



# Food for Thought ... on Evidence-Based Toxicology

Thomas Hartung

Johns Hopkins University, Bloomberg School of Public Health, Dept. Environmental Health Sciences, Doerenkamp-Zbinden Chair for Evidence-based Toxicology, Center for Alternatives to Animal Testing, Baltimore, USA

I would like to devote the first article of this series that I am writing after my move to the US to the topic that has become my chair's designation, i.e. evidence-based toxicology. This topic is really close to my heart. It all began in 1993, when my friend Edmund "Edi" Neugebauer co-edited the book "Handbook of Mediators in Septic Shock" (Neugebauer and Holaday, 1993). To the best of my knowledge, this was the first book to apply principles of evidence-based medicine (EBM) not only to clinical studies but also to animal studies and *in vitro* experimental work. So, when I started at ECVAM in 2002 and was developing ideas for possible directions to take, my list included evidence-based toxicology, i.e. the translation of EBM (for an introduction to this see, for example, Mayer, 2004<sup>1</sup>) to toxicology, as a most interesting option. I was most fortunate that a Ph.D. student of mine at that time, Sebastian Hoffmann, who is a statistician by training, not only joined me to go from Konstanz to ECVAM, but agreed to refocus his thesis under my supervision to developing the concepts of an evidence-based toxicology (EBT). His thesis "Evidence-based *in vitro* toxicology," submitted in January 2005 to the University of Konstanz (Hoffmann, 2005), is the first extensive publication on EBT, as far as I am aware. Sure, there were a few previous attempts to link EBM and toxicology (Buckley and Smith, 1996), especially in the discussion around the toxicity of amalgam (Dodes, 2001). Also, Phil Guzelian and co-authors (Guzelian et al., 2005) independently developed a concept for EBT, though they took a different approach, focusing on causation and not on method evaluation (see below).

One might ask whether these "Food for Thought" articles are not the opposite of evidence-based. They are by purpose personal, provocative, not really peer-reviewed, not evaluated by statistics, broad in scope, etc. That is correct, but they also do not pretend to produce new knowledge but try to challenge common beliefs and stimulate new thinking. By this, they might, however, occasionally initiate a systematic review in the spirit of EBM and EBT. With this disclaimer in mind, let's get into evidence-based science.

## Consideration 1: EBM tries to solve some problems that are strikingly similar to those of toxicology

In a nutshell, EBM was born from the need to somehow handle the flood of information in medicine and to sort the available evidence in an objective manner, which includes traditional approaches and new scientific developments of variable quality. More than half a million papers included in MedLine per year of an estimated more than 2 million in medicine every year (Hunt, 1997) address questions relevant to the life sciences and therapy – no way could individual physicians overlook all this information. Overall use of information retrieval systems occurs just 0.3 to 9 times per physician per month, whereas physicians have 2 unanswered questions for every 3 patients (Hersh and Hickham, 1998). "It is astonishing with how little reading a doctor can practice medicine, but it is not astonishing how badly he may do it." (Sir William Osler, 1849-1919). MedLine (<http://www.ncbi.nlm.nih.gov/pubmed/>), the most popular resource, covers over 5,000 journals with an estimated 8,000 citations entered per week (Wilczynski et al., 2004). For example, entering the search term "toxicology" for the time period since 2003 results in 28,500 article hits in PubMed, a database not even covering all relevant publications in the biomedical field.

Instead of expecting individuals to determine what is the best evidence for a specific question or approach at a given time, high-quality reviews available at a central repository should represent a primary resource of information. This requires agreed quality standards, so that the individual physician can rely on the information received. And here the key difference between evidence-based and traditional ("narrative") reviews sets in: Leaving aside conflicts of interest, which might impact on any article in toxicology (Claxton, 2007), most reviews represent a story told by (knowledgeable) authors who present their personal views on their topic of interest in a more or less well disguised manner. They tend to select their own papers and those that fit the story line of their review. The literature included is largely what has been accumulated over time and shaped the

<sup>1</sup> Notably, Greenhalgh in 2005 estimated 500 textbooks and 15,000 journal articles devoted to different angles of the basics of EBM.



opinion of the author(s). This is sometimes distinguished from EBM as “expert-based” or “eminence-based,” to make clear that individual experts are speaking here. The systematic review, which is the first main tool of EBM, proceeds differently: The sources (typically MedLine and other literature databases) and a search strategy, (which decides which papers shall be considered and which not), are defined upfront. Before collecting the actual articles, the procedure for information analysis is defined. Ideally, these search and analysis strategies are peer-reviewed to safeguard objective and efficient processes. The analysis of the collected evidence requires weighing the quality (the second main tool of EBM) of individual pieces of evidence and summarising these as objectively as possible. The latter often involves the third major tool of EBM: meta-analysis. Meta-analysis describes statistical approaches to combine results from different studies. These studies will differ in key parameters. By either factor analysis or stratification of all data by one parameter after the other, the influential parameters can be identified. Then, where possible, an overall quantitative answer to the well-defined research question is deduced.

Obviously, toxicology has a similar problem of information flooding and coexistence of traditional and modern methodologies, as well as various biases (Wandall et al., 2007). It is most difficult to find and summarise the relevant information for any given major question. This has been nicely illustrated by Christina Rudén (Rudén 2001a; 2001b): She showed the divergence in judgment and limitations of analysis for 29 cancer risk assessments carried out for trichloroethylene – 4 assessments concluded that the substance is carcinogenic, 6 said it is not, and 19 were equivocal. The main reason for this divergence was a selection bias in the materials considered, i.e. an average reference coverage of only 18%, an average citation coverage of most relevant studies of 80%, as well as an interpretation difference of most relevant studies in 27%, and the lack of study/data quality assessment not documented in 65% of the assessments.

The similarities between the problems of toxicology and clinical medicine, and especially the similarities between making a diagnosis in medicine and deciding on whether a substance is hazardous (Hoffmann and Hartung, 2005), prompted us to think about whether EBM tools could be suitable for toxicology. The term EBT was coined, which led to some misunderstandings, such as, “We have always used evidence!”, Sure – as have physicians when treating their patients, but this evidence was to a large extent the result of subjective collection and interpretation. Often, the standardisation and formalisation of processes and committees disguises the nature of our decision-making. Some people correctly speak of the “art of toxicology” (though it is more a craft) – this much better reflects the intuitive and individual components. Certainly, it is an applied science, in which compromise and pragmatic decisions are necessary, but we should be clear on where we have to take such shortcuts, otherwise we will soon forget the limitations of our decisions and make them gold standards, textbook knowledge, and the unquestioned basis for further decisions (read-across, QSAR, new use and exposure scenarios, reference for validation, etc.). Many of the now highly respected test guidelines for methods

or classifications and labelling of substances are based on GOB-SAT (“good old boys sat around a table”).

A major step toward the formation of an EBT movement was the first International Forum toward an Evidence-based Toxicology in 2007 ([www.ebtox.org](http://www.ebtox.org)). Unfortunately, the Forum only formulated a declaration (Box 1) and ten defining characteristics (Box 2) of EBT, but not a consensus definition. An interesting starting point for this might be the following translation of the definition of EBM given by Greenhalgh (2006):

“Evidence-based toxicology is the use of mathematical estimates of the harm of agents, derived from high-quality research

### Box 1 The Declaration of Como

We, the undersigned participants of the First International Forum Towards EBT, commit to the further development, refinement and application of Evidence Based Toxicology (EBT) as described in the Defining Characteristics agreed during the Forum.

We invite the scientific community and other stakeholders to join with us in this effort.

Como, October 2007

### Box 2 10 defining characteristics of

- promotes the consistent use of transparent and systematic processes to reach robust conclusions and sound judgments
- addresses societal values and expectations and is socially responsible
- displays a willingness to check the assumptions upon which current toxicological practice is based to facilitate continuous improvement
- recognises the need to provide for the effective training and development of professional toxicologists
- acknowledges a requirement for new and improved tools for critical evaluation and quantitative integration of scientific evidence
- embraces all aspects of toxicological practice, and all types of evidence of which use is made in hazard identification, risk assessment and retrospective analyses of causation
- ensures the generation and use of best scientific evidence
- includes all branches of toxicological science: human health assessment, environmental and ecotoxicology and clinical toxicology
- acknowledges and builds upon the achievements and contributions of Evidence Based Medicine/Evidence Based Health Care
- fosters the integration of expert judgment with best possible external evidence

on methods and groups of substances, to inform decision making on the regulation and use of substances and the treatment of exposed patients.”

**Consideration 2: There are four very different areas of application of EBT: method evaluation, quantitative combination of different studies on a given substance, causation of a health effect, and clinical toxicology adopting EBM**

In fact, some miscommunication occurred at the first EBT forum, because different people had different ideas about what EBT should be about. These four (Fig. 1) have crystallised so far:

- 1) Similar to the evaluation of therapeutic options or even closer diagnostic means (see our article on the similarity of clinical diagnostics and toxicology (Hoffmann and Hartung, 2005)), different tools of toxicology need to be evaluated to identify their usefulness, their limitations, and to compare options. This is somewhat in the realm of validation and was thus our starting point into EBT, but would in contrast to validation typically not mean setting up studies but analysing existing information.
- 2) The need to combine, possibly quantitatively, information from various sources is a typical problem for regulators. Often one study is identified as the lead study and the others are used as additional information. Many perceive this as unsatisfying, but objective approaches to combine study results are lacking. Meta-analysis as used in the clinical field is most promising here. Ellen Silbergeld’s structured reviews are forerunners for this approach (Navas-Acien et al., 2006; 2007).
- 3) Causation was the starting point for Phil Guzelian and colleagues for suggesting an EBT approach (Guzelian et al., 2005). This addresses the question whether we can trace a certain health effect, back to a toxicant, such as lung cancer to smoking. What are the quality criteria and logical steps for such a proof? This is closely linked to legal arguments (Rodricks, 2006).
- 4) A field that is embarking into EBM quite independently is clinical toxicology, where guidance for the treatment of intoxicated patients, etc., has to be found. There are already some guidance documents, which claim to be EBM-based (Dargan et al., 2002; Wallace et al., 2002). For the purpose of this article, I will not focus on these developments of clinical EBT, but possible synergies and overlaps should be considered for further activities.

Most probably, EBT can well live with all these different spins, but certainly resources are limited and priorities need to be set. It is noteworthy that some collaborators not only want to focus on one aspect but sometimes heavily oppose the others. My key desire to evaluate current test methods (especially animal test methods) in the most objective way to open up the field for new approaches (see below) (Hartung, 2008b; Hartung, 2009b) is not shared by all, since it means challenging current practices and thus the results of former risk assessments. Others challenge the aspect of causation, since this might make it more

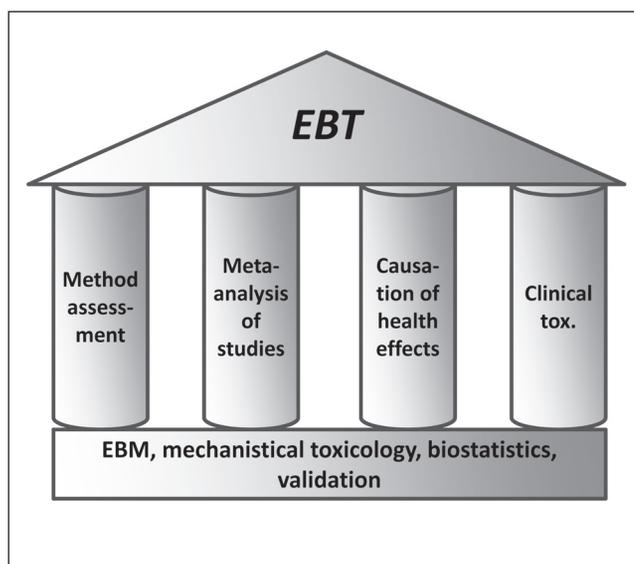


Fig. 1: Building the “temple” of Evidence-based Toxicology

difficult to ban certain substances if we raise standards of how to prove causation – just the opposite of a precautionary approach. The controversy between Christina Rudén (Rudén and Hansson, 2008) and Phil Guzelian centres on this suspicion. It would indeed be detrimental if, for example, the health effects of tobacco smoke were challenged by sophisticated intellectual arguments on available evidence. Noteworthy, already in the first paper (Guzelian et al., 2005) proposing EBT, while suggesting to raise the standards of evidence provided, the authors did argue that this would not affect the judgment on smoking. However, we will need to find the right balance between scientific proof of causation and the need to take protective measures also in the absence of final evidence. Anyway, it would be good to be clear about where we have evidence and where we act in a precautionary manner – otherwise we risk closing the books too soon. If everybody had taken for granted that stomach ulcers are caused by stress and too much acid, *Helicobacter pylori* might never have been identified as a cause.

The idea that EBT might be employed to hinder public health measures is somewhat frightening. It reminds us of health care providers who argued that certain therapies were not EBM-based to refuse their reimbursement. We need to be clear about the golden rule: Absence of evidence is no evidence of absence. There will always be many more things that are true than we can prove (as even proven mathematically by Gödel in his incompleteness theorems, 1931). In a growing scientific field the availability of adequate reviews will always lag behind the generation of new knowledge. And especially those questions not yet addressed are not “non evidence-based” but “not yet evaluated.” And those questions for which insufficient evidence exists to draw a conclusion will have to live with exactly this statement or a preliminary judgment based on expert consensus. This is, by the way, another common misunderstanding: EBM does not exclude consensus – on the contrary, consensus processes are the fourth major tool of EBM – but this needs to be attained by sufficiently transpar-



ent, documented, and objective processes. There is a wealth of literature on these methodologies, such as the Delphi process and the nominal group technique (Fink et al., 1984; Williams and Webb, 1994; Jones and Hunter, 1995).

### Consideration 3: We often confuse weight-of-evidence with evidence-based

The term “weight-of-evidence” is commonly used to describe a process of making a decision based on different pieces of information, each not definitive or even contradictory. In the absence of clear procedures, this is a highly subjective process. In many aspects, this is just the contrary of an evidence-based approach. The term comes from the legal field, where it means the measure of credible proof on one side of a dispute as compared with the credible proof on the other, particularly the probative evidence considered by a judge or jury during a trial (Farlex legal dictionary). The weight of evidence is based on the believability or persuasiveness of evidence. Believability is certainly not an EBM criterion. Certainly this confusion is another problem of adopting terms in common use as scientific terminology – there was some beauty in deriving scientific terms from ancient Greek and Latin: They would, if at all, only after years become commonly used and thereby confused, more broadly interpreted, etc.

Noteworthy, there is also a well-defined approach to weight-of-evidence in the field of Bayesian statistics (Good, 1985), which offers enormous opportunities, but this is not the way the term and the process are used in toxicology.

### Consideration 4: The basic toolbox – we need a scoring tool, meta-analysis and an internet portal

When confronted with information retrieved from literature, etc., we always have the problem of how to assess its quality and relevance for the question to be addressed. While relevance can normally be judged, quality of the evidence is much more difficult to assess. However, exactly this is required in order to

systematically exclude studies of low quality or weigh the different pieces of evidence when combining them.

In EBM, low reliability would, for example, be attributed to case reports, while randomised, controlled multi-centre trials would represent the highest quality achievable. An equivalent might be single experiments reported compared to formal validation studies. Although detailed scoring might be desirable (such as 100 points for the blinded, controlled multi-laboratory study down to 1 point for information of the type “my grandma always said this is dangerous”), this is most difficult to achieve and probably also unnecessary. EBM works very well with a system based on few classes, i.e. 5 levels of evidence with 3 subclasses for level 1 and 2 as well as 2 subclasses for level 3 (<http://www.cebm.net/index.aspx?o=1025>). Actually, a similar system, known as the Klimisch scores, exists in toxicology, (Klimisch et al., 1997). However, the criteria to assign these scores are ill-defined. As a central contribution to EBT we have therefore started a project to develop such criteria with a contractor and an expert advisory panel (Hoffmann, 2008; Schneider et al., 2009). Two sets of criteria for *in vivo* and for *in vitro* studies, respectively, were put forward and tested in two rounds with raters. Though the variability of results leaves room for improvement, it is already a major step in the right direction. The necessary continuation of this project is under discussion, but a sponsor will need to be identified. Hopes lie with the chemical industry, because of the obvious impact of such an approach for REACH: The legislation requires taking into account all existing information on a given substance. Such information can be strengthened and weak information excluded if such criteria are available.

Similarly, methods to combine different studies, like meta-analysis in clinical medicine (Hunt, 1997), are urgently needed in toxicology (Fig. 2). The problem is how to combine studies, which were designed by different people, in different places at different times. Although the British mathematician Karl Pearson already systematically combined different studies as early as 1904, it was only Gene V. Glass, in 1976, who initiated the concept of the meta-analysis (Hunt, 1997). Essentially, this describes the five-step process of (1) formulating the problem,

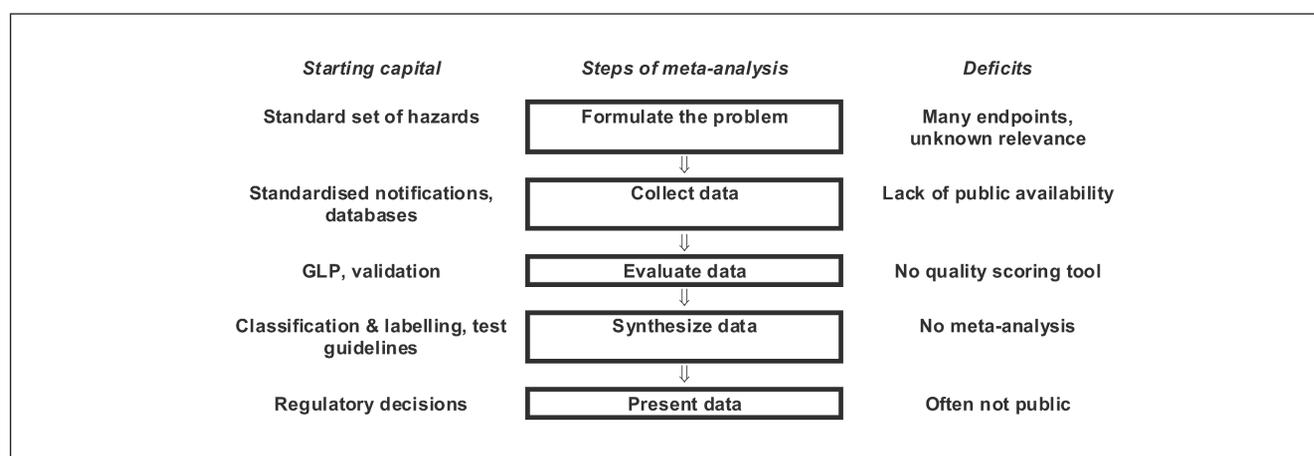


Fig. 2: Application of meta-analysis to toxicology – what do we have in hand and what is lacking?

(2) collecting the data, (3) evaluating the data as to validity and usability, (4) synthesizing the data, and (5) presenting the data.

For (3) scoring tools are typically applied. The synthesis of data (4) is usually done on the basis of probabilities. Here, odds-ratios are typically used: The odds ratio is a measure of effect size, describing the strength of association between two variables (e.g. drug and outcome). It is commonly interpreted as relative risk, i.e. a ratio of the probability of the event occurring in the exposed group versus the non-exposed group., e.g. treated vs. non-treated, treated with drug A vs. drug B, or of a person/animal exposed to a toxin vs. non-exposed. It seems quite easy to adopt this to toxicological studies. Instead of a black/white result, i.e. having a certain hazard or not, we would need to work with probabilities, which means, in essence, expressing the uncertainty of our results. This is first of all the uncertainty of the method itself and then its combination with other methods and its performance in a study. While we do have some statistics with which we can describe the probability of an outcome of a single experiment to be true, we lack them for combinations of different experiments, to bring us to an overall conclusion.

Notably, meta-analysis does not always yield a result, but it often identifies which variable of different studies impacts on the uncertainty and thereby defines what to address in future studies to account for the differences.

In the end, EBM is a marriage of medicine with statistics. It aims to assign numbers (probabilities, odds ratios, effect sizes, significances) to relevant medical questions. Not every physician is ready to embrace this approach. They do not need to go as far as Benjamin Disraeli (1804-1881), who is known for the quote, "There are three kinds of lies: lies, damned lies and statistics." However, many consider the individual patient as not adequately reflected by means and averages from population-based studies. This is a fundamental misunderstanding – general guidance must only be applied if there is no reason to deviate for a specific case. If there is additional information, we must take this into consideration. This will hold true for toxicology as much as it does for clinical medicine. Not all chemicals are equal, but a generalised evidence-base rule is more likely to be correct than assumptions, prejudice, and superstition.

The Cochrane Library (<http://www.cochrane.org/>) solves many of these problems for clinical medicine. Since 1996 (<http://www.update-software.com/publications/Cochrane/history.pdf>), systematic reviews (Mulrow, 1995), prepared and maintained by the Cochrane Collaboration (Dickersin and Manheimer, 1998; Dickersin et al., 2002), have been published in the Cochrane Library, along with bibliographic and quality-assessed material on the effects of healthcare interventions submitted by others. It consists of a regularly updated collection of evidence-based medicine databases, including the Cochrane Database of Systematic Reviews. This database includes systematic reviews of healthcare interventions that are produced and disseminated by the Cochrane Collaboration. The Cochrane Library is published on a quarterly basis and made available both on CD-ROM and the Internet. It is the best single source of reliable evidence about the effects of healthcare. The review abstracts are available free of charge, but the full text is unfortunately only available to subscribers.

As a first activity, we have commissioned a scoping study for an Internet portal (Kinsner-Ovaskainen, 2009). It aims to create a starting point for a platform to build up a library of structured/systematic reviews and to organise EBT activities. The website [www.ebttox.org](http://www.ebttox.org) was already set up for this and serves to inform on the initiative hosted by the Joint Research Centre. However, financing of the further development of the Internet portal will depend on safeguarding the respective funding. No organisation has so far taken on the responsibility of furthering this.

### **Consideration 5: Retrospective validation is a type of EBT**

When confronted with the need to validate a large number of tests for the purposes of the 7<sup>th</sup> amendment and REACH, we discussed opportunities to speed up processes and save resources without impairing the scientific rigour at ECVAM. One opportunity identified and then formalised in the modular approach to test validation (Hartung et al., 2004) was the introduction of retrospective validation. The idea was simple: Most tests entering validation have already been in use for a while – why do we not take into account all this information but start validation studies from scratch as if nothing were known? The term "retrospective" should distinguish this approach from prospective, new studies, borrowing terms from epidemiology and clinical study designs. In addition, we also suggested combining existing data with newly generated data to complete the information needs for validation.

The first and successful test case was the validation of the micronucleus test (Corvi et al., 2008; [http://ecvam.jrc.it/publication/ESAC25\\_statement\\_MNT\\_20061128\\_C.pdf](http://ecvam.jrc.it/publication/ESAC25_statement_MNT_20061128_C.pdf)). Similar to a structured review, criteria for the inclusion and exclusion of existing studies were defined and then applied by the mutagenicity taskforce. Data from the remaining studies were combined, which sometimes required going back to raw data to apply the same data analysis. The success was striking: in only one and a half years and without any additional experimental studies the information needs for validation could be satisfied.

The approach has since then been applied to other areas, such as eye irritation. Data were again collated in a type of meta-analysis, and our colleagues from ICCVAM/NICEATM then attempted their analysis (<http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox.htm>). However, instead of using the increased power of combining studies, they decided on a single protocol and eliminated all data that were obtained with variants of the protocol. This is forgoing the advantage of meta-analysis, where minor variations with no impact on results are ignored (after they are properly checked). The recently completed reanalysis of the data in the sense of a true meta-analysis is an exciting perspective to broaden the use and acceptability of these methods. OECD acceptance for severe irritants is a big step forward these days, but especially the cosmetics industry needs accepted methods for mild irritants for the provisions of the 7<sup>th</sup> amendment (Hartung, 2008c).



### Consideration 6: The problem of data availability

There is a key difference between EBM and EBT: There is enormous pressure to publish results in clinical medicine, especially of clinical trials, where, for example, major journals have agreed on registration and data availability ([http://www.icmje.org/clin\\_trial.pdf](http://www.icmje.org/clin_trial.pdf)). This is very different in toxicology, where we have little incentive for publication of regulatory test results (Hartung, 2008a): Data are considered proprietary, most are negative anyway – typically not very attractive for publication. REACH will make a lot of this information publicly available in the future – a leap ahead for the development of modelling approaches or the validation of alternative methods.

It is difficult to understand why this information is withheld at all. Should we not be entitled to know about possible health hazards of the chemicals we are exposed to? We might need a “freedom of information act on substances we are exposed to.” The standard argument is that this is sensitive business information, but what makes it sensitive? Competitors can only use it when regulators accept the data source. Thus, if not intended by the respective legislation (mandatory data-sharing), control of abuse of public information is easy. Anyway, customers should have the right to know the basis for letting a product enter the market.

However, such considerations do not change the principal problem that a lot of extremely valuable knowledge on toxicity tests is not in the public domain. Thus, voluntary or legislative efforts to make data available will be critical to the success of EBT. The databases of the US National Toxicology Program ([http://ntp-apps.niehs.nih.gov/ntp\\_tox/index.cfm](http://ntp-apps.niehs.nih.gov/ntp_tox/index.cfm)) and the US Library of Congress (ToxLine or ToxNet, <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLIN>) are crucial for this, and there is no European match so far.

### Consideration 7: The more complex toxicology becomes, the more we need EBT

Centralised evaluation of methods and the provision of objective tools become ever more important, since overlooking the field of toxicology is getting more and more difficult: More people, more (types of) substances, more health concerns, more methods, more legislation – who can keep track?

- More people: The large programmes like REACH are bringing a new generation of (often inexperienced) toxicologists into the process. No education scheme has foreseen that Europe will need several hundred toxicologists for the chemical agency, national regulatory agencies and industry.
- More (types of) substances: Not only existing chemicals are in the focus of REACH and other programmes, but the acceleration of chemical synthesis strategies or new types of products (biologicals, nanoparticles, genetically modified organisms, cell therapies, medical devices, etc.) broaden the scope of safety evaluations (Hartung and Koeter, 2008; Hartung and Leist, 2008; Leist et al., 2008).
- More health concerns: Research is constantly putting forward new possible health threats: endocrine disruption, respiratory

sensitisation, developmental neuro- and immunotoxicity, antigenicity, and viral safety of cell therapies.

- More methodologies: The life sciences are driven by technological developments, such as organotypic cell cultures, molecular biology, analytical chemistry, computer modelling, etc. They all offer promise for new approaches in regulatory toxicology but will require standardisation, validation, accumulation of experience and regulatory frameworks to interpret them. This becomes further complicated when old or new methodologies are integrated into testing strategies and when mechanistic toxicology is embraced.
- More legislation: Not only Europe is pushing for novel health and environmental safety regulations (food, pesticides, cosmetics, GMO, chemicals, drugs, etc.). In global economies, we have to keep track of developments in other relevant economic areas as well (Bottini et al., 2007; Bottini and Hartung, 2009).

Central, trustworthy evaluations of methods and approaches represent a key resource to counter this increasing complexity. EBT can therefore make a major contribution to the feasibility of programmes if it stays out of the political compromises, which, for example, limit the role of the OECD in promoting new scientific approaches. Scientific rigour and transparency are the benchmarks for becoming the repository of best evidence available for a given question.

### Consideration 8: EBT as the door opener to a new toxicology

If, for a moment, we leave aside all the difficulties of attaining the new approaches, we should ask, why should it be easier to get these accepted in regulatory frameworks than it was for alternative methods of the past? They should undergo validation for sure, but here the problem starts: if we continue validating against current practices, we will not really move ahead (Hoffmann et al., 2008). We will add or, at best, replace a patch of the patchwork of toxicology. This does not allow us to overcome inherent limitations, since we define the methods of the past as our gold standards, which must be met.

The big risk is that, even when the usefulness of new approaches can be established, for example showing that certain substances missed or misclassified can suddenly be detected properly, the new methods will be considered to be valuable additional information, but not as substitution of the traditional approach. Certainly, for a while, methods will need to coexist, but this should be done with the clear understanding that after a certain time period a decision is going to be taken on whether to replace or not.

These are only a few of the challenges lying ahead (Hartung, 2009a). EBT holds much promise, but only if the shortcomings of current approaches are shown in a most objective way – ruling out any major doubt on the quality of the assessment – will we have a basis for validating against something else and substituting something better. We have discussed earlier in this series that there are psychological and economic forces at work, quite apart from the limitations of current approaches (Bottini

and Hartung, 2009). The only non-disputable basis for change is sound science – the conceptual framework the different players have been trained on and are committed to.

### Consideration 9: How to gain a critical mass?

The essential question for an open movement like EBT is whether a sufficiently high number of colleagues is willing to invest in the future. Only if a larger number of high quality studies are made available to attract further contributions, a portal that promises quality can develop and therefore be used. Something like Wikipedia<sup>2</sup>, but with quality control as the essential component, because this constitutes the appeal of EBT. And this is exactly the problem: quality has its price. It is essential that some sponsors be found right at the starting phase of such a project. This might then create a chain reaction. The recent creation of our Transatlantic Think Tank of Toxicology (t<sup>4</sup>, see ALTEX 4/2008, page 361), with the main goal of furthering high-quality studies such as systematic reviews, is a first but small step in this direction. We will have to show that we will be able to develop the quality assurance necessary to find acceptance by the scientific community without creating barriers for contributions.

We had a good start with the First International Forum Towards Evidence-Based Toxicology in Como, but one and a half years have passed since then, and we are only now publishing the proceedings (Griesinger et al., 2009). Nobody is to blame for this, considering all the restructuring within the main sponsor, ECVAM, but we have lost our momentum. Claudius Griesinger, Agnieszka Kinsner-Ovaskainen, and Sandra Coecke, the remaining members of the original EBT team, do their best to keep the ball rolling, but they need support from their hierarchy to maintain this key role for EBT. The initiative has been embraced by EuroTox, the SOT, and others. The EBT section in the journal *Human and Experimental Toxicology*, which already published the two initial papers (Guzelian et al., 2005; Hoffmann and Hartung, 2006), is a good start; a special issue of the journal *Toxicology* (co-edited by Claudius Griesinger and Alan Goldberg) is another. It will require a number of ambassadors to spread the idea. For sure, we at the new Doerenkamp-Zbinden Chair for Evidence-based Toxicology at Johns Hopkins University are committed to moving this idea ahead, to bringing in our accumulated experience with alternative methods, and to making a contribution to a real paradigm shift in toxicology. The fact that JHU also hosts the US Center of the Cochrane Collaboration, which helped with EBT from the start, is a big opportunity. The new toxicity testing strategy of US EPA published on 25 March 2009 (<http://www.epa.gov/osa/spc/toxicitytesting/>, see also this issue of ALTEX, page 149) is the basis for a paradigm-shift in toxicology, and I hope that the understanding that we cannot create something new

without learning our lessons from the past will grow in our scientific field. “Those who cannot remember the past are condemned to repeat it”<sup>3</sup> – this quote from George Santayana, a Spanish-born American author of the late nineteenth and early twentieth centuries, drives the point home.

### References

- Bottini, A. A., Amcoff, P. and Hartung, T. (2007). Food for thought ... on globalization of alternative methods. *ALTEX* 24, 255-261.
- Bottini, A. A. and Hartung, T. (2009). Food for thought ... on economic mechanisms of animal testing. *ALTEX* 26, 3-16.
- Buckley, N. A. and Smith, A. J. (1996). Evidence-based medicine in toxicology: where is the evidence? *Lancet* 347, 1167-1169.
- Claxton, L. D. (2007). A review of conflict of interest, competing interest, and bias for toxicologists. *Toxicol. Industr. Health* 23, 557-571.
- Corvi, R., Albertini, S., Hartung, T. et al. (2008). ECVAM Retrospective Validation of in vitro Micronucleus Test (MNT). *Mutagenesis* 23, 271-283.
- Dargan, P. I., Wallace, C. I. and Jones, A. L. (2002). An evidence based flowchart to guide the management of acute salicylate (aspirin) overdose. *Emerg. Med. J.* 19, 206-209.
- Dickersin, K., Manheimer, E., Wieland, S. et al. (2002). Development of the Cochrane Collaborations's central register of controlled clinical trials. *Evaluat. Health Professions* 25, 38-64.
- Dickersin, K. and Manheimer, E. (1998). The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials. *Clin. Obstet. Gynecol.* 41, 315-331.
- Dodes, J. E. (2001). The amalgam controversy – an evidence-based analysis. *J. Am. Dent. Assoc.* 132, 348-356.
- Fink, A., Kosecoff, J., Chassin, M. and Brook, R. H. (1984). Consensus methods: characteristics and guidelines for use. *Am. J. Public Health* 74, 979-983.
- Good, I. J. (1985). Weight of evidence: a brief survey. *Bayesian Stat.* 2, 249-270.
- Greenhalgh, T. (2006). *How to read a paper – the basics of evidence-based medicine*. 3rd edition. Oxford: Blackwell Publishing Ltd.
- Griesinger, C., Hoffmann, S., Kinsner-Ovaskainen, A. et al. (2009). Foundations of an Evidence-Based Toxicology. Proceedings of the First International Forum Towards Evidence-Based Toxicology. Conference Centre Spazio Villa Erba, Como, Italy. 15-18 October 2007. *Human Exp. Toxicol.* (in press).
- Guzelian, P. S., Victoroff, M. S., Halmes, N. C. et al. (2005). Evidence-based toxicology: a comprehensive framework for causation. *Human Exp. Toxicol.* 24, 161-201.

<sup>2</sup> Actually, toxipedia does exist (<http://toxipedia.org/>), but is not linked to principles of EBT.

<sup>3</sup> I also like the first part of the quote, which made less sense in our context: “History is a pack of lies about events that never happened told by people who weren't there.”



- Hartung, T. (2009a). Fundamentals of an evidence-based toxicology. *Human Exp. Toxicol.* (in press).
- Hartung, T. (2009b). A toxicology for the 21st century: Mapping the road ahead. *Tox. Sci.* (in press).
- Hartung, T. (2008a). Towards a new toxicology – evolution or revolution? *ATLA* 36, 635-639.
- Hartung, T. (2008b). Food for thought ... on animal tests. *ALTEX* 25, 3-9.
- Hartung, T. (2008c). Food for thought ... on alternative methods for cosmetics safety testing. *ALTEX* 25, 147-162.
- Hartung, T. and Koeter, H. (2008). Food for thought ... on alternative methods for food safety testing. *ALTEX* 25, 259-264.
- Hartung, T. and Leist, M. (2008). Food for thought ... on the evolution of toxicology and phasing out of animal testing. *ALTEX* 25, 91-96.
- Hartung, T., Bremer, S., Casati, S. et al. (2004). A Modular Approach to the ECVAM principles on test validity. *ATLA* 32, 467-472.
- Hersh, W. R. and Hickam, D. H. (1998). How Well Do Physicians Use Electronic Information Retrieval Systems? A Framework for Investigation and Systematic Review. *JAMA* 280, 1347-1352.
- Hoffmann, S., Edler, L., Gardner, I. et al. (2008). Points of reference in validation – the report and recommendations of ECVAM Workshop. *ATLA* 36, 343-352.
- Hoffmann, S. (2008). Development of a scoring tool to assess inherent quality of toxicological data. Abstracts of the 45th Congress of the European Societies of Toxicology. *Toxicol. Lett.* 180 Suppl. 1, S18.
- Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Human Exp. Toxicol.* 25, 497-513.
- Hoffmann, S. and Hartung, T. (2005). Diagnosis: Toxic! – Trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Tox. Sci.* 85, 422-428.
- Hoffmann, S. (2005). Evidence-based in vitro toxicology, dissertation, University of Konstanz.
- Hunt, M. M. (1997). *How Science Takes Stock: The Story of Meta-Analysis* (1st edition). New York: Russell Sage Foundation Publications.
- Jones, J. and Hunter, D. (1995). Qualitative Research: Consensus methods for medical and health services research. *Brit. Med. J.* 311, 376-380.
- Kinsner-Ovaskainen, A., Griesinger, C., Hoffmann, A. et al. (2009). An online portal to evidence-based toxicology. *Human Exp. Toxicol.* (in press).
- Klimisch, H.-J., Andreae, M. and Tillmann, U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. *Regulat. Toxicol. Pharmacol.* 25, 1-5.
- Leist, M., Hartung, T. and Nicotera, P. (2008). The dawning of a new age of toxicology. *ALTEX* 25, 103-114.
- Mayer, D. (2004). *Essential evidence-based medicine* (1-381). Cambridge: Cambridge University Press.
- Mulrow, C. (1995). Rationale for systematic reviews. *Brit. Med. J.* 309, 597-599.
- Navas-Acien, A., Silbergeld, E. K., Streeter, R. A. et al. (2006). Arsenic exposure and type 2 diabetes: a systematic review of the experimental and epidemiological evidence. *Environ. Health Perspect.* 114, 641-648.
- Navas-Acien, A., Guallar, E., Silbergeld, E. K. and Rothenberg, S. J. (2007). Lead exposure and cardiovascular disease—a systematic review. *Environ. Health Perspect.* 115, 472-482.
- Neugebauer, E. A. and Holaday, J. W. (1993). *Handbook of Mediators in Septic Shock* (1-608, 1st edition). Florida, USA: CRC-Press, Boca Raton.
- Rodricks, J. V. (2006). Evaluating disease causation in humans exposed to toxic substances. *J. Law Policy*, 39-63. [https://www.brooklaw.edu/students/journals/bjlp/jlp14i\\_rodricks.pdf](https://www.brooklaw.edu/students/journals/bjlp/jlp14i_rodricks.pdf)
- Rudén, C. and Hansson, S. O. (2008). Evidence-based toxicology: “sound science” in new disguise. *Int. J. Occup. Environ. Health* 14, 299-306.
- Rudén, C. (2001a). Interpretations of primary carcinogenicity data in 29 trichloroethylene risk assessments. *Toxicol.*, 169-225.
- Rudén, C. (2001b). The Use and Evaluation of Primary Data in 29 Trichloroethylene Carcinogen Risk Assessments. *Regulat. Toxicol. Pharmacol.* 34, 3-16.
- Schneider, K., Schwarz, M., Burkholder, I. et al. (2009). “ToxR-Tool”, a new tool to assess the reliability of toxicological data. (submitted).
- Wallace, C. I., Dargan, P. I. and Jones, A. L. (2002). Paracetamol overdose: an evidence based flowchart to guide management. *Emerg. Med. J.* 19, 202-205.
- Wandall, B., Hansson, S. O. and Rudén, C. (2007). Bias in toxicology. *Arch. Toxicol.* 81, 605-617.
- Wilczynski, N. L. and Haynes, R. B. for the Hedges Team (2004). Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Medicine* 2, 23-27.
- Williams, P. L. and Webb, C. (1994). The Delphi technique: a methodological discussion. *J. Adv. Nursing* 19, 180-186.

### Acknowledgements

I would like to express my gratitude to those who have undertaken with me the first steps toward EBT – especially my former team Sebastian Hoffmann, Claudius Griesinger, Agnieszka Kinsner-Ovaskainen, and Sandra Coecke – but also the friends and colleagues on both sides of the Atlantic who got us started.

### Correspondence to

Prof. Thomas Hartung  
 Johns Hopkins University  
 Bloomberg School of Public Health  
 Center for Alternatives to Animal Testing  
 615 N. Wolfe St. E7035  
 Baltimore, MD, 21205, USA  
 e-mail: THartung@jhsph.edu