., Koppelstaetter, C., Pfaller, W., et al. (2004). Assessment of a new cell culture perfusion apparatus for in vitro chronic toxicity testing. Part 2: Toxicological evaluation. *ALTEX 21*, 61-66.

Khetani, S. and Bhatia, S. N. (2006). ScienceDirect.com – current opinion in biotechnology – Engineering tissues for in vitro applications. *Curr Opin Biotechnol 17*, 524-531.

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

Koppelstaetter, C., Jennings, P., Ryan, M. P., et al. (2004). Assessment of a new cell culture perfusion apparatus for in vitro chronic toxicity testing. Part 1: Technical description. *ALTEX 21*, 51-60.

Kwik, G. G., Trent, D., Borio, L., et al. (2007). The FDA animal efficacy rule and biodefense. *Nature Biotechnology 25*, 1084-1087.

Lappin, G. and Garner, R. C. (2008). The utility of microdosing over the past 5 years. *Expert Opin Drug Metab Toxicol 4*, 1499-1506.

Leist, M., Efremova, L., and Karreman, C. (2010). Food for thought ... considerations and guidelines for basic test method descriptions in toxicology. *ALTEX 27*, 309-317.

Leist, M., Hasiwa, M., Daneshian, M., et al. (2012). Validation and quality control of replacement alternatives – current status and future challenges. *Toxicol Res*. doi: 10.1039/C2TX20011B

Mundae, M. K. and Ostör, A. J. (2010). The long road of biopharmaceutical drug development: from inception to marketing. *QJM 103*, 3-7.

NRC (2007). *Toxicity Testing in the 21st Century – A Vision and a Strategy*. Washington, DC, USA: The National Academies Press.

NRC – National Research Council, Committee on Animal Models for Assessing Countermeasures to Bioterrorism Agents (2011). *Animal Models for Assessing Countermeasures to Bioterrorism Agents* (1-153). Washington, DC, USA: The National Academies Press. http://dels.nationalacademies.org/Report/Animal-Models-Assessing-Countermeasures/13233

Ni, M., Tong, W., Choudhury, D., et al. (2009). Culture on MEMS platforms: A review. *Int J Mol Sci 10*, 5411-5441.

OECD (2004). Advisory document of the working group on GLP – the application of the principles of GLP to in vitro studies. *Series on Principles of Good Laboratory Practice and Compliance Monitoring 14*, 1-18.

OECD (2005). Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. *OECD Series On Testing And Assessment 34*, 1-96.

Opal, S. and Cross, A. (1999). Clinical trials for severe sepsis – past failures, and future hopes. *Infect Dis Clin North Am 13*, 285-297.

Peck, R. W. (2007). Driving earlier clinical attrition: if you want to find the needle, burn down the haystack. Considerations for biomarker development. *Drug Discov Today 12*, 289-294.

Pessina, A., Albella, B., Bayo, M., et al. (2003). Application of the CFU-GM assay to predict acute drug-induced neutropenia: an international blind trial to validate a prediction model for the maximum tolerated dose (MTD) of myelosuppressive xenobiotics. *Toxicol Sci 75*, 355-367.

Robinson, W. (2007). Innovative early development regulatory approaches: expIND, expCTA, microdosing. *Clin Pharmacol Ther 83*, 358-360.

Robinette, B. L., Harrill, J. A., Mundy, W. R., et al. (2011). In vitro assessment of developmental neurotoxicity: Use of Microelectrode arrays to measure functional changes in neuronal network ontogeny1. *Front Neuroeng 4*. doi: 10.3389/fneng.2011.00001

Shintu, L., Baudoin, R., Navratil, V., et al. (2012). Metabolomics-on-a-chip and predictive systems toxicology in microfluidic bioartificial organs. *Anal Chem 84*, 1840-1848.

Tamimi, N. A. M. and Ellis, P. (2009). Drug development: From concept to marketing! *Nephron Clin Pract 113,* c125-c131.

van der Greef, J. and McBurney, R. N. (2005). Innovation: Rescuing drug discovery: in vivo systems pathology and systems pharmacology. *Nature reviews. Drug Discov 4*, 961-967.

van Vliet, E., Stoppini, L., Balestrino, M., et al. (2007). Electrophysiological recording of re-aggregating brain cell cultures on multi-electrode arrays to detect acute neurotoxic effects. *Neurotoxicology 28*, 1136-1146.

Zhang, C., Zhao, Z., Abdul Rahim, N. A., et al. (2009). Towards a human-on-chip: culturing multiple cell types on a chip with compartmentalized microenvironments. *Lab Chip 9*, 3185-3192.

Zurlo, J. (2011). Evidence-based toxicology collaboration kick-off meeting. *ALTEX 28*, 152.

Correspondence to

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