Clear path: towards an evidence-based toxicology (EBT)
PS Guzelian, MS Victoroff, C Halmes and RC James

Proceedings of the 1st International Forum Towards Evidence-Based Toxicology

Greeting
CL Galli

Preface
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

The Scientific Steering Committee
Declaration of Como

Provisional Definition of an Evidence-Based Toxicology

1 Fundamentals of an evidence-based toxicology
1.1 Opening statement
T Hartung

1.2 Aspects of test assessment
S Hoffmann

1.3 Consensus, opinion, and evidence-based science – three methods of reaching conclusions in toxicology
P Guzelian

1.4 Comparing medicine with toxicology – a mapping of knowledge creation, concepts and basic epistemology
C Griesinger

2. Evidence-based medicine – a possible model for evidence-based toxicology?

2.1 Translation of evidence-based medicine into practice
EAM Neugebauer

2.2 Evidence-based health care and the Cochrane Collaboration
RW Scherer

3. Core problems and case studies
3.1 Key challenges of toxicology
ME Cebrián

3.2 Quantification of uncertainty within and between species and the role of uncertainty factors
L Edler

3.3 Trovafloxacin: a case study of idiosyncratic or iatrogenic liver toxicity – molecular mechanisms and lessons for pharmacotoxicity
J Borlak

4. Toxicological decision-making on hazards and risks – status quo and way forward
4.1 Current concepts and schemes of science-driven toxicological decisionmaking – an overview
C Portier

4.2 Applying an evidence-based approach: arsenic as a health risk
E Silbergeld

4.3 In-vitro genotoxicity tests to detect carcinogenicity: a systematic review
L Müller

4.4 The proposed replacement of the no observed adverse effect level with benchmark dose levels in food risk assessment
L Edler

4.5 Evidence-based individual toxicological analysis
P Guzelian

4.6 Toxicogenomics and biology-based modeling framework for health risk assessment
D Sarigiannis

4.7 Biological modeling as a method for data evaluation and integration in toxicology
HA Barton

4.8 Current schemes for decision-making in toxicology
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

4.9 Current information sources for hazard identification
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

5 Steps toward an evidence-based toxicology
5.1 Evidence-based tools in toxicological basic research
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

5.2 Evidence-based tools in toxicological hazard identification
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

5.3 Evidence-based tools in toxicological decision-making
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

5.4 Possible improvement of information sources on hazard and risk
C Griesinger, S Hoffmann, A Kinsner, S Coecke and T Hartung

5.5 Standardization: an efficient tool to support evidence-based toxicology
A Pirlet

5.6 An online portal to evidence-based toxicology
A Kinsner-Ovaskainen, C Griesinger, S Hoffmann, S Coecke, G Bowe, C Campana and T Hartung

6 Conclusions
I Kimber

7 Annex

7.1 Suggested Charge Questions for Break-out group work

http://het.sagepub.com
Human & Experimental Toxicology

Scope The Editors will consider papers on all aspects of experimental and clinical studies of functional, biochemical and structural disorders. Attention is paid to experimental work and studies of basic toxicological and biological mechanisms and to clinical reports of important cases.

Human & Experimental Toxicology is indexed by Biological Abstracts, BIOSIS, Cambridge Scientific Abstracts, Chemical Abstracts, CSA Neurosciences Abstracts, Current Citations, Current Contents, Current Contents Life Sciences, Elsevier BIOBASE/Current Awareness in Biological Sciences, EMBASE/Excerpta Medica, Index Medicus, Index Veterinaris, ISI Alerting Services, Medical Documentation Service, Nutrition Abstracts and Reviews Series A: Human & Experimental, Pollution Abstracts with Indexes, Research Alert, Science Citation Index, Science Citation Index Expanded (also known as Scisearch), Social Sciences Citation Index.

Editorial Manuscripts (original plus 2 copies) and all editorial correspondence should be sent to:

UK, Europe and Far East:
Professor Kai Savolainen
Editor-in-Chief
Human & Experimental Toxicology
Finnish Institute of Occupational Health
Department of Industrial Hygiene and Toxicology
Topeliuksenkatu 41 aA
00250 Helsinki
Finland
E-mail: leila.ahlstrom@ttl.fi

Americas:
Dr A. Wallace Hayes
Editor for the Americas
Human and Experimental Toxicology
Department of Environmental Health
Harvard School of Public Health
298 South Main Street
Andover, MA 01810
USA
Tel: +1(978) 749-3085
Fax: +1(978) 409-1154
E-mail: awallacehayes@comcast.net

Publisher
All business correspondence, advertisement enquiries, supplement enquiries, reprint orders and permission requests should be addressed to Human & Experimental Toxicology, Sage Publications, 1 Oliver's Yard, 55 City Road, London, EC1Y ISP.

WWW
Human & Experimental Toxicology’s homepage is www.sagepublication.com. Editor’s details, scope, subscription prices, reprint ordering, sample copy ordering, supplement information and contact names for the publishing, production, marketing and advertising departments are available.

Subscriptions

<table>
<thead>
<tr>
<th>Subscription Type</th>
<th>US</th>
<th>Europe</th>
<th>Rest of the World</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institutional</td>
<td>$1,141</td>
<td>£796</td>
<td>£803</td>
</tr>
<tr>
<td>Individual</td>
<td>$560</td>
<td>£390</td>
<td>£393</td>
</tr>
</tbody>
</table>

Members of the British Toxicology Society and the Society of Toxicology are entitled to a reduced individual subscription rate. All subscriptions include print copies and online access. For 2007 orders email subscription@sagepub.co.uk

Orders and payment
Correspondence concerning subscriptions and back issues should be addressed to: SAGE, 1 Oliver's Yard, 55 City Road, London EC1Y ISP, UK. Tel: +44 (0)20 7324 8500; Fax: +44 (0)20 7324 8600; Email: subscriptions@sagepub.co.uk; website: www.sagepublications.com

Periodicals postage is paid at Rahway, NJ.

USA Postmaster: please send address corrections to Human & Experimental Toxicology, c/o Mercury Airfreight International Ltd Inc, 365 Blair Road, Avenel, New Jersey NJ 07001.

Copyright © 2007 SAGE Publications, Sage Publications, 1 Oliver’s Yard, 55 City Road, London, EC1Y ISP.

All rights reserved. No part of the publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without the prior permission of SAGE Publications.

UK Apart from any use permitted under UK copyright law, this publication may only be reproduced, stored or transmitted, in any form, or by any means with prior permission in writing of the publishers or in the case of reprographic production in accordance with the terms of licences issued by the Copyright Licensing Agency.

USA Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by SAGE Publications, provided that the base fee of $20.00 per copy of each article is paid directly to the Copyright Clearance Center Inc, 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy licence by CCC, a separate system of payment has been arranged. The fee code for users of the Transactional Reporting Service is: [0960-3271/06 $20.00].

For all other use, permission should be sought from the publisher.

SAGE Publications is a member of the HINARI programme http://healthinternetwork.net/

Organised by WHO, HINARI enables developing countries to gain access to one of the world’s largest collections of biomedical and health literature. Over 2900 journal titles are now available to health institutions in 113 countries, benefiting many thousands of health workers and researchers, and in turn, contributing to improved world health. By choosing to publish in Human & Experimental Toxicology your paper will be available to developing countries who may otherwise not have had access.
Human & Experimental Toxicology

Editor-in-Chief
Professor Kai Savolainen
Human & Experimental Toxicology
Finnish Institute of Occupational Health, Department of Industrial Hygiene and Toxicology,
Topeliuksenkatu 41 aA, 00250 Helsinki, Finland
E-mail: leila.ahlstrom@ttl.fi

Editor for the Americas
Dr A. Wallace Hayes
Harvard School of Public Health
298 South Main Street,
Andover, MA 01810, USA
Tel: +1(978) 749-3085
Fax: +1(978) 409-1154
E-mail: awallacehayes@comcast.net

Field Editor for Evidence-based Toxicology
Philip Guzelian, M.D.
University of Colorado
USA

Editorial Board
Dr S.M. Barlow, Independent Consultant, UK
Professor Nursen Basaran, Hacettepe University, Turkey
Dr D.N. Bateman, Royal Infirmary of Edinburgh, UK
Dr Barbara D. Beck, Gradient Corporation, USA
Dr Linda S. Birnbaum, US Environmental Protection Agency, USA
Dr Bas J. Blaauwboer, Utrecht University, The Netherlands
Professor Alan Boobis OBE, Imperial College, London, UK
Dr Paul Borm, Heinrich-Heine Universitaet Dusseldorf, Germany
Professor Edward J. Calabrese, University of Massachusetts, USA
Dr Daniel A. Casciano, National Center for Toxicological Research, USA
Professor J.K. Chipman, University of Birmingham, UK
Dr Emanuela Corsini, Università degli Studi di Milano, Italy
Professor E.E Creppy, Université de Bordeaux, France
Dr Michael L. Dourson, Toxicology Excellence for Risk Assessment, USA
Professor Erik Dybing, Norwegian Institute of Public Health, Norway
Dr S.J.S. Flora, Defence Research & Development Establishment, India
Dr Sidney Green, Howard University College of Medicine, USA
Professor Helmut Greim, Technical University of Munich, Germany
Dr Mary Gulumian, NIOH, South Africa
Professor Philip S. Guzelian, University of Colorado, USA
Dr E.S. Harpur, Sanofi-Synthelabo Research, USA
Dr Kenneth Hartigan-Go, University of the Philippines, Philippines
Dr P.W. Harvey, Covance Laboratories, UK
Dr Salmaan H. Inayat-Hussain, Universiti Kebangsaan Malaysia, Malaysia
Dr Thomas A. Lewandowski, Gradient Corporation, USA
Dr Lionel D. Lewis, Dartmouth Hitchcock Medical Center and Dartmouth Medical School, USA
Dr T.C. Marrs, Independent Consultant, UK
Dr John O’ Brien, Danone Vitapole, France
Professor B.K. Park, University of Liverpool, UK
Dr C. Pease, Unilever Colworth, UK
Professor Chada S. Reddy, University of Missouri, USA
Professor Ivonne M.C.M. Rietjens, University of Wageningen, The Netherlands
Professor Tetso Satoh, Chiba University, Japan
Dr B. Schilter, Nestle Research Centre, Switzerland
Dr Roel Schins, IUF - Environmental Health Research Institute, Germany
Dr Bryan K. Shipp, Pfizer Inc., USA
Dr G.B. van der Voet, Leiden University Medical Centre, The Netherlands
Dr William J. Waddell, University of Louisville, USA
Professor F.A. de Wolff, Leiden University Medical Centre, The Netherlands
Clear path: towards an evidence-based toxicology (EBT)

Evidence-based medicine (“EBM”), a decision-making process that uses the best available objectively assessed knowledge as a substitute for authority-based opinions, has transformed the practice of medicine. EBM has been hailed as one of the 15 greatest medical milestones since 1840.1 EBM is now regularly taught in medical schools and has been accepted by most major health-related organizations, worldwide.2 As a method to reduce the effects of bias and excessive reliance on expert judgment to make decisions, evidence-based logic (“EBL”) has been adopted by many nonmedical disciplines, including businesses, engineering, baseball,2 and even advertising.3

Recognizing the opportunities for use of EBL in toxicology, we developed a comprehensive framework for an evidence-based toxicology (“EBT”).2 Our article explored the principles of causation determination, showing how use of EBM/EBL could identify evidence-based causation conclusions for toxicology and, thus, improve risk communication. Contemporaneously, Hoffman and Hartung4 proposed use of a quantitative EBT to develop statistical and probabilistic methods for analyzing and validating the predictive value of toxicity tests. We are gratified that the interest generated by these two articles led to the first International Forum Towards Evidence-Based Toxicology, held in October 2007. Proceedings of the First International Forum Towards Evidence-Based Toxicology, Conference Centre Spazio Villa Erba, Como, Italy, 15–18, October 2007. The proceedings are presented in this issue.5 Attending this three-day meeting made it clear that many participants had limited familiarity with EBT, and yet its logic and its intended transparency seemed promising. Much of the commentary was encouraging. However, at least two almost reflexive and somewhat emotional objections were vocalized by some attendees. These were as follows:

1) Toxicologists already use “evidence” in reaching causation decisions and regulatory decisions. Acknowledgment of EBT might linguistically imply that pre-EBT toxicological opinions had not been formed with the use of evidence and scientific reasoning. In short, EBT is nothing new except the terminology.

2) Toxicologists often do not have available human experimental data (e.g., randomized clinical trials), as is common with EBM. Therefore, requiring actual experimental results for toxicology is too demanding and unrealistic. We should continue to reach conclusions as we always have (derived from a lesser standard of evidence) because we are doing the best we can do.

For those familiar with EBM and EBT, the simple answer to the first concern is that while all authoritative (expert) opinions are based at least in part on evidence (unless they are pure speculation), not all opinions are evidence-based conclusions. Sometimes, toxicologists or a regulatory agency like the United States Environmental Protection Agency (USEPA), just like some physicians, formulate an opinion and cite selected scientific data in support.2,6 EBT is the reverse. EBT conducts a complete and transparent rule-based analysis of the data and then formulates the conclusion, one that should be reproducible by others familiar with the method. The need for EBT seems clear. Toxicologists can be irrational in their assessment of chemical risks.7 It has been reported that toxicology causation decisions reached in the recent past did not exhibit such evidence-based characteristics. See for example Refs. 8–13.

The second concern, lack of definitive required data in toxicology, stands the matter on its head. Causal conclusions are not dimensional entities adjustable by accommodating them to the availability of evidence. Data should be judged by the standard, and not vice versa. When there are sufficient amounts and types of evidence, we may be able to conclude that we have discovered a scientific fact. In some

Correspondence to: PS Guzelian (Clinical Professor of Medicine), University of Colorado Health Sciences Center, Clinical Toxicology (Private Office), Centennial, CO 80112-3500, USA. Email: phil.guzelian@uchsc.edu
fields, such as the social sciences or economics, limited opportunity for experimentation inherently makes it difficult to reach cause and effect conclusions or to quantify outcomes about complex processes.\footnote{First, Dr Rudén’s forebodings were accompanied with charges of our philosophic dishonesty, linguistic manipulations to deceive, secretly conspiring with the tobacco industry to dismantle expert judgment in regulatory activity so as to endanger public health, and inappropriate commingling of medicine with toxicology. None is true. Second, she declined both a public invitation\footnote{Dr Rudén is not a physician and, according to her background and publication record, has little experience with making contributions to original basic science research. Perhaps Dr Rudén did not read our paper completely, inasmuch as she omits mention of large sections that readily address virtually all of the arguments she advances. Many attendees at the International Forum Towards Evidence-Based Toxicology meeting found useful information when directed to appropriate sections of Guzelian, et al., 2005\textsuperscript{2}.} and a personal invitation offered by the HET field editor to submit an article or commentary on EBT. Instead, Dr Rudén chose to publish her commentary in the International Journal of Occupational and Environmental Health, a publication that now communicates little in the way of scientific find-

laced with a number of errors of fact and reasoning, self-contradictions, and contains, inappropriately for scientific discussion, personal polemics. Her evident errors of understanding many parts of our article may be explainable.\footnote{Dr Rudén’s “good science” sounds quite similar to the phrases “sound science” and “Good Epidemiological Practices,” which she vilifies because she claims they are associated with the tobacco industry.} Less obvious is why she seems to be quite unfamiliar with (or willing to disregard) her own work. We cited Dr Rudén for being the first author we encountered to show that toxicological risk assessments and causation determinations (whether affirmative or negative) by various scientific agencies did not always follow “good science”\textsuperscript{c} principles. Examining the evidence in 29 published “cancer risk assessments” done for the same chemical (trichloroethylene) by authoritative groups, Dr Rudén showed that the datasets used were varyingly incomplete and concluded that the reviews were biased in their data selection, data interpretation, and data evaluation.\textsuperscript{8–10} Her stinging criticism of toxicology for producing inconsistent, uncritical, and incomplete causation analyses for various chemicals, for example,\textsuperscript{8–12,23–25} is not easily reconciled with her agitated denouncement of EBT, which offers a remedy for the very flaws she had uncovered in her review of these 29 “risk assessments.”

Still, at least some of Dr Rudén’s misconceptions\textsuperscript{22} also surfaced at discussions held at the International Forum Towards Evidence-Based Toxicology (though presented more decorously). Therefore, we address salient (but by no means all) of Dr Rudén’s admonishments about EBT to give additional clarifications to the principles we presented previously.\textsuperscript{26}

To begin with, Rudén and Hansson simply cannot seriously object that our use of the term “epistemic” to mean “known” is somehow misleading, when their own definition is “of or relating to knowledge...
or degree of acceptance”. More important, among “nomological possibilities,” we included knowledge of the substance/disease causal relationship (epistemic) as being a necessary feature of a toxicologic “risk” because that is how “risk” is understood in common parlance. For example, bankruptcy from embezzlement is a risk, a traffic accident from bus driving is a risk, hepatitis from consuming 15 g of acetaminophen is a risk, but being abducted by a space alien is not a risk. Notice that for these three risks, the predicted frequency of the occurrence of the harm in the future can be readily estimated from the scientific observations that established the identify of the risk in the first place (i.e., its causal relationship to a stimulus).

Strangely enough, Dr Rudén herself attempts to differentiate between known hazards and presumed hazards for human health based on animal data or mechanistic data in her proposed Cancer Risk Assessment Index (CRAI) system; she further differentiates between known hazards and presumed relationship to a stimulus).

The use of animal data is based on the assumption that toxic and carcinogenic effects seen in mammalian species are relevant to humans” [emphasis added]. Finally, she has noted that such “uncertainties” are inherent to variations and biases in the collection and interpretation of data. So, although perhaps more philosophically grounded, our designation of risks and uncertainties seems to correspond rather nicely with Dr Rudén’s favored phraseology.

Still, we understand that risk assessors’ argot defines “risk” differently, by referring to all nomological possibilities as chemical harms. Therein lies the problem for risk communication. To forestall chemical harms to the public, risk assessors often simply assume as true the underlying substance/disease causal relationship for all doses under consideration even when the data do not permit such knowledge (i.e., the proposition is uncertain). They assume that some “unknown risks” (to use the language of Rudén and Hansson) are true (i.e., they are known). However, the frequency of such an assumed risk is unknown and cannot be predicted. Rather than best estimates, the common risk assessment “risk calculations” actually are attempts to identify a range of doses at which the range of possible frequencies of occurrence of an adverse effect is acceptably small, if not zero. Indeed, when tested empirically with human epidemiology, “risk calculations” in regulatory risk assessments from animal data are not predictive of human health. In short, risk assessments derived from animal cancer bioassays are known not to be human risks by the very agencies promulgating them. Such risk estimates do not, can not, and are not intended to tell us what exposures are human risks. They attempt to identify exposures that are not risks, or, more completely, exposure levels unlikely to later be shown to be a risk.

Rudén and Hansson condemn EBT. They believe it is a stealth scheme to undermine regulatory risk assessments and expunge current protections of

\[1\] “The rule that extreme improbabilities have to be neglected [to identify scientific certainties]... becomes sufficiently explicit only [upon detailed examination]. ... I do not deny the possibility that improbable events might occur. I do not, for example, assert that the molecules in a small volume of gas may not, perhaps for a short time spontaneously withdraw into a part of the volume, or that in a greater volume of gas spontaneous fluctuations of pressure will never occur. What I do assert is that such occurrences would not be physical effects because... they are not reproducible at will. Even if a physicist happened to observe such a process, he would be quite unable to reproduce it, and therefore would never be able to decide what had really happened in this case, and whether he may not have made an observational mistake.”

\[2\] According to the USEPA, “[a]n established procedure does not yet exist for making “most likely” or “best” estimates of risk within the range of uncertainty defined by the upper and lower estimates. If data procedures become available the Agency will also provide...”

\[3\] For example, the USEPA guidance for noncancer risk assessments states that “[t]he RIF and RIC can be used to estimate a level of environmental exposure at or below which no adverse effect is expected to occur.” but that “[i]n general IRIS values cannot be validly used to accurately predict the incidence of human disease or the type of effects that chemical exposures have on humans” (emphasis added) (http://www.epa.gov/iris/limits.htm). For cancer risk assessments the USEPA states, “... the linearized multistage procedure leads to a plausible upper limit to the risk ... Such an estimate, however, does not necessarily give a realistic prediction of the risk. The true value of risk is unknown, and may be as low as zero.”

---

\[d\] All causal propositions for risk (e.g., consuming benzene in drinking water for a lifetime at 10 ppb causes leukemia in humans) are either true or false. All propositions that are not impossible are “nomological possibilities.” Some will be demonstrated scientifically to be true and are then termed risks. The others, the as yet unresolved possibilities, are termed “uncertainties.”

\[e\] Instead of the contrasting terms “risks versus uncertainties,” the former and latter are sometimes designated by such adjectival modifiers, respectively, as actual risks, real risks, or known risks versus possible risks, potential risks, proposed risks, hypothetical risks, or unknown risks. Labels aside, the concept of underlying causation is the same.
public health from chemical harms.\textsuperscript{22} On the contrary, our framework for EBT\textsuperscript{2} does not even attempt to address the usefulness, propriety, or limitations of regulatory risk assessment/risk management procedures or policies.\textsuperscript{1} Also, Rudén and Hansson chose not to share with their readers our published position, almost to the converse, that we believe it may be prudent for preventative purposes to act as if some chemicals present health risks even when scientific knowledge is inadequate\textsuperscript{39,40} and that, understandably, regulators assume the hazard does occur in humans to meet the intended purpose of regulatory action.\textsuperscript{40}

What we do state is that toxicological risk assessments can present a problem for risk communication; we state this because, not surprisingly, the press and public think of “risk calculations” as predictions of occurrence of “actual risks” (epistemic). Consider this frightening entry entitled “Air, Cancer Risk Linked In Some Houston Areas” in the Houston Chronicle\textsuperscript{41}:

The results showed that some people living near industrial plants are being exposed to concentrations of pollutants that over a lifetime would increase the risk of cancer. Using data collected by air quality monitors in Texas City ... found levels of 3 hazardous chemicals that, if inhaled continuously during a period of 70 years, would likely create an additional 29 to 199 cancer cases per million.

(emphasis added)

Simply, rewriting the latter sentence as a nomological “risk” statement results in:

Using data collected by air quality monitors in Texas City ... found levels of 3 hazardous chemicals that, if inhaled continuously during a period of 70 years, would limit the uncertainty of additional theoretical cancer cases to fewer than 29 to 199 per million.

(emphasis added)

The latter, a literal and more accurate description of the risk assessment, would be less likely to unduly alarm the public by inappropriately implying a prediction that extra cancer cases will occur. So, just whose interests are served by perpetuating an opaque methodology and nomenclature that can so easily mislead the public? Eschewing Rudén and Hansson’s political arguments, we support any thoughtful discussions on nomenclature that improve risk communication regarding known and uncertain health hazards.

Quite unlike Rudén and Hansson’s characterization of EBT as eccentric and out of touch with mainstream toxicology,\textsuperscript{22} EBT fits perfectly with the evolution of toxicology from a largely applied science of conducting hazard testing into a modern experimentalism employing the latest advances in the basic sciences of biochemistry and cellular and molecular biology. The evidence-based requirements for acceptance of experimental results in toxicology by top-flight science journals should not be conflated with regulatory efforts to identify all potential hazards of a chemical, the first of the four steps in the National Academy of Science’s (NAS) risk assessment’s paradigm.\textsuperscript{42} However, contrary to the views of Rudén and Hansson and others heavily invested in the process, the formulaic approach to making a regulatory “Hazard Identification”\textsuperscript{43} is easily distinguishable from the original NAS description of “The determination of whether a particular chemical is or is not causally linked to particular health effects”. Indeed, even the original NAS document recognized in a gesture to policy over science that such a determination is “often restated in terms of effects in laboratory animals or other test system” The most recent restatement of risk assessment goals by the NAS goes further by stating, explicitly, that

‘once the available evidence, either epidemiologic or experimental, is judged sufficient to establish that a given finding of toxicity or carcinogenicity is potentially relevant to humans ... the committee sees no reason for [USEPA] to spend time and resources to fine-tune the hazard classification . . . .’

(emphasis added).\textsuperscript{44}

The committee’s report makes it clear that the regulatory Hazard Identification step has not in the past, is not now, and most assuredly should not in the future be considered the same as an evidence-based conclusion of causation.\textsuperscript{1} Hence, EBT would

\textsuperscript{1} On this issue we did state that “[o]f course, many regulatory agencies, public advocacy groups, crisis managers, and safety experts advise that certain actions (such as evacuations, waste cleanups, restricted product usage, etc.) be taken to reduce “risks” even if the possibility of harm is only nomological. Such pragmatic actions may be defended by policy considerations such as reference to the Precautionary Principle: It is better to be safe than sorry. However, treating uncertainties as if they were risks, out of an abundance of caution, is wholly distinguishable from stating that those nomological possibilities are epistemic and represent risks.”\textsuperscript{22}

\textsuperscript{1} The committee’s reasoning makes sense because whether the causal substance/disease relationship is proven by an expensive and time consuming evidence-based review or assumed as true because it is “potentially relevant,” the frequency (numerical “risk” estimate) calculations are the same.
have no necessary effect on the various classification schemes Dr Rudén devotes so many pages to discussing because, as her own work has shown, such “Hazard Rankings” are designed to meet a specific regulatory goal. “Hazard Rankings” are seldom evidence-based, nor do they necessarily need to be. Often missing is an overt acknowledgment of the difference.

We see a role for EBT in helping with risk communication because the categorical names for the Hazard are potentially misleading to the press and general public. Consider the category of “probable” human carcinogen. Although most English speakers might think this means that the chemical is highly likely to be a human carcinogen (i.e., an “odds-on” favorite), nothing could be further from the truth in the world of risk assessment. Though the International Agency for Research on Cancer (IARC) has, without qualification, used this label for decades, recently IARC added unceremoniously to their monographs this belated disclaimer:

The terms probably carcinogenic and possibly carcinogenic have no quantitative significance and are used simply as descriptors of different levels of evidence of human carcinogenicity, with probably carcinogenic signifying a higher level of evidence than possibly carcinogenic.45

(emphasis added)

What is meant by level of evidence is not explained by IARC or, for that matter, by Rudén and Hansson. Not to be left out of this effort to retranslate the category’s cognomen, the USEPA recently stated:

Although the term ‘likely’ can have a probabilistic connotation in other contexts, its use as a weight of evidence descriptor does not correspond to a quantifiable probability of whether the chemical is carcinogenic . . . As stated previously, the use of the term ‘likely’ as a weight of evidence descriptor does not correspond to a quantifiable probability.46

(emphasis added)

So, it seems that all these years, “likely” never really meant “likely”. Although such regulatory classification schemes are developed for the narrow use of regulatory risk management, their widespread dissemination by the press and others to the general public without any clarification of this specific jargon can become a misleading communication about the true human risk.

Dr Rudén, a champion for reliance on animal test results for regulatory decisions, feels that EBT would deprive the regulator of a vital source of nonhuman data for “state-of-the-art” toxicological risk assessments.22 Her forecast of such a privation is fanciful flogging of a dead horse. The differences between causation and regulatory decision-making have long been recognized by knowledgeable toxicologists, physicians, and epidemiologists28,30–35,44–51 and has been the basis for the discussions of many articles related to the areas of toxicological risk assessments. Many, if not most, regulatory decisions will not (cannot) be evidence-based, just as not all physicians’ decisions can be evidence-based due to lack of data or of data review.2

Hence, our disagreement with Dr Rudén is that rather than accept the limitations of animal results for what they are and for the specific regulatory need they serve, she attempts,22 instead, to elevate by sheer weight of rhetoric the predictive qualities of animal testing (although as noted earlier, in characteristic self-contradiction she does concede with one line that the many limitations of extrapolating animal data are well known among toxicologists (see page 304, column 2, paragraph 2).22 Although a detailed rebuttal of her argument is beyond the scope of this commentary, Dr Rudén’s unfamiliarity with EBL limits her understanding of simple concepts like test predictivity. For example, she states, without evidence, that animal studies “reduce the risk of false negatives at the price of an increased number of false positives.” In fact, animal results translate poorly into clinical practice52 and the leading cause of limitation or withdrawal of drugs is serious adverse reactions missed by extensive preclinical animal testing.53 From a broader perspective, at a time when there are widespread efforts to find more useful alternatives to animal toxicity testing54 or to develop advanced human investigative systems such as molecular epidemiology, Dr Rudén’s intoned advocacy for heavy reliance on animal testing alone seems dated. Although Dr Rudén retrogresses, the NAS, is looking forward and states that “[the] new vision and strategy for toxicity testing in the 21st century . . . would be based primarily on human biology instead of animal biology, and would require anywhere between substantially fewer animals and virtually no animals.55 Indeed, the United States Food and Drug Administration, in recently announcing its intention to use a full dress Evidence-Based Review System for evaluation of causal health claims for dietary supplements, was unequivocal on this point: “Lacking any data from human studies, animal and in vitro studies alone would not adequately support a health claim. Animal and in vitro studies can be used to generate hypotheses, investigate biological plausibility of
hypotheses, or to explore a mechanism of action of a specific food component through controlled animal diets; however, these studies do not provide information from which scientific conclusions can be drawn regarding a relationship between the substance and disease in humans.”

Dr Rudén’s advocacy for increased reliance on animal testing absent human data for proof of causation is an echo from a time long past.

Rudén and Hansson’s self-admitted befuddlement that “it was initially a riddle for us” regarding the similarities between EBM and EBT is readily explained by the way in which clinicians make patient management decisions. Simply, both EBM and EBT are concerned with causation. Both are concerned with knowing what we know at a point in time about propositions of causation for chemicals and their effects (see also EBM and EBT similarities discussed in this issue). For physicians to weigh the risks versus the benefits of a treatment, they first must know both what beneficial and adverse effects are caused and what are the predicted frequencies at which each will occur. Otherwise no weighing can occur. If the answers are not known, the physician must revert to authority (and often has to). Rudén and Hansson’s distinction between toxic and therapeutic effects is one of values, not chemicals. Chemicals are apolitical. They simply act. Humans decide if the effect is desirable or not. Poisoning the formation of prostaglandin H2 and thereby inhibiting the normal platelet function of adhering to endothelial basement membrane sounds toxic, but actually underlies a widespread use for aspirin administration, that is, to prevent recurrent myocardial infarction. Once the status of knowledge has been determined, both medicine and toxicology then make management decisions that consider other factors our article intentionally does not discuss. For example, physicians must take patient values into account even when it is possible to make an evidence-based decision. Similarly, for the regulatory management of known or uncertain hazardous chemicals, such factors as costs, feasibility of remediation or controls, and the size of the population at issue may be important. There are roles for both facts and value judgments in public policy and individual therapeutics, alike. Transparency requires explicit designation of which is operative at a given time.

Finally, having become convinced that our views of EBM and risk assessment are “bizarre,” Dr Rudén apparently performed a transcontinental psychological analysis into our motivations. She deduced that we are secretly trying to advance the evil interests of the tobacco industry. On the contrary, we would have acknowledged sponsors for our EBT article, but we had none, including tobacco companies. Moreover, Dr Rudén either did not read or does not care about the many pages in our article devoted to a searching analysis of cigarettes and lung cancer as the foremost example of reaching an evidence-based conclusion of risk even when the database contains solely observational studies but no human experimental trials. Selectively setting aside such unaccommodating facts, Dr Rudén then seized on “sound science” which, she states, is a code phrase of the tobacco industry used to oppose regulatory action unless full scientific proof of harm is in hand. This subversive political agenda was abandoned when “sound science” became thoroughly discredited (we are not told how, by whom, or why). Seeing a rough equivalence between her interpretations of “sound science” and her own concocted implications of EBT for risk management, Dr Rudén reasoned that we must have developed EBT as a new, more palatable slogan for the same obstructionist principle. Finally, after apparently carrying out extensive research on toxic tort litigations in the United States (not an unremarkable accomplishment for a Swedish scientist who acknowledges no assistance from attorneys in this regard), she asserts that because she believes one of us consulted for a tobacco company, briefly, more than 25 years ago, the loop closes on her supposition that we designed EBT as a pretext to help sell tobacco.

Ironically, Dr Rudén chose, perhaps without thinking the matter through, one of the strongest known examples of the limitations of animal extrapolations. Human observational evidence led to causation even though the animal evidence for cigarette smoking gave a false negative result.

How did Swedish scientists acquire such professional interests about U.S. civil litigations? Regardless, Dr Rudén makes false or misleading assertions about Dr Guzelian’s practice of consulting medical toxicology. Dr Guzelian has testified under oath that he has no documentation of a contract for consulting services to Phillip Morris and was never was paid $100,000 in Phillip Morris consulting fees for even one year, much less yearly. Pertinently, Dr Guzelian has never testified on behalf of a tobacco industry client. For more than 20 years, Dr Guzelian has consulted for, and at times has testified on behalf of, lawyers representing plaintiffs and those representing defendants in personal injury matters, as well as parties involved in regulatory matters, product development, criminal matters, and others all involving the possible toxic effects of chemical agents.

Such sleuthing was so unnecessary! A simple collegial call or email to anyone of us could have cleared up for Dr Rudén this “puzzle,” thereby freeing her from her binding
Probably not intended as a compliment, she goes on to criticize one of us for using the same high standard of evidence and analysis (characterized and published as EBT) when formulating expert opinions for use in a legal matter about causation of disease by a chemical. This assertion, actually, is correct. We most certainly do attempt to use EBT whenever possible. We cannot imagine doing otherwise! To have one standard when trying to be accepted in public by scientific peers but then to adopt a different, lower, biased standard for the more cloistered world of the courtroom would be unprofessional and, probably, unethical. Furthermore, how better to avoid the influences of the bias that Dr Rudén attempts to accuse us of, than to openly commit to an EBT approach and, thereby, be held to opinions guided by an objective, rule-based analysis of all the best evidence.

In reality, none of us had ever heard of “sound science” as a specific approach to epidemiology analyses before reading the Rudén and Hansson commentary. Nonetheless, we became curious to learn just how the tobacco industry tried to hoodwink the toxicology community into believing cigarettes were safe. The only article Rudén and Hansson cite is a long recounting of people, strategies, meetings, and other events during the 1990s concerning tobacco regulation and litigation (our names do not appear). The only scientific material presented is a list of 15 Good Epidemiology Practices supposedly advanced by the tobacco company as “sound science,” but whose origin was the Chemical Manufacturers Association who developed a framework for judging the quality of a study. Having written extensively about the use of epidemiology in EBT, we scanned the Good Epidemiology Practices principles and found that they set forth the following guidelines:

1) Provide a clear definition of all objectives and hypotheses
2) Pay attention to control group selection
3) Describe all statistical techniques
4) Adhere to the study protocol
5) Train and monitor those administering questionnaires and surveys
6) Analyze results as per study protocol
7) Provide adequate description of the raw data
8) Odds ratios of two or fewer should be treated with caution particularly with wide confidence intervals
9) Be careful in use of meta-analysis
10) Include observations inconsistent with the main body of the data
11) Publish all completed research, regardless of outcome
12) Be aware of interpretation problems if reporting many hypotheses tests not specified by the study protocol
13) Statistically significant association does not in itself provide direct evidence of causal relationship
14) Encourage graphic display of results and figures
15) Use rigorous scientific objectivity when reporting epidemiological results, including defects in study design, conduct and analysis.

What a let down. We have to admit we are hard-pressed to find fault with any of these guidelines, most of which are found in prominent textbooks of epidemiology. We find it unfortunate that Rudén and Hansson give the reader no indication as to which of these guidelines they find offensive, or the combination of which leads to the claimed “thorough discrediting of sound science.” Dismissing a piece of science or scientific guidelines/principles only because of its source rather than its content, like Rudén and Hansson’s entire declamation against EBT, is unprofessional, reveals bias, may be used to produce unsupportable opinions, and indeed, is neither evidence-based nor objective.

In summary, the last 2 years have seen a remarkable interest develop in exploring the applications of EBT. With continued discussion and explanation, as went on at International Forum Towards Evidence-Based Toxicology, unwarranted suspicions will diminish. Most encouraging, interest in EBT is not limited just to academic circles but also is appearing in the clinic to improve the care of acutely poisoned patients and, already confounding the naysayers, is appearing in policies of regulatory agencies who, in reality, are well served also by access to the best evidence objectively assessed. The NAS
committee, reviewing the USEPA’s dioxin reassessment, put it succinctly:\footnote{26}

Furthermore, the EPA Reassessment continues to rely on the approach that diverse human data collected across disparate studies of different types and inherent strengths can be interpreted with confidence without applying the more formalized tools of evidence-based medicine. Thus, the EPA Reassessment (as well as Institute of Medicine [IOM] committee report) relies largely on committee-based, consensus evaluation of the available data rather than on specifically commissioned, rigorous analyses constructed according to established criteria that both formally evaluate the strengths of the available evidence and integrate, by quantitative systematic review, the data across available studies.

Encouraged by such developments, we look forward to continued exploration (but not to unfounded personal attacks) in the pages of Human and Experimental Toxicology of the future of EBT.\footnote{26}

**Disclosures and Acknowledgements**

PS Guzelian, Christine Halmes, and Robert C James have for a fee consulted for or have testified on behalf of (or both) parties in regulatory or litigation matters in which toxicity from a drug or environmental chemical was at issue. We thank Christopher Guzelian for helpful discussions about evidence-based logic.

**References**


38 Federal Register 1985; 51: 33998.


40 Garabrant, DH, James, RC. Trichloroethylene and cancer in humans: recognizing the need for an evidence based analysis. Toxicology 2005; 212: 80–84.

41 Cappiello, D, Snyder, M. Air, cancer risk linked in Cancer in humans: recognizing the need for an evidence based analysis. Toxicology 2005; 212: 80–84.


46 Guidelines for carcinogen risk assessment. EPA/630/P-03/001F. March, 2005.


PS Guzelian¹, MS Victoroff², C Halmes³ and RC James⁴

¹University of Colorado Health Sciences Center, Clinical Toxicology (Private Office), Centennial, Colorado, USA

²Recall, Inc., Centennial, Colorado, USA

³Halmes and Associates, Centennial, Colorado, USA

⁴Terra, Inc., Tallahassee, Florida, USA
Greeting

CL Galli\textsuperscript{1,2}

\textsuperscript{1}Department of Pharmacological Sciences, Faculty of Pharmacy, University of Milan, Milan, Italy; and \textsuperscript{2}Federation of European Toxicologists and European Societies of Toxicology (EUROTOX)

I was very happy to be present at the first International Forum towards Evidence-Based Toxicology, not only in my function as president of EUROTOX, but also as a practicing scientist and risk assessor.

It is clear to me that there is room for improvement in toxicological assessment, especially with regards to in-vivo animal models. Current uncertainty over the assumptions made in using such models makes clear-cut risk assessments difficult.

For example, there is a great need to improve our knowledge of the pharmacodynamics and pharmacokinetics of many substances. Understanding the mechanism of action of these substances would bring us nearer to a good scientific basis for risk assessment.

In the future, we need an approach that relies less on assumptions and more on evidence. Evidence-based toxicology may provide the route to such an approach.
Preface

C Griesinger¹, S Hoffmann², A Kinsner¹, S Coecke¹ and T Hartung³

¹European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; ²TÜVRheinland, BioTech GmbH, Köln, Germany; and ³Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

“A wise man, therefore, proportions his belief to the evidence.”
(“Enquiry Concerning Human Understanding” by David Hume, 1748)

The First International Forum Towards Evidence-Based Toxicology

By all means of appraisal, the First Forum Towards Evidence-Based Toxicology, held from October 15 to 18 in Cernobbio (Como), Italy was a great success: more than 170 scientists from Europe, Africa, America, and Asia debated the emerging concept of an evidence-based toxicology at an inspiring location – Villa Erba at the shores of lake of Como, former home of the great visionary of Italian cinema, Luchino Visconti.

But beyond this impressive participation, the forum timely provided additional momentum toward the paradigm shift we are currently witnessing in toxicology¹,²: there is growing agreement that especially regulatory toxicology needs to review its toolbox. This implies finding new methods and strategies to replace traditional methodologies that may rely on traditional assumptions to a significant extent and, more importantly, appear to be out of register with the progress in basic toxicological and life sciences. Toxicology needs to incorporate means of adapting to scientific progress. However, such adaptation must not be based on wishful thinking or political or societal pressure, but instead on the careful evaluation of evidence.

The way forward in this continuous process of improvement is the increased use of more sophisticated and biologically relevant test systems whose relevance and reliability have been established by evidence-based approaches. This should include the development of methodologies to combine in an objective and where possible quantitative manner evidence from various sources, either through methods of meta-analysis pioneered by evidence-based medicine (EBM) or through the quantitative scoring of data using predefined quality/reliability criteria in the context of data integration. The next step must be the use of such evidence-based methodologies within Integrated Testing Strategies (ITS) that allow the intelligent and combinatorial use of available and newly generated empirical data and mathematical or modeling data.³ The resulting decision-making should be more transparent regarding the procedures used as well as sources and extent of certainty or uncertainty involved. Such evidence-based (un)certainties will allow to improve the frameworks used to generate information (e.g., ITS) in the first place. And, more importantly, the establishment of more explicit, conscientious, and thus transparent criteria for data integration and decision-making will allow the same criteria to be applied to the next “case.” This should improve the consistency of toxicological decision-making across institutions, countries, and perhaps even global areas.

Luckily, toxicology does not need to reinvent the wheel – the basic methodologies and logic of evidence-based approaches are simple and are already well-established in other disciplines, notably clinical medicine where so-called EBM plays nowadays a prominent role. Although emphasis of toxicology is on prevention rather than on intervention (cure) and although toxicology has to act on the basis of precaution to protect human and environmental health, we are deeply convinced that the structural similarities in many “crafts” whose decision-making is based on the life sciences (e.g., medicine, nutrition, toxicology) are so strong that evidence-based approaches are indeed widely applicable to situations where all available information needs to be evaluated to determine whether it qualifies as evidence in a specific context.

Organization of the forum

The forum was organized by the European Centre for the Validation of Alternative Methods, part of the
Institute of Health and Consumer Protection (IHCP) of the European Commission’s Joint Research Centre (EC JRC). However, the organization would not have been possible without the support of the Scientific Steering Committee (listed below) and, in particular, the committee’s core group: Drs Lutz Edler, Ian Kimber, Edmund Neugebauer, and Ellen Silbergeld whom we cordially thank for their support during the preparatory phase of the forum.

The forum was not conceived as the usual speaker/audience type of conference in the traditional sense, but – nomen est omen – as an opportunity for the toxicological community to meet and openly exchange views, to speak candidly about toxicology, its underlying assumptions, its possible deficiencies, and potential improvements of toxicological practice. Consequently, presentations by members of the Scientific Steering Committee and invited speakers were intercalated with breakout group work involving all participants of the forum. The resulting spirit and work atmosphere were overwhelmingly frank, open, and critical but also constructive and forward-looking.

Perhaps, the genius loci of Villa Erba eventually worked: at the end of the forum, most participants appeared to share a similar vision as reflected in the tangible results of the conference: the Declaration of Como (signed by about one-third of the participants) and a concise list of Ten Defining Characteristics of an Evidence-Based Toxicology. These were endorsed by the participants during the last plenary session of the forum. These Ten Defining Characteristics served as the basis for a short provisional definition or mission statement of evidence-based toxicology, which was developed by the Scientific Steering Group after the meeting.

Goal of the forum

The overall objective of the forum was to take stock of current toxicological practices and to identify whether toxicological practice including toxicological decision-making may benefit from evidence-based approaches. The overall goal was to initiate and broker scientific debate on a promising avenue for a modern toxicology: evidence-based approaches.

Today, evidence-based approaches are successfully used in several disciplines, especially in applied sciences or crafts that transcend basic research (which is or should be evidence-based) and intend on acts involving decision-making based on the integration of various types of knowledge or information. Evidence-based approaches are derived from the very logical tools we all know from good science but have developed into a distinct methodology in their own right comprising critical appraisals, systematic reviews, and meta-analysis and, to start with, clearly phrased hypotheses/questions. The latter being, as every scientist can tell, a pivotal prerequisite for getting clear answers whether through experimentation or logical approaches.

An example remote from toxicology is evidence-based best practices in pedagogical work. The closest and most relevant example for our purpose here is EBM/evidence-based health care (EBHC) and the work of the Cochrane Collaboration. EBM, although not completely uncontroversial within the medical scientific community, has reached near-paradigmatic status as the methodology of choice for the conscientious and unbiased assessment of possible causal relationships and risks in medicine and the merit of specific medical acts: whether therapeutic and, perhaps to a lesser extent, preventive interventions can be regarded as plausible knowledge, as a plausible and reasonable benefit for the average patient group or for the individual patient.

The underlying leitmotif of all disciplines that may benefit from evidence-based approaches is uncertainty. The general uncertainty associated with all knowledge that we generate, be it on mechanisms of action or risks, has important reverberations when it comes to applying this knowledge in practical terms, for example, for selecting the right treatment for a specific patient or for classifying a substance as a carcinogen. In toxicology, sources of uncertainty associated with scientific information, if not correctly identified and dealt with, may contribute to variability in toxicological practice leading to discordant and therefore unreliable appraisals. Obviously, uncertainty of knowledge is also a prime concern of epistemology: what is actually knowledge and how can we gain it? Where do propositions and opinions end and where does knowledge start? It is certainly not by coincidence that concepts of evidence as qualifiers for information fulfilling criteria of consistency and plausibility, as well as the concept of various degrees of evidence or confirmation, were first explored in philosophy of science and epistemology already in the early 20th century.

In summary, we are deeply convinced that the fundamental similarities shared by live science-based crafts (e.g., medicine, toxicology, and also nutrition) warrant the exploration of the usefulness of the practical tools of those evidence-based approaches that have proven useful and trust-building in clinical medicine and health care. These similarities include concepts such as uncer-
tainty of hazards and risks, causative relationships between specific factors (e.g., xenobiotics, pharmaceuticals, food), and adverse effects as well as the probability and effectiveness of specific acts that intend on curing or preventing adverse human or environmental health effects.

Organization and results of the breakout group work

Breakout groups were organized around three themes: 1) “Taking stock”; 2) “Possible use of evidence-based approaches in toxicology”; and 3) “Towards a definition of EBT.” Each breakout group was provided with suggested charge questions (see Table 1) but was ultimately free in determining the course of their debate and thus the results generated.

Structure of the forum proceedings

The proceedings are organized in deviation from the chronological order of presentations at the forum. The individual contributions as well as the results of the breakout group work were regrouped according to five thematic sections.

- In Section 1 “Fundamentals of an evidence-based toxicology,” key problems of toxicological practice are put into context of evidence-based approaches. In his opening statement, Thomas Hartung briefly summarizes the need for renewing toxicological practice and exemplifies this on the basis of two key problems, that is, the impact of prevalence and the “validation dilemma” encountered in the area of method validation. Sebastian Hoffmann presents key problems of toxicological test method assessment, fundamental to sound data generation in toxicology and presents possible avenues of evidence-based approaches. Philip Guzelian explores the differences between consensus-, opinion-, and evidence-based decision-making in toxicology and highlights the need for more transparency in toxicological practice. Claudius Griesinger briefly analyzes the basic similarities and differences between knowledge creation in medicine and toxicology and explores concepts of causation, probability, and evidence in the context of epistemology. This leads over to the next section.

- Section 2 “Evidence-based medicine – a possible model for evidence-based toxicology?” provides a concise overview about the use and frameworks of evidence-based approaches in medicine. Edmund Neugebauer focuses on EBM and specifically how results of evidence-based approaches are put into practice through consensus-based guidelines. Roberta Scherer explains the nuts and bolts of evidence-based approaches as used by the Cochrane Collaboration.

- Section 3 “Core problems and case studies” provides a brief overview over some challenges of toxicology that are felt most pressing. Mariano Cebrian briefly introduces the issue of uncertainty, while Lutz Edler explores current approaches to tackle uncertainty, stressing the importance of uncertainty factors for risk assessment. Jürgen Borlak finally presents an

<table>
<thead>
<tr>
<th>Theme</th>
<th>Breakout group question</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Taking stock</td>
<td>Group 1: “What are the current schemes for decision making in toxicology?”</td>
<td>Chapter 4.8: Current schemes for decision making</td>
</tr>
<tr>
<td></td>
<td>Group 2: “What are the strengths and weaknesses of current information sources for hazard identification?”</td>
<td>Chapter 4.9: Current information sources for hazard identification</td>
</tr>
<tr>
<td>2. Use of evidence-based approaches in toxicology</td>
<td>Group 3: “Which evidence-based tools might be useful in toxicological basic research?”</td>
<td>Chapter 5.1: Evidence-based tools in toxicological basic research</td>
</tr>
<tr>
<td></td>
<td>Group 4: “Which evidence-based tools might be useful in toxicological hazard identification?”</td>
<td>Chapter 5.2: Evidence-based tools in toxicological hazard identification</td>
</tr>
<tr>
<td></td>
<td>Group 5: “Which evidence-based tools might be useful in toxicological decision-making?”</td>
<td>Chapter 5.3: Evidence-based tools in toxicological decision-making</td>
</tr>
<tr>
<td></td>
<td>Group 6: “How can we improve the toxicological toolbox?”</td>
<td>Chapter 5.4: Possible improvement of information sources on hazard and risk</td>
</tr>
<tr>
<td>3. Towards a definition of evidence-based toxicology</td>
<td>Group 7: “Working towards a definition of an evidence-based toxicology”</td>
<td>Defining characteristics of EBT (endorsed by the forum participants) Provisional definition of evidence-based toxicology (finalized post hoc by members of the Scientific Steering Committee)</td>
</tr>
</tbody>
</table>
interesting case study demonstrating how current pharmacotoxicological safety assessments failed to detect grave side effects of a drug.

- Section 4 “Toxicological decision-making on hazards and risks – status quo and way forward” brings together speaker contributions as well as breakout group work. Chris Portier discusses current knowledge-based decision-making in the field of carcinogenicity. Ellen Silbergeld discusses a systematic review on the possible causal link between arsenic and type 2 diabetes mellitus and cardiovascular diseases. Phil Guzelian explores the usefulness of a new emerging EBM concept, evidence-based individual decision-making (EBID) for the purposes of toxicology. Lutz Edler summarizes the pros and cons regarding the current debate concerning the replacement of no adverse effect levels (NOELs) with benchmark dose levels (BDLs) in food risk assessment. Lutz Müller returns to test method assessment and presents the result of a systematic review of in-vitro genotoxicity tests for detecting carcinogenicity. The benefits of modeling approaches for evidence-based evaluations are introduced by Dennis Sarigiannis and Hugh Barton. Finally, the section is closed by results from breakout group work discussing current schemes of decision-making as well as strengths and weaknesses of current information sources for hazard identification (breakout group theme 1).

- Section 5 “Steps towards an evidence-based toxicology” captures mainly the debates of breakout groups on themes 2 and 3. Group work in this section was concerned with possible ways of improving the toxicological toolbox, discussions on possible defining characteristics of an evidence-based toxicology as well as the usefulness of evidence-based tools in 1) basic research, 2) hazard identification, and 3) decision-making procedures.

Moreover, André Pirlet presents the merits of standardization in the context of evidence-based approaches and Agnieszka Kinsner summarizes the concept of a web-based portal to evidence-based toxicology, developed by European Centre for Validation of Alternative Methods (ECVAM).

Finally, in the conclusions, Ian Kimber succinctly summarizes both necessity and promise of an evidence-based toxicology.

We hope that these proceedings will help demonstrating the tremendous value of evidence-based approaches for toxicological practice. We cordially thank all participants of the forum for their valuable contributions. Our gratitude goes also to Dr Wallace Hayes, Harvard School of Public Health in his function as Editor of Human & Environmental Toxicology (HET) for his excellent cooperation. Lastly, we would like thank the Institute for Health and Consumer Protection for making this meeting possible and gratefully acknowledge additional financial support received from the Doerenkamp-Zbinden Foundation.

References

The Scientific Steering Committee

• Yasunobu Aoki  Bureau of Science and Technology Policy, Cabinet Office, Japan
• Hugh Barton  ORD National Center for Computational Toxicology, EPA, USA
• Jürgen Borlak  Fraunhofer Institute for Toxicology and Experimental Medicine, Germany
• Herman van Cauteren  Johnson & Johnson, Belgium
• Sandra Coecke  European Commission Joint Research, Ispra, Italy
• Mariano Cebrian  Center for Research and Advanced Studies, Mexico
• Lutz Edler  German Cancer Research Center, Germany
• Steven Eisenreich  European Chemicals Bureau, EC, DG JRC, Italy
• Kevin Gaido  The Hamner Institutes for Health Sciences, USA
• Bernard Goldstein  University of Pittsburgh, USA
• Claudio Griesinger  European Commission Joint Research, Ispra, Italy
• Eliseo Guallar  Johns Hopkins University, Bloomberg School of Public Health, USA
• Ursula Gundert-Remy  Bundesinstitut für Risikobewertung (BfR), Germany
• Philip Guzelian  University of Colorado Health Sciences Centre, USA
• Christine Halmes  Halmes and Associates, USA
• Thomas Hartung  European Commission Joint Research, Ispra, Italy
• Wallace Hayes  Harvard School of Public Health, USA
• Sebastian Hoffmann  European Commission Joint Research, Ispra, Italy
• Robert Kavlock  ORD National Center for Computational Toxicology, EPA, USA
• Martin Kayser  BASF Aktiengesellschaft, Germany
• Ian Kimber  Syngenta Health Assessment and Environmental Safety, UK
• Agnieszka Kinsner-Ovaskainen  European Commission Joint Research, Ispra, Italy
• Herman Koeter  European Food Safety Authority (EFSA), Italy
• Edmund Neugebauer  University of Witten/Herdecke, Germany
• Pierluigi Nicotera  MRC Toxicology Unit, UK
• Christopher Portier  NIEHS, USA
• Testuo Satoh  Chiba University, Japan
• Kai Savolainen  Finnish Institute of Occupational Health, Finland
• Roberta Scherer  Johns Hopkins Center for Clinical Trials, USA
• Joel Schwartz  Harvard University, USA
• Lauri Scott  Proctor & Gamble, Belgium
• Ellen Silbergeld  Johns Hopkins Bloomberg School of Public Health, USA
• Bill Stokes  NIEHS, USA
• Ray Tice  NICEATM, USA
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

Defining characteristics of EBT

The First International Forum Towards an Evidence-Based Toxicology agreed the following defining characteristics of an evidence based approach towards toxicology:

Evidence-Based Toxicology

1. promotes the consistent use of transparent and systematic processes to reach robust conclusions and sound judgements
2. addresses ethical values and expectations and is socially responsible
3. displays a willingness to check the assumptions upon which current toxicological practice is based to facilitate continuous improvement
4. recognises the need to provide for the effective training and development of professional toxicologists
5. acknowledges a requirement for new and improved tools for critical evaluation and quantitative integration of scientific evidence
6. 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
7. ensures the generation and use of best scientific evidence
8. includes all branches of toxicological science: human health assessment, environmental and ecotoxicology as well as clinical toxicology
9. acknowledges and builds upon the achievements and contributions of evidence-based medicine and evidence-based health care
10. fosters the integration of expert judgment with best possible external evidence
Provisional Definition of an Evidence-Based Toxicology

The Scientific Steering Committee of the 1st International Forum towards Evidence-Based Toxicology proposes the following provisional definition of Evidence-Based Toxicology:

Evidence-based toxicology (EBT) seeks to promote the conscientious, judicious, explicit and transparent use of evidence derived from structured approaches aiming at the objective, complete and quantitative evaluation of all information to arrive at robust and sound decisions regarding human and environmental safety.
1 Fundamentals of an evidence-based toxicology

1.1 Opening statement

T Hartung

Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Toxicology and effective risk assessment depend on scientific and technological information and should constantly adapt to these advances in this information. However, toxicology still largely relies on traditional assessment methods that were established decades ago and that have changed little despite scientific and technological progress. Consequently, safety assessments are often based on tests of unknown relevance and reliability and whose predictive validity has never been assessed objectively. The margin of safety for many assessment methods is also unclear. Furthermore, changes in our understanding of disease, the types of product now requiring safety assessment (including novel biotechnological products), legislative frameworks [in particular, the Registration, Evaluation and Authorisation of Chemicals (REACH) legislation in Europe], and public expectations pose significant challenges for the entire field.

The way in which data generated by toxicological tests are used to assess potential hazards and risks, and subsequent decision-making in risk management, may also be improved. These issues mainly concern the transparency and consistency of data interpretation and the integration of multiple data sets. Faced with increasing amounts of information being created in research at an ever-accelerating pace, we urgently require efficient tools that enable us to make full and exhaustive use of this information in a structured, reliable, and transparent manner.

The concerns in toxicology may be exemplified by two key problems. The first is that posed by the impact of prevalence. In other words, if one tests for a rare toxicity, even the most sensitive tool will produce many false-positives. For example, if one is testing for a toxicity occurring in only 1 in 10,000 cases (0.01% prevalence), a test with 99.9% accuracy will produce 1 true positive result in 10,000 tests but also 10 false-positives. The test could therefore only be relied upon only in 1/11 cases, that is, the positive predictive value (PPV) of this test would be approximately 9% (Table 1). In practice, most toxic effects are not that rare and toxicological tools are never 99.9% accurate. If one assumes that biological test systems have an accuracy of 80–90% and that most toxicities have a prevalence of 1–10%, positive results are only accurately detected with a single test 4–50% of the time.

The impact of prevalence may be countered by properly evaluating our tools and systematically combining them to compensate each of their deficiencies, for example, by combining sensitive and specific tests for a particular toxicity. However, to do this, we need to validate our tools, assess their performance, and systematically compose combinations of them. The resulting combinations will then need to be validated.

A second key problem in toxicology may be called the “validation dilemma.” This is posed by the fact that the point of comparison for any novel toxicological tool will be a traditional, poorly assessed methodology. The challenge faced in toxicology is how to objectively assess the value of new tools.

Problems such as these make it necessary to ensure that safety assessment methods continually reflect best scientific practice and technological advances. Therefore, it is important to ensure that structures are available that will encourage, facilitate, and support a process of critical appraisal and...
renewal of the toxicological repertoire available for safety assessment. Part of this process is to embrace approaches that ensure the best possible scientific evidence is applied to judge product safety and likely risks to human and environmental health.

Over the last two decades, a new concept has arisen in clinical medicine, termed evidence-based medicine (EBM). This approach seeks to ensure that any therapeutic decision is based on the best scientific evidence available in an impartial, transparent, and structured manner. A suite of assessment methods and decision-aiding tools, as well as structures for their implementation in the decision-making process, have been developed. Collectively these have served to significantly strengthen the scientific basis of medicine.

Similarly, toxicology might embrace an approach that we may name evidence-based toxicology (EBT), after EBM. In this, the best scientific evidence would be systematically applied to assess testing tools, and results generated by these tools would be assessed in a transparent, structured manner to make clear, dependable decisions on product safety and likely risks to humans and the environment. Tools in EBT may include practices pre-existing in toxicology, such as the validation of methods, which has evolved over the last 15 years, and tools of quality assurance, such as Good Laboratory Practice and Good Cell Culture Practice guidelines. EBT may involve the validation of alternative methods, but it is not intended to restrict EBT to this narrow focus. Other tools such as systematic reviews and meta-analyses, which are not often used in toxicology, may also be taken from EBM.2,3

Toward this end, toxicological experts from around the world convened at Villa Erba, Como, Italy in October 2007 for the First International Forum Towards Evidence-Based Toxicology. The proceedings of this successful meeting follow in this supplement. Considerable debate was generated in response to the stimulating lectures given by experts in toxicology and EBM, and the conference ended in the drafting of several consensus statements, including a provisional definition of EBT. It is hoped that the meeting will be the start of a sustained EBT movement. Readers of these proceedings, both toxicologists and other stakeholders, are invited to join the declaration and contribute toward the shaping of a toxicology fit to meet the safety challenges of today and tomorrow.

References


Table 1 Positive predictive values (PPV; %) for combinations of test accuracy and prevalence

<table>
<thead>
<tr>
<th>Prevalence of toxicity (%)</th>
<th>Test accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99.9%</td>
</tr>
<tr>
<td>0.01</td>
<td>9.1</td>
</tr>
<tr>
<td>0.1</td>
<td>50.0</td>
</tr>
<tr>
<td>1</td>
<td>91.0</td>
</tr>
<tr>
<td>10</td>
<td>99.1</td>
</tr>
</tbody>
</table>

Adapted with permission from Hoffmann and Hartung.1
1.2 Aspects of test assessment

S Hoffmann
TÜVRheinland, BioTech GmbH, Köln, Germany

Introduction

Hazard assessment comprises the steps of hazard identification, characterization, and evaluation. The rationale and scientific basis of current hazard identification methods are increasingly coming under scrutiny. An indication for this is, for example, that the Organisation for Economic Co-operation and Development (OECD) has published a guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. Moreover, further developments and emerging technologies might change testing methods over time, thus potentially changing the outcomes of hazard identification and, consequently, of hazard assessment. This article focuses on concerns over hazard identification, using data-poor examples, where assessments are often based upon a single test.

Test assessment and validation

The safety level provided by a toxicological test may be thought of as lying on a continuum. At either end of this continuum are “no-test” situations where all substances would be regarded by default as being either nontoxic or toxic (Figure 1). The real-world situation exists somewhere in the middle of this continuum, with tests being underprotective or overprotective to varying unknown extents. Due to the precautionary approach in hazard assessment, most tests are often assumed to be more overprotective than underprotective, that is, they are associated with false positive data to some degree. This unknown level of conservatism creates two problems. First, because this conservatism is not quantified, a degree of uncertainty is added to decisions based on hazard identification results. Second, because the strengths and weaknesses of a given test may be obscured by conservatism, there is no clear basis from which to improve this test.

For informed decision-making, these problems must be confronted. The risks of false decisions should be identified and quantified as precisely as possible. The strengths and weaknesses of a given test should be known such that the test may be improved or replaced in the light of scientific or technological advances. For these reasons, testing methods need to be assessed. A further reason for assessing tests is to gauge their suitability for new testing needs. Also, it is becoming increasingly necessary to transparently communicate to stakeholders the reliability of tests and, in particular, the relevance of data thereby obtained.

Tests may be assessed retrospectively or prospectively. Retrospective assessment, that is, after the relevant decision has been made on the basis of a test result, may be performed by recording reported adverse events or through active surveillance after the test has entered routine hazard identification. This approach has the inherent risk of allowing intervention only after adverse events have occurred. This risk, however, is in practice reduced by degrees of conservatism implemented both in the test itself and in the decision-making, that is, hazard assessment. Prospective test evaluation, which safeguards the appropriateness of a test before it is routinely used, can effectively be performed through a process of validation. In formal validation of test methods, the scientific basis and reliability (reproducibility) of a test system and the predictive

Correspondence to: Sebastian Hoffmann, PhD, TÜV Rheinland BioTech GmbH, Nattermannallee 1, 50839 Köln, Germany. Email: sebastian.hoffmann@de.tuv.com

www.het.sagepub.com

© 2009 SAGE Publications 10.1177/0960327109105757
capacity of an associated prediction model are independently assessed, usually by comparison against a reference test.

However, validation has two limitations. First, it cannot be used to assess a test’s suitability for new testing needs because a comparative approach cannot be used. Second, and more importantly, if it is not known how well the reference test (often the respective currently regulated test) identifies hazards for humans or the environment, it may be impossible to accurately assess the validity of any new test. This has been termed the “validation dilemma.”

Steps toward an evidence-based toxicology

In moving toward greater transparency and quantification and reduction of uncertainty in risk assessment, which could be considered aspects of an evidence-based toxicology (EBT), certain steps are proposed.

First step is that the degree of conservatism in current tests should be assessed to determine the validity of the decisions based on these tests. In some cases, these tests were established decades ago and their true performance is unknown. They represent only surrogates for true human responses because factors such as differences in interspecies and intraspecies may cause the wrong conclusions to be drawn. For example, modeling the effects of interanimal variability on the predictive capacity of the Draize test for skin irritation resulted in a sensitivity of 94.1% and a specificity of 99.7%. Comparing the Draize test data with respective results from the highly relevant human patch test for skin irritation showed a specificity of approximately 60% and a sensitivity of <100% (i.e., up to 40% of compounds that induced skin irritation in the rabbit did not induce skin irritation in humans). These data could be used to quantify the conservatism of the Draize test.

The second step is to continuously develop methodology and methods of validation, for example, with regard to reducing or solving the validation dilemma. With regard to the validation dilemma, it includes accounting for the imperfectness of the reference test by using diagnostic test assessment tools from evidence-based medicine (EBM). Also, EBM-style systematic reviews may be used for retrospective validation whether data are already available. Other approaches include approximating human–animal differences by refining uncertainty factors of intraspecies and interspecies, and using human data as much as possible.

Hoffmann and Hartung identified parallels between toxicological and medical diagnostic test assessment with regard to patients versus toxicological agents, reproducibility, personnel issues (e.g., test transfer or biases), and relevance. On the basis of these parallels, it was proposed that EBM methodology is incorporated into toxicological test validation. As a follow-up of this proposal, the European Centre for the Validation of Alternative Methods (ECVAM) is currently developing a quality assessment tool with which to evaluate the reliability of toxicological test data.

In drawing up this proposition, certain statistical aspects of test assessment were noted. First, in assessing predictive capacity, specificity and sensitivity are mutually dependent upon each other. Without changes to test protocols, increasing the one can only be achieved by decreasing the other. Second, the use of predictive values allows the incorporation of prevalence (in the study or defined population) into test assessment to better describe test performance regarding specific (regulatory) purposes/applicability domains.

Conclusions

Test assessment is crucial in toxicological decision-making. It is proposed that test assessment is performed in a systematic, structured, and informative manner as one integral facet of an EBT. This would entail the validation of existing tests and the continuous development of validation methodology, as well as periodical reassessment of routinely applied tests according to internationally accepted validation procedures, for example, as those of the OECD, ultimately opening up avenues to adapt to scientific progress.

References

1.3 Consensus, opinion, and evidence-based science – three methods of reaching conclusions in toxicology

P Guzelian

Section of Medical Toxicology, University of Colorado School of Medicine, Denver, Colorado, USA

Introduction

Evidence-based medicine (EBM) has clear aims and methods: Where possible, decisions on diagnosis, treatment, prevention, or causation are made by conscientious, explicit, and judicious use of the best evidence as determined from a systematic, objective, and unbiased review of accumulated human knowledge.1 “Thus, the EPA Reassessment (as well as Institute of Medicine [IOM] committee report) relies largely on committee-based, consensus evaluation of the available data rather than on specifically commissioned, rigorous analyses constructed according to established criteria that both formally evaluate the strengths of the available evidence and integrate, by quantitative systematic review, the data across available studies.”6 In other words, EB science is a method of using evidence; it is not whether or not one uses evidence in support of a statement. In this context, this article discusses the use of clinical judgment and consensus, and the use of causation in risk analysis.

Clinical judgment and consensus

In medicine before the 1970s, physicians tended to make patient-based decisions according to “clinical judgment,” that is, intuition derived from education, experience, expert consensus, journal information, and colleagues’ clinical judgment. It is now generally agreed, following the introduction of EBM, that medical decisions should be based not only on subjective judgment or consensus but also on a systematic review of all the evidence and its quality. The concept of EBM has now been adopted by many medical societies and journals.

Expert judgment is a synonym for the opinion of one or more specialists. Opinions always have some degree of subjectivity although their holders necessarily believe it to be true. Consensus may be viewed as merely the opinion of a group and similar concerns apply. In science, opinions have consistently been shown to be poor predictors of actuality. In medicine, it has been repeatedly shown that interventions believed to have a beneficial outcome based on pathophysiological evidence alone often did not have the effect expected by specialists. For example, doxazosin prescribed as a treatment for blood pressure actually increased mortality.2

In toxicology, there are often no similar means for testing hypotheses. However, one opportunity arose when many American workers became seriously ill as a result of their occupational exposure to kepone (chlordecone). The critical effect for setting a safe environmental level chosen for this compound by the Food and Drug Administration was proteinuria on the basis that it persisted in rats when exposure was reduced such that all other effects were undetectable.3 The industrial accident provided the means to test this assumption.

In the toxicological study, 32 men were followed up for 9 years and ≈4000 urinalyses were performed. Over this period, the body content of kepone was mostly 2–3 orders of magnitude higher than the lowest observed adverse effect level dose in chronically treated rats. However, proteinuria was never observed in these patients. As a result of this study, the effects of kepone in rats could be classified into true and false positives. Most effects, including proteinuria, were false positives. The sensitivity of the rat as a predictor of human toxicity with regard to kepone could therefore be calculated as 88%. The specificity, at 11%, was much lower.3
Such evidence underlies the need for rigorous EB methodology with which to interpret toxicological data and guide decisions. In a further example, Rudén examined the evidence in 29 published cancer risk assessments on the same chemical (trichloroethylene) to assess how consistently authoritative groups used good scientific principles to reach conclusions of cancer causation. She found that no group assembled the complete database. The number of articles cited ranged from 5 to 81% of the available literature. Similarly, the numbers of key experiments (29–100%) and key epidemiology studies (38–100%) cited ranged widely. She also found that there was a high degree of biased interpretation, and the fractions of erroneously judged positive and negative studies reached almost 30%. If EB rules were followed, these findings should have been a lot more uniform because bias resulting from judgment would have been restricted to minimal levels.

Causation and risk assessment

It has been suggested that the key to EB science is the idea of causation. In the specific case of an EBT, it is the idea of an agent causing harm, an integral part of risk. Causation may be defined as a preceding factor that was necessary for an event to occur at the time and in the manner in which it did occur.

In formulating a causal proposition, there are known impossibilities. For example, it is not possible to receive toxicological injury from a chemical one that does not come into contact with. Scenarios that are not impossible are termed nomological possibilities. Among these nomological possibilities, there are propositions that may be either correct (these are called risks) or incorrect (termed uncertainties). The nomological possibilities known to be correct are known causal relations, for example, between paracetamol (acetaminophen) and hepatic injury. In these cases, there is strong, specific evidence, which, whether reviewed in a systematic way would strongly indicate one outcome.

Uncertainties, nomological possibilities, whose truth are unknown, may not be quantified. In the example of a dose-cancer relationship for certain chemicals, the risks are only known for the area of the graph populated by data (Figure 1). At lower untested doses, there is an area of uncertainty. The causal relationship is not known. When cancer risk assessments are performed, extrapolations, based on assumptions, are made into areas of uncertainty. It may be assumed that low doses pose no risk or that every dose has a risk. However, the true relation over this dose range is unknown.

Such concerns over assumptions, and lack of transparency and rigor are becoming widespread. In its assessment of the US Environmental Protection Agency’s (EPA’s) 10-year reassessment of dioxin, the US National Research Council stated “the EPA risk assessment continues to rely on diverse human data, collected across different kinds
of studies, of different strengths, and could not be interpreted without applying formalized rules of EBM. The EPA assessment relies on committee-based consensus evaluation of the available data rather than specifically commissioned rigorous analysis constructed according to established criteria that both formally evaluate the strength of the available evidence and integrate in a systematic review the data from across all studies. It is proposed that an evidence-based toxicology would increase the rigor and accuracy of risk assessment thereby answering such criticisms.

Conclusions

Expert opinion and consensus are fallible. The only means of reaching sound scientific conclusions is to use rigorous EB methods to assess all the evidences. Risks for which the causal relationships are unknown should not be expressed in risk assessment.

References

1.4 Comparing medicine with toxicology –
a mapping of knowledge creation, concepts
and basic epistemology

C Griesinger

European Centre for the Validation of Alternative Methods (ECVAM), In vitro Methods Unit, Institute for
Health and Consumer Protection (IHCP), European Commission Joint Research Centre, Ispra, Italy

Introduction

The main concerns of this forum on evidence-based
toxicology (EBT) are 1) how toxicological practice
may be improved, 2) whether toxicology may
profit from evidence-based tools (especially from
evidence-based medicine; EBM), 3) how EBT may
be defined, and 4) how a practicable EBT may be
achieved. To determine whether toxicology can
determine from EBM it is first necessary to evaluate
basic concepts and processes of knowledge-based
decision-making in both disciplines. Because any
debate concerning evidence is about the way we
transform propositions into knowledge, the article
moreover will, very briefly, touch upon epistemolog-
ical key concepts such as causation, knowledge, and
evidence, which will be briefly discussed on the
background of hallmarks of epistemological debate.

Knowledge creation for decision making:
basic parallels between medicine and
toxicology

The process of knowledge creation in toxicology and
medicine shows structural similarities and differ-
ences (Figure 1). Both disciplines are “crafts”¹ inas-
much as they transcend basic research (for pure
knowledge creation) but intend on the application
of knowledge through concrete acts of intervention
and/or prevention. For a clear mapping of the steps
of knowledge-based decision-making in both disci-
plines, the process has been broken down into four
sequential steps (Figure 1). In reality, knowledge cre-
atation in both disciplines is clearly far more complex.
Experimental or theoretical basic research is
“open-ended:” it primarily intends on acquiring
new knowledge of biological phenomena and
observable facts without aiming at any particular
application/use. It may include more targeted basic
research to understand mechanisms of pathogenesis
(medicine) or toxicogenesis (toxicology) driven by a
possible application, for example, a curative or pre-
ventive act. Knowledge acquired through basic
research subsequently leads to routine testing of
e.g., causative) hypotheses in view of decision mak-
ing (“purposeful knowledge creation”) with a partic-
ular application or act (curative/preventive) in view.
Such purposeful knowledge creation relies on
highly standardized experimental methods. In med-
icine, purposeful knowledge creation includes diag-
nostic testing, the assessment of diagnostic tests, as

Correspondence to: Claudius Griesinger, Sandra Coecke,
Agnieszka Kinsner, In vitro Methods Unit, Institute for Health
and Consumer Protection (IHCP), European Commission Joint
Research Centre, Ispra, Italy.
Email: claudius.griesinger@ec.europa.eu

© 2009 SAGE Publications 10.1177/0960327109105759
well as efficacy, and safety testing in clinical trials. In toxicology, this includes hazard testing and validation of testing systems, the equivalent of medical diagnostic test assessment. Such “routine” hypothesis testing (e.g., is patient X infected with hepatitis B? Is trichloroethylene a carcinogen?) is followed by decision making through the integration of various sets of data leading to an act of prevention or intervention intended to reduce the probability of adverse effects.

**Causation and probability – key concepts in toxicology and medicine**

In both disciplines, *causation* and *probability* are key concepts associated with knowledge creation. In medicine, hypotheses on causation include *inter alia* whether an agent or a genetic predisposition causes a disease or adverse effect (pathogenesis), or whether specific curative (e.g., drugs) or preventive interventions (e.g., diets, surgical interventions) decrease the probability of specific disease-related adverse effects. In toxicology, causative hypotheses mainly concern whether and how a xenobiotic substance may cause a toxic (adverse) effect and include “toxicogenesis”: a description of the steps involved in the origination of toxicity, that is, the actual perturbation of the normal homeostatic function of a biological system to an extent that adverse effects become evident. Similar concepts are subsumed under the terms “mode of action” or “toxicity pathways”. Causation in toxicology also is concerned with whether specific acts of prevention (e.g., risk reduction measures) may prevent adverse effects occurring. Regarding causation, there are parallels but also important differences between medicine and toxicology (Figure 1): medicine is concerned with both, disease and hazard, whereas toxicology primarily with specific xenobiotic hazards. Most importantly, medicine is concerned with both, curative and preventive acts, whereas toxicology’s main concern are effective preventive acts, thus involving decision making on the basis of precaution with a high margin of safety.

Probability in both disciplines deals with both *risks* and *effectiveness* (efficacy) but also causation (although in that context more often referred to as uncertainty, see section “Epistemological cornerstones concerning causation, probability, and evidence”, on “uncertain”/“probable” knowledge about causation). A risk is the probability of an adverse effect occurring. Examples of risks in medicine are the probability of a disease/condition occurring, and drugs/interventions having adverse effects themselves. In toxicology, risks include the probability of a xenobiotic substance causing a toxic effect or risk reduction measures having adverse effects, either because they may be insufficient or because they may have intrinsic harmful effects. Pharmacotoxicology is where the two disciplines overlap. Lastly, effectiveness refers to the probability of an intervention/prevention either having a beneficial effect or lowering at least the probability of the adverse health effect to occur. Decision making is performed by evaluating the balance between risks and effectiveness. Interventions or preventions are then made accordingly. Importantly, knowledge about causation is achieved through experimentation and is always probable and of limited certainty (see section “Epistemological cornerstones concerning causation, probability, and evidence”).

Given the importance of causation and probability, a brief exploration of the terms (probable) knowledge, causation, and evidence may be helpful in the present context. From the vast body of philosophical thought on epistemology and causation, only two hallmarks can be briefly considered in this study.

**Knowledge creation and epistemic justifications: evidence**

Knowledge may be viewed as an accurate mental representation of the truth: knowledge ideally is somehow “tracking the truth” and, in this sense, is essential for formulating theories that describe causal relationships assumed to be real/true. According to a classical definition by Plato, true knowledge is a “justified true belief.”2,3 To put it very simple we approximate truth by mental representations and concepts, which we consider as knowledge of the real world based on a network of justifications that we can provide for our propositions. Consequently, propositions that are not sufficiently justified are not regarded knowledge. One traditional way to test whether certain propositions or beliefs are true is to assess whether they are justifiable by specific reasons (“epistemic justification”). Reasons that sufficiently justify propositions (making their relationship to the truth “evident”) are called “evidence.” However, justifiability alone is insufficient to declare a belief as true knowledge.4 One way of addressing this is to allow only those justifications which are linked to predefined standards/criteria regarding specific requirements of quality and plausibility, for example, coherence with existing knowledge.
Epistemological cornerstones concerning causation, probability, and evidence

The modern theory of causation began with David Hume and his fundamental work “Treatise of human nature” (1739), especially the section “On Knowledge and Probability.” With probability, Hume does not mean the statistical probability (e.g., frequency) of events. Instead, Hume uses the term probability in an “epistemic” sense, referring to uncertain knowledge such as obtained from empirical data. He showed that knowledge about causation cannot be created by logical reasoning alone. Beforehand, the connection between cause and effect had been assimilated to that of ground and consequent in logic. Instead, he realized that by employing empirical observations all knowledge, including that about causation, remains probable knowledge.2,3

Another more recent epistemological cornerstone, particularly relevant in this context, is the 20th century school of logical empiricism (logical positivism) and here mainly the group of philosophers known as “Wiener Kreis”/“Vienna circle.” Their philosophy was concerned with the clarification of sentences and their logical relationships as modules to express mental concepts of truth.5 One of the most prominent proponents of the Vienna Circle, Rudolf Carnap, formulated the “verification principle” only verifiable sentences are meaningful. The only admissible evidence for verifications is empirical data in combination with logical rules and probability theory. Only this leads to new and consequent in logic. Instead, he realized that causation cannot be created by logical reasoning and empirical data. He showed that knowledge about knowledge cannot be ultimately verified. Consequently, Popper suggested his in his publication (in German) “Logik der Forschung” (1934) that science should use a methodology based only on falsification. Popper rewrote this work later in English and published it in 1959 as “The Logic of Scientific Discovery.”

Defining evidence – an attempt

Epistemological evidence is inseparable from the concept of justification and thus verification and falsification (note that the prepositions most commonly used with evidence are “for” and “against”). Although “information” or “data” (e.g., on/about an object) are more observational, neutral terms, evidence is ultimately a term of valuation and judgement. It directly relates to confirmation and justification of a specific assumption, belief, proposition or hypothesis. The term evidence is only meaningful in the context of a relationship between a proposition and the “truth”. Information on its own cannot justify an assumption, it needs to be assessed and used for that purpose. Only through the act of evaluating whether information stands criteria of quality and coherence with the context-dependent relevant network of existing knowledge (plausibility), information may qualify as evidence for or against a proposition. Any unevaluated information is thus in a “pre-evidential” state. Evidence may be regarded as a qualifying attribute attached to a specific piece of information found to be acceptable as a justification for a specific proposition. Evidence is information that succeeds in connecting a proposition to its confirmation as true knowledge. Evidence “translates” a proposition into (true) knowledge.

In summary, a very general definition of evidence for a given proposition A may be phrased as follows: Evidence is any piece of information that increases the probability of that proposition A being true. More specifically, “scientific evidence” may be defined as any information/data that, for specific predefined criteria of data quality and plausibility with regard to existing knowledge, provides a sufficiently strong justification for/against the assumption that a specific hypothesis is true (falsifying or verifying it). Thus, information that qualifies as evidence functionally plays the role of an “epistemic justification” justifying or rejecting a specific proposition/hypothesis with a certain probability. Clearly, whether a certain piece of information qualifies as evidence is highly context-dependent and concerns the specific epistemic relationship between a

current context of evidence-based approaches in toxicology:

- Classificatory – evidence confirming a hypothesis
- Comparative – evidence supporting a first hypothesis more than an alternative hypothesis
- Quantitative – evidence supporting a hypothesis to a given degree
particular piece of information and a specific proposition: a particular piece of information may qualify as evidence in the context of hypothesis A, but may fail to qualify as such in the context of hypothesis B.

Conclusion

What criteria does toxicological information have to fulfill to be regarded as evidence for a hypothesis/proposition so that we may regard the hypothesis as probable knowledge? Especially when dealing with more complex decision-making scenarios in toxicological practice, the available information is often sparse and stems from various data sources that are difficult to integrate. Decisions on whether a piece or set of information constitutes “evidence” for the verification/falsification of hypotheses with relevance to public or environmental health may be very difficult. Toxicological methods of knowledge creation and knowledge-based decision-making will benefit from an examination of the standards used for translating information into evidence in complex decision-making situations. Adequate evidence-based tools such as more structured, standardized approaches allowing the conscientious and complete evaluation and objective integration of information through the use of predefined criteria, and evidence levels will improve toxicological decision-making, making it more transparent and thus consistent.

References

2. Evidence-based medicine – a possible model for evidence-based toxicology?

2.1 Translation of evidence-based medicine into practice

EAM Neugebauer

Institute for Research in Operative Medicine, Medical Faculty, University of Witten/Herdecke, Cologne-Merheim, Germany

Introduction

Evidence-based medicine (EBM) was defined by David Sackett, one of its founders, as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.” The practice of EBM involves integrating individual clinical expertise with the best available external evidence from systematic research. The aims of EBM are behavioral changes through evidence-based recommendations, resulting in improvements in health care quality, improvements in clinical research, decreased costs, and the supply of both health care employees and patients with the best relevant information. Toxicology may benefit from the incorporation of some tools and processes used by EBM. EBM may be characterized by the interaction between the physician, the patient, and research evidence. The three areas of EBM must be balanced appropriately. It would be wrong to over-rely on the physician’s clinical expertise, which is fallible, but also it would be wrong to over-rely on research evidence. Evidence may be biased due to flaws in study design, performance and analysis, or may be inapplicable due to inappropriate outcomes being used. Research evidence may also be insufficient due to lack of data or may not be applicable due to local circumstances. Finally, the needs of the patient should not be neglected. Patients have their own varying requirements – such as better quality of life, prevention of social stigma, or better self-management – which may differ from those of the physician. The task of the physician is to integrate the best relevant evidence and the values of the patient in the decision-making process. This article will focus on how the best medical practice may be brought to the patient.

Using guidelines to introduce EBM into practice

A critical change that was introduced by EBM into medical practice is the way in which evidence is collated. A specific question is formulated, which directs the search of the evidence. The relevant evidence is collected, systematically evaluated, and then acted upon. In the past, this process was performed as needed by individual physicians or small medical groups; today, this process is performed by professional societies in the production of national and international medical practice guidelines. Guidelines are produced through an integration of published literature and expert knowledge combined with good consensus methods. A guideline based on an incomplete or biased evaluation of the literature can lead to inappropriate recommendations. Guidelines, therefore, exhibit the highest validity if they are based on systematic evaluations of the evidence available. Nevertheless, the application of an adequate consensus technique is as important as the evidence base. A formal and transparent consensus process should be applied, using clear-cut procedures. Throughout the production process,
the best attainable standards of methodology must be maintained.

Practical considerations

The first step in producing evidence from published literature is the formulation of answerable questions to conduct a systematic, transparent literature search. Each question must include terms corresponding to the relevant patient group, the intervention, the control (or comparison), and the outcome. It has to be shown that all relevant information was included; the information that was excluded should also be shown. The evidence is then graded by experts, using an established evaluation scheme, according to the level of evidence, 1–5, 1 being the highest. Consequently, recommendations are made, levels of evidence translating to different grades of recommendation, A–C (or D), A being the highest. Finally, the grades are modified according to factors such as sample size, presence of control, risk–benefit ratio, and ethics. The clinical recommendations are then translated into guidelines, consensus being reached through a transparent and defined method (Box 1).

Quality assurance and implementation of guidelines

Concern over the quality of guidelines led to the convening of an EU group to produce the Appraisal of Guidelines – Research and Evaluation (AGREE) statement (2001), “a validated, international consensus instrument to evaluate guidelines.” The AGREE Instrument awards points for quality and rigor in six areas: scope and purpose, stakeholder involvement, rigor of development, clarity and presentation, applicability, and editorial independence. Guidelines scoring the most points are thought the best to improve the health care of patients.

There are many practical problems in the implementation of guidelines. Lack of knowledge or acceptance of the guidelines may block their implementation. There may also be concerns over the applicability, large-scale effectiveness, or benefit derived from the guidelines. For example, there may be no hospital benefit from a set of guidelines because they counter hospital policy. To oppose these problems, it is important that there is a broad publication policy in a number of different formats. They must also be locally adapted, and they must also have a demonstrated connection to quality assurance indicators.

Finally, the development of guidelines should always be accompanied by an evaluation of effectiveness. It is very important to develop quality indicators and evaluate whether the guideline has achieved expected levels of compliance, clinical effectiveness, and cost-effectiveness. In Germany, for example, there is a health services research fund available for estimating the effect of the implementation of guidelines on changes in health care.

Conclusions

In summary, EBM introduced into medicine the search and use of the best available external evidence, the matching of this evidence with the knowledge of the physician about the patient, and the explicit integration of the patients’ values into the decision-making process. Today, best practice is brought to the physician and the patient through the adoption of national and international guidelines. The EBM processes of knowledge acquisition, including transparency of information sources, retrieval and appraisal, and production of guidelines may be used as templates for similar processes in an evidence-based toxicology approach.

References


5 Council of Europe. developing a methodology for drawing up guidelines on best medical practices (Rec 2201/13). 2002 (ISBN 92-871-4788-4) and ZÄFQ 2002; 96 (Suppl. III).


2.2 Evidence-based health care and the Cochrane Collaboration

RW Scherer

John Hopkins Center for Clinical Trials, John Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Introduction

The purpose of this article is to describe evidence-based medicine (EBM), and more generally, evidence-based healthcare (EBHC), especially as practiced by the Cochrane Collaboration in the production of systematic reviews. EBHC may serve as a model or template for an evidence-based toxicology (EBT) and the first task of toxicologists in moving toward an EBT may be in identifying what EBHC tools, concepts, and practices are applicable in their discipline.

A brief introduction to evidence-based health care

EBM was defined as the integration of the best research evidence with clinical expertise and patient values.1 The driving force behind it was the realization that health care decisions were not being based on current valid scientific evidence. Traditional sources of information – journals, experts, textbooks – do not provide current information when a patient is present and a care decision must be made. Also, although much research is published, only a fraction is valid, important, and applicable to decision-making. Furthermore, constant changes in technology and treatment options make it difficult for a practitioner to keep up-to-date and choose the best decision for a specific circumstance.

EBHC is a process with several steps to deal with a patient situation: 1) frame the question to guide the search for information and eventual decision-making; 2) find the best evidence to answer this question by systematically investigating carefully selected sources; 3) critically appraise the evidence using rigorous criteria and standardized assessments; 4) apply the evidence to patients’ healthcare, involving integration of the evidence with stakeholders’ views and expertise.2

Framing the question

The clinical questions posed in EBHC may involve incidence or prevalence of a condition, therapeutic interventions, screening, prevention, diagnostic accuracy, prognosis, harm, etiology, or cost. Of these, harm is most relevant to toxicology. The type of question asked determines the kind of evidence needed to address it. Controlled clinical trials provide the best evidence for answering many questions, including those related to therapeutic interventions, harms, or diagnostic accuracy. However, questions related to incidence, prevalence, or etiology are best answered using evidence from surveys or follow-up studies. The question also informs the EBHC process. Well-formulated questions determine the criteria used to search for the evidence, to select studies, and to abstract data.

Although decision-making may take into account evidence from many different kinds of study designs, in EBHC, the evidence is weighted based on the potential for confounding factors. The kinds of evidence used for intervention questions may be viewed as a pyramid showing these study designs (Figure 1). In this scheme, systematic reviews of randomized trials and randomized controlled trials are preferred evidence for decision-making in EBHC. This type of evidence rating is used by many organizations producing clinical guidelines.

The Cochrane Collaboration

The Cochrane Collaboration aims to facilitate well-informed decisions about healthcare by preparing, maintaining, and promoting the accessibility of
systematic reviews of the effects of health care interventions. As part of this process, the Collaboration also identifies randomized and controlled clinical trials, trains individuals to conduct systematic reviews and associated tasks, and assesses other reviews.

A Cochrane systematic review is a review of existing, up-to-date knowledge that uses explicit, scientific methods. They are produced using a structured and transparent process, involving a comprehensive search for relevant studies and explicit methods of appraisal. Reviewers summarize results of similar but separate studies, either qualitatively or quantitatively (through meta-analyses).

Cochrane systematic reviews are published quarterly in The Cochrane Library, which also houses many databases holding related information. The Cochrane Database of Methodological Reviews, for example, includes reviews on methods used to perform systematic reviews.

The Cochrane Collaboration has many guiding principles underlying its strengths. The first is international collaboration in producing systematic reviews and studying methods of producing them. The reviews therefore have international scope and are subjected to an extensive peer review process. A second principle is to build on the enthusiasm of others. This is important as most work is performed by volunteers. Further principles are aimed at ensuring review quality by minimizing bias (e.g., industry funding is avoided) and improving the quality of work by appraising the methods used. Other principles are aimed at ensuring relevance, keeping up-to-date (reviews are updated every 2 years), avoiding duplication, and maintaining continuity. Finally, the Collaboration promotes access and participation, including active consumer involvement.

Nevertheless, limitations remain. First, the Collaboration currently publishes reviews of health care interventions only. The Collaboration is currently developing methods to perform diagnostic accuracy reviews and umbrella reviews (comparing multiple interventions for the same health care condition). Second, review topics are currently chosen unsystematically although the Collaboration recently provided internal funding to examine methods of prioritization. Finally, the independence of each Cochrane Review Group limits complete standardization of methods (e.g., search methods, updating, review length, and complexity of writing).

The structure of the Cochrane Collaboration

The Cochrane Collaboration is organized into a number of entities. National Cochrane Centers support contributors within their assigned regions, act as a regional focus for Cochrane activities, are the main contact point for the public, provide training, hand-search regional journals, and promote accessibility of The Cochrane Library. Cochrane Review Groups, designated by specific diseases or medical specialty, prepare and maintain reviews of interventions to prevent, treat, and diagnose health problems. Cochrane Fields represent populations, groups, or types of care that overlaps multiple Cochrane Review Group areas. Methods Groups coordinate reviews on the methodology of conducting reviews; provide advice, training and support; conduct methodological research; and monitor review quality. The Collaboration recognizes the value of including consumers at every stage of review development to help make the reviews more usable as an end product. The Consumer Network therefore encourages consumer involvement in writing or reviewing reviews, hand searching journals, or preparing plain language summaries. Finally, the Steering Group is the governing body.
of the Cochrane Collaboration and sets policy and strategies.

The Cochrane Colloquium, held annually, features an introduction to the Collaboration, training workshops, and a scientific meeting. Communication across the Collaboration and with the public also occurs via a webpage, http://www.cochrane.org, providing information and resources.

**Conclusions**

EBHC is the integration of best research evidence with clinical expertise and patient values, and is a dynamic and flexible process. The Cochrane Collaboration is an international organization dedicated to help people make well-informed health care decisions on treatment interventions. The Collaboration’s strengths include an international collaborative effort and rigorous methodology. The process of EBHC may provide a model for EBT with the structure and functions of the Cochrane Collaboration providing a model for the organizational and methodological development of EBT.

**References**

3. Core problems and case studies

3.1 Key challenges of toxicology

ME Cebrián

Sección de Toxicología, Centre for Research and Advanced Studies (CINVESTAV), Mexico City, Mexico

Toxicologists today face a number of key challenging uncertainties. One problem is that, although dose–response curves may be constructed over a standard dose range (e.g., for carcinogens in animal models), the data may not extend to lower doses often encountered in the environment. The risks of relevant doses of a compound may therefore not be accurately known because it is not known how to correctly extrapolate the dose–response curves.1

A second challenge relates to uncertainty factors, elements commonly criticized in risk assessment. Inter- and intraspecies differences in reactions to toxins caused by variations in the pharmacokinetics and pharmacodynamics of a substance are accounted for arbitrarily by these factors2,3. A more evidence-based approach should be taken.

A third problem is that in the characterization of risks, most assessments lack an accurate exposure assessment. A toxicological effect is a function of both dose and duration of exposure. At low doses, a long duration of exposure may be necessary to induce a toxic effect, and this has been inadequately explored in human toxicology and in human health.

References

3.2 Quantification of uncertainty within and between species and the role of uncertainty factors

L Edler

Department of Biostatistics, German Cancer Research Centre, Heidelberg, Germany

Introduction

Uncertainty and uncertainty factors (UFs) are important topics in toxicology, which likely become relevant for evidence-based toxicology (EBT). In general terms, the aim of assessing uncertainty is to evaluate the limits of our knowledge. In particular, uncertainty analysis (UA) addresses the type and extent of toxicological knowledge used, model selection, choices of modeling assumptions, statistical options during modeling, and statistical and computational assumptions during analysis (technical uncertainty). In contrast to UA, the purpose of UFs is specific, namely, to convert a defined risk-characterizing parameter into a relevant risk management or health risk guidance parameter in the presence of uncertainty. This article discusses uncertainty and UA methods as well as UFs and their refinement as specific tools.

Uncertainty

In discussing uncertainty, one must be precise about the definitions of risk, variability, and uncertainty. An agreed definition of risk (e.g., in food risk assessment) is the probability and severity of an adverse effect or event occurring to humans or the environment following exposure to risk sources under defined conditions (e.g., see Renwick, et al.). Probability in this definition should not be confused with uncertainty, as probability of an event should not be confused with the variability of a parameter.

Informed decision-making requires evaluation of uncertainty as an integral part of risk assessment. Uncertainty may be characterized by defining the overall structure of the process, enumerating the alternatives within this structure that may influence the outcome, assessing the likelihood of the alternatives, and assessing the influence of likely alternatives on the final outcome or decision, carried over from risk assessment to risk management.

Uncertainty occurs as structural, statistical and/or technical uncertainty (e.g., see Edler). Structural uncertainty occurs when the model itself bears uncertainty, either due to incomplete or insufficient knowledge about biological, physiological or toxicological mechanisms, or due to the existence of more than one plausible model of a specific phenomenon. Statistical uncertainty corresponds to statistical variation. Technical uncertainty, for example, reflects measurement error. All three sources of uncertainty should be accounted for in risk assessment.

Uncertainty factors and chemical-specific adjustment factors

UFs are numerical values capturing as much as possible of the uncertainty, when risk guidance values have to be determined. Specifically, UFs are extrapolation factors based on changes of toxicokinetic or toxicodynamic data from animals to humans (inter-species) and between humans (intraspecies). UFs have been primarily used to convert the no observed adverse effect level (NOAEL) into a risk guidance value. Most acceptable daily intake (ADI) values in food risk assessment have been obtained in this way.

Default UF values are 10 for both inter- and intra-species UFs, being multiplied to give a total value of 100. Although much used, they lack a clear scientific basis and have been the subject of much argument. Attempts have been made to replace the default UFs by incorporating scientific data by chemical-specific adjustment factors (CSAFs)
This approach – first adopted as consensus committee opinion – is now widely accepted and included in guidelines. In this scheme, each 10-fold factor is subdivided to allow for differences in toxicokinetics and toxicodynamics. Chemical-specific information, if known, then replaces default assumptions. For example, if the toxicodynamic interspecies difference is known and already accounted for, the default 2.5 value source of uncertainty for this is removed, leaving an interspecies factor of 4.0 instead of 10 as the remaining UF value.

The strength of the approach is that chemical-specific toxicological data are used for better understanding uncertainty and for refinement of its assessment, leading to factors different from default UFs, in most cases leading to reduced values. In that sense, usage of CSAFs reduces uncertainty. Limitations are that the approach depends on the validity of the subdivision of effects into toxicokinetic and toxicodynamic ones and on the availability of appropriate data upon which the specific numerical subfactors are based. Mostly, CSAFs are based on the comparison of mean parameter values (interspecies) and on the difference between means and percentiles (intraspecies). The approach requires in-vivo toxicokinetic data and/or in-vitro measurement of elimination, ideally combined with a physiologically based toxicokinetic model and/or in-vivo or in-vitro toxicodynamic data using equipotent doses or concentrations. Interspecies CSAFs are then calculated from the difference between the mean values for the appropriate parameters in the test species and humans. Intraspecies CSAFs can be obtained from statistical data on relevant toxicological endpoints measured on a population level. The factor for human variability can so be calculated as the ratio between the mean or median for the population and the parameter value corresponding to a predefined and accepted proportion of the population covered. It should be noted that those data may not exist in specific risk assessment. On the one hand, this limits the utility of the approach; however, it exhibits data gaps to be resolved for the reduction of uncertainty.

CSAFs can also be adjusted for sensitive subpopulations, such as children or ethnic subgroups. Adjustment factors are determined from percentile differences of the general population and the sensitive group based on chemical- and population-specific data. The chemical-specific subfactor depends upon the proportion of the population, which needs to be covered, the presence of potentially at-risk subgroups, and the variability within the main population and the subgroup. The approach of Renwick and Lazarus develops the adjustment factor such that it covers a predefined proportion of the whole population (the healthy adult and sub-group distributions). This analysis takes into account differences between the main population and the subgroups for the mean parameter estimates and the coefficient of variation.

Further refinements have been proposed, including the use of specific pharmacokinetic or physiologically based toxicokinetic model data to replace the kinetic component of the interspecies UF, and the use of hierarchical probabilistic models to derive UFs from population distributions, rather than the means.

Conclusions
Uncertainty should be fully assessed in meaningful risk assessments. The use of CSAFs or other approaches to replace default UFs require compound-specific data in order to reduce uncertainty. In accordance with the principle of integrating all available toxicological information, usage of CSAFs can contribute EBT.

References


3.3 Trovafloxacin: a case study of idiosyncratic or iatrogenic liver toxicity – molecular mechanisms and lessons for pharmacotoxicity

J Borlak1,2

1Fraunhofer Institute of Toxicology and Experimental Medicine (ITEM), Hannover, Germany; and 2Centre of Pharmacology and Toxicology, Medical School of Hannover, Hannover, Germany

Introduction

Serious hepatic injury is the leading reason for drug removals and restrictions, and idiosyncratic drug toxicities, most frequently targeting the liver, remain a challenge to the medicine industry and regulatory authorities. Such toxicity is frequently dose-independent and not related to the pharmacology of the drug. Also, idiosyncratic adverse drug reactions (ADRs) usually cannot be predicted during preclinical or clinical drug development. The inability of animal models to detect these events may lie with interspecies differences or the lack of accurate disease models. Furthermore, clinical trials are often insufficiently large to detect rare events such as idiosyncratic ADRs. They typically only become obvious when drugs are used in large human populations, for example, postapproval.

Idiosyncratic toxicities may be iatrogenic (resulting from medical treatment). There are a number of factors that may lead to iatrogenic toxicity, including genetic polymorphism associated with the enzymes regulating the pharmacokinetics of the drug, an allergic response to the drug or one of its metabolites, and increased toxicity consequent to inflammation induced by bacterial infection. Therefore, idiosyncratic drug toxicities are in some degree host-dependent. Drug-induced mitochondrial injury may also be a factor. In addition, it is thought that drugs with the potential to cause idiosyncratic drug toxicity may regulate common physiological or biochemical processes. In practice, idiosyncratic reactions probably occur because of a number of factors, both biological and environmental. Understanding the responses to drugs that cause such toxicities, at the molecular level, may provide the means both to elucidate the mechanisms of toxicity and to predict idiosyncratic toxicity during drug development.1

Hepatic injury related to trovafloxacin

Trovafloxacin, a fluoroquinolone antibiotic targeted against bacterial DNA gyrase and topoisomerase IV, was developed by Pfizer, Inc. and was approved by the United States Food and Drug Administration in 1997. During its development, no cases of hepatic failure were reported in approximately 7000 patients to whom it was administered. After approval, more than two million people received trovafloxacin. In a short period of time, 150 cases of liver toxicity were reported, including 14 cases of acute liver failure. Trovafloxacin was withdrawn from the European market in 1999 and remains under very strong restrictions in the United States.2,3

Trovafloxacin is not the only drug of its class to be associated with such idiosyncratic toxicity. Post-marketing surveillance of other fluoroquinolones (e.g., trovafloxacin, temafloxacin, and grepafloxacin) revealed serious ADRs, including potentially lethal livertoxicity and cardiotoxicity, associated with their use – also not identified during drug development and resulting in withdrawal. However, no such significant ADR was associated with older fluoroquinolone agents, including ciprofloxacin, ofloxa-cin, norfloxacin, and levofloxacin.2,3

Also, such hepatic toxicity could not be detected in rat in-vivo assays by trovafloxacin administration alone.2 Experimental studies should therefore
address why fluoroquinolones differ so profoundly in causing hepatotoxicity. It was hypothesized whether toxicogenomics could predict these differences or other unwanted drug effects.

Profound modulation of genes by trovafloxacin in human hepatocytes

Hepatic cell toxicity associated with trovafloxacin becomes apparent in primary human hepatocyte cultures incubated with the drug at peak physiological concentrations for 120 h. This was not apparent with ciprofloxacin, a drug not associated with idiosyncratic hepatic toxicity. Experiments using DNA microarrays were performed to assess the genomic responses to trovafloxacin underlying this toxicity, and identify which kinds of metabolic pathways are perturbed, providing the evidence to generate testable hypotheses.

Trovafloxacin treatment of cultured human hepatocytes had a profound effect, eliciting >1000 gene regulations (Figure 1). The genes affected were associated with important physiological functions, including metabolism of fatty acids, lipids, and carbohydrates; and synthesis of proteins, cholesterol, and steroids. Genes associated with DNA repair and RNA transcription were affected, indicating that the drug inhibits mammalian as well as bacterial topoisomerases. Mitochondrial function (transport of glucose, electrons, etc) was also significantly disturbed, as were signal transduction pathways (e.g., growth factors, growth factor receptors, receptor tyrosine kinase signaling, and downstream processing of these signals). Trovafloxacin was, therefore, able to affect basic components of cellular biology. Finally, trovafloxacin significantly downregulated anti-inflammatory signals and upregulated proinflammatory signals (e.g., leucotriene B4 and ligands of tumor necrosis factor).

Hepatocyte regulatory protein analysis

To reduce the complexity of the gene expression data, the regulatory proteins controlling the transcription of the genes affected by trovafloxacin were studied using bioinformatic techniques. Genomic footprinting of transcription factor (TF) binding sites reduced the complexity of over 1000 trovafloxacin-regulated genes to approximately 20 master regulatory proteins. Most of these TFs are involved in regulating fatty acid and glucose metabolism and mitochondrial function. There was at least partial concordance with previous data.

One key TF that was affected was hepatic nuclear factor-4α (HNF-4α). This protein is 1 of 30 HNFs belonging to five families. HNFs are responsible for regulating approximately 5000 metabolic processes throughout the liver. Furthermore, HNF-4α is also an important developmental protein for the liver. Any modulation of HNF-4α would, therefore, be expected to have profound effects on hepatocyte physiology.

Western blot analysis revealed that HNF-4α expression was severely reduced by trovafloxacin treatment of hepatocytes. Electromobility shift assays also showed that DNA-bound HNF-4α expression was inhibited. Moreover, comparison with other fluoroquinolones (e.g., ciprofloxacin and levofloxacin) demonstrated that this inhibition of HNF-4α expression was unique to trovafloxacin.

The implication of these data is that trovafloxacin, by inhibiting the expression of HNF-4α, suppresses the function of a network of genes that govern major metabolic processes, lipid and carbohydrate metabolism, and mitochondrial biology.

Furthermore, these studies show that human hepatocyte cultures are an important tool for understanding and detecting idiosyncratic ADRs. The European Medicines Agency (EMEA) guideline on the early detection of liver toxicity, taking account of such findings, is a major breakthrough in the prevention of ADRs.

Conclusions

The clinical trials used to develop drugs are insufficiently large to predict rare events, such as idiosyn-
Idiosyncratic ADRs, which are also not detected with conventional animal studies. Primary human hepatocyte culture studies provided important information on a mechanism of idiosyncratic liver toxicity, and may have been used to predict the excessive hepatotoxicity of trovafloxacin, and of other drugs.

References


4 Meier, T, Reymann, S, Niehof, M, Borlak, J. Unraveling trovafloxacin’s molecular basis of hepatotoxicity (submitted).

4 Toxicological decision-making on hazards and risks – status quo and way forward

4.1 Current concepts and schemes of science-driven toxicological decision-making – an overview

C Portier1,2

1National Institute of Environmental Health Sciences, North Carolina, USA; and 2Office of Risk Assessment Research, Principal Investigator, Environmental Systems Biology, North Carolina, USA

Introduction

This paper addresses ways in which evidence is currently being used to make decisions. Within risk assessment, evidence is currently used for two basic purposes: hazard identification and characterization. A number of guidelines have been published to assist the toxicologist in how to use evidence in determining the existence or nonexistence of a hazard. Examples of these guidance frameworks will now be described.

Hazard identification

The process produced by the International Agency for Research on Cancer (IARC) for evaluating evidence for hazards falls into three carefully guided steps.1 First, individual studies for the putative hazard are assessed for scientific quality, statistical power, control of confounders and sources of bias, and presentation and clarity. Second, the evidence is then weighed in the areas of cancer in humans, cancer in animal models, and mechanistic (and other) evidence, and a judgment is made in each of these three areas (Figure 1).

For the areas of cancer in humans and cancer in animal models, the evidence falls into one of four categories. “Sufficient evidence” is used when the evidence suggests a causal relation between the agent and carcinogenicity. “Limited evidence” is used to describe evidence suggesting an association between the agent and carcinogenicity, but not one strong enough to make a causal inference. “Inadequate evidence” is used when the strength of the evidence is insufficient to reach a conclusion about a causal association. The category “evidence suggesting a lack of carcinogenicity” is used when a causal relation between the putative hazard and carcinogenicity is not detected although the quality of the evidence is high enough to detect one if present. Different criteria exist for different kinds of evidence. Mechanistic and other evidence is judged as being “weak,” “moderate,” or “strong” and whether the mechanism is likely to be operative in humans.

Finally, an overall evaluation is made on the putative hazard, ranging from group 1 (carcinogenic to humans) to group 4 (probably not carcinogenic to humans). Conclusions are reached primarily by different combinations of human and animal evidence ratings. For example, if the human evidence is of “limited” strength, but the animal evidence is “limited,” “inadequate,” or “suggests a lack of carcinogenicity,” the agent is classed as group 2b, a possible human carcinogen. Mechanistic evidence is also used in the overall evaluation and may be pivotal.
if the human evidence is not conclusive. For example, if an agent has evidence that is “inadequate” in humans and “sufficient” in animals, strong evidence demonstrating that the mechanism operates in human tissues may cause the agent to be moved from group 2b to 2a, a probable human carcinogen.

Hazard quantification

The US Environmental Protection Agency (EPA) runs the Integrated Risk Information System (IRIS) program, which is the repository for those risk assessments run by the EPA and that set quantitative exposure standards in the United States. The IRIS process results in a hazard being characterized in terms of estimates of daily oral or inhalation exposure to the human population that is likely to be without an appreciable risk of deleterious effects over one lifetime. There are a number of steps that may be involved in the process:

- Use of data from human studies in health assessments
- Accounting for less-than-lifetime exposure durations
- Qualitative and quantitative use of mode of action data in noncancer and cancer assessments
- Benchmark dose modeling and selection of the benchmark response
- Evaluation and use of physiologically based pharmacokinetic modeling
- Accounting for life-stage and subpopulation susceptibility in uncertainty factors
- Use of data-derived uncertainty factors
- Characterization of uncertainty in noncancer and cancer analysis
- Use of time-to-tumor modeling for cancer assessments

The important point in this discussion is that the process is carefully guided, with at least one guidance document existing for each step.

Challenges

Despite these rigorous guidance frameworks, risk assessment faces three main challenges: the first is the large and increasing number of substances that have not been subject to risk assessment, and the recent legislation both in Europe and in the United States that requires they be assessed. Moreover, the increasing number of substances in the human environment interact with each other in ways that must also be assessed. Also, novel biological and physical agents, such as genetically modified food and nanomaterials, pose unknown risks. The second challenge is the expanding knowledge base in toxicology. It is not yet fully understood how all the available data may be used optimally, particularly in regard to evaluating toxicity mechanisms. The third challenge is that the science of toxicology is continuously improving and that the working practices of toxicologists have to adapt to accommodate these improvements.

In response to such challenges, the United States National Toxicology Program (NTP) aims to change their toxicological practice, as described in their 25-year “roadmap” for the 21st century. Their aim is “to move toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused upon a broad inclusion of target-specific, mechanism-based biological observations.” In their roadmap, they describe a three-tiered system. The first level would involve screening many in an inexpensive, easily mechanized, high-throughput process that would uncover many mechanistic links and build a scientific base from which to make decisions about
the use of animal models. The second level would involve medium-throughput screening in integrated living organisms to yield complex inter-related mechanistic information. The final level would consist of the definitive evaluation, involving bioassays yet to be designed. The first level of testing is currently being validated.

Conclusions

Hazard identification and characterization are governed by the carefully guided evaluation of evidence as prescribed by national and international organizations. However, toxicology faces great challenges and must change to face these challenges. It is likely to change to a largely predictive science, driven by bioinformatics. Intensive research will be needed to fulfill this goal.

References

2 Environmental Protection Agency. The integrated risk information system. www.epa.gov/iris [accessed 20.4.2009].
4.2 Applying an evidence-based approach: arsenic as a health risk

E Silbergeld

Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

Introduction

Evidence-based medicine (EBM) offers a number of tools applicable to toxicology, including the systematic review. This article presents an account of using systematic reviews to assess arsenic as a health risk.

Arsenic is a known human carcinogen (group 1, IARC 2008).1 High-level environmental exposure to arsenic is associated with drinking water and is a major public health concern. However, control of arsenic in drinking water is expensive. The goal of these analyses was to evaluate the evidence associating arsenic exposures with type 2 diabetes mellitus or cardiovascular disease (CVD); to integrate analyses of epidemiological, clinical, and experimental studies; and to reach conclusions on the risks of those outcomes associated with environmental arsenic.2,3

The nature of systematic reviews

The systematic review process can be defined as follows. First, there are pre-existing conditions for the exclusion of information, for example, reviews and editorials, which are considered inadequate to be considered as evidence as well as publications dealing with the same cohort. Second, there is a specified evidence search strategy, with defined and revealed databases and search terms, allowing replication by others. Third, there are pre-existing criteria for weighing information and turning it into evidence.

In deciding search strategies and weighing criteria, the nature of the evidence must be considered. All stakeholders should decide what constitutes evidence in the context of the review according to rigorous criteria. The purpose of the review, the relevant decision-making it is expected to aid, should also be considered.

The fourth characteristic of the systematic review process is the use of standard analytical procedures. Finally, the conclusions are drawn up as a qualitative ranking or quantitative set of conclusions in a standardized manner.

The searching and weighing of evidence

It is important to recognize that the evidence-based approach does not replace the importance of weighing or other acts of judgment; rather, this approach makes the elements and criteria of judgment more explicit. In these analyses, the search terms and databases were described and revealed. It was decided to consider experimental as well as human evidence. The rationale was that biomarkers, often used as preclinical or intermediate endpoints in epidemiological studies, are often developed from experimental studies. The exclusion criteria for experimental studies were nonmammalian cells, noncellular systems, single-dose (i.e., nonchronic) animal studies, and studies of arsenic-containing drugs.

The vast majority of papers failed the exclusion criteria. For the type 2 diabetes mellitus review, 1029 references were identified but 994 were excluded. Only 19 epidemiological studies and 29 experimental studies were included. Similarly, for the CVD review, 1217 papers were identified and 1162 excluded. In total, 29 epidemiological studies were included.2,3

The criteria for weighing epidemiological evidence were those of Longnecker et al., who adapted EBM randomized, controlled, trial criteria for observational studies.3 These were validity and standardization of the outcome diagnosis; individual exposure information; adequate information on other risk factors for outcome or exposure; standard and
similar methods of data collection for all participants; and appropriate response rate (particularly among control or nonexposed patients), intensity of ascertainment, and loss to follow-up, independent of exposure status.

Information on which studies met what criteria was published as checklist tables in the reviews. The few human studies admitted had many limitations. For both reviews, no epidemiological study met all criteria and most lacked adequate information on the diagnosis or on individual exposure. More seriously for the CVD review, most cohorts were not adequately followed, so it was not known what the total yield of the outcome might have been in the study populations.2

Integration, analysis, and presentation of the data

In a systematic review, the data may be presented by comparing a common metric in individual studies or by pooling the raw data in a meta-analysis. In these analyses, common metrics were found in terms of exposure and outcome. In all cases, exposure had to be inferred from levels in drinking water as there were no individualized exposure data. For studies that reported more than two categories of exposure, the study outcomes were then expressed in terms of adjusted relative risk and plotted against exposure. Relative risk data may also be presented on a “forest plot,” as used by the Cochrane Collaboration (Figure 1).

It was difficult to integrate the experimental and mechanistic studies into the reviews. The endpoints used were only surrogates for CVD or diabetes mellitus. Also, the experimental studies used much higher doses than those in the human studies, and the data were contradictory and difficult to compare across studies even across those using the same dose range in the same species. Furthermore, there were relatively little mechanistic data available from human studies. Finally, the mechanisms by which arsenic induces CVD or diabetes mellitus are unknown.

---

![Figure 1](image-url)  
*Figure 1  Forest plot showing relative risk of cardiovascular disease (CVD) for patients exposed to arsenic in drinking water.*
Conclusions of the systematic reviews

Significant limitations were common to all studies, especially those related to exposure information in both reviews, and outcome assessment in the CVD review. For the CVD review, it was concluded that evidence from highly exposed populations was “consistent with a role for arsenic in atherosclerosis”; however, information from lower exposures was “inadequate to answer this question.” No estimation of the magnitude of the effect or of the dose−response relationship could be made.¹ For the type 2 diabetes review, the results were inconclusive, but there was suggestive evidence to associate high-level exposures and an increased risk of diabetes. No dose−response analysis could be conducted due to lack of individual exposure data.²

Conclusions

Systematic reviews have many strengths. First, transparent methods are used throughout. Another strength is that study limits are clearly indicated with consistent criteria across all studies. Also, quantitative assessment tools, such as odds ratios, may be used. Finally, the process is efficient and rapid.

However, such reviews also have limitations. Most retrieved studies were excluded, but as the analysis still used inadequate studies, it is possible that equally useful data were discarded. Also, the methods are not readily applicable to experimental studies, which cannot be directly integrated. It may take much time to develop standard methods to integrate experimental data. Finally, the quality of the human data is limited by the lack of controls in epidemiological studies.

Generally, there are strong reasons for introducing EBM tools into toxicology. The urgency of decision-making has increased with an increased burden of work from many different regulatory regimes, as well as the concerns of the public. At the same time, many decisions are difficult and controversial. For these reasons, it is important to ensure fairness, transparency, and participation in decision-making. However, it is also true that there are already many rules for translating information into evidence and it could be questioned whether a new and very difficult process should be embarked upon.

References

4.3 In-vitro genotoxicity tests to detect carcinogenicity: a systematic review

L Müller

Toxicology Project Leader, Non-Clinical Safety, F. Hoffmann La Roche, 4070 Basel, Switzerland.

Introduction

Systematic correlation analysis of genotoxicity cell assays and rodent cancer assays was introduced by the National Toxicology Programme (NTP) in the 1980s. The NTP concluded that the Ames test had low sensitivity (~45%) but good overall concordance and that adding other assays improved sensitivity but lowered specificity. Since this work was started, papers re-evaluating the NTP data have been published, including criticisms that the chemical set used was irrelevant because it was small or unrepresentative (e.g., pharmaceuticals), and that it contained few noncarcinogens thereby casting doubt on the performance of the assays. Furthermore, there have since been numerous changes to the genetic toxicology assays and increased pressure to reduce animal use. This article describes a recent re-evaluation of the ability of various genetic toxicology in-vitro tests to predict rodent carcinogenicity and proposed changes to international guidelines governing their use. The outcome of this work is important in the context of improving in-vitro tests for detection of genotoxic potential and to design possibly more predictive in-vitro tests than currently conducted.

Specificity and sensitivity of in-vitro assays

In a concordance study, the Ames test and the mouse lymphoma assay, in-vitro chromosome aberrations test, and micronucleus test were evaluated in their ability to characterize 554 carcinogens and 177 noncarcinogens. In this study, there was a wide separation between the degrees of specificity and sensitivity of individual tests (Figure 1). The Ames test had the lowest sensitivity but the highest specificity of the assays tested. Specificity was poor for all mammalian cell tests.

Sensitivity could be improved by combining tests but only at the expense of specificity. The low specificity of two or three combined in a battery, as currently used to test pharmaceuticals and other chemicals for human use, may be associated with a high false positive rate.

These findings were confirmed by Matthews et al. (2006), using a wider database that included many pharmaceuticals and using conservative weight of evidence criteria (favoring sensitivity over specificity) to classify chemical carcinogens. Therefore, these conclusions apply to most chemicals, including new pharmaceuticals. One implication of high false positive rates associated with testing batteries is the possibility that potentially useful pharmaceutical compounds are unnecessarily being discarded. Hence, the design of more predictive in-vitro genotoxicity assays continues to be a worthwhile effort.
Relative predictivity

Data have also been analyzed in terms of relative predictivity (RP). In such analyses, RPs are the ratios of carcinogens to noncarcinogens giving positive or negative genotoxicity results, and from sets of positive or negative results indicate how more likely the chemical is carcinogenic than noncarcinogenic, or vice versa. An RP of 2.0 was set as the significant limit.

The positive RP for the Ames test was approximately 2.5, higher than for any other single test. However, negative predictivity was consistently low in all tests. No individual test had both positive and negative RPs of >2.0. The prediction of carcinogenic potential from test batteries was only considered meaningful if the RP directions from the assays were consistently negative or positive. When this was achieved, only two batteries (the Ames plus MLA test battery and the Ames, MLA, plus CA test battery) showed positive and negative RPs of >2.0.10

Revision of the ICH genotoxicology testing battery

Such evidence has led the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) to conclude that the test battery for genotoxicity should be changed. Its steering committee recently decided to establish a maintenance process of the file guidelines for pharmaceuticals (from October 2006). A revised guideline has been agreed by the ICH expert working group and is currently in the consultation process.11

One proposal put forward is to design a test battery consisting of a combination of the Ames test and two in-vivo tests thereby discontinuing with mammalian in-vitro tests under these conditions. This would increase specificity, but potentially decrease sensitivity, in the testing battery. In the context of reduction, refinement, and replacement, the in-vivo tests would ideally be integrated into a standard rodent toxicity study, whereas, currently, in-vivo genotoxicity tests are normally performed as stand-alone acute high-dose tests.

Another proposal includes retaining the mammalian in-vitro cell tests but reducing the highest tested concentration of an agent to 1 mM instead of the currently used 10 mM top concentration for noncytotoxic compounds. According to a recent industry survey of compounds testing negative in the Ames test, genotoxic effects were indicated in mammalian in-vitro cell tests in 40–60% of compounds only at concentrations of >100 μM and a further 10–20% of compounds at concentrations of >1 mM (unpublished industry survey). Such high concentrations are normally unrealistic for pharmaceutical exposure and it can be argued that positive test results obtained at such high concentrations only are normally not relevant for human risk.

Unlike for most chemicals, genuine risk assessments may be performed with pharmaceutical compounds because these agents are used at high purities often already at early stages on clinical development (normally >98%) and because there is extensive animal and human pharmacokinetic, metabolism and exposure information available. This information can be used to assess whether there is need to explore genotoxic activity at very high concentrations.

Human exposure for 313 small molecule drugs were reviewed, using data from Goodman & Gilman’s The Pharmacological Basis of Therapeutics (2002)12 (L Müller, unpublished). For approximately 70% of drugs, peak steady state plasma concentrations range between 10 ng/mL and 5 μg/mL. Few compounds – mostly antibiotics, antiviral agents, and chemotherapy compounds – achieve concentrations greater than 10–100 μg/mL. Notably, this category does not include most modern pharmaceutical agents. As the molecular weight for most compounds is in the range of 200–400, a plasma concentration of 10 μg/mL is equivalent to approximately 25–50 μM. If it is accepted that genotoxicity is driven by peak exposure, it is reasonable that the upper limit of tested concentrations in mammalian in-vitro cell tests should be reduced from 10 mM to at least 1 mM since concentrations beyond 1 mM will never be achieved anywhere in the human body with proper use of a pharmaceutical.

Conclusions

The specificity of existing in-vitro mammalian cell assays for genotoxicity is a matter for concern. In a period of increasing pressure to replace in-vivo tests with in-vitro assays, increased demands for in-vivo studies to verify positive in-vitro data are likely, especially for compounds requiring more defined risk analysis, such as pharmaceuticals. This is particularly troublesome in industries where in-vivo testing procedures are restricted. Moreover, as human exposure to pharmaceuticals is generally below 100 μM, in many cases, positive results achieved at concentrations above 1 mM are of doubtful human relevance.
In moves toward what could be considered evidence-based use of in-vitro genotoxicology tests for pharmaceutical compounds, the ICH expert working group has proposed that an upper limit of 1 mM should be used in mammalian cell genotoxicity tests and that the stand-alone in-vivo genotoxicity test should be replaced by integrating its endpoint into a standard rodent toxicity study. Further, European Centre for Validation of Alternative Methods (ECVAM) has designed a specific program to look at possibilities to further improve the predictive performance for in-vitro genotoxicity tests.

References


10 Kirkland, D, Aardema, M, Müller, L, Hayashi, M. Evaluation of the ability of a battery of 3 in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. *Mutat Res* 2006; 608: 29–42.


4.4 The proposed replacement of the no observed adverse effect level with benchmark dose levels in food risk assessment

L Edler

Department of Biostatistics, German Cancer Research Centre, Heidelberg, Germany

The no observed adverse effect level (NOAEL) has been the traditional point of departure used in food risk assessment. However, as reference point for estimating health-based guidance values, it has serious deficits. For example, it depends on the design of the respective study, in particular the sample size, and does not control the risk at the NOAEL dose. The benchmark dose (BMD) approach avoids such fallacies and has been recommended to replace the NOAEL. It has been used widely by the US Environmental Protection Agency and occasionally by the European Food Safety Authority, the Joint Food and Agriculture Organization of the United Nations, and the World Health Organization Expert Committee on Food Additives. In simple terms, it is applied by selecting the data, selecting the best fitting model, applying statistical linkage, estimating the model parameters, and implementing the outcome given a benchmark response level in the observable range of effects. The BMD makes full use of the dose–response data, is a full scale modeling approach, and tends to require more and higher quality data than the NOAEL. It requires a minimum of statistical knowledge and dedicated software (although not complex), and the reasons for its use must also be transparently explained to all stakeholders. Those additional challenges to the risk assessor are however more than compensated by being a more science-based method than the NOAEL.

Reference

4.5 Evidence-based individual toxicological analysis

P Guzelian

Section of Medical Toxicology, University of Colorado School of Medicine, Denver, Colorado, USA

Introduction

From the viewpoint of evidence-based science, guidelines are arrived at, whenever possible, through a systematic and objective review of the evidence. From the viewpoint of evidence-based science, guidelines are arrived at, whenever possible, through a systematic and objective review of the evidence. One model of EBM has been defined as the integration of research evidence (general propositions), patients’ preferences and actions, and clinical state and circumstances by clinical expertise. However, the design of this model ensures that the decision over whether a patient should receive a particular intervention, even when there is compelling evidence to support it, reverts largely to the clinical judgment of the individual clinical expert. As a key aim of EBM is to replace subjective judgments with the use of clear evidence-based guidelines in clinical decision-making, this model cannot be viewed as a step toward this goal.

In a 2005 paper, David Eddy introduced the term “evidence-based individual decision-making” (EBID). In his view, an EBM-based decision is a combination of an EB-based guideline proposition plus the decision to apply that guideline to an individual patient made in a methodical, objective basis. This paper discusses how, using a similar model, toxicological analysis may be applied to the individual patient.

Evidence-based individual toxicological analysis

In toxicology, we face the problem of determining the cause of harm in individuals. For example, arsenic is known to be a cause of skin cancer. For the individual patient with skin cancer, it must be decided whether arsenic is the cause of the disease in that person. Rather than use clinical judgment or opinion, a methodology with an evidence-based objective approach is preferred. Based on the example of EBID, an evidence-based methodology – evidence-based individual toxicological analysis (EBITA) – of deciding causation of harm in individual patients is proposed. This methodology applies readily adaptable principles in EBM to toxicology.

The EBITA approach has three steps: collection of the data, collection of the knowledge, and the synthesis of this data and knowledge into a series of objective and categorical steps to aid individual decision-making in toxicology (Box 1).

In the first step, data relevant to specific causation are collected, that is, information is gathered that

<table>
<thead>
<tr>
<th>Box 1 Steps in an evidence-based individual toxicological analysis. Adapted from Sackett, et al.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Collect the data</td>
</tr>
<tr>
<td>• Source</td>
</tr>
<tr>
<td>• Exposure</td>
</tr>
<tr>
<td>• Dose</td>
</tr>
<tr>
<td>• Diagnosis</td>
</tr>
<tr>
<td>2) Collect the knowledge</td>
</tr>
<tr>
<td>• Frame the question</td>
</tr>
<tr>
<td>• Assemble the literature</td>
</tr>
<tr>
<td>• Appraise the literature</td>
</tr>
<tr>
<td>3) Synthesis of data and knowledge</td>
</tr>
<tr>
<td>• General causation (qualitative toxicity/literature precedence)</td>
</tr>
<tr>
<td>• Exposure and dose</td>
</tr>
<tr>
<td>• Time course</td>
</tr>
<tr>
<td>• Alternate causes</td>
</tr>
<tr>
<td>• Coherence (mechanism/biological plausibility)</td>
</tr>
</tbody>
</table>

Correspondence to: Philip Guzelian, MD (Clinical Professor of Medicine (retired)), Section of Medical Toxicology, University of Colorado School of Medicine, Denver, Colorado, USA.
Email: pguzelian@mac.com
will allow the toxicologist to decide whether the individual had the disease cause. Four sets of data are needed. The source of the agent, governing how it is present in the human environment, is an important consideration, for example, whether arsenic is present as a free metal or as an insoluble complex in treated wood. Other data needed are exposure, dose, and the formal identification of the disease in question (a medical diagnosis).

In the second step, information relevant to general causation is collected, that is, information about the toxicological agent is gathered that will allow the toxicologist to decide whether it is likely to be the cause of the disease. The first stage in this step, as in EBM, is to frame the question. To obtain relevant, pertinent information as a result of the search, the question must be framed using specific delimiters important to toxicology, as suggested in Box 2.

These delimiters are important in framing the question because they restrict what literature to include and what literature not to include in the analysis. For example, it is important to delimit the question to acute, subacute, or chronic effects of an agent, as evidence about one type of effects is likely to be irrelevant to another. To take a further example, it is also important to specify the route of exposure, for example, inhaled chromium may be carcinogenic but swallowed chromium is likely not to be.

In the third step, that data and knowledge are synthesized in a specific causation analysis for that individual, that is, a decision on whether that agent caused the disease under question in that individual. A proposed set of five criteria, derived from causation criteria commonly referred to as Hill’s criteria, are applied to the individual patient. The criteria are general causation (the strength and consistency of the literature), the amount of exposure that individual had and what dose they may have received from that exposure, the time course of the disease (onset and resolution, toxicokinetics), whether there are other explanations that better account for the patient’s illness, and coherence (whether the proposition is biologically plausible).

A very similar set of causation principles were proposed by David Sackett (1985) in the context of deciding whether an individual suffering an adverse effect developed it because that individual received a specific drug. These criteria, developed for use in EBM, are practical principles readily applicable to toxicology and support the view that relevant toxicological information in humans are often available and usable for decision-making.

Conclusions

The principles discussed above constitute an evidence-based process allowing the specific causation of a toxicological disease to be determined. It is proposed that this series of specific steps may form the basis of a practical, readily applicable process that may be used by the toxicologist to determine the causation of putative toxicological disease in individual patients.

References


4.6 Toxicogenomics and biology-based modeling framework for health risk assessment

D Sarigiannis

Human Exposure to Environmental Stressors and Health Effects, European Commission – Joint Research Centre, Institute for Health and Consumer Protection, Physical and Chemical Exposure Unit, Ispra, Italy

Introduction

A major challenge in modern risk assessment is to integrate and put in context increasing amounts of information from advanced analytical techniques coming from many different disciplines. One solution is to use biologically based models and toxicological discovery techniques using gene expression microarrays. This article describes how toxicogenomics – in which the activity of a toxin on living tissue is characterized by profiling its effects on genetic material – in combination with biologically based models may be used to assist risk analysis in a framework that brings together more biological evidence of early effects to support the formation of plausible hypotheses on mechanism of toxic action.

The Joint Research Centre approach

Biological structures and associated behavior may be described mathematically, encoded in computer programs, and then simulated. However, to integrate biological data and place those into context, in a way that effectively supports risk assessment, contextual frameworks are needed. These could be supplied within biologically based computational models.

The Joint Research Centre uses an integrated multilayer computational approach, toxicogenomics, linking genomic to proteomic analysis. The process starts by using whole genome discovery systems to analyze study tissue from animals or humans. This is followed by gene identification, quantitative polymerase chain reaction (PCR) validation, and statistical evaluation. Subsequent gene pathway analysis using PANTHER1 is then integrated with proteomics and metabonomics, using bioinformatic systems, to produce biomarker and systems toxicology models that may be used in risk assessment and screening of chemicals.2 This approach complements traditional toxicological studies both in vivo and in vitro and provides the mechanistic underpinning for integrating biomonitoring, epidemiological, and clinical data in an evidence-based toxicology framework. It bears particular potential when tackling complex problems, which require better mechanistic understanding of toxic action, such as the assessment of chemical mixture toxicity.

Genetic modulation by indoor air mixtures plus polycyclic aromatic hydrocarbons

This toxicogenomic approach was used to assess the toxicological effects of indoor air mixtures (IAMs). Indoor air is contaminated by mixtures of volatile organic compounds, including aromatic compounds, aldehydes, and terpenes.3 Real-life exposure is, however, compounded by polycyclic aromatic hydrocarbons originating from outside sources. IAMs were therefore compared with ambient air, that is, indoor air plus PAHs.4

Human A549 lung cell and HaCaT skin cell cultures were exposed to air mixtures, and gene expression analysis of the whole genome was used to assess subsequent genetic effects. Hierarchical cluster analysis of the expression data sets showed that pure IAM-modulated gene expression differently than the mix of IAM plus PAHs.

The number of genes in A549 cells modulated by the air mixtures was small relative to the whole gene set analyzed. Compared with control cells,
approximately 15,000 genes were expressed equally when exposed to pure IAM or to IAM plus PAHs. However, only 325 genes had at least a fourfold change in expression compared with control when cells were exposed to pure IAM. The corresponding number for IAM plus PAH exposure was 289. Of these genes showing high levels of modulation, 148 were modulated by both air mixtures. The differences and commonalities in gene expression induced by the two air mixtures may be modeled as oxidative stress response pathways, and the biological mechanisms induced by these exposures may be depicted visually. These differences and commonalities in gene expression were then validated using PCR in selected gene panels. Distinct differences in the biological underpinning of eventual physiological responses were apparent.

**Genetic modulation of IAM components**

The next step in this study was to examine the effects of different classes of IAM contaminant: aromatic compounds, aldehydes, and terpenes. It was found that air mixtures containing each of these classes of compound exhibited different levels of cytotoxicity in different cell lines (keratinocytes-HaCaT, lung cells-A549, and hepatocytes-HepGe2). The pure IAM mixture depressed cell survival by approximately 10–60%. Aromatic compounds alone had little effect on cell survival in all three cell lines. However, aldehydes alone suppressed cell survival by approximately 35% in HepGe2 cells and terpenes alone decreased cell survival by approximately 20–30% in A549 and HepGe2 cells.

Gene expression analysis revealed differential modulation of expression dependent upon the class of IAM contaminant. PCR validation verified the gene expression data showing that 21 genes of 47 tested were modulated by the compounds.

**Modeling the toxicological effects of environmental exposure to airborne contaminants**

The AIRMEX (European Indoor Air Monitoring and Exposure Assessment) Project evaluated outdoor and indoor concentrations, and personal exposure, relating to organic contaminants in air in selected cities throughout continental Europe. There were much greater levels of indoor pollution and personal exposure apparent in southern Europe. In addition to the quantitative differences between regions, there were also qualitative differences:

<table>
<thead>
<tr>
<th>Mixture A (%)</th>
<th>Mixture B (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(northern Europe)</td>
<td>(southern Europe)</td>
</tr>
<tr>
<td>20 benzene</td>
<td>10 benzene</td>
</tr>
<tr>
<td>40 toluene</td>
<td>60 toluene</td>
</tr>
<tr>
<td>10 ethylbenzene</td>
<td>10 ethylbenzene</td>
</tr>
<tr>
<td>30 xylenes</td>
<td>30 xylenes</td>
</tr>
</tbody>
</table>

In northern Europe, indoor air contained relatively more benzene; in southern Europe, there was relatively more toluene. The gene modulatory effects of different doses of these mixtures were evaluated in A549 cells. The number of genes having at least twofold change in expression showed a dose–response relationship for the two air mixtures. Gene modulation was highest for Mix A at all doses. Hierarchical cluster analysis revealed clear clustering distinctions in gene expression data after exposure to Mix A or B.

Following this, data were used to construct a physiologically based pharmacokinetic model for the quaternary mixture including a detailed description of benzene metabolism which was then validated against human population data. The predicted and actual data coincided well.

From such analyses, it was possible to estimate internal concentrations and biologically effective doses for the active substances in different tissues. For example, xylene and toluene venous concentrations in the alveoli were estimated to be markedly higher than those for benzene and ethylbenzene. However, ethylbenzene concentrations were higher than those of benzene. This is similar for other tissues. Further analysis showed that tissue concentrations of individual components of the mixture depended upon the presence of the other components. This system dynamics modeling was then coupled with pathology modeling. The resultant model for lifetime risk of cancer from benzene exposure was found to fit previously published data well.

Finally, hierarchical population modeling was performed for use in a health risk assessment, taking into account variation in the pathological, physiological, and exposure parameters, plus measurement errors. In this context, Markov chain Monte Carlo is a useful tool to convert individual risk estimates to population risks, taking uncertainty explicitly into account. Using this tool, the mean value \(15 \times 10^{-6}\) was found to be approximately half the World Health Organization estimate \(33 \times 10^{-6}\), which is calculated using a linear dose–response function with no threshold (see Figure 1). However, over the 95th percentile, the risk estimate overlaps with this value.
Conclusions

Biologically based dose–response models built on the basis of better understanding of the biological mechanisms resulting in the observed toxicity can provide more reliable and conservative population health risk estimates for environmental exposure to VOC mixtures than linear dose–response models based on high-dose extrapolations. Current health protection legislation should take into account exposure scenarios representing realistic conditions, namely, account for the interaction effect of chemical mixture components in the environment, in the workplace, and in consumer products.

References

4.7 Biological modeling as a method for data evaluation and integration in toxicology

HA Barton

National Center for Computational Toxicology, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina, USA

Introduction

Evidence-based methods rely on a range of analytical methodologies to assess and characterize the degree of consistency of scientific data. Although there has been a focus on statistical (or semiquantitative ranking) methods, biological models have also been recognized as valuable approaches to analyze and integrate data from multiple study designs, including in-vitro and in-vivo assays, often representing different measured endpoints. Biological models are explicit mathematical representations, in varying levels of detail, of biological hypotheses and knowledge of physical systems. They are used to generate and test hypotheses of alternative biological descriptions and to make predictions. Predictions may be made for improving risk and safety assessments or for evaluating uncertainties. Predictions may also be made to characterize population variability, reflecting such factors as polymorphisms or life stages.

Contexts of biological models

Biological models may be used in a number of toxicological contexts. One is the source-to-outcome continuum. Exposure to an agent from the environment leads to transport of that agent to body tissues, governed by the pharmaco- or toxicokinetics of that agent. Concentrations of the agent in tissue, referred to as tissue dosimetry, elicit a response via pharmaco- or toxicodynamic processes.1 Toxicodynamic processes are also referred to as the mode of action, which specifies the nature of the interactions between the chemical and the body that lead to toxic responses.2 Early cellular changes are typically associated with adaptive stress responses, but as the system is perturbed, further toxicity pathways can be initiated leading to cellular injury and morbidity.3

The context of toxicity pathways is central to the National Research Council’s (NRC) recently published vision of toxicity testing in the 21st century.3 In this, the NRC proposed moving away from using 1) animal data, except for targeted testing for specific purposes; 2) high-dose toxicity studies; and 3) histopathology as the dominant adverse end-point. Instead, they proposed moving toward using data from (largely human-derived) in-vitro models of toxicity pathways combined with computational modeling to predict molecular perturbations of toxicity pathways in humans at environmentally relevant exposure levels. In this vision, computational models are integral to data analysis and interpretation. Acceptance and success of such an approach will be highly dependent upon developing evidence that it adequately predicts human toxicological responses.

Evaluating consistency of data and integrating diverse data

Vinyl chloride-induced angiosarcoma

Vinyl chloride is a trans-species carcinogen associated with liver angiosarcoma in rodents and humans.4 The comparative potency of vinyl chloride across species was assessed by quantitatively comparing its carcinogenic potency using data from inhalation and oral rodent bioassays and from human epidemiological studies.4 A physiologically based pharmacokinetic (PBPK) model was developed to support target tissue dosimetry for cancer risk assessment. The human risk of angiosarcoma...
was estimated using a low-dose linear analysis from the model-predicted dose metric of the metabolism normalized for liver volume. The results were expressed as 95% upper confidence limit risk estimates (per million) for lifetime exposure to 1 part-per-billion vinyl chloride in air.

Risk estimates from the epidemiology studies were in the range of 0.4–4.2. These approximately coincided with those from the animal inhalation and diet studies (range 1.1–5.17). The risk estimates from the rat oil gavage study (8.7–15.7) were the only outliers. On the basis of this interspecies consistency and other considerations, the US Environmental Protection Agency (EPA) decided that no further interspecies extrapolations were required for its risk analysis of this compound. These analyses demonstrate the power of biological modeling, in this case PBPK modeling to evaluate the consistency across data in different species based upon internal dosimetry. Such methods are likely to be essential for cross-species comparisons in a framework for evidence-based toxicology.

**β-Chloroprene-induced lung tumors**

In a similar PBPK biomodeling analysis, the comparative potency of β-chloroprene to cause lung tumors across animal species was assessed. Analysis revealed a strong relationship between exposure and response for mice, a relationship for Fischer rats, but no apparent relationships for hamsters or Wistar rats. In an attempt to obtain concordance between different test animals for the same range of exposures to β-chloroprene, a dose metric of predicted lung metabolism was taken. A concordant relationship was then found between response and the internal dose, that is, the dose of chemical metabolized per gram of lung tissue per day. The mice responded most strongly as a result of much higher internal doses. However, an unexplained difference between rat strains was still apparent.

**Endocrine system modeling: gene-to-tissue response in the prostate**

In this example, a biological model was created that described not only pharmacokinetics but also pharmacodynamics, specifically regarding the testicular-hypothalamic-pituitary axis and prostate function. Understanding the regulatory mechanisms of this axis is important for assessing the reproductive effects of environmental and pharmaceutical androgenic and antiandrogenic compounds. The objective of this study was to model prostate dynamics for intact and castrated rats.

A biological model for the dynamics of androgen synthesis, transport, metabolism, and regulation of the adult rodent ventral prostate was developed. The model described the pharmacokinetics of testosterone, 5α-dihydrotestosterone (DHT), and luteinizing hormone (LH). Also included were pharmacodynamic feedback loops regulating testosterone and LH. The model simulated maintenance of the prostate as a function of hormone concentrations and androgen receptor (AR)-mediated signal transduction. The processes involved in prostate size and function include cell proliferation, apoptosis, fluid production, and 5α-reductase activity, each process being controlled through the occupancy of a representative gene by androgen-AR dimers (Figure 1).

After calibration, the model accurately simulated the castration-induced regression of prostate fluid secretion and mass observed experimentally. The model also accurately predicted serum testosterone, DHT, and AR levels following castration. This model brings together quantitatively in-vitro and
in-vivo studies of a variety of endpoints in a way that few other methods can.

**Systems biology modeling**

Systems biology is the modeling of entire biological systems as quantitative frameworks for studying biological processes spanning from genes and proteins through to cells, organisms, and populations. It can be very complex, requiring teams of investigators with complementary skills and knowledge, but can also be very informative. Development of systems biology approaches represents a major frontier of biological research, which needs to be considered in the development of evidence-based toxicology.

**Conclusions**

Mathematical models help to evaluate consistency among data sets and integrate diverse kinds of data and information making them ideal candidates for methods to assist in evidence-based toxicological analyses. They are finding many uses in toxicology and are playing increasingly important roles in risk assessment. However, their use faces a number of challenges. First, there are problems of acceptance in the toxicological community, particularly regarding public decision-making. Second, because of their mathematical and computational base, making the processes transparent to other toxicologists is often difficult. Finally, there are numerous technical challenges, in particular, regarding systems biology for response processes and characterizing uncertainty in model outputs.

**References**

4.8 Current schemes for decision-making in toxicology

C Griesinger¹, S Hoffmann², A Kinsner¹, S Coecke¹ and T Hartung³

¹European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; ²TÜVRheinland, BioTech GmbH, Köln, Germany; and ³Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Breakout group results

In an evaluation of current schemes for decision-making, several areas were identified for improvement. First, regulatory requirements are often inconsistent (e.g., there is no human exposure data requirement for the assessment of agrochemicals). There should be a drive toward standardization so that substances or materials are assessed to a uniform and comparable standard. Second, decision-making could be improved by increasing the usage of reported data. Formal voluntary data submission processes should be set up, together with schemes for incorporating the data into the decision-making process. Pharmacovigilance schemes from the pharmaceutical industry may serve as models. Third, the ways in which data are evaluated and integrated may be improved. Klimisch¹ scores provide a good basis for assessing data quality, but additional information is required on the adequacy of studies regarding specific questions. Despite the importance of improved standardization regarding data evaluation (e.g., by more elaborate standards or scoring systems), it is important to allow for a certain flexibility in the way data are interpreted. Fourth, there is no current link between advances in scientific knowledge and the processes of decision-making. Processes such as formalized decision analyses are needed. Finally, uncertainty is not addressed explicitly in decision-making schemes and uncertainty factors need to be less arbitrary and more evidence-based.

There were specific recommendations regarding ecotoxicology. Ecosystem health is a growing concern, and decision-making is increasingly based on in-situ diagnostics of low ecological levels of exposure using biomarkers of pollution. A strength of this approach is that biomarkers are measurable in individual organisms, providing early indications of ecotoxicity. Weaknesses include methods of extrapolating mechanistic evidence to the levels of populations and ecosystems.

Current decision-making schemes such as those of the International Agency for Research on Cancer (IARC), the United States Environmental Protection Agency (EPA), and the Organization for Economic Cooperation and Development (OECD), and the European Union Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation are science-based but do to a large extent depend on authority-led judgment. A comparison of the strengths and limitations of individual decision-making schemes may provide the basis for developing an evidence-based toxicological approach.

Reference

4.9 Current information sources for hazard identification

C Griesinger¹, S Hoffmann², A Kinsner¹, S Coecke¹ and T Hartung³

¹European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; ²TÜVRheinland, BioTech GmbH, Köln, Germany; and ³Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Breakout group results

The types of information available differ for existing and new chemicals. For established chemicals, exposure and adverse effect data may be readily available, for example, from scientific literature or poison control centers. Such information is often unavailable for unapproved new substances for which no practical experience exists.

Toxicological information may be gained prospectively or retrospectively. Prospective information may be gained from in-vivo (including human), in-vitro, or in-silico studies. Data may be obtained using standardized test protocols or novel mechanistic techniques, for example, toxicogenomics. Retrospective information includes epidemiological, post-marketing surveillance, and occupational health data and also information from nonhuman sentinel species studies, for example, fish, wildlife, or livestock.

Before collecting and/or evaluating toxicological information, it is important to assess its source. The quality of the data and the quality control measures used for generating the data should be assessed. An assay should be validated for its reliability and (predictive) relevance, and its mechanistic basis (relevance) should be understood. Finally, the appropriateness of the statistical methods used should be evaluated.

Sources of information should also be evaluated for their relevance to the question being addressed, for example, taking into account interspecies variability. For example, weaknesses of in-vitro systems include the inability to predict pharmacokinetic parameters, whereas in-silico models have this capability but do not provide empirical information. Epidemiological studies should be assessed for population variability and uncertainties over dose and exposure.

Other factors to consider include whether the original data are available for examination and audit. All information should be evaluated for variability and inconsistencies and their potential causes. Laboratory testing (e.g., toxicogenomics) should adhere to validated, explicit test protocols. It is important also to consider the demonstrated applicability domain limitations of test methods based on physical/chemical properties. Lastly, publication bias should be considered.

Recommendations for evidence-based toxicological assessments include increasing the availability of standardized databases of toxicological information. Integrated testing strategies, including informatic methods to integrate and analyze data from different methodologies, are also needed. Finally, the scientific community should develop and disseminate test methods and strategies for assessing the safety of novel materials, for example, nanomaterials.
5 Steps toward an evidence-based toxicology

5.1 Evidence-based tools in toxicological basic research

C Griesinger1, S Hoffmann2, A Kinsner1, S Coecke1 and T Hartung3

1 European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; 2 TÜVRheinland, BioTech GmbH, Köln, Germany; and 3 Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Breakout group results

Basic research is critical to toxicology in that it contributes to the understanding of how biological systems respond to perturbation (mechanistic studies) and to advances in toxicity testing methodologies (e.g., toxicogenomics and in-silico models). Basic research also advances the understanding of inter- and intra-species variability and improving the assessment human population responses (biomarker studies). The identification and validation of biomarkers (relevant to both exposure and outcome assessment) also draws upon basic research. Similarly, basic research is critical to the process of phenotypic anchoring for new computational and toxicogenomic methods.

Evidence-based considerations can enhance the translational relationship between basic research and toxicology. These considerations include criteria for evaluating research data, the standardization of methodologies (e.g., platforms, procedures, genotypic characterization of cells, and quality control), the selection of relevant doses or concentrations to toxicological problems, and the provision of parameter information (e.g., free chemical concentrations in cell cultures).

A number of tools and approaches relevant to basic research in an evidence-based toxicology were identified. With regard to data quality, criteria are required to evaluate evidence regarding relevance and quality and standardized data analysis and presentation. Data quality should also be evaluated by the level of experimental design detail available (in journals or otherwise) and relevance to the toxicological question. Regarding publication, editorial advice and good referee practice should also be encouraged.

An evidence-based toxicology movement should also aim to improve data availability. Original data should be available whenever possible, public-access databases should be encouraged, and publication bias should be addressed (including the publication of negative results).

To apply basic research to new toxicological methods, it is important to understand variability and uncertainty, to apply criteria to data with respect to relevance to humans, and to design or select the most relevant experimental models.

Available evidence-based tools include established validation criteria (e.g., those developed by the Organization for Economic Cooperation and Development) and guidelines such as those for Good Cell Culture Practice.
5.2 Evidence-based tools in toxicological hazard identification

C Griesinger1, S Hoffmann2, A Kinsner1, S Coecke1 and T Hartung3

1European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; 2TÜVRheinland, BioTech GmbH, Köln, Germany; and 3Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Breakout group results

Hazard identification depends on the gathering and evaluation of evidence. In this context, it is important to differentiate between approaches based on expert opinion and “weight of evidence” and evidence-based approaches using predefined methods and criteria. In moving toward an evidence-based approach, clear guidance for decision-making, specific for each class of agent, should be developed. Quality criteria specific for each type of hazard identification tool need to be defined, with levels of quality acceptable for each. General guidelines exist, but they are not sufficient to evaluate evidence quality in a clear and transparent manner. Further specific guidelines are required to evaluate the summed evidence. In developing these guidelines, it is important to consider existing guidance documents.

Careful thought should be given to designing studies to produce the evidence needed. A number of evidence-based hazard identification tools were identified. These include in-silico, in-vitro, in-vivo, and ex-vivo (analyzing animal or human tissue after treatment) tools. In-vitro and in-vivo tools themselves include classical and novel methodologies, (e.g., transcriptomics, proteomics, and metabolomics). Clinical sample data may be used for exposure assessment, metabolomics, and biomarker analysis. These may be representative of patients, the population at large (including susceptible subgroups), workers, or consumers. Market-based information sources include rapid alert systems and other “hot lines,” poison centers, occupational health data, epidemiological studies, consumer protection organizations, postmarketing surveillance, and hospital data. For all such tools, the purity profile of the agent in question should also be included in the assessment. Evidence-based guidelines should take into account all of these tools and should be able to incorporate both retrospective and prospective data and data from susceptible (sensitive) population subgroups.

Evidence-based medicine is a useful template for evidence-based approaches to toxicology, but there are differences between these two branches of science. The evaluation of the differences and commonalities should be part of the process of establishing an evidence-based toxicology.
5.3 Evidence-based tools in toxicological decision-making

C Griesinger¹, S Hoffmann², A Kinsner¹, S Coecke¹ and T Hartung³

¹European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; ²TÜVRheinland, BioTech GmbH, Köln, Germany; and ³Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Breakout group results

The scope of an evidence-based toxicology (EBT) is likely to cover prospective and retrospective systematic evaluations of in-vitro, in-vivo, and human data for risk assessment and management and to provide high-quality evidence relevant to human health and environmental issues. EBT toolboxes should therefore contain data-generating systems, methods for qualifying data, biological and ecological models, data mining methods (able even to integrate physical and chemical information at different levels), and expert judgment, possibly integrated into formal decision-making processes.

In EBT, high importance would be attached to framing the correct question, and both the relevance of questions and criteria of relevance would be important. Criteria for the inclusion and exclusion of studies should be developed, including those for novel methodologies, such as toxicogenomics. Methods to validate assays predicting human effects would also need to pass formal standards, designed using models from other scientific disciplines if not already available. Weighing schemes need to be developed and quality criteria for these weighing schemes established. Their use should address compound class-specific methods and common scaling to integrate in-vivo and in-vitro dose–response data. Finally, guidelines would be needed and prominently published on a broad consensus basis.

It may be possible to break down the process of EBT-based decision-making into modules, including information retrieval, information generation (accounting for data gaps, acute vs chronic effects, rare vs frequent effects, and species and strains of animal), data scoring, data integration, data interpretation, uncertainty assessment, decision-making, and decision documentation as suggested by the charge questions for this breakout group. Evidence-based tools could be applied to each of these modules independently, but, in general, the modules should be assessed in an interconnected manner, depending on the endpoints relevant to risk assessment, and multilayered tests should be applied conditionally to the observations.
5.4 Possible improvement of information sources on hazard and risk

C Griesinger\(^1\), S Hoffmann\(^2\), A Kinsner\(^1\), S Coecke\(^1\) and T Hartung\(^3\)

\(^1\)European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; \(^2\)TUV Rhenland, BioTech GmbH, Köln, Germany; and \(^3\)Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Breakout group results

To improve current information sources, it is first necessary to improve testing methods. This includes adding high-throughput biological profiling for new chemicals in a tiered approach. Here, the high-throughput screening methods used in the pharmaceutical industry, involving an integrated in-silico, in-vitro, and in-vivo approach, may serve as an initial template. In-vitro systems may be improved (e.g., by developing human stem cell applications), and in-vivo/in-vitro test batteries may be optimized (e.g., through minimizing the use of acute toxicity testing and using transgenic mice appropriately).

Testing procedures should focus on human safety and should also be able to predict rare risks relevant in humans, such as idiosyncratic liver toxicity, without using animal models. Novel methods (e.g., toxicogenomics) should complement classical tools. Gold standards for each endpoint should be defined using species-specific predictive biomarkers, and validation criteria, for both tools and decision-making, should be developed. An emerging concept is microdosing, the evaluation of pharmacokinetic and pharmacodynamic endpoints in humans by administering nonactive doses of substances.

Information sources could also be improved by building shared, standardized data banks, especially between regulatory authorities or pharmaceutical companies where parallel experimental and clinical data are available. There is a need for increased banking and integration of data across different methodologies and for availability to this data. This would require a relaxation of patent protection.

Tools for decision-making should be efficient, cost-effective, rapid, reliable, transparent, consistent, and standardized. Legislation should be less prescriptive in scope. Decision-making would also be aided if purity and characterization standards were harmonized internationally. Information should be freely available to the public, using simple terminology. Finally, regulators should be willing to accept new tests and paradigms.

An initiative is suggested to thoroughly analyze and identify areas in toxicology that are already evidence-based and those that need improvement and might benefit from standardized and evidence-based approaches.
5.5 Standardization: an efficient tool to support evidence-based toxicology

A Pirlet

CEN Management Centre, Brussels, Belgium

International Standards, for example, ISO standards, are established norms or requirements that specify uniform technical and scientific criteria, methods, processes, and practices. Standards are usually laid down in formal documents that summarize the relevant methodologies (e.g., standardized reliable testing methods) and promote interoperability and certification of entities that operate in accordance with such standards. A practicable approach toward standardization and the drafting of workable standards is the so-called “standard integrated approach.”

In this approach, the first step is awareness-building: the goal is to identify common difficulties or weakness of current practice relevant for a group of stakeholders and state these problems explicitly. Once the group of stakeholders is aware of overlapping deficiencies of current practice, these can be prioritized to form objectives. The next step is to decide by consensus on the most efficient ways to reach these often complex objectives. This is typically done by a thorough analysis of a repertoire of components, which typically contribute to a successful implementation of standardized procedures. These may, in the case of increased standardization in toxicology, include research and development, extra legislation, standardization of procedures for decision-making on hazards and risk, and other assorted complementary measures, including education and standardized training.

A more standardized approach in toxicological practice would increase the consistency, explicitness, and conscientious use of toxicological information sources for decision-making on hazards and risks and thus supports an evidence-based approach in toxicology.

Reference

5.6 An online portal to evidence-based toxicology

A Kinsner¹, C Griesinger¹, S Hoffmann¹-², S Coecke¹, G Bowe¹, C Campana³ and T Hartung⁴

¹European Centre for the Validation of Alternative Methods (ECVAM), In Vitro Methods Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy; ²TÜVRheinland BioTech GmbH, Köln, Germany; ³Ecoh Media Srl, Somma Lombardo, Italy; and ⁴Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

To support Evidence-based Toxicology (EBT)-related activities and facilitate the practical implementation of EBT, European Centre for the Validation of Alternative Methods (ECVAM) has developed a concept for a web-based “portal to EBT,” an internet platform to search/retrieve topics related to evidence-based approaches such as events, other relevant web-pages, and key documents. Most importantly, the EBT portal is designed to allow the virtual interaction of working groups developing systematic reviews or appraisals pertaining to specific toxicological hypotheses. Experts and working groups will be able to discuss topics in the area of evidence-based hazard and risk assessment and to deposit, search, and exchange relevant information in a structured manner. The portal will allocate tools of efficient electronic document management, which will ease the creation or revision of review documents or any other type of text document. Besides these functions, the portal is foreseen to serve as a publicly available searchable internet library for documents produced in the context of EBT, such as critical appraisals, systematic reviews, meta-analyses, guidelines. Finally, it will provide a reviewed collection of links to gain systematic and structured access to topical Web sites and other resources on issues related to toxicological safety assessments (e.g., ongoing research programs on toxicological hazard assessment, decision making, risk analysis).

To serve above-mentioned public-based and expert-based functions, the portal will be divided into a working zone, restricted to selected groups of active and contributing users, and a public zone, open to all users. The password-protected working zone is designed to allow interaction of expert groups engaged in specific projects with the aim of creating a final EBT document (e.g., a systematic review, a meta-analysis, a statement). This area will be organized according to the different projects and will have a structured workflow and a repository of working documents. The finalized documents will be published in the public area. The public zone will be focused on information retrieval from finalized and approved projects. Thus, core tools for public users will be advanced search methods as well as other tools including online forms, surveys, a newsletter, and a collection of relevant links. One feature foreseen is “connectivity mapping” between documents or topics, essentially a visual searching tool. The “connectivity map” is a graphical plot that visualizes the degree of connectedness of one particular topic (e.g., a study on a particular substance and its potential adverse effects regarding a particular human/environmental health endpoint) to other topics (e.g., the same substance’s effects on other human/environmental health endpoints). By the use of the connectivity map, users can graphically follow up related documents starting from one particular systematic review or critical appraisal.

The EBT portal is intended to constitute a single well-structured and easy-to-use virtual entry point to a living and growing community committed to fostering evidence-based toxicology. The EBT portal will ease the collaboration of experts distributed across the globe in their common goal of increasing the transparency of toxicological hazard and risk assessment. Moreover, it will directly contribute to increased transparency by enabling free access to evidence-based evaluations of toxicological problems and may thus contribute to a growing awareness of the sources of variability and uncertainty in toxicological hazard and risk assessment.

Correspondence to: Agnieszka Kinsner, European Centre for the Validation of Alternative Methods (ECVAM), In-Vitro Toxicology Unit, Institute for Health and Consumer Protection, European Commission Joint Research Centre, Ispra, Italy.
Email: agnieszka.kinsner@jrc.it

© 2009 SAGE Publications 10.1177/0960327109105782
6 Conclusions

I Kimber

Faculty of Life Sciences, University of Manchester, Manchester, UK

The start of a movement for evidence-based toxicology (EBT) represents an exciting opportunity to renew and reinvigorate toxicology. During 3 days, relevant examples were displayed and discussed. These illustrated not only shortcomings and challenges the discipline faces, but also tools and concepts that are now available and which could be exploited to improve the toxicological sciences. For example, it was shown that the there is room for improvement in the sensitivity and specificity of tests used in risk assessment. However, it was also argued that, by definition, tests used for safety assessment will never be perfect and that they should be regarded as tools that if used appropriately and when coupled with expert judgment will aid in developing effective risk assessments. It was also proposed that in this context it is vital that the limitations of test methods are understood and acknowledged. Evidence-based toxicology would encourage and facilitate the development of more transparent and effective decision-making in safety assessment, particularly regarding the design, validation, and the acceptance of new and improved approaches to safety assessment.

This First International Forum Towards EBT should serve to signal the beginning of a new movement in toxicology, involving a continual process of self-examination and improvement.

To this end, delegates at the Forum reached a consensus over the defining characteristics of EBT – shaped into a definition by the Scientific Steering Committee – and a declaration.
7 Annex

7.1 Suggested Charge Questions for Break-out group work

Theme 1: “Taking stock”

Break Out Group 1: “What are current schemes for decision making in toxicology?”

(The results of this group are summarised in section 4.8 under “Current schemes for decision making in toxicology”).

1. What are current decision-making schemes in toxicology?
2. What are the different types of decision making schemes (‘science-based’, ‘authority-based’, ‘evidence-based’, other?), and what are the differences between them?
3. List the decision-making schemes that you know (environmental and human health). Address the follow issues
   a. Evidence levels/quality scoring: Do these schemes provide guidance how to weigh (or assign weight to) different information/evidence in order to reach a decision.
   b. Data / evidence integration: Do these schemes provide guidelines of how integrate data or evidence from various information sources of variable quality and reliability?
   c. Acceptance of schemes: What is the level of acceptance of these schemes of decision-making (global, national, institutional…)?
   d. Documentation: Are there documentation guidelines of how decisions were reached?
   e. Uncertainty: are there guidelines for quantifying the certainty or uncertainty of the information that was used in reaching the decision?
4. Summarise the results in a table listing the different schemes, their endpoints, where they are used (institutes, countries…) and their strengths / weaknesses.

Break Out Group 2: “What are the strengths & weaknesses of current information sources for toxicological hazard/risk analysis?”

(The results of this group are summarised in section 4.9 “Current information sources for hazard identification.”)

1. Prepare an inventory of information sources.
   a. Distinguish between
      ■ existing information and ■ methods for information generation. Address whether methods provide new knowledge (about causal relationships = hazards) or whether are particularly useful for formation/rejection of hypotheses about potential hazards.
   b. What are the pros and cons of the results of all information sources (e.g. ■ the relevance to the human situation ■ the suitability of a method to lead to clear C&L decisions ■ the objectivity of the information gained from each particular source)
c. Are there differences with regard to regulatory, scientific, societal acceptability of different information sources? Why is this, and how could this be changed?
d. On what factors does the reliability of information generated by test methods for depend on? Are some methods useful only for either hazard or risk assessment and why is this?
e. What are the mechanism to assess the performance and reliability of test methods?

2. Summarise this results in a table

**Theme 2: “Use of evidence-based approaches in Toxicology”**

*Break Out Group 3: “Which evidence based tools might be useful in toxicological basic research?”*

(The results of this group are summarised in section 5.1 “Evidence-based tools in toxicological basic research”)

1. **General issues**
   a. What is the objective of any basic research, independent of the scientific discipline?
   b. What does “evidence“ or “evident” mean? How does evidence relate to ■ opinion / belief ■ hypothesis ■ knowledge ■ proof ■ justification ■ information / data
   c. How is the term evidence being used in basic research in general and toxicological basic research in particular?

2. Prepare a table listing (1st column) evidence-based tools (for instance from EBM) versus different areas or steps of knowledge creation in toxicological research – indicate and score usefulness (10) or unsuitability (1) of the evidence-based tools for toxicological basic research.

3. Can you think of additional new tools that are evidence-based and perhaps provide benefits to toxicological basic research?

4. Can you anticipate new technologies, methods or concepts in basic toxicological research? What are their likely impacts on toxicology as a whole? Would such new technologies need new evidence-based tools so that they can be used for end-products of toxicology (decision-making for human health and environment)?

*Group 4: “Which evidence based tools might be useful in toxicological hazard identification?”*

(The results of this group are summarised in section 5.2 “Evidence-based tools in toxicological hazard identification”)

1. General issues: What does “evidence“ or “evident” mean? How does evidence relate to ■ opinion / belief ■ hypothesis ■ knowledge ■ proof ■ justification ■ information / data

2. What are the experimental tools used for hazard identification and what existing information sources can inform about intrinsic hazardous properties of substances?
3. Distinguish between
   a. (evidence-based) assessment of a method/tool for hazard identification and
   b. (evidence-based) assessment of information for hazard identification derived from a) methods/tools or b) existing information
4. What are the similarities between test method assessment in medicine (diagnostic test) and that of predictive tests in toxicology (=regulatory toxicology, pharmacotoxicology, toxicology of vaccines etc.)?
5. Prepare a table listing evidence-based tools (for instance from EBM) versus different tools for hazard identification – indicate and score usefulness (10) or unsuitability (1) of the evidence-based tools for assessing the evidence levels of hazard information.

**Group 5: “Which evidence based tools might be useful in toxicological decision-making?”**

(The results of this group are summarised in section 5.3 “Evidence-based tools in toxicological decision-making ”)

1. How does “Toxicity-testing strategy” (NAS document) relate to a general “decision-making framework (DMF)” or “Intelligent Testing Strategy” as used in REACH?
2. Can we break down the process of toxicological decision-making into the modules of ■ information retrieval, ■ information generation, ■ data scoring/weighing or ■ data analysis according to predefined evidence levels, ■ data integration, ■ data interpretation and – if applicable – extrapolation, ■ uncertainty calculation/description, ■ decision-making and ■ decision documentation?
3. Can evidence-based tools be applied to each of these modules independently or do they have to be assessed in an interconnected way, in their totality? Can you conceive be evidence levels for each of the modules and prepare a table summarising these?
4. How are decision making schemes or frameworks assessed and adapted to progress in science? What mechanisms would be needed to guarantee appropriate renewal?

**Break Out Group 6: “How can we improve the toxicological “toolbox” (i.e. methods and decision-aiding tools)?”**

(The results of this group are summarised in section 5.4 “Possible improvement of information sources on hazard and risk ”)

1. What are desirable properties of such a “toolbox”? e.g. ■ efficient ■ cost effective ■ rapid ■ reliable ■ transparent ■ consistent ■ standardized without becoming a dull tick-box approach ■ traceable with regard to the information used ■ intelligent ■ flexible ■ protective ■ acceptable in terms of societal expectations ■ …
2. To which extend complies the current approach with this?
3. What is needed to work toward such a toolbox? ■ focused test development ■ improved communication of all involved parties/stakeholders, such as regulators, scientists… ■ inclusion of new technologies ■ assessment of tools ■ explore the needs ■ analysis of current status ■ constant review and adaptation ■ …
Theme 3: “Towards a Definition of Evidence-Based Toxicology”

Group 7: “Working towards a definition of an evidence-based toxicology?”

1. What are the hallmarks of toxicology when compared to other life sciences?
2. What may define an evidence-based toxicology versus current toxicological practice?
3. How could evidence-based approaches provide added value?
4. What are the minimum characteristics that may define an evidence-based toxicology?
5. What could be the aims and objectives of an evidence-based toxicology?
Human & Experimental Toxicology

Instructions to Authors

Human & Experimental Toxicology is an international journal which covers all types of chemical harm to living systems. Papers are published on all aspects of experimental and clinical studies of functional, biochemical, immunological and pathological toxic processes and disorders in vivo and in vitro, on their causal mechanisms, and on toxicity in human and animal patients and on its treatment.

Original research papers, short communications and letters are welcomed. Editorials are invited but they should first be discussed with an editor. Announcements of meetings are published.

Review articles
The editors also invite review articles on relevant and current topics in toxicology. Each review should contain a summary, an introduction to the topic, a careful review of the relevant literature, conclusions of the reviewed literature and recommendations, if applicable. The journal will publish both minireviews (about 10 typewritten, double-spaced pages) and full-length, more thorough reviews (not to exceed 50 typewritten, double-spaced pages). The reviews, either minireviews or full-length reviews, should provide a timely view on the topic covered. Figures, figure legends, tables and references are included within the allotted pages.

These reviews should follow the references, figure, figure legend and table system used for original articles published in Human and Experimental Toxicology.

All papers reporting results of investigations. Investigations in patients and healthy volunteers must indicate that informed consent was properly obtained and that the work was approved by the relevant Research Ethics Committee (IERB).

Reports of experiments involving animals must indicate that national legislation or the equivalent local standard for protecting animal welfare has been followed.

The Editor may alter manuscripts to make them conform to the stylistic and bibliographic conventions of the journal.

Submission of manuscripts
Authors should send two copies of their manuscript and illustrations and a third photocopy of the illustrations as well as an electronic version on floppy disk or CD to:

UK: Europe and Far East:
Professor Kai Savolainen, Editor-in-Chief,
Human & Experimental Toxicology,
Finnish Institute of Occupational Health,
Department of Industrial Hygiene and Toxicology,
Topeliuksenkatu 41 A, 00250 Helsinki,
Finland.
E-mail: leila.ahlstrom@ttl.fi.

Americas:
Dr A. Wallace Hayes, Editor for the Americas,
Human and Experimental Toxicology
Department of Environmental Health,
Harvard School of Public Health
298 South Main Street, Andover, MA 01810
USA
E-mail: awallacehayes@comcast.net

The final, accepted version of the manuscript should also be submitted on disk. The disk should be formatted and prepared using IBM, IBM-compatible or Apple computers, and should be clearly labelled with the file name and the software name and release number.

One author should be clearly designated as correspondent, to whom proofs will also be sent, and the full postal address, telephone/fax numbers and email of this person should be given. Neglecting to include this information may result in delay of publication.

The manuscript should be accompanied by a signed statement to the effect that it is not currently under consideration for publication elsewhere and that the content has not been anticipated by previous publications.

Copyright
Illustrations, tables, or quotations taken from other publications are copyright protected and can only be reproduced with written permission from the copyright owner. Send author’s and publisher’s written permission with the manuscript.

Presentation
Manuscripts should be double-spaced (including references) and should be structured into sections as follows: Title page; Summary and keywords; Introduction; Methods; Results; Discussion; Acknowledgements (details of supporting grants, permission for reproduction of copyright materials, etc.): References. Each section should begin on a new sheet. Other subsection headings within these main sections may be used but should be limited. The main title should be explicit and specific (80 character limit) and an abbreviated (running) title (40–50 character limit) should be given. The summary should form a synopsis of the work, at the same time underlining the contribution of the research to the understanding of toxicology; it should be less than 200 words in length. Keywords (3–6) must be supplied.

Style
For style information and lists of standard abbreviations, consult the American Medical Association Manual of Style: A Guide for Authors and Editors, 9th Edition (Baltimore: Williams & Wilkins, 1998). Use metric units throughout the manuscript. The use of footnotes is not permitted. Single quotation marks should be used throughout. Words to be italicised should be underlined. Whenever possible drugs should be referred to by their generic name rather than by a proprietary name. When a proprietary name must be used, it should begin with a capital letter to indicate this.

References
References should be numbered consecutively in the order that they are first mentioned in the text, and these numbers should be inserted above the line each time the author is cited, e.g., ‘...as others have done.2,3’ The full list of references at the end of the article should be arranged in numerical order. Each reference should include the names and initials of all authors, but if there are more than six the first six authors may be followed by et al., the title of the article, the unabridged title of the journal, the year, the volume, and first and last page numbers. Titles of books should be followed by the place of publication, the publisher, and the year. References to ‘unpublished observations’ or ‘personal communications’ should be mentioned in the text but not included in the list of references. The titles of periodicals should be abbreviated. The style is as follows:


Illustrations
Artwork should be supplied as EPS files, or TIFF files at a minimum resolution of 300 dpi for superior reproduction. Lettering should be planned for 50% reduction; text should be readable after reduction. Illustrations should be referred to as Figure 1, Figure 2 etc.

Tables
Tables ideally should not have more than 80, and certainly not more than 100, characters to the line (counting spaces between columns as 4 characters) unless absolutely unavoidable. Each table should be on a separate page with its caption. All tables should be specifically referred to in the text.

Proofreading and E-prints
Authors will receive proofs, which they should correct and return as soon as possible. Access to 25 free e-prints will be provided; the corresponding author will receive one complimentary copy of the journal for each contributor.

SAGE Open
If you wish your article to be freely available online immediately upon publication (as some funding bodies now require), you can opt for it to be included in SAGE Open upon payment of a publication fee. Manuscript submission and refereeing procedure is unchanged, but on acceptance of your article you will be asked to let SAGE know directly if you are choosing SAGE Open. For further information please visit http://www.sagepub.co.uk/sageopen.sp

If funding has been provided for your research, convention dictates that you acknowledge the source and grant information in the Acknowledgements section of your article. Please note that in some cases, grant funders require authors to attribute the funding source, describe the role of the study funding body and include the relevant grant reference numbers.

Business correspondence and enquiries relating to advertising, subscriptions, back issues or reprints should be addressed to: SAGE Publications, 1 Oliver’s Yard, 55 City Road, London, EC1Y 1SP.