Science and the Law: Analytical Data in Support of Regulation in Health, Food, and the Environment

Science and the Law: Analytical Data in Support of Regulation in Health, Food, and the Environment

William G. Town, Editor
Kilmorie Clarke Ltd.
London, United Kingdom

Judith N. Currano, Editor
University of Pennsylvania
Philadelphia, Pennsylvania, United States

Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

ACS Books Department

Preface

The ACS symposium from which this book is derived aimed to explore the interaction between science and the law by examining various case studies and by focusing on the use of analytical data in support of regulation of health, food, and the environment. It was held at the Fall ACS Meeting in Philadelphia in August 2012, organised with the ACS Chemical Information Division's program but co-sponsored by a number of other ACS technical divisions. The symposium was a full day symposium with ten speakers and most of their presentations (with one exception) have been included in this book and, where possible, have been expanded to mini-reviews and updated with new material. Two chapters have been added to expand and further define the context of the book.

We are extremely grateful to the authors of all the chapters for their patience and hard work in making this book possible. In addition, we recognise and salute the hard work of Rachel Ashton for her careful copy editing and of Rachel Deary, at ACS Books, for keeping the project on track and in both cases for going the extra mile when necessary. Without the hard work of these two, the book would not have been completed. We also thank our partners and families for their forbearance and understanding during the conception and delivery of this book.

Judith N. Currano
Chemistry Library, University of Pennsylvania
231 S. 34th Street, 5th Floor
Philadelphia, PA 19104-6323

William G. Town
Kilmorie Clarke Ltd.
24A Elsinore Road
London SE23 2SL, UK

Editors' Biographies

William G. Town

William G. (Bill) Town is President of the Kilmorie Consulting Division of Kilmorie Clarke Ltd. Dr Town obtained chemistry degrees at the University of Birmingham and the University of Lancaster in the UK and worked at the Universities of Sheffield and Cambridge before joining the European Commission at JRC Ispra in Italy. It was here, as leader of the ECDIN and EINECS projects, that his interest in the interaction between Science and the Law was first raised. His long career has spanned chemistry, databases, software, and publishing. In 2000, he was Chair of the ACS Chemical Information Division.

Judith N. Currano

Judith N. Currano received a Bachelor of Arts degree in chemistry and English from the University of Rochester, where she performed independent research under Prof. Robert K. Boeckman, and a Master of Science degree in library and information science from the University of Illinois at Urbana-Champaign. For the past fifteen years, she has been the head of the Chemistry Library at the University of Pennsylvania, where, in addition to providing chemists with traditional library services, she teaches a graduate-level course in chemical information and trains students in research and publication ethics.

Chapter 1

Looking Forward: Science-Based Policy-Making

William Town*

Kilmorie Clarke Ltd., 24A Elsinore Road, London SE23 2SL, UK
***E-mail: bill.town@kilmorie.com**

In this chapter I present a brief analysis of some of the problems faced today by scientists and policy-makers. For example, the ways in which they perceive each other or interpret each other's work might influence policy decisions, and the lack of scientific knowledge common among politicians and civil servants can affect the communication and understanding of science. I provide some pointers to understanding scientific evidence. Finally, I note some of the new directions highlighted during the symposium that will substantially affect science-based policy-making.

Introduction

My personal interest in the use and communication of science in policy-making derives from the time I spent in the Joint Research Centre of the European Union in the late 1970s and early 1980s. I initially worked on designing and building a database (*1–3*) to support policy-making related to chemicals in the environment. In addition, I later worked on the regulation of dangerous substances. I started as a narrowly focused structural chemist but rapidly had to learn about many other subjects, such as toxicology, metabolism, environmental degradation and so on, through talking with colleagues and experts. This broadening of my knowledge served me greatly when I later began to interact directly with scientists working in other Directorates General of the European Commission, especially with those who dealt with consumer protection, labelling of dangerous substances, and transport, which again were new areas to me.

My experience of policy-making processes and drafting new regulations came with the negotiation of the sixth amendment of the European Directive

on Dangerous Substances (now replaced by the Regulation on Classification, Labelling and Packaging of Substances and Mixtures). This legislative vehicle was used to introduce the testing of new chemical substances. The determination of whether a substance was new to the European market was based on the European Inventory of Existing Commercial Chemical Substances (EINECS) (*4*). Once the legislation had been adopted, I was given responsibility for the technical unit charged with building EINECS from reports submitted by industry through national contact points.

The sixth amendment was drafted by experts within the European Commission and reviewed and amended by a group of experts (mostly scientific civil servants) from European Union member states and from the European chemical industry, represented by the European Chemical Industry Council. There was also an expert group that considered detailed proposals for the completion of EINECS. This process, which lasted a couple of years, gave me direct experience of the interactions between national interests, science, politics and industrial considerations.

My direct involvement in these processes ceased in 1983, when I left the European Commission and returned to the UK. My interest, however, never waned and recently led me to explore these topics again with a series of American Chemical Society symposia on science and the law. The initial focus was on analytical data and analytical methods, arising from awareness that new instrument technology was changing detection limits and could increase the potential for regulation of chemical substances. I was repeatedly surprised by how important legislation about food, air quality and transport seemed to be decided without input from the scientists with the most knowledge and information. The approach seems too often to be 'here is the policy we want; find the science to support it'. Although we have all heard about so-called bad science or bad pharma, the impact of bad policy-making is much less discussed. Of course, science does not always have clear answers, and the role of factors beyond the reach of science, such as fear, hype, ignorance, resentment, or economic and political advantage, cannot be dismissed.

Policy-Making in Problem Areas

In the modern developed world many 'problems' are of our own making. For instance, science has extended life expectancy: children born now are expected to reach 100 years old. Consequences, such as increasing incidence of dementia and the costs of dealing with an increasingly older and infirm population, however, remain unresolved. The most common causes of preventable death have changed from disease to substance addictions (e.g. tobacco, alcohol and drugs) and pollution caused by industry and transport (*5*). Modern civilization and technology have also challenged the environment, and climate change is an increasing concern, with natural habitats, and thereby many species, being destroyed or under threat. To resolve these problems, scientists must engage in the policy-making process and initiate a revolution in how public policy is framed.

Involvement of Scientists in Policy-Making

At the 2012 Annual Meeting of the American Association for the Advancement of Science held in Vancouver, Canada, a European panel appealed to scientists to help improve communication of science at all levels, particularly in emotive areas or those difficult to explain, such as nuclear energy, crop innovations, and tobacco harm reduction (*6*). The symposium was organized by Brussels-based SciCom – Making Sense of Science, and it brought together prominent Europeans who were leading efforts to find science-led solutions to address citizens' priorities and concerns. The key attributes they purported were integrity, to uphold the inherent honesty of scientific enquiry and debate; openness, to improve transparency of the work and to declare any special interests; clarity, to speak in terms the public can understand; and engagement with the public, to demonstrate that duty to society is taken seriously. To strengthen the relationship, scientists must avoid thinking that they are policy-makers and must remain independent. Additionally, the science coming out of industry needs to be trusted more (*7*), which will be achieved by challenging spin and encouraging publication of all data relevant to products.

In another recent SciCom workshop, Addictions & Their Brain Reward Systems (*8*), expectations of various stakeholders were discussed.

- Science and policy were understood to form a crucial relationship

    ◦ Science is a fundamental pillar of knowledge-based societies
    ◦ Science can help provide the evidence base for sound public policy
    ◦ The dialogue between science and policy is never straightforward but remains a special relationship

- The scientific community was expected to take into account the following:

    ◦ The integrity of science needs to be positively asserted and defended
    ◦ Social sciences must be included to improve understanding of how the public might react or adapt to lifestyle challenges
    ◦ Scientists must learn to use established communication channels for providing policy advice more effectively, especially on life-or-death issues

- The policy-making community was expected to take into account the following:

    ◦ Policy-makers must be receptive to scientific advice, even when this advice is uncomfortable
    ◦ For the science and policy relationship to work, policy-makers have to challenge scientists to deliver on their public investment

- ◦ Policy-makers should consult more widely and learn from best practices and pitfalls encountered elsewhere

- • Public, industry, and interest groups were expected to take into account the following:

  - ◦ The public plays a critical role in determining what positions policy-makers will take
  - ◦ Industry is the largest investor in science and has every right to have its voice heard
  - ◦ Interest groups similarly have every right to have their voice heard as guardians of the common good or legitimate sector interests

- • Two key recommendations for what needs to happen were made

  - ◦ Scientific advice must be taken at all stages of the policy-making cycle
  - ◦ Policy-making must take account of the speed of scientific development and anticipate change

In a perfect world, scientific knowledge would be used to improve policy and have a positive impact on the lives of many. In the real world, however, politicians and civil servants frequently think that they can formulate sound policies without input from external experts. One suggestion to improve matters is to encourage more scientists to become involved in politics. Although this proposal has merit, substantially increased political involvement on the part of scientists is improbable. Also, perhaps, the role of chief scientific advisers could be expanded and numbers of positions increased to enable participation in political processes. Neither approach, however, deals with the core problem of scientific ignorance among politicians. Teaching science to politicians is an attractive idea, but unrealistic, at least in the short term, as most policy-makers read few scientific papers or books. Rather, relevant research is interpreted for them by advisers or external advocates who frequently represent powerful lobbying groups. In this context, the immediate priority is to improve policy-makers' understanding of the nature of science, not least its imperfections, and their ability to intelligently interrogate experts and advisers and understand the quality, limitations, and biases of the evidence. These interpretive skills are more accessible than those required to understand the fundamental science itself, and would fit more comfortably with broad skill set required of most politicians.

Concepts To Aid the Interpretation of Scientific Evidence

Sutherland and colleagues (*9*) suggested a number of concepts for understanding science that I have adapted below with civil servants, politicians, policy advisers, and journalists—in fact, anyone who might have to interact with science or scientists—in mind. Although science aims to discover what underlies

the patterns in the world around us, it is frequently hampered by unpredictable variations, and trends are more common than definitive differences or changes. Each variation might have many explanations, and, therefore, a major challenge in research is to tease apart the importance of different processes and to understand their relations and effects. For example, if two conditions differ at various time points, can it be shown unequivocally that either is related to or affecting the other, or are they being affected independently by an unknown influence? There may be innumerable sources of variation, from widespread changes to local processes to chance events. It might not be possible to rule all of these in or out as contributing factors. Thus, the informative value of findings needs to be judged according to established factors.

Error and Bias

Any measurement can be subject to some degree of error. Frequently, repeated measurements yield slightly different results due to observational error, loss of calibration of instrumentation, sampling errors, and other factors. Sound scientific reports should clearly describe any preventive measures put in place to ameliorate these risks. Likewise, possible errors should be explored in relation to unexpected findings and put into context with previous work in the field. Study methods should provide information on the tools and instruments used. These should have the appropriate precision to avoid quoting an unjustified degree of accuracy, which can be illustrated by stating the margin of error for key findings. This will aid interpretation of value ranges that seem large, while the real differences are quite small, or vice versa.

Extreme patterns with very wide ranges of values or a few outlying values in data can be problematic to interpret if they differ substantially from previously-reported outcomes which might themselves be in error. Such extreme patterns, however, are frequently anomalies attributable to chance or error. Ideally in such cases, similar findings obtained under comparable conditions should be sought.

Despite careful experimental design or measurement, bias can affect what information is available and how it is presented. If patterns of data within and/or across studies vary notably (or, perhaps, vary too little), consideration must be given to whether they are subject to methodological error, bias, or chance. Bias can derive from multiple sources. For instance, studies that report statistically significant results are more likely to be published than non-significant results, which potentially skews the magnitude of problems or the effectiveness of solutions represented in the literature. Other sources of bias are the expectations of the researchers, participants in human studies, or sponsors.

Bias is often unintentionally introduced. Confirmation bias might arise if scientists unconsciously look for or report evidence that supports a favored theory and/or insufficiently critique their results. Base-rate bias, also known as the base-rate fallacy, is the unintentional supplementation of irrelevant features for statistical fact. For instance, if one condition changes in the presence of a second condition, researchers occasionally neglect to report whether studies indicate the condition might also have changed in the absence of the second condition.

When policy-makers assess scientific reports, they should look for the following desirable features, which might help to minimize the risk of basing decisions on data subject to error and bias. An important method to avoid bias is random assignment to experimental groups. If the study population has been poorly sampled or group allocation has not been done randomly, there is a good chance that significant differences will be seen in baseline characteristics (e.g. size, quality, age). A double-blind approach to assessment is ideal to prevent bias related to the researchers' expectations of an intervention's effectiveness, but it is not appropriate or possible for many circumstances.

Average values obtained from large numbers of observations are more likely to be informative than those taken from fewer observations. One reason is that high numbers of observations are likely to lessen the impact of natural variation and measurement error on final outcomes. Another reason is that scientists improve their knowledge with accumulating evidence, which might lead to helpful alterations in methods of data collection or assessment. Systematic review and meta-analysis, which calculate average values from multiple studies, should, therefore, be sought to obtain an overarching picture of a specific topic.

Whether correlation and/or causation have been assessed is important to identify. It is tempting to assume that, if patterns are seen in a specific circumstance, the relation is causal, but this is not always the case. The effects could be coincidental or the result of both patterns being affected by a third known or unknown factor (a confounding variable). Decisions about the strength of the data in relation to policy should take into account whether testing was done to assess whether relations are significant, correlative, or non-significant trends.

The relevance of a study depends partly on how much the experimental conditions reflect those in real life. Extrapolation of data beyond the scope of the research carried out can risky. That is, patterns found within a given range do not necessarily apply outside that range. For example, there are limits to the generalizations that one can make from animal or laboratory experiments to humans.

Studies and experiments that have been carefully designed to answer a prespecified hypothesis or hypotheses are likely to be less subject to bias than those that have not. If it is unclear whether the data answer any specific questions or whether hypotheses have been drafted to fit the findings, then this should play an important part in deciding what is considered acceptable for policy-making. There is always a danger of confirmation bias.

Finally, scientists, as with many other groups, have a vested interest in promoting their work, not only for status, but, importantly, to secure further research funding. However, this could lead to selective reporting of results to support a point of view or, occasionally, an exaggeration. Peer reviewed work is generally viewed as the most reliable, but the peer review system is not infallible, and systems differ between journals (open; single blind, where reviewers can see the author details; and double blind, where reviewers and authors are given no details of each other). When faced with more than one paper on a specific topic with similar peer review reports, the journal editors' choice might take into account additional features, such as positive findings or newsworthiness.

Policy-makers should, therefore, look for multiple, independent reports with similar findings in different sources.

Control and Repeatability

Wherever possible, studies should include controls. Without controls, it might not be possible to determine whether an intervention or condition has had an effect, and the relevance of findings becomes complicated to interpret. Controls also help to avoid the effects of confounding variables on results. Studies that are designed to be repeatable by the same or different research groups in matched populations are most likely to be useful because they will have to have been carefully planned and described. If studies are not easily repeatable, they are at risk of leading to pseudo-replication, where the combination of study design and analysis technique becomes inappropriate to test the hypothesis, which can lead to unwarranted faith in results and possibly even lead to harm. The results of multiple experiments with similar methods are also well suited for comparison in systematic reviews or meta-analyses, which potentially have greater statistical power than the studies individually.

The degree of statistical significance of findings affects the relevance of results. Significance is a measure of how likely a result is to occur by chance. Thus probability of p=0.01 means there is a one in 100 likelihood that the effect of an intervention was due to chance. If a small number of observations are made, there is a chance that the study will have insufficient power to detect a difference between populations or conditions. This supports the idea that larger studies provide more relevant results than small studies. Nevertheless, if results are not significant, it does not mean that there was no underlying effect. Rather, it means that no effect was detected and it is impossible to determine whether or not there is an underlying effect.

Interpretation of Data

Effect size indicates the direction and magnitude of an effect of an intervention and the degree of difference between intervention groups. Small responses are less likely to be detected than larger responses. However, a study with many replicates might yield a result that is significant but the overall effect size is small and, perhaps, unimportant. Thus, comparison of effect sizes can be helpful when considering a range of studies of the same intervention but with different sample sizes or designs, which prevents the direct comparison of statistical significance. This approach avoids the misinterpretation of the meaning of significance tests, which can result in ineffective or misguided recommendations being made by policy-makers.

Future Directions

Predicting the future is always challenging, and in a topic as broad as science and the law it is almost impossible. The American Chemical Society symposium on which this book is based covered topics as diverse as food science, the monitoring of imported pharmaceutical materials, and data retrieval, as it is appropriate to let the views of experts in the wide range of areas affected prevail. In many areas, progress is likely to be incremental and driven by technological advances. However, in some areas, major changes are likely to occur as a result of disruptive technologies or ideas. For these there are future trends that are worth highlighting, namely, non-animal methods for toxicological testing and the evolution of harm reduction, particularly with regard to tobacco product regulation.

Food Science and Technology

In his chapter "Ensuring that nutrition and health claims in the European Union on foods and food (dietary) supplements are justified and scientifically substantiated", David Richardson notes that multidisciplinary applications, such as biotechnology, genomics, microbiology, physical chemistry, engineering, sensory science, and toxicology are being widely used in food analysis, quality testing, safety analysis, and shelf life. Science and evidence-based approaches are being used around the world to underpin regulatory developments in nutrition and health claims (*10*). Additionally, the development of functional foods and ingredients is contributing to processing and preserving raw materials from agriculture, horticulture, fisheries, and aquaculture.

He comments that research is required to link the scientific data with claims that are truthful and meaningful to consumers. Likewise, to aid policy-making, the strengths and limitations of the different sources of scientific evidence needs to be assessed, as most experimental designs other than randomized, controlled intervention trials seem unable to distinguish whether observed differences are due to the intervention or other factors. Much of the state-of-the-art human nutrition knowledge is currently based on epidemiological evidence, yet these data underpin most of the national and international dietary recommendations in Europe.

Regulatory developments in the food industry will include mandatory nutrition labelling, nutrient profiles, front-of-pack labelling about components (e.g. fats and sugars), the setting of maximum safe levels of vitamins and minerals in fortified foods and dietary supplements. Various methods of chemical and biochemical analyses will, therefore, be required to ensure regulatory compliance.

Surveillance of Pharmaceutical Imports

Jason Rodriguez *et al*, in their chapter, address rapid screening methods for pharmaceutical surveillance. They describe several portable technologies that have been developed to introduce rapid methods for point-of-entry testing on imported medicines. This approach is helping the United States Food and Drug Administration to monitor supply chain integrity and ensure the availability of

safe and effective drugs. These major initiatives, made possible by technological innovation in field deployable instrumentation, have enabled screening on an increasing number of products before they reach consumers. As technology continues to improve, the program will be able to add products and tailor screening to the areas of most need.

Harm Reduction

According to Delon Human (*11*), one of the public-health triumphs of the 20th century was prevention of the spread of the HIV. This change was due partly to the decrease in transmission between intravenous drug-users, through harm reduction methods such as education and implementation of needle-exchange programs. Although the term "harm reduction" is strongly associated with HIV, it is really an age-old practice. A proportion of the population has and always will engage in risk behaviors and, therefore, methods are developed to mitigate the potential dangers and associated health risks without achieving total abstinence. Modern examples of harm reduction are vehicle safety features (e.g. seat belts, helmets, banned use of cell/mobile phones while driving, etc), campaigns to promote responsible drinking of alcohol, the introduction of non-combustible nicotine products, measures to improve infection control (e.g. hand washing), and safe sex (use of condoms).

Although harm reduction prevents disease, disability, and premature deaths, it is not universally accepted or practiced as part of public health. For instance, a quit-or-die approach to drug addiction is widely upheld. This abstinence-only approach often leads to stigmatization and dehumanization of consumers and demonization of the substances concerned, whereas constant negative messaging on the risks of lifestyle choices are not always clearly understood or acted on. Ten years after Europe's first smoking ban was put in place in Ireland, for example, the Chief Medical Officer's report shows a 3–4% increase in smoking in the general population. Alcohol awareness campaigns face similar hurdles. Science and policy-making clearly need to work together to find effective solutions. Consumers' attitudes and the ethical elements of harm reduction need further assessment, and consumers, patients, interest groups, and others should be included in debates. Fortunately, the growth of HIV/AIDS consumer activism has highlighted the need for the involvement of consumers and patients in the debate.

In this book, the harm reduction issue is addressed in the chapter by Christopher Proctor and colleagues. Non-combustible tobacco products, such as Swedish snus (an oral tobacco pouch) or electronic cigarettes, have been available for some time, but regulation of these is not always clear, so they explore what science would be most helpful to support the development of regulations on nicotine and toxicants.

Alternative Toxicology Methods

Helena Hogberg and Thomas Hartung highlight in their chapter the regulatory challenges being faced in all chemical industries in moving away from animal toxicology testing towards testing with *in vitro* and *in silico* models. They

comment that exciting technical advances are enabling the development of the next generation of safety tests that can visualize at the molecular level what is happening when substances harm tissue. Other major new technologies are miniature 'organs-on-chips' to test drugs and virtual experiments created with computer modeling. Although most animal tests cannot yet be replaced by single, directly-corresponding alternative tests, they suggest that with strategic thinking that integrates a variety of tools animal experiments might ultimately be replaced entirely. They also note that alternative tools could allow quick and low-cost development of products and manufacturing.

Hogberg and Hartung conclude that the new methods to maintain high safety standards are crucial. Although the culture of validation is thought by many to cling to excessive regulation, acceleration in the advancement of technologies is improving methods of quality assurance and encouraging establishment of the best practices, systematic reviews, and meta-analyses to create a new regulatory science that is evidence-based, humane, and predictive for human risk.

Conclusions

In this book, and in the symposium on which it was based, it was only possible to provide a small number of case studies that illustrate the interaction between science and policy-making. What is clear is that this interaction is becoming increasingly important as we face problems, such as climate change, pollution of air and water, and the need to ensure that the food and medicines (drugs) we consume are safe and unadulterated. Ethical considerations, such as the minimization of the use of animals in toxicology testing, are leading us to develop alternative methods that do not involve animals. We are challenged by living in a world that has unprecedented access to information, which is sometimes of questionable value, and, therefore, the skills of information retrieval and evaluation are becoming increasingly important. Communication and assessment of scientific information is as important as the science itself, especially when our policy-makers, politicians, and media specialists lack scientific backgrounds. We need scientists with broad bases of scientific knowledge, communication and assessment skills, and a willingness to be involved in the political process. Equally we need policy-makers, civil servants, and politicians who have been trained in scientific methods. It is incumbent on our universities and academic societies to awaken and foster an interest in science-based policy-making.

References

1. Nørager, O.; Town, W. G.; Petrie, H. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 134.
2. Penning, W.; Town, W. G. *Clin. Toxicol.* **1981**, *18*, 1169–1181.
3. Hushon, J. M.; Powell, J.; Town, W. G. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 383.

4. Geiss, F.; Del Bino, G.; Blech, G.; Nørager, O.; Orthmann, E.; Mosselmans, G.; Powell, J.; Roy, R.; Smyrniotis, T.; Town, W. G. *Toxicol. Environ. Chem.* **1992**, *37*, 212.

5. Yoon, P. W.; Bastian, B.; Anderson, R. N.; Collins, J. L.; Jaffe, H. W. *MMWR, Morb. Mortal. Wkly. Rep.* **2014**, *63*, 369–374.

6. Gilligan, A. European scientists call for greater integrity, openness, clarity and public engagement. http://www.eurekalert.org/pub_releases/2012-02/s-ms-esc020912.php (accessed July, 2014).

7. Glover, A. Finding and element of Trust. *Financial Times*, Oct. 17, 2013.

8. Evidence-based policy versus policy-biased evidence: the challenge of feeding scientific advice into policy-making. addictions & their brain reward systems: key findings and recommendations. http://www.sci-com.eu/home/docs/scicom-science-in-the-public-interest-vol2.pdf (accessed June 30, 2014).

9. Sutherland, J. S.; Spiegelhalter, D.; Burgman, M. *Nature* **2013**, *503*, 3353.

10. Richardson, D. P.; Affertsholt, T.; Asp, N. G.; Bruce, A.; Grossklaus, R.; Howlett, J.; Pannemans, D.; Ross, R.; Verhagen, H.; Viechtbauer, V. *Eur. J. Nutr.* **2003**, *42* (suppl 1), I96–I111.

11. Human, D. In *Evidence-Based Policy Versus Policy-Biased Evidence: The Challenge of Feeding Scientific Advice into Policy-Making. Addictions & Their Brain Reward Systems: Key Findings and Recommendations*; Science in the Public Interest; SciCom: Brussels, Belgium, 2014; Vol 2, pp 24–28.

Chapter 2

Hunting and Gathering: Locating and Evaluating Information on the Cusp between Science and Legislation

Judith N. Currano[*,1] and Kenneth R. Foster[2]

[1]Chemistry Library, University of Pennsylvania, 231 S. 34th Street, 5th Floor, Philadelphia, Pennsylvania 19104-6323
[2]Department of Bioengineering, University of Pennsylvania, 240 Skirkanich Hall, 210 S. 33rd Street, Philadelphia, Pennsylvania 19104-6392
[*]E-mail: currano@pobox.upenn.edu

The scientific and legal issues surrounding health and the environment tend to elicit strong emotional responses from scientists, legislators and the general public, with each group frequently citing "hard facts" to bolster its assertions. In this chapter, we present useful techniques for locating the data behind the issues and discuss how to employ evaluation criteria and line drawings to determine the reliability of the information retrieved. We close with two case studies that demonstrate the techniques presented, one dealing with fracking in the Marcellus Shale, and the other on the health effects of low-level radiofrequency energy. We conclude with a summary of criteria that professionals use to evaluate the reliability of scientific papers, which differs in some respects from those suggested here.

Introduction

A county official is trying to find a way to address public concerns about the adverse effects of fracking in the Marcellus Shale. She does not have a strong background in science, and she is concerned that any information she presents to the county residents will be greeted with suspicion. Therefore, she creates a citizen's committee to review the issue, naming her staff assistant its technical

advisor. The assistant also lacks a background in the appropriate areas of science, yet he must help the committee members obtain and evaluate reliable information about the topic to aid them in coming to conclusions about the safety of the procedures.

A city is facing public protests about the installation of smart meters—wireless-enabled utility meters that incorporate radiofrequency transmitters similar in principle to Wi-Fi access points. At the last public meeting on the issue, members of the public brought several reports by expert scientific committees that came to diametrically opposed conclusions about the possible health effects of the radiofrequency energy emitted by the meters. An elected member of the city council wants to understand the issue, evaluate the reliability of these reports, and determine which is the most reliable.

The intersection of science, public policy and law is fraught with political tension. Policy-makers at all levels of government are asked to make decisions about food, fuel and environmental issues, but they rarely have strong scientific backgrounds and are largely unaccustomed to dealing with technical material. Even if they have had some scientific training, they may not be accustomed to evaluating the reliability of material found in the scientific and technical literature. In our experience, students in various areas of science and engineering tend to accept the results of Internet searches without being careful to evaluate the potential reliability of the sources. These might range from attorneys' or advocates' Web sites to the Web sites of national health agencies, not to mention the very wide range of quality of technical journals that are indexed in Google Scholar.

We describe an approach to evaluating the reliability of scientific and technical reports related to socially controversial issues that we developed for an engineering ethics class at the University of Pennsylvania, Philadelphia, PA. This process can be used in any form of ethical decision making and includes four main steps: 1) determining the scope of technical information needed for background for the case; 2) searching for the material using appropriate information resources; 3) evaluating the retrieved information to determine its credibility; and 4) subjecting the search results to ethical analysis and arriving at an informed decision grounded in ethical principles and based on an accurate appraisal of the facts. We present effective methods of searching for and evaluating information that can be used equally well when instructing students in finding and filtering literature on a subject of choice and when researching a controversial subject. Finally, we close with a demonstration of the process using the aforementioned case studies of fracking in the Marcellus Shale and the safety of radiofrequency transmitters.

Searching the Literature

Determining the Scope of the Query and Selecting Sources of Information

When beginning a literature search, one must first determine the scope of the query to be performed. In general, when searching for information, there are two basic approaches: the directed approach and the exhaustive approach. The directed approach is best used when attempting to locate a specific piece of information, as

this approach attempts to maximize the precision of the search although potentially sacrifices some recall. In other words, while all the results will contain the exact piece of information requested other relevant results might be missed. Exhaustive queries, on the other hand, attempt to locate all possible relevant information on the topic at hand by use of general keywords and by searching a wide variety of resources. In other words, they sacrifice precision in favor of recall. The results of an exhaustive query will require more evaluation and review than the results of a directed search

In conducting a thorough review on a controversial topic, one should attempt to retrieve as much material as possible given time constraints, but at the same time not retrieve so much as to overwhelm the reader. This will involve the use of a series of search tools, rather than simple Internet searches via popular search engines. To be comprehensive, researchers should always try to answer the following questions:

- What is being reported in the news?
- What are scientists saying in the primary (research journal articles, patents, dissertations, reports, etc.), secondary (databases and catalogs) and tertiary (review/technical/trade literature)?
- What is the law and what case history already exists?
- What are experts (commissions and panels) and other interested parties (interest groups) saying?

Researchers must thoroughly evaluate all information retrieved to attempt to determine the most relevant and reliable pieces on which to base a decision. Clearly, some types of information will need a greater degree of scrutiny than others. We present sample criteria and a proposed method of applying them later in the paper. However, before one can evaluate any information, one must locate it.

Query Design

Building a query is a three-step process that begins with an articulation of the information need. This can take the form of a simple question or a statement, such as, "I am looking for information about the use of gold in cancer therapies." Researchers who are not savvy in query design frequently forget that most information systems parse queries differently from individuals and popular search engines. A search through a database or catalog for the phrase, "gold cancer therapies," could locate any of the following: records that contain this exact phrase; articles that include all or any of the words, exactly as they are typed; or references that include all or any of the words along with variants, such as plurals or alternate verb tenses. Very few systems will locate records that use synonyms in exchange for the terms typed. As a result, to begin a search, we find it most effective to break a query into concepts; in the example above, the important concepts are gold, cancer and therapies.

The next step in constructing the query is to spend some time brainstorming a list of words that the author of a reference of interest might use to describe each

concept. In some cases, this takes the form of a list of synonyms or interchangeable terms, whereas in other cases one may need to consider alternate or related terms. One can keep track of potential search terms using a chart that has each concept at the head of a column of synonyms or related terms (see Case 1, Figure 1, for an example of such a chart).

Once one has listed all of the interchangeable terms, one must combine them in a way that the search system of choice will understand. This final step of the process is that with which most researchers are most familiar. Nevertheless, it is important to note that, for best results, one should examine all the possible search terms in the context of the resource to be employed to determine whether they are too specific, too general or will retrieve too many unrelated results.

When carrying out the search itself, the best approach will take into account the structure of the resource and the unique search techniques that it employs. While many electronic resources still employ basic Boolean operators ("AND", "OR" and "NOT"), several, including Chemical Abstracts Service's SciFinder, use natural language algorithms that retrieve more-relevant results when prepositions are placed between terms in lieu of the "AND" operator. Likewise, truncation (placing a wildcard symbol at the end of a word stem to locate variants) is available in most resources, although an increasing number is introducing some form of auto-stemming (also termed automatic truncation) into search algorithms. Auto-stemming should be used with care when performing an exhaustive query because the search systems rarely indicate which word variants they use and which they omit from a query. For example, at the time of writing this chapter, a search in Web of Science for "catalyst" retrieved the terms "catalyst" and "catalysts", but omitted "catalysis", "catalyzed" and "catalytic". Finally, the search system may allow the user to search through a controlled vocabulary of index terms that describe the articles contained in the database. Use of these index terms can be helpful when there is a term that exactly describes one of the query concepts because it will retrieve highly relevant articles that focus on the topic of interest. The danger of using only controlled vocabulary when searching is that the terms are not always up to date and, therefore, might not retrieve cutting-edge references. That would not be a problem when researching controversial issues, on which a great deal has, presumably, already been written.

Evaluating Information

Basic Criteria

By definition, an issue is controversial if there are many different ways in which the available information can be interpreted. Controversial issues inspire great emotion on at least one side of an issue. As a result, one must carefully assess all pieces of information, scientific or otherwise, to determine whether or not they are credible and appropriate for use in a given situation. The following criteria can form the basis for a procedure to determine whether or not information is worthy of consideration.

Accuracy

Accuracy refers to the degree with which the results of a measurement agree with the true value of the quantity being measured; precision (a different concept in science) refers to repeatability of a measurement. Although accuracy may seem like an obvious criterion for evaluation, it is one of the more challenging to gauge, particularly if the field of study is unfamiliar. Scientists are notoriously prone to overestimating the reliability of their own results (*1*).

One may judge the accuracy of a result by comparing it with established findings; however, novel scientific findings may differ drastically from previously published results. Assessment of the external validity of a study requires examination of the study design, which can be challenging for non-experts in a subject area.

Data Integrity

Data presented to a lay audience is typically filtered by the writer, who might have only cursory understanding of the subject matter being discussed (and who may also have some stake in the issue). Careful readers will check the original sources of a report to ensure that any results quoted are accurately described. Even within a primary scientific article, however, one should carefully examine the methods by which data were obtained, as well as the original source of and methods used to acquire any repurposed data.

Authority

The authority of a source refers to the qualifications of the authors and the institutions with which they are affiliated. When gauging authority, one should consider the authors' credentials and reputation in the field in which he or she is writing. Reputation can be based on the soundness of past work, institutional affiliation, amount and sources of funding, and the degree of impact that previous works have had on the field. When evaluating an institution, one should consider similar things: whether or not the institution is known for the type of science under consideration, the reputation and body of past work that came out of the institution, and whether or not the institution has any biases or agendas that could unduly slant the report of the science. For example, a report on alternative medicine by the World Health Organization is likely to be more authoritative than one from an association concerned with promoting alternative medicine; the latter potentially has greater expertise but also has a stake in the acceptance of the data.

Quality of Source

Like authors and institutions, information resources gain reputations over time. Readers turn to journals that have a history of publishing excellent and exciting science before reading the so-called second tier titles. News outlets gain reputations for being "leftist" and "right-leaning" based on the types of stories and angles that they report. Researchers new to a field may require some guidance from more-senior individuals in selecting the best sources of reliable information. If such assistance is unavailable, however, one can examine the nature and potential agenda of the publisher, as well as the degree of review to which articles are subjected before publication.

Traditionally, an important criterion of reliability has been whether or not a paper was published in a peer-reviewed journal. The application of this criterion, however, has become increasingly challenging in recent years, with the appearance of more than a thousand new open-access journals that claim to be peer reviewed but, in fact, apply low or non-existent standards to the contributions that they accept (http://scholarlyoa.com/). The choice of information retrieval system can help in separating selective, scholarly journals from their less-stringent counterparts. Databases such as the Science Citation Index (Web of Science) and PubMed have human editors who evaluate sources and choose only to index those that they find appropriate for their audience. On the other hand, search engines like Google Scholar use computerized algorithms that crawl the World Wide Web and collect any literature available, which results in many more links to material of unknown or dubious authority.

Bias

Bias, or slant, is the trickiest criterion to evaluate because of the many forms in which it can enter into a scientific report or review. All scientists perform research with certain expectations for outcomes and must design studies while avoiding falling prey to confirmation bias (i.e. the tendency of scientists to find in a study what they anticipated before they set out to do it). Even experts in a field can find the detection of confirmation bias in a study extraordinarily difficult. Because of confirmation bias, a scientist can put more emphasis on results that he or she anticipates to find or, at the worst, may ignore or even delete conflicting evidence. A well done study will have features, such as blinding, to try to avoid major sources of bias. A careful reader will assess the validity of the study design, certainly a difficult task for non-scientists (and even for scientists themselves). One can counteract lack of experience in the area in question through extensive and diverse reading and consultation with experts in the field. When dealing with so-called hot-button topics or extremely controversial and politicized areas, other, more easily identified forms of bias will also rear their heads. Sources of funding, obvious personal or political agendas and relationships between parties are some of the more obvious indicators of bias in a study. Many journals now require

disclosure statements indicating potential conflicts of interest, but this practice is not universal.

Timeliness

Timeliness is a potentially problematic criterion by which to evaluate information, and students and untrained researchers frequently fall into its trap. Researchers frequently consider newer information is to be "better" than older information, and, in some areas of science, this is very likely to be correct. In fields where advances frequently disprove previously held views and where the quality and speed of equipment improves rapidly, the latest research tends to be the most accurate or precise. However, in fields such as organic chemistry, a reaction that ran in near-quantitative yield in 1848 is still likely to produce excellent results. As a result, more recently published items are not necessarily better, even though they may discuss new research topics. An older, possibly definitive, article on an established area of science will be more reliable than, for example, a student essay on the topic. One needs to weigh publication dates against other criteria when coming to a conclusion about the trustworthiness of a source.

References

Checking the references of an article can help researchers to determine how well rounded the claims in it are likely to be, show the level of background research done by the author and, thus, give clues as to its accuracy. This is, however, not necessarily the case; many authors fill their bibliographies with 'relevant' references that are included for scientific-political reasons or simply because it is expected of them. One should apply the same evaluation criteria to the references that the author cites as to the article itself. Even if the references indicate that an article presents one-sided claims, one should not necessarily discard it. The limited point of view of the references should form a single part of a researcher's evaluation of the integrity of the research, balanced against the other criteria.

Impact

The impact of an article is generally measured by the number of times it has been cited and is taken by many researchers to be an indication of its quality. However, a paper might be cited many times for reasons other than its scientific merit; it might present controversial findings or even be spectacularly incorrect or inaccurate, as in the case of the 1989 paper by Pons and Fleischmann that announced the "discovery" of cold fusion, which has been cited nearly 800 times to date. Whatever the reason, when researching a controversial area, researchers should be aware of all of the high-impact articles available because they will help to demonstrate how individuals and groups come to make the claims that they do.

Impact also comes into play when evaluating the quality of journals. The most popular measure of a journal's quality based on citation metrics is the Thomson Reuters Impact Factor, derived annually by calculating the average number of citations per article in the 2 previous years. Non-scientists should be wary of accepting impact factor at face value, as several previously unknown organizations have begun issuing their own impact factors, generally based on unknown criteria, to journals for a fee, which muddies the water considerably (*2*).

Relevance

The piece of information under scrutiny could be the most reliable and sound piece of information in the world, but, if it is not relevant to the case at hand, it should be filed away for future reference. The choice of resource and search query can greatly aid in the identification of relevant information.

Use of Line Drawings To Help Evaluate Sources

Line drawing is a useful way to apply these criteria. This method, which is a version of casuistry or case-based reasoning, originated in applied ethics (*3*). To begin, one chooses a negative and a positive paradigm (*4*). In this application of the technique, the negative paradigm should be a source whose reliability is highly questionable, while positive paradigm is one generally acknowledged as reliable. One then identifies a number of important features, drawn from the criteria previously discussed, that distinguish the reliable from unreliable source (*4*). Each feature has a separate line assigned to it, with the most positive possible quality of that feature on the left end of the line and the most negative possible quality on the right (e.g. research findings are currently relevant vs. outdated data; the journal has an impact factor assigned by a reputable company vs. no impact factor; the author has a solid record of scholarly publication in a relevant field vs. few quality publications in a relevant field). One then evaluates the source under scrutiny by indicating on the line diagram whether it falls closer to the negative or the positive quality for each feature. After evaluating each individual feature, one must balance the relative importance of the features to determine an overall reliability ranking for the source in question. Thus, an expert panel assessment of a controversial topic under the auspices of an official body might be marked closer to the positive paradigm case, whereas a blog by an unknown writer on the subject would be marked closer to the negative paradigm.

The line drawing approach provides a rough judgment about the overall reliability of the source. As described, it is simply a method for keeping score, but, as an educational device, it forces an individual to consider which sources he or she deems reliable and unreliable by use of a defined set of criteria. This approach is hardly definitive as the importance of criteria in assessing the reliability of a source or the choice of paradigm cases is relative, but, for the average student, it is far preferable to conducting an Internet search and accepting the results without question.

Examples

Our students used the searching, evaluation and line-drawing techniques to great effect in their classwork, and we believe that other less savvy users of the scientific literature could benefit from these techniques. The following examples represent two potential uses of the procedure at the intersection of science and politics.

Case 1: Fracking in the Marcellus Shale

In this hypothetical case, a county official is being bombarded by protests from municipal officials, county residents, and citizens' advocacy groups against a proposal to begin using fracking techniques to extract natural gas from the Marcellus Shale. She has created a citizens' committee headed by one of her staff assistants and has charged the assistant with locating all available information about fracking and shale gas so that the group can try to understand the science behind this technique and weigh the claims made by the various interested parties. The assistant performs an exhaustive search in order to present his committee with as many different sources and viewpoints as possible.

Since the assistant has little background in this area of science, his first step is to visit a few scientific dictionaries and encyclopedias to determine potential terms and sub-concepts that he might use in his search. The dictionaries range in subject coverage from general, such as *AccessScience* (McGraw-Hill) and *Van Nostrand's Scientific Encyclopedia* (Wiley), to subject dictionaries in engineering (*Kirk-Othmer Encyclopedia of Chemical Technology*, Wiley), earth and environmental science (*Encyclopedia of Earth*, NCSE, or the *Dictionary of Earth Sciences*, Oxford University Press), and chemistry (*Hawley's Condensed Chemical Dictionary* and *Ullmann's Encyclopedia of Industrial Chemistry*, both Wiley). From these searches, he learns that fracking is a shortened form of the term hydraulic fracturing, which is a method of extracting oil and natural gas from sandstone by cracking the rock using a mixture of sand and water (*5*). After reading a few more general encyclopedia entries, he feels that he has enough vocabulary to begin constructing queries. His next step is to identify the types of literature that he will need to search. He decides that he will need to retrieve scientific publications that describe the process and name the byproducts of fracking; find out what is being reported in the news in order to understand what the citizens are hearing and what their concerns might be; apprise his committee of appropriate legislation, regulations, and case law; and understand and balance the claims of groups that have an interest in the issue.

Searching the Scientific Literature

The staff assistant decides to start with the scientific literature in order to help improve his understanding of the science behind hydraulic fracturing before moving on to more subjective parts of the literature. To achieve exhaustive

coverage of the scientific literature, he chooses resources that index findings in chemistry, engineering, geology, and environmental science, as well as some multidisciplinary databases. He knows that there will be a large number of repeat articles that are indexed by all of the databases, but he is confident that each tool will have some unique content based on its disciplinary focus. In order to minimize redundancy, he exports his results into a reference management system so that he can quickly identify and remove duplicate records retrieved by the various tools. He identifies the following potentially helpful search systems. These examples (presented in alphabetical order) represent tools in the specific disciplines that the staff assistant has selected for his query and do not constitute a comprehensive list of available information systems.

Chemical Abstracts (https://www.cas.org/), which is searchable via SciFinder or STN, covers the chemical literature from 1907 to the present. It indexes over 10,000 journals, as well as patents, technical reports, books, book chapters and dissertations. In addition to searching by topic, using either keywords or standard index terms (a controlled vocabulary), the assistant can also retrieve information about specific substances involved in the process by linking to *Chemical Abstracts* through a search of the CAS REGISTRY.

Compendex (http://www.engineeringvillage.com/) is the electronic version of the print *Engineering Index*. It indexes over 6,000 sources, such as journals, trade magazines, conference proceedings, and technical papers published in all areas of engineering since the late 1800s. It employs a controlled vocabulary that, once the assistant has located a single article of interest, can quickly guide him to other, related papers about the technology that currently exists to perform hydraulic fracturing.

Environment Abstracts (http://search.proquest.com/envabstractsmodule) includes material from 950 journals, conference papers, proceedings, governmental and non-governmental reports and other sources dealing with environmental issues and energy. The assistant will use this database to supplement his knowledge of the energy possibilities provided by shale gas and the environmental effects of fracking.

GeoRef, from the American GeoSciences Institute (http://www. americangeosciences.org/georef/georef-information-services), indexes over 3,500 journals in geology and also includes records for books, maps, reports and publications of the United States Geological Survey. This is where he will turn to find scholarly material about shale formations and the geologic aspects of the process.

Web of Science (http://thomsonreuters.com/thomson-reuters-web-of-science/) is one of many general science databases that can be used to find information on this topic. Advantages to using Web of Science are that it includes information dating back to 1900 from over 9,000 journals in all areas of science and that it has extensive tracking of citations, making it very easy to link forwards and backwards in time and quickly acquire related information. Sorting by times cited, a feature once unique to Web of Science but now available in many tools, will also allow the assistant to focus on articles that have had the largest impact.

The staff assistant will begin his search through the scientific literature by seeking general reviews, also called review articles, which summarize

the literature on various aspects of his topic and guide him to key pieces of experimental (primary) information. He will then supplement his understanding by performing his own searches through the primary literature. As he searches, he will use the terminology obtained from the articles he retrieves to refine and alter his search strategies, making his retrieval more comprehensive.

Searching News Sources

Once the assistant has a slightly greater understanding of the science behind hydraulic fracturing, he can turn to the news sources to find out what the public is hearing about the issues. News stories are another good source of vocabulary and terminology because they are written for lay people. They are, however, seldom written by scientists or individuals knowledgeable in the area of science concerned, which is a cause for some caution. The assistant decides that, to provide the best possible information from the news sources, he will look at national, regional and local newspapers and news wires. The following resources aggregate a large number of news sources: LexisNexis (see http://www.lexisnexis.com/en-us/home.page); Factiva (http://www.dowjones.com/factiva/);Library PressDisplay (http://www.proquest.com/products-services/newspaperdirect_pd.html); Access World News from NewsBank (http://www.newsbank.com/schools/product.cfm?product=24).

An advantage of using aggregator sources is that they frequently include full text for the news articles they index. The assistant will supplement these searches with searches of the local newspapers' archives and some general World Wide Web searches. He decides to limit by date to focus first on the latest, breaking news. He is aware that he will need to focus on data integrity; news sources frequently cite the results of studies without giving detail on how the studies were conducted. He plans to check the reporters' references and read the relevant scientific articles, regulations, laws and case law cited.

Finding Appropriate Legislation, Legal Precedents and Case Law

The legal aspects to his research are the most challenging for the staff assistant because legal information is complicated and best negotiated with the assistance of a professional. The assistant plans to visit the local state university's law library to obtain help, but through conversation with one of the librarians, he has already assembled a few good starting points. He has been advised to begin his search by locating a secondary source, such as a law review, bar journal, or treatise. These sources present background information, including historical context; descriptions of how the relevant laws were developed; important case histories; and relevant laws, statutes and regulations. They are extremely well documented and will lead him directly to the appropriate primary information (*6*). The librarian recommended some additional resources: LexisNexis; Georgetown

University Law Library's *Free and Low-Cost Legal Research Guide* (*7*); and *Locating the Law: A Handbook for Non-Law Librarians* (*8*).

One excellent way to begin searching for legal information is to locate a legal research guide written by staff at a university law library. There are two basic ways to do this: the first is to visit the home page of such a library and search for links to the research guides, and the second is to perform an online search in a general search engine. A search of "'hydraulic fracturing' AND 'research guide'" yields a hydraulic fracturing guide from the Pace Law Library (*9*) and the Marcellus Shale Resource Area from the Agricultural Law Resource and Reference Center at the Pennsylvania State University (*10*).

Discovering Hidden Agendas and Interests

Once he has a greater understanding of both the science and the law, the staff assistant decides to try to determine the interested parties in this area and to locate some commissions and reports that might provide useful information. The best way to approach this type of work is to perform a simple search on the topic using a general Internet search engine, and the staff assistant believes that a simple search for "fracking Marcellus Shale" (without quotation marks) in Google will yield relatively good results. Although the method is simple, the staff assistant knows that he will need to spend a great deal of time sifting through the resulting information and evaluating the claims made.

Building the Query

Having selected the tools that he wishes to employ, the assistant next turns his attention to locating the information itself, starting with a statement of information need. In this case, it is very general and can be expressed as, "find everything available relating to fracking in the Marcellus Shale". The statement is deceptive in its simplicity; on the surface, "fracking the Marcellus Shale" does seem to be a single concept. In actuality, it involves two large concepts, "fracking" and "Marcellus Shale" that in turn encompass several other concepts. The reason for fracking in shale formations is to produce natural gas, so this must be one of the terms employed. Some of the more contentious issues surround environmental effects, so those must be examined as well. Finally, proponents of shale gas indicate that this is a very efficient manner of extracting natural gas, and, therefore, efficiency should also be added to the query. The assistant arrives at a final list of five concepts and uses the results of his dictionary and encyclopedia research to generate a list of interchangeable terms for each, which will be used together in various combinations during the searches (Figure 1).

Natural Gas Production	Fracking	Environmental Effects	Efficiency	Marcellus Shale
Methane production	Hydraulic fracturing	Water contamination	Optimization	Shale gas
Natural gas well production	Hydrofracking	Methane migration	Economics	Shale gas formations
Unconventional natural gas development		Air emissions		Names of other, similar deposits
Natural gas development (NGD)		Impact assessment		
Shale gas				

Figure 1. A list of five core concepts related to fracking in the Marcellus Shale; each concept heads a list of search terms that could be used to describe it.

For his first search, the staff assistant selects a source that contains scientific journal articles and decides to focus the query on all aspects of the retrieval of natural gas through hydraulic fracturing of shale deposits. After consulting the help files for his chosen tool, he discovers that the resource does not use controlled vocabulary, has an autostem feature that can be deactivated, and employs Boolean operators for combining terms. In addition to the standard "AND", "OR" and "NOT" operators, the database has an adjacency operator, "ADJ/n" that can be used to search for two words with a specified number of words (n) between them. Bearing this in mind, the aide submits the following search.

(((methane OR ("natural gas" ADJ/1 well)) ADJ/3 produc*)*
OR ("natural gas" ADJ/3 develop) OR NGD OR (shale ADJ/3 gas*)) AND*
(frack OR hydrofrack* OR (hydraulic ADJ/1 fractur*))*

Quotation marks are used to exact phrases; however, truncation is not permitted within phrases, making the use of adjacency operators very attractive. As in mathematics, parentheses can be used to tell the database which operations should be performed first. Unlike in mathematics, many databases have no specific order of operations for Boolean operators, and the use of parentheses, therefore, is critical when mixing operators in the same query. For example, consider the search statement fracking AND "natural gas" OR "shale gas" The database could parse this query in two different ways: fracking AND ("natural gas" or "shale gas") or (fracking AND "natural gas") OR "shale gas". The first example yields the results that the researcher wants—those that contain the term fracking and either the term natural gas or the term shale gas. The second search, however, find the desired articles—those containing the term fracking and the term natural gas—but the Boolean "OR" operator would serve to combine these results with any results containing the term shale gas. As a result, the system returns many false hits.

Evaluating the Results

Once the staff assistant has located all the information that he wants to present to the committee, he realizes that the output is so huge and that he will need to provide the committee members with some guidance on how to approach them. He wishes to identify a few key pieces from each category, and is looking for the most reliable information available. Thus, he employs a series of line drawings in order to judge the material. Using the criteria for information evaluation described earlier in this paper, he determines the most appropriate criteria for each type of resource. For example, when evaluating the scholarly scientific literature, he bases his decision mainly on the criteria of timeliness, authority and quality of source, with some consideration given to impact as a proxy for accuracy and data integrity (Figure 2). When looking at news stories from the popular press, quality or reputation of the source, references, and bias are his top criteria, and, when dealing with information from advocacy groups, bias, references and data integrity take center stage.
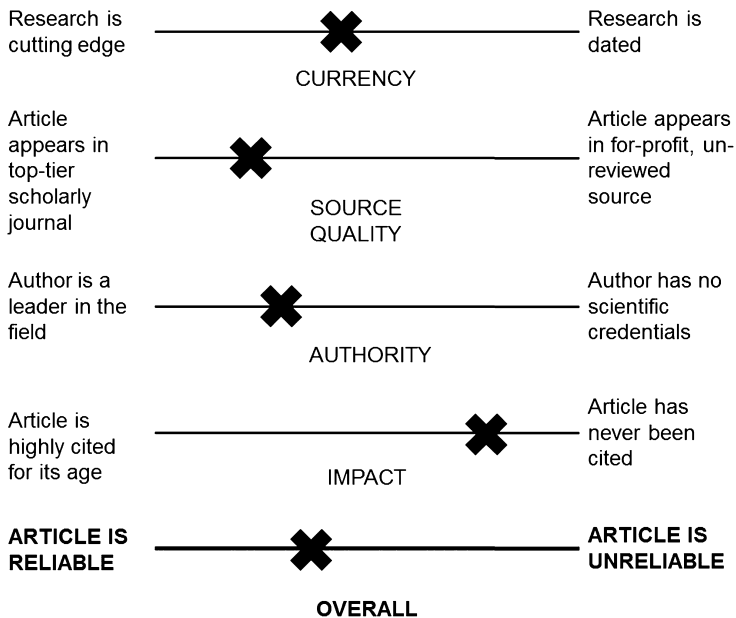
Research is cutting edge	✖ CURRENCY	Research is dated
Article appears in top-tier scholarly journal	✖ SOURCE QUALITY	Article appears in for-profit, un-reviewed source
Author is a leader in the field	✖ AUTHORITY	Author has no scientific credentials
Article is highly cited for its age	✖ IMPACT	Article has never been cited
ARTICLE IS RELIABLE	✖ **OVERALL**	**ARTICLE IS UNRELIABLE**

Figure 2. Example of line drawing analysis as a method of evaluating the reliability of an article from a scientific journal.

Case 2: Health Effects of Low-Level Radiofrequency Energy

This case derives from a recent series of (actual) email exchanges between KRF and a local government official in another state. The local electric utility had announced plans to install wireless-enabled electric utility meters (popularly called smart meters) on customer's homes. These meters incorporate low-powered radiofrequency (RF) transmitters that are similar in principal to the Wi-Fi access points found in many people's homes. These enable the utility company to record hour-to-hour usage of electricity. The goal is to enable time-of-use pricing, which would help to smooth out the peaks and valleys of electric use in the region and improve the efficiency of the power grid. Local citizens were protesting out of concern that the meters would create health risks to them.

The government official had been swamped with advocacy documents submitted by the concerned citizens, which included materials ranging from newspaper clippings to lengthy reviews of the topic with widely varying conclusions by scientists. For instance, a 2013 report of 74 pages by six physicians on biological and health effects of RF transmissions concluded that installation of smart meters would increase health risks in the community (*11*). A 2012 report of nearly 1,500 pages by 29 authors argued that low-level RF energy is very hazardous to health but did not specifically discuss smart meters. Nevertheless, it advocated strict precautionary measures to minimize exposure to low-level RF energy (*12*). A 378-page systematic review of the scientific literature related to possible health effects of radiofrequency energy report by the International Commission on Nonionizing Radiation Protection (ICNIRP) (*13*, *14*), prepared in 2009 by 29 scientists from several European countries and the United States, supported existing exposure limits for RF energy, but did not directly address smart meters. The government official asked KRF for expert guidance on the matter.

Since these reviews all represented the collective efforts of groups of experts, we used line drawing analysis to evaluate them. To represent the extreme limits of reliability of group reports, irrespective of differing conclusions, we chose a 2012 review by the independent Advisory Group on Non-ionising Radiation (AGNIR) as the positive paradigm because it had been undertaken by a carefully selected panel under the auspices of the United Kingdom Health Protection Agency (*15*). As the negative paradigm we choose a 2014 statement by the American Academy of Environmental Medicine (AAEM) (*16*). The AGNIR report was prepared by 10 internationally known scientists in the field (*17*). The individuals who prepared the AAEM report are not identified, but AAEM consists of environmental health physicians and has a history of advocacy on environmental health issues including the possible health effects of low-level radiofrequency energy. The AGNIR review and AAEM statement were strikingly different in quality. The AGNIR review devoted extensive space to analyzing the studies and identified potential limitations in study design that may have compromised the reliability of their conclusions. By contrast, the AAEM report simply cited studies from a much larger literature that supported its conclusions.

To select the features of reliability, we adapted criteria developed by Luc Verschaeve, (a specialist in this subject area at the University of Antwerp,

Belgium) from his 2012 review of 33 expert reviews on health effects of radiofrequency energy (*18*). Verschaeve graded each review according to a set of criteria (Table I).

Table I. Ten criteria for evaluating expert group reports. Adapted with permission from reference (*18*). Copyright 2012 InTech.

Topic for analysis	Criteria
Expert group	1. Procedure for selection of members and presence or absence of declarations of interest 2. Composition, complementarity and expertise of expert group members 3. Possibility to include minority statements in cases of disagreement
Methods used in the evaluation of the scientific data	4. Peer-reviewed publications, transparent procedure for selection of data 5. Method employed
Criteria for evaluation of scientific data	6. Transparent and clearly described criteria 7. Attention to the number of participants/animals/cells considered in the studies 8. Attention to potential bias and confounding factors 9. Attention to dosimetry 10. Evaluation of study methods and experimental set up used in the studies under consideration

To simplify this example, we have adapted the three main classes of criteria used by Verschaeve, adapting his Table I. For composition of expert group we looked at whether the group members were selected according to clearly identified criteria and represented a range of points of view, or whether the group was self-selected with all member having similar points of view. We included provision for minority reports in cases of disagreement and added whether the purpose of the group was advocacy. For methods used for evaluation of the scientific data, we considered how the papers included in the reviews were chosen: whether papers were selected to bolster the conclusions of the panel and whether the search for papers was limited to peer-reviewed literature. The AAEM report stated no criteria for selection of papers and based its conclusion on a very small number of studies (n=31). The AGNIR review was far more comprehensive, in fact exhaustive, and was based on hundreds of cited articles. Additionally, some of the AAEM papers were old (published before 1990), whereas the AGNIR review information identified up to the time that the review had been completed. For the quality of reviews, we used the criteria of whether studies were evaluated for validity (quality of exposure assessment, evidence of appropriate blinding and use of controls, statistical analysis in support of conclusions) or whether the conclusions of the study authors were accepted uncritically. Additionally, we looked at whether the reviews attempted to compare studies with respect to consistency of findings.

The evaluations for the three example sources are presented here. The evaluations were either the opinions of KRF or were based on the criteria outlined by Verschaeve.

The 2013 review (evaluated by KRF) was prepared by a self-selected group of six physicians that had been formed to address community concerns about smart meters. No information was available on how the papers were chosen and evaluated, and the report had little or no critical evaluation of studies with respect to possible limitations in study design. The report was written from a strong advocacy perspective (Figure 3).

Dart et al. 2013
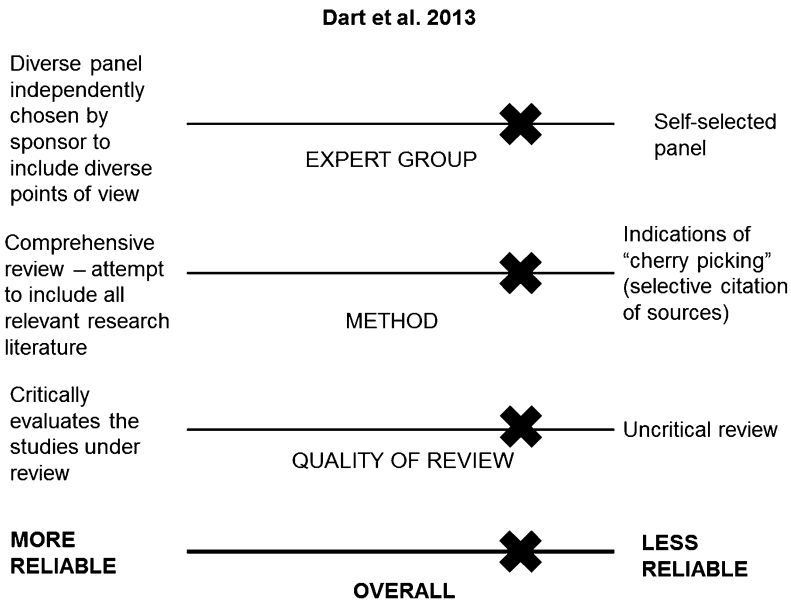
Figure 3. Line drawing analysis for the 2013 sample review, using reference (11).

The 2012 review (evaluated with the criteria proposed by Verschaeve) was a lengthy report, by a group of 29 self-selected scientists. It was comprehensive (more than 1,000 papers were cited) but not critical, with abstracts being reprinted verbatim from the studied papers without further comment (Figure 4). The review was explicitly intended for advocacy purposes: "to document the reasons why current public exposure standards for non-ionizing electromagnetic radiation are no longer good enough to protect public health." The report has been criticized by health agencies for lack of balance and selective reporting (*19*).

The 2009 ICNIRP review (evaluated with Verschaeve's criteria) was conducted by a publicly funded organization that works in collaboration with the World Health Organization. The lengthy report reviews the literature on health effects of RF energy, focusing on work published since its previous review in 1998. The review is comprehensive and critical (Figure 5), and the report was subject to extensive peer review.
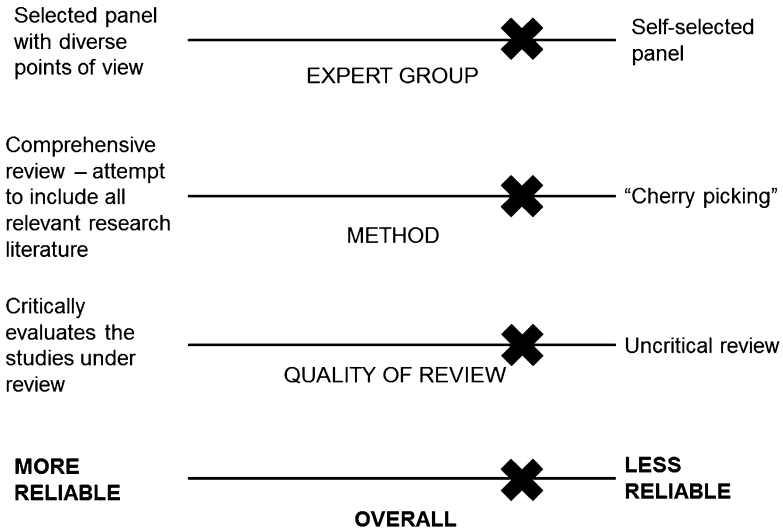
BioInitiative Report 2012

Selected panel with diverse points of view ————————✖———— Self-selected panel
EXPERT GROUP

Comprehensive review – attempt to include all relevant research literature ————————✖———— "Cherry picking"
METHOD

Critically evaluates the studies under review ————————✖———— Uncritical review
QUALITY OF REVIEW

MORE RELIABLE ————————✖———— **LESS RELIABLE**
OVERALL

Figure 4. Line drawing analysis for the 2012 example review, using reference (12)

ICNIRP 2009

Selected panel with diverse points of view ✖———————————— Self-selected panel
EXPERT GROUP

Comprehensive review – attempt to include all relevant research literature ✖———————————— "Cherry picking"
METHOD

Critically evaluates the studies under review ✖———————————— Uncritical review
QUALITY OF REVIEW

MORE RELIABLE ✖———————————— **LESS RELIABLE**
OVERALL
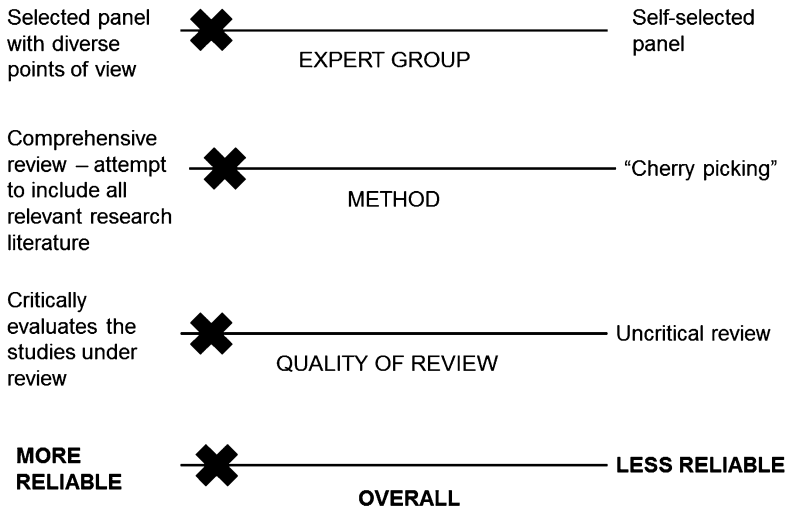
Figure 5. Line drawing analysis for 2009 ICNIRP example review, using references (13) and (14)

30

There are some caveats related to our analysis. First, the quality assessments of the three example reviews were not done according to one set of criteria. Second, the purpose of the documents varied. The 2012 and 2013 reviews are clearly advocacy documents, whereas the ICNIRP report was prepared to inform a revision of exposure limits and its goal was not explicitly advocacy. Second, reliable does not necessarily mean correct. A spirited advocacy document can be unreliable by the measures discussed in this chapter, but in the end could be more correct than a so-called reliable one.

Finally, the three example reports we present in this case study were easy to analyze, as they were all at the extremes of reliability. Finer discernment is possible between reviews that are closer in reliability, but that would require more analysis than the brief comments presented here. Nevertheless, the approach described can help to sensitize laypeople of the issues involved in deciding whether a technical review is reliable.

It is interesting to consider the perspective of scientists themselves in evaluating scientific sources. This was explored in a recent project by the Sloan Foundation that surveyed professional scientists to ascertain how they evaluated the reliability of scientific reports, the final report of which was released at the end of 2013 (*20*). "According to interviewees", the report concludes, "the top five reasons for choosing/trusting a citation were: 1) the author was known to the researcher; 2) the journal or conference proceedings were known to the researcher; 3) the reference was a seminal work in the field; 4) the reference supported their methodology; 5) the research group/institution was known to the researcher". These criteria generally resemble the ones outlined earlier in this chapter but indicate that professional scientists weight their evaluation of papers much more heavily on the basis of their personal knowledge of the authors of the paper and the institutions with which they are associated. Unfortunately, decision makers in the public sphere do not have this stock of personal experience upon which to rely.

Conclusion

We have presented a procedure for finding and evaluating information on controversial fields. After searching for scientific articles, news stories, appropriate legal and regulatory information, and interest group Web sites and publications, we recommend the use of line drawing, a common tool employed in ethical decision making, in order to evaluate the information. This approach is particularly critical in contentious fields, where conflicting information is frequently used to bolster the claims of interest groups. While we developed this approach for use by mid-level engineering students in their coursework, we believe that it would also be highly effective for members of the public performing personal research into contentious subjects, individuals working in legal or political areas, and even scientists working outside their fields of expertise. Only with proper retrieval and subsequent evaluation of the information can one begin to locate the grains of truth among the sound bites.

References

1. Ioannidis, J. P. A. *PLoS Med.* **2005**, *2*, 696–701.
2. Beall, J. Look out for bogus impact factor companies. http://scholarlyoa.com/2013/08/06/bogus-impact-factor-companies/ (accessed June, 2014).
3. Harris, C. E., Jr.; Rabins, M. J. In *Frontiers in Education, 1993. Twenty-Third Annual Conference. 'Engineering Education: Renewing America's Technology', Proceedings*; Grayson, L. P. Ed.; Institute of Electrical and Electronics Engineers: New York, 1993.
4. Harris, C. E., Jr.; Prichard, M. S.; Rabins, M. J. *Engineering Ethics: Concepts and Cases*, 4th ed.; Wadsworth: Belmont, CA, 2009.
5. Access Science Editors. Hydraulic fracturing (frackin). http://www.accessscience.com/content/hydraulic-fracturing-fracking/BR0808131 (accessed June, 2014).
6. Dong, A. University of Pennsylvania, Philadelphia, PA. Personal communication, 2012.
7. Venie, T. Free and low cost legal research guide. http://www.law.georgetown.edu/library/research/guides/freelowcost.cfm (accessed June, 2014).
8. SCALL Public Access to Legal Information Committee. *Locating the law: a handbook for non-law librarians*, 5th ed., revised; Southern California Association of Law Libraries: Beverley, CA, 2011.
9. Rucinski, T. Hydraulic fracturing (hydrofracking). http://libraryguides.law.pace.edu/hydrofracking (accessed June, 2014).
10. Penn State Law. Marcellus Shale. http://law.psu.edu/academics/research_centers/agricultural_law_center/resource_areas/marcellus_shale (accessed June, 2014).
11. Dart, P.; Cordes, K.; Elliott, A.; Knackstedt, J.; Morgan, J.; Wible, P.; Baker, S. Biological and health effects of microwave radio frequency transmissions: A review of the research literature. A report to the staff and directors of the Eugene Water and Electric Board. http://skyvisionsolutions.files.wordpress.com/2013/11/paul-dart-md-lead-author-report-to-eweb-june-2013.pdf (accessed June, 2014).
12. BioInitiative Working Group. BioInitiative 2012: a rationale for biologically-based exposure standards for low-intensity electronagnetic radiation. http://www.bioinitiative.org (accessed June, 2014).
13. International Commission on Non-Ionizing Radiation Protection. *Health Phys.* **2009**, *97*, 257–258.
14. International Commission on Non-Ionizing Radiation Protection. Exposure to high frequency electromagnetic fields, biological effects and health consequences (100 kHz-300 GHz). http://www.icnirp.de/documents/RFReview.pdf (accessed June, 2014).
15. Advisory Group on Non-ionising Radiation. Health effects from radiofrequency electromagnetic fields: Report of the independent Advisory Group on Non-ionising Radiation. http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317133827077 (accessed June, 2014).

16. American Academy of Environmental Medicine. Electromagnetic and radiofrequency fields effect on human health. http://aaemonline.org/emf_rf_position.html (accessed June, 2014).
17. Advisory Group on Non-ionising Radiation (AGNIR). http://www.hpa.org.uk/Topics/Radiation/RadiationAdvisoryGroups/Advisory GroupOnNonIonisingRadiation/#establishment (accessed June, 2014).
18. Verschaeve, L. Evaluations of internal expert group reports on the biological effects of radiofrequency fields. In *Wireless Communications and Networks - Recent Advances*; Eksim, A., Ed.; Intech: Rijeka, Croatia, 2012; pp 523–546.
19. Foster, K. R.; Trottier, L. Picking cherries in science: The Bio-Initiative Report. http://www.sciencebasedmedicine.org/picking-cherries-in-science-the-bio-initiative-report/ (accessed June, 2014).
20. Tenopir, C.; Allard, S.; Levine, K.; Volentine, R.; Christian, L; Boehm, R.; Nichols, F.; Christensen, R.; Nicholas, D.; Watkinson, A.; Jamali, H. R.; Herman, E.; Thornley, C. Trust and authority in scholarly communications in the light of the digital transition. http://ciber-research.eu/download/20140115-Trust_Final_Report.pdf (accessed June, 2014).

Chapter 3

Environmental Databases: A Trip down Memory Lane and New Journeys into the 21st Century

Frederick W. Stoss*

Science and Engineering Information Center, Arts & Sciences Libraries, Room 228-B Capen Hall, SUNY University at Buffalo, Buffalo, New York 14260-1672
***E-mail: fstoss@buffalo.edu**

In this chapter I take an historical look at the scientific, technical, engineering and medical bibliographic databases (e.g. BIOSIS Previews, Compendex-Plus, Engineering Index, GeoBase, GEOREF, MEDLINE, SciFinder, Scopus and the Web of Science). These resources, along with other scientific, technical, engineering, and medical databases, formed an initial core of bibliographic information retrieval systems from which environmental content could be retrieved. Various value-added analyzing and refining functions in current searching platforms are discussed. A number of subject-specific databases (e.g. Pollution Abstracts, Ecology Abstracts, Sustainable Science Abstracts, BuildingGreen and GREENR) that focus their content within the more broadly defined environmental information are discussed. I also feature several newer, non-traditional environmental databases: Carbon Dioxide Information Analysis Center (Oak Ridge National Laboratory), Love Canal Collections, MapCruzin', Global Change Master Directory, the National Science Digital Library, Scorecard and Right to Know Network.

Introduction

A declining number of people in the library profession were around during the first wave of automated information retrieval in the 1960s and 1970s. We recall (for the amusement of our younger cohorts) the 'good old days' when electronic information systems were new and using them required skills only found in 'factory trained' science librarians. Searches were done with telephone-coupled connections between computers operating at a 128 Bd. Short literature searches were performed via teletype terminals and printed on long sheets of newsprint, whereas larger batch searches were printed in hard copy and mailed to the requestor with remarkable speed. The luminaries in the field were Vannevar Bush, Roger Summit, Eugene Garfield, Alvin Weinberg, and Carlos Caudra. The early providers of electronic scientific, technical, engineering and medical (STEM) databases were Dialog, ORBIT, and MEDLARS.

People in need of literature searches set up appointments with librarians or information specialists who performed the searches in direct consultation with researchers, faculty, and students. Librarians knew the searching commands and codes and how to develop appropriate search strategies, and they became the gateway for delivery of information. They attended special training workshops at the annual meetings of the American Society for Information Science, the Medical Library Association, the Special Libraries Association and the American Library Association. Those were the days, and they were memorable ones.

Researching the Environment

To understand how searches for information on the environment are achieved, the concept of "environment" must be defined. In its broadest sense the word "environment" is used to describe the chemical, physical, and biological conditions, resources, attributes, and interactions affecting the survival of life forms, as well as the technical, social, and cultural aspects resulting in those changes and reactions to them. In the context of this chapter, "environment" represents the biotic and abiotic factors and interactions found in the places where we work, study, live and play.

Nearly 40 years ago Marta L. Dosa, now Professor Emerita in the School of Information Studies at Syracuse University, Syracuse, NY, described "environmental information" not as a physical entity or as a concept of data components, but rather as a "process that transfers data and information from source to user in any field of knowledge or activity applicable to environmental problem solving" (*1*). She proposed six basic characteristics of environmental information.

- It reflects the interdisciplinary nature of research and professional work
- It shows the differences between people in how they perceive problems, propose solutions, and assign priorities
- It displays peaks and valleys in public–policy attention to those problems, resulting in uneven funding of research information services and collection development
- It is dispersed among the literature in almost all types of information resources, including indexing and abstracting services, directories, specialized bibliographies, government documents and statistical sources
- It requires users to learn how to interact with information resources and systems to determine the most useful search terms and definitions
- It is represented by the ongoing proliferation of new information services and systems that mandate sound techniques for searching and evaluation

Although the methods of searching for environmental information have changed substantially, the process and characteristics described by Professor Dosa remain applicable. Before discussing the new environmental databases, however, it is useful to understand from whence they evolved.

The Beginnings of STEM Databases

Table I shows the chronological introduction of some of the best-known STEM abstracting and indexing services. The predecessors of the electronic databases of the late 20th and early 21st centuries began well before information became available in machine readable formats.

Compilation of environmental information began with the organization of the chemical literature, specifically pharmaceutical chemistry, with a listing of the factual information published in German and foreign sources. If one uses that axiom of the German-Swiss scientist Philippus Aurreolus Theophrastus Bombastus von Hohenheim, known more popularly as Paracelsus, that "dose makes the poison" (roughly reduced from his statement, "All things are poison and nothing [is] without poison; only the dose makes that a thing is no poison." (2)), we find that pharmacy science and toxicology are intermingled in their study of the effects of chemicals on human health and the concepts related to environmental health. Impacts and interactions of humans with chemicals provide a convenient place from which to initiate an examination of compiling information about the environment.

Information bases continued to grow beyond the pharmaceutical sciences, especially in the fields of bacteriology and microbiology, and information services for the applied (engineering), physical, life, and biomedical sciences followed. Primary research was systematically organized in volumes, bibliographies, and catalogues to make it easier to stay abreast of advances in research and development.

Table I. Major scientific abstracting and indexing services, by year of inception

Database[a]	Year started
Pharmaceutisches Central-Blatt (book, full-text available on Hathi Trust Digital Library from Jahrg. 24, 1853)	1830
Engineering Index (Compendex+)	1884
Index Medicus (MEDLINE & PubMed)	1879
Science Abstracts (INSPEC)	1898
Chemical Abstracts (SciFinder, STN)	1907
Applied Science &Technology Index (H.W. Wilson)	1913
Biological and Agricultural Index (EBSCO)	1916
Biological Abstracts (BIOSIS Previews)	1926
Abstracts of Bacteriology	1917
Botanical Abstracts Botanical Abstracts	1918
Publications Bureau/NTIS	1945/1973
Excerpta Medica (EMBASE)	1947
Science Citation Index, et al. (Web of Science)	1961
GeoRef	1966
AGRICOLA	1970
Biology Digest	1974
GEOBASE (Elsevier)	1986?
Scopus	2004

[a] These services and databases form the basis of environmental content and all can be found today in online formats.

Medical and biological research in the second half of the 19th century saw growth in interest in the connection between humans' interactions with their external conditions. Most notable were the relationships between air pollution and water contamination and adverse health effects (e.g. inhalational anthrax, asthma, chicken pox, influenza, measles, small pox, tuberculosis, cholera, dysentery, *Escherichia coli* infections, and typhoid fever). Unsurprisingly, therefore, the STEM databases reflect the topical coverage of these subjects, including further extensions to toxicology and environmental health, with the emergence of sanitation and hygiene, the discharge and dumping of industrial and municipal wastes, and public health concerns. Developments related to incorporating new environmental resources into existing databases, and in creating a new generation of databases specific to the environment required establishing new lexicons of indexing terms and concepts to retrieve relevant items in the primary literature.

Publication of the classic reference book, *Standard Methods for the Examination of Water and Wastewater* (*3*), serves as an outstanding example of the evolving science for analyzing and researching a wide variety of physical, chemical, and biological aspects of water sanitation and water quality practices. The book's origins are in the meetings of the American Association for the Advancement of Science, where it was proposed in the late 1800s that a book codifying a number of water quality tests and procedures be published. First published in 1905, this reference now provides details of numerous analytical procedures and has grown ever since. Its present day online version is, for all practical purposes, a mini-database of hundreds of procedures, tests and standards.

Computerization

The burst of technological achievements in computers and automated data and information systems after World War II and the competitiveness in scientific and technical arenas brought on by the escalating Cold War and the Space Race (which unintentionally stimulated the environmental movement in the 1960s and 1970s) provided a setting ripe for the development of computerized data and information retrieval systems. These changes sparked a revolution in the delivery of information and enabled researchers to stay up to date with a rapidly increasing proliferation of scientific and technical information.

During this period, the idea of providing a computerized version of the United States National Library of Medicine's *Index Medicus* (Table I) was conceived. The MEDLARS database was created and made available in 1964, and was the first publicly available computerized information storage and retrieval system. MEDLARS Online (shortened to MEDLINE) was unveiled in 1971 as an information retrieval system searchable from remote settings. It was mostly accessed through the libraries of medical centers and hospitals in the United States. A rush of database development followed, with teams beginning independent ventures that provided a rich setting for online retrieval of information from all disciplines.

Research, education, public policies and concerns about topics related to the environment were hastened owing to several factors, most notably public attention to and understanding of air, water, and land pollution and the general deterioration of environmental quality. Several major actions sparked these interests, such as publication of Rachel Carson's landmark book, Silent Spring in 1962 (*4*), and the creation of the Environmental Defense Fund in 1967, which led to the passing of the National Environmental Policy Act of 1969 (*5*).

Other events around the same time contributed to the growing concerns for protection of the environment and the resources found in it. Three notable events of this era were key in giving rise to environmental awareness and activism. The first Earth Day in April, 1970, sparked interest among the college youth and drew attention to environmental issues. In December of the same year, the relatively unheralded reorganization of the Federal government, during the administration of President Richard Nixon, led to the creation of the Environmental Protection Agency and the National Oceanic and Atmospheric Administration in the

Department of Commerce. Finally, in 1971, Greenpeace, one of the most influential environmental nonprofit organizations in history, was created.

Another factor influencing the growth of information related to the environment was a surge of legislative initiatives at state and Federal levels, beginning in the 1960s and 1970s, which increased regulatory authority over a multitude of environmental issues and greatly increased support of basic research across scientific and technical disciplines. The increase in basic research contributed to a surge in the publication of results in the primary literature that were begging for new and innovative ways to be retrieved and disseminated.

Various environmental databases that emerged in the second half of the 20th century are listed in Table II. Over time, increased attention was given to niche subjects and sub-disciplines of the traditional STEM disciplines, and database contents became more specific (e.g. ecology, aquatic biology, ocean science, health and safety, and sustainability). Several of these niche databases were parsed from larger collections, which enabled smaller customer bases to bring smaller packages of relevant environmental information to their libraries at affordable prices.

Routes of Retrieval

The first three decades of online information retrieval, especially in academic settings, were the domain of librarians and information specialists specifically trained in computerized information searches. End-users had to make appointments with librarians to present their information queries. Librarians would ask questions to refine concepts and select keywords, develop adequate search strategies, and produce bibliographic downloads to meet the user's needs. Often this service was provided under a cost-recovery model.

Things began to change in the mid-1980s when science librarians, such as Arleen Sommerville at the University of Rochester, gave upper-division undergraduate and graduate students instruction in chemical literature and information skills and retrieval, including the use of online searching, and conducted workshops with the chemistry department faculty. This move enabled select students and faculty to do their own searching and set an important precedent for the importance and acceptance of online information retrieval in aiding research productivity.

Sommerville (*6*) provided a detailed history of searching for chemical information and her early end-user training experiences. In her conclusion she wrote, "Increased availability of information in computerized form will require continual updating of training and curricular materials. Successful efforts to provide knowledgeable and enthusiastic instructors, relevant and timely curricular materials, and affordable information sources will require the joint commitment of chemistry faculty, librarians, professional societies, publishers, and computer-searching organizations." Much like Marta Dosa's comments about environmental information nearly a decade earlier, there was a prophetic ring to Sommerville's landmark work.

Table II. Environmental science and pollution management databases

Database	Year started
Meteorological & Geoastrophysical Abstracts	1950
Oceanic Abstracts (ProQuest)	1964
Bacterial Abstracts (CSA/ProQuest)	1966
Water Resources Abstracts (CSA/ProQuest)	1967
Pollution Abstracts (CSA/ProQuest)	1970
EIS: Digests of Environmental Impact Statements (to Pollution Abstracts, 2010)	1970
Aquatic Sciences & Fisheries Abstracts (CSA/ProQuest)	1971
Environment Abstracts (CSA/ProQuest)	1971
Environmental Periodicals Bibliography (CSA/ProQuest)	1972
Bibliography (see Environmental Index) Health and Safety Science Abstracts (CSA/ProQuest)	1973
TOXLINE (NLM/NIH, see TOXNET)	1974
Ecological Abstracts (Elsevier)	1974
Ecology Abstracts (CSA/ProQuest)	1975
Toxicology Abstracts (CSA/ProQuest)	1978
Acid Rain Abstracts (to Environment Abstracts, 1991)	1985
TOXNET (NLM/NIH)	1985
Aquatic Pollution and Environmental Quality (CSA/ProQuest)	1990
Environmental Engineering Abstracts (CSA/ProQuest)	1993
Risk Abstracts (CSA/ProQuest)	1985
Health and Safety Science Abstracts (CSA/ProQuest)	1973
Sustainability Science Abstracts (CSA/ProQuest)	1995
BuildingGreen (BuildingGreen, Inc.)	2001
Environment Complete	2006
GreenFile (EBSCO)	2008
GREENR (Gale/CENGAGE)	2010

By the late 1980s and early 1990s another major technological phenomenon was taking place: the migration of information databases into digital online platforms. Users across disciplines and in different lines of work and demographic settings rapidly embraced the technologies of the Internet and the World Wide Web. A perfect storm of enhanced environmental content, end-user searching, and the new technologies gave unprecedented access to data and information

in many forms. By the mid-1990s researchers, educators and their students, policy-makers and decision-makers, and staff in private and government settings had a plethora of options for viewing articles in journals and popular and trade magazine articles, technical reports, bibliographies, reference works, and data files via their own desktop and laptop computers. From the 2000s onwards, these resources have been marketed for access via smartphones and tablet devices.

Online Data Searches

Methods for Massaging Results

Various bibliographic databases now sport impressive numeric and reference (handbook-like) components. For instance, Chemical Abstracts Service (CAS) augmented its database with the introduction of the SciFinder tool, which enabled access to physical, chemical, biological, and regulatory data and information. Another exciting feature added to database platforms was the ability to download a set of references from a search and have the ability to further analyze, categorize, refine, filter, or other otherwise manipulate the retrieved information to improve the precision of results. Again, it was CAS, in the mid-1990s, that began implementing value-added features that enabled early and robust post-search . By use of its "Analyze" feature, CAS search results can be sorted by multiple characteristics: author name, CAS Registry Number, Chemical Abstracts section title, company or organization, database (CAPlus or MEDLINE), document type, index term, CA Concept heading, journal name, language, publication year, or supplemental term. The analyzed search results are broken down into histograms to show, in ranked order, the frequency of different document types within the entire output.

The second feature to massage the retrieved output is the "Refine" feature. With this tool, the search results can be narrowed by the application of filters, such as research topic, author, company name, document type, publication year, language, and database. These results can be further broken down by subtype. For example, when refining a search by document type, the results may be organized by biography, book, clinical trial, commentary, conference (presentation or paper), dissertation, editorial, historical, journal (article), letter, patent, preprint, report, and review.

Use of these refine and analyze tools on the SciFinder platform shows patterns and trends in publications and helps to identify key individuals or institutions leading research in specific fields, core journals in publishing on a specific topic, and so on. For example, if you do a search on the topic of acid rain (also called acidic precipitation or acidic deposition), refine the results to dissertations as an indicator of active research, and analyze these documents by publication year, the results show a rapid growth in the number of dissertations published from the mid-1970s to the early 1990s. During this period the National Acid Precipitation Assessment Program was running and directed much of the Federally sponsored research in academic settings in the United States.

SciFinder's "Categorize" feature uses standard indexing terminology to organize searches to selected category headings. For instance, a search for the

drilling fluids used in the process of hydraulic fracturing (more commonly called "fracking") can be displayed in the broad category of substances, or, with greater granularity, by use of subtypes, such as potassium chloride, calcium chloride, polyalkylenes, amines, alcohols, polyethylene glycol, polymers, cellulose, guar gum, sand and so forth. Selecting the most frequently cited index terms from the candidate lists provides retrieved items that can be analyzed by CAS Registry Number for a more precise listing of associated chemicals.

Innovation is often said to stimulate competition, and it is worth noting that these dynamic and robust features for massaging search results now appear in some form or another in other major database access platforms, including the Web of Science, Compendex Plus, EBSCO, Gale/CENGAGE, Scopus, and Wiley, and have also found their way into library online public access catalogs.

Advanced Data Manipulation Features

The ability to seamlessly navigate from database search to full-text downloads of primary literature (and, increasingly, access to primary data files, supplemental materials, and other aspects of a research project's execution) begs the question of what is on the next horizon?

There are two relatively 'new kids on the block' that are forging novel pathways for access to environmental information. The leading driver is the enhancement of database content with additional data and information sources, especially those provided in new formats (e.g. blogs podcasts, and digital audio and visual clips). These additional resources allow users to tap the growth and depth of niche markets and the new strategies developed for information and data retrieval, and, more importantly, identify and make accessible new points of access via social media.

BuildingGreen

The first 'new kid on the block' is BuildingGreen from BuildingGreen, LLC (http://buildinggreen.com). This database was developed to meet the needs stimulated by 21st century concepts of "green" (also broadly defined as "energy-smart" or "environmentally friendly") and sustainable design in architecture, interior and exterior design, planning and construction settings for new construction and renovation and occupancy of older buildings. Content is gathered with designers, architects, and contractors in mind (as well as students in technical, college and university programs in engineering, architecture, planning, and design). The basic core of BuildingGreen is a traditional approach of providing print and digital bibliographic resources. This whole-system approach covers all aspects of building associated with maximizing energy efficiencies and minimizing adverse environmental impacts.

The database's content ranges from scholarly, trade and news articles to books of interest and to products and specifications from the Construction Specifications Institute (CSI). Detailed project case studies provide examples of green construction practices, and results derived from the High-Performance

Building Database provide hyperlinks to more than 2,000 green products that assist in achieving the once coveted and now increasingly required Leadership in Energy and Environmental Design (LEED) Certification by the Green Building Council. To enhance the database, BuildingGreen.com LIVE is a fully archived blog with coverage of projects, programs, persons, and places where technologies and strategies, policies, regulatory compliance issues, and funding opportunities can be found. This hybrid model illustrates a new wave of database development geared to enterprise activities.

Global Reference on the Environment, Energy, and Natural Resources (GREENR)

Gale/CENGAGE Learning provides one of the most innovative environmental databases to come to market in several decades. Launched in 2010, GREENR ramped up the concepts of timely information delivery and value-added features to fuse traditional and new means of accessing information.

As the database's name suggests, and owing to the development history of Gale Research Company, the content is global and it is not surprising to see major features and content areas of GREENR emerging from Gale's past. There are so many unique features to GREENR it is difficult to find a place to begin. Users search for content from nearly 180 countries of origin, and geography is the primary means of retrieval. Search results are shown by location on interactive maps, which adds visualization by country as a strategic search capability. It is something of a surprise to see an enhanced, integrated world map serving as a primary search interface for online information.

GREENR's users have a large number of other means to search content; they can browse nearly 200 major issues (e.g. acid rain or wildlife) or enter search terms. Information profiles and resources from 15 major organizations with additional external links are also available for searching:

- Center for International Environmental Law
- Greenpeace
- Intergovernmental Panel on Climate Change
- International Union for Conservation of Nature
- United Nations Development Programme
- United Nations Educational, Scientific and Cultural Organization UNESCO)
- United Nations Environment Programme
- United Nations Food and Agriculture Organization
- Urban Environment
- United States Environmental Protection Agency
- World Bank
- World Business Council for Sustainable Development
- World Health Organization
- World Trade Organization
- World Wildlife Fund

Display of GREENR results reflects the robust nature of this media-rich resource, with an individual entry displaying issue-specific or topic-specific links to academic journals, associations, case studies, getting-involved advocacy activities, images, magazines, news, overviews, podcasts, presentations, primary sources, references, related portals, special libraries and research centers, statistics, videos, Web sites and blogs, and a world map, with searches limited or filtered by 24 document or format types. From the wide variety of content and unique and extremely user-friendly search capabilities to the retrieval and presentation options available, GREENR offers not only impressive quality, but provides more than a glimmer of hope for what potential lies ahead in the development of environmental databases online.

Some New Directions

In this section I present seven additional resources that reflect some of the areas from which new ideas might emerge and others might fade and fall by the wayside. There are, of course, hundreds, if not thousands, of other highly specific environmental databases from Federal and state agencies, GAP, international organizations and institutions, and other government and non-governmental agencies addressing a multitude of issues related to the environment, natural resources, and the human dimensions of living and interacting within Earth's ecosystems.

I selected six of these resources as examples of places from which important data and information may be identified and extracted. They are not necessarily databases per se, but they represent new and innovative methods of packaging and presenting data and information related to critical environmental issues that make it easier to extract and utilize the information in more broadly defined research, education and policy settings. Their potential for data and information delivery may stimulate future data collections, services and management. A seventh database is included for two reasons. First and foremost was its development as a major search tool that provided easy and unambiguous access to local data on the use, storage, and disposal of hazardous chemical wastes. The second reason I include this resource is an example of what happens when dedicated funding is lost for the continuing support of a good idea and the execution of data delivery services, and an important database is relegated to the status of high-quality, historical data.

Carbon Dioxide Information Analysis Center (CDIAC)

CDIAC (http://cdiac.ornl.gov) is one of the major and oldest climate change programs operating at Oak Ridge National Laboratory. It is the primary data and information analysis center of the United States Department of Energy's Global Change Research Program, and it houses the World Data Center for Atmospheric Trace Gases. Data are held in the following areas: fossil-fuel CO_2 emissions, trace gas emissions, atmospheric trace gases, oceanic trace gases,

carbon cycle, terrestrial carbon management, vegetation response to CO_2 and the climate, climate, and land-use and ecosystems. CDIAC's data are obtained through agreements with principal investigators who collect them. CDIAC then assists with the thorough evaluation, quality control, and quality assurance checks necessary before distribution. This attention to detail builds trust in the quality of the data and reduces uncertainties when applying these data to further scientific, technical, or policy and decision-making efforts.

Love Canal Collections

Events in 1975 and 1976 in Love Canal, a blue-collar neighborhood in the City of Niagara Falls, NY, contributed to the most detailed and excruciating examples of the management of hazardous chemical wastes from scientific, medical, legal, policy, and risk communication perspectives. The 1980 signing of the Comprehensive Environmental Response, Compensation, and Liability Act, commonly known as Superfund, provides the means for cleaning up abandoned or uncontrolled hazardous waste sites throughout the entire United States. Documenting this historic episode in the annals environmental history was no easy task, but the State University of New York University at Buffalo, NY, acquired and consolidated comprehensive collections of resources into the Love Canal Collections (http://library.buffalo.edu/specialcollections/lovecanal/) in the University Archives. There are ten special collections comprising the research of graduate students and faculty, resources from the Love Canal Area Revitalization Agency and a large number of newspaper clippings (listed in a searchable database) provided by the Ecumenical Task Force of the Niagara Frontier, books, technical reports, and many photographs from six organizations or institutions and by nine individual photographers. Additionally, full-text newspaper articles from the *Niagara Gazette* are available online.

MapCruzin'

Mapcruzin' (http://www.mapcruzin.com/) was started in 1996 as a repository for high-quality geographic information system (GIS) resources and initially focused on cartographic, sociodemographic and geographic data records on the Internet related to toxic chemical facilities. Mapcruzin' has since expanded its services to include access to various GIS shape files, software, maps and other resources. It has also greatly expanded its coverage to include nearly 50 additional environmental topics, including environmental justice, right-to-know issues, climate change, toxic release inventory (TRI) and other community environmental issues. Environmental GIS emerged as a major analytical tool that organized and displayed spatiotemporal information for researchers and policy makers dealing with chemical contamination. This breakthrough technology became an important tool for STEM data analyses. Mapcruzin's expansion from its original mission reflects the growth of GIS capabilities to assess environmental

and ecological impacts from naturally occurring and man-made perturbations and to measure and model changes to environmental conditions.

Global Change Master Directory (GCMD)

The broad study of global environmental change, including climate change, has generated vast repositories of scientific data and information in areas related to atmospheric chemistry and physics, ocean chemistry, land use and changes (including agriculture), spectral and other data related to global environmental changes in the Earth and geosciences and with human dimensions and interactions in the environment. The GCMD (http://gcmd.nasa.gov/) was designed by data and information managers at the Goddard Space Flight Center with funding from the National Aeronautics and Space Administration, under the United States Global Change Research Program, and presently lists more than 29,000 data-generating and gathering programs worldwide, including more than 3,400 datasets from the World Data Centers/World Data System. Major (*7*) details the precise and thorough use of metadata for records in the GCMD, which is one of its strengths and reflects the growing trend for the value-added attributes of metadata analyses related to geospatial data for its primary users. GCMD data are accessed through portals and directories. These can be searched by keyword; by specific data sets, services and tools (software, models, etc.); or by instrumentation or data platforms. The GCMD is freely accessible by researchers, educators, students, and any other people seeking resources related to Earth systems dynamics.

National Science Digital Library

The National Science Digital Library (http://www.nsdl.org) was initially created with funding from the National Science Foundation to serve as a high-quality repository of STEM resources for teachers, educators and students to supplement pre-service and in-service instruction of teachers. It was also intended as a resource-sharing network for grade-appropriate classroom instruction throughout school (formal K-12, college, and university) and in non-formal education settings. Users can browse collections by broad categories or search with specific keywords and have results sorted by education level and by resource type (e.g. assessment materials, audiovisual resources, datasets, and instructional or reference materials). Science literacy maps connect topics and issues across the areas of instruction in language arts, science, mathematics, and social studies and help teachers integrate those topics into their lesson plans.

Environmental Fate and Properties Data

As public awareness, research agendas, and regulatory compliance regarding environmental conditions and chemical hazardous wastes grew in the 1960s and through the 1970s, the need for systematic organization of physical, chemical, and

certain biological properties data also grew. The Environmental Science Center ([ESC] http://www.srcinc.com/what-we-do/environmental/), a division of SRC (formerly the Syracuse Research Corporation), began collecting and organizing toxicological and environmental fate data in the late 1970s and maintains a series of databases comprising its Environmental Fate Database (EFDB) for the United States Environmental Protection Agency (EPA) in linking properties data to bibliographic records.

Over the years SRC has led the way for innovative, timely developments, including its inventory of environmental data products listed below. The ESC Web site provides greater details and points of contact for more information on these products from its home page. These are the primary data products and services available:

- AIM: enables screening of chemicals for pollution prevention on the basis of chemical structure in the absence of testing data (free download, with EPA)
- Commercial Explosives Database: provides the EPA and the United States Customs and Border Patrol with essential chemical properties data for the rapid identification of explosives at ports of entry
- Database of SMILES Notations: Simplified Molecular Information and Line Entry System describing chemical structures (part of SRC EPI Suite)
- EPI Suite: chemical and physical properties, environmental fate and transport (free download, with EPA)
- ECOSAR: ecological structure-activity relationship model for predicting aquatic toxicology data when no toxicity test data are available (free download, with EPA Office of Science and Technology Policy)
- Endocrine Disruptor Priority Setting Database: for the EPA Endocrine Disruptor Program
- Fate Pointer Database: links chemical names and structures to their presence in any of the 18 SRC data sources
- GeoSIM: a GIS chemical exposure model
- Hazardous Substances Data Bank: environmental fate and chemical and physical properties for the National Library of Medicine TOXNET toxicity database
- IEUBK Model for Lead in Children: Integrated Exposure Uptake Biokinetic predictive modeling tool examining multi-media lead exposures in children and levels of lead in blood (free download, with EPA Superfund)
- Monte Carlo-based Multi-media Fate Model (aka GLAD multi-media model): statistical and probabilistic models of half-lives in environmental (ecological) compartments based on properties (free download, for the Great Lakes Atmospheric Deposition [GLAD] Program)
- NSCaT Database: toxicity of nanoparticles and substances
- Response Status Database: tracks data and information for use by the EPA for chemicals produced in or imported into the United States in quantities of more than 1 million lb)

- TSCATS Database: links information from unpublished resources to health and environmental effects and fates, by industry, under the Toxic Substances Control Act
- TSCA 8(e)/FYI Initial Screen Database: assists EPA in managing data submissions by industries disclosing harmful injury or health effects required by Section 8(e) of the Toxic Substances Control Act
- TRI-CHIPs Database: Toxic Release Inventory Chemical Hazard Information Profiles

SRC also has developed several databases and services to assist manufacturers to establish alternative processes that reduce environmental damage and to develop and maintain a green chemistry presence (http://www.srcinc.com/what-we-do/environmental/pollution-prevention-and-green-chemistry.html). These data products and services are provided in conjunction with the EPA and include the following:

- Alternative Assessments
- Safer Products Labeling Program
- Safer Chemical Ingredient List
- Sustainable Futures

ScoreCard and Right-to-Know Network

The most important and most powerful resource of environmental data freely accessible by the public was the Toxics Release Inventory (TRI), which was established by Section 313 of the Emergency Planning and Community Right-To-Know Act of 1986, as part of the Superfund Amendments and Reauthorization Act of 1986. TRI provides facility data inventories of chemical releases into the environment and reports on waste management practices for specified chemicals in specified industries and business. The Chemical ScoreCard (http://scorecard.goodguide.com/general/tri/tri_gen.html) was initially developed by the Environmental Defense Fund, and quickly became a prominent tool for users, because of its ease of use and comprehensive coverage of annual TRI reporting data. Precise and detailed records of chemicals at the facility, neighborhood, community, county, and state levels could be retrieved. The chemical inventories could be compared over time and with data from other facilities and regions. Sadly, the most recent data files in Chemical ScoreCard are from 2004. Its importance, however, remains the historical accounting of chemicals in communities across the country.

Current TRI data are available from the Right-to-Know Network (http://www.rtknet.org/), which picked up the work of Chemical ScoreCard and has added new features. This network is supported by five major databases: TRI, Spills and Accidents, Risk management Plan, Hazardous Wastes (via the Biennial Reporting System) and Hazardous Waste—Violations and Permits (via the Resource Conservation and Recovery Act Information System). A database for searching by city is undergoing beta-testing and is designed to update the Master

Search feature. An additional place to search for TRI data is the EPA's EnviroFacts TRI-data port (http://www2.epa.gov/toxics-release-inventory-tri-program).

References

1. Dosa, M. L. In *Library Searching: Resources and Strategies with Examples from the Environmental Sciences*; Morris, J. M., Elkins, E. A., Eds.; Norton Publishers: New York, 1978.
2. DeCavalho, J. P. Paracelsus. http://www.toxipedia.org/display/toxipedia/Paracelsus (accessed June, 2014).
3. *Standard Methods for the Examination of Water and Wastewater*, 22nd ed.; Rice, E. W., Baird, R. B., Eaton, A. D., Ciesceri, L. S., Eds.; American Public Health Association, American Water Works Association and Water Environment Federation: Washington, DC, 2013
4. Carson, R. *Silent Spring*; Houghton Mifflin: Boston, MA, 1962.
5. National Environmental Policy Act of 1969. http://www.epa.gov/compliance/basics/nepa.html (accessed July, 2014).
6. Somerville, A. N. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 314–323.
7. Major, G. *Sci. Technol. Libr.* **2003**, *23*, 21–36.

Chapter 4

Regulatory Toxicology: Progress in Law

Francois Busquet,[1] Michele Palopoli,[2] and Thomas Hartung[*,1,2]

[1]Center for Alternatives to Animal Testing Europe, University of Konstanz, Universitaetststr. 10, 78464 Konstanz, Germany
[2]Center for Alternatives to Animal Testing, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, Maryland 21205, United States
[*]E-mail: THartun1@jhu.edu

The legislative contexts for product regulation and animal welfare legislation in the European Union and the United States of America are very different. They offer very disparate opportunities for accommodating novel and alternative methods for regulatory toxicology. In this chapter we present a summary of laws and political decision processes, which is complemented by a description of recent developments. The Center for Alternatives to Animal Testing (CAAT), in the United States at Johns Hopkins School of Public Health, and in Europe at the University of Konstanz, Germany, is among the few voices of science directly informing policy-makers through policy programs on scientific opportunities. These opportunities should be accommodated in legislation and the developments should be parallel on both sides of the Atlantic. The example of CAAT's policy activities is used to show how scientific advocacy can impact on policy making.

Introduction

To defend against contamination of the environment and protect public health, we need to deploy the best science in toxicology and biomedical research. This approach has so far required the use of millions of animals every year to assess the safety of substances and products. No reliable data are available for the United States, but recently extrapolated European data suggest that 5–10 million

animals are used for this purpose worldwide every year. Programs, such as the Registration, Evaluation, Authorisation and Restriction of CHemicals (REACH), the European Chemical legislation from 2006, a possible reauthorization of the United States Toxic Substances Control Act (TSCA) from 1976, and new programs for nanoparticles, will increase this number. In contrast to basic research and drug discovery, which are very much driven by scientific and economic considerations, the regulatory use of animals is stipulated by policy and legislation. Such regulatory testing accounts for 25% of all animal use and has a lighthouse function for other areas because it is endorsed by validation and international harmonization. The toolbox of toxicology is remarkable in that, despite scientific progress, it represents a continuously growing number of primarily animal tests that have changed little since their introduction decades ago.

Laboratory animals are generally used to screen for health effects in humans and, at best, the relevance of any finding is afterwards assessed with modern mechanistic studies. Humans, however, are not 70 kg rats, and the need to revamp regulatory toxicology is increasingly being recognized. The major driving forces are the need for improved public health protection and animal welfare, as well as the steep costs in time and money associated with animal research. Further, animal models are limited in their ability to predict human health effects and inherently yield low throughput in the current system. Novel testing concepts must be based on the rapidly expanding understanding of how substances harm humans—that is, the pathways of toxicity. This concept was voiced prominently in the National Research Council's 2007 document *Toxicity Testing in the 21st Century – a Vision and a Strategy* (Tox 21) (*1*). This report has created an atmosphere of departure in toxicology; it has opened the door to revise current practices and reduce animal usage dramatically. The Johns Hopkins Center for Alternatives to Animal testing (CAAT US) is closely involved in setting this vision into action. CAAT US aims for paradigm and culture shifts to enable the use of modern, humane science for public health. Figure 1 illustrates the activities of CAAT US in the overall context.

CAAT US also steers a number of research activities, most prominently a National Institutes of Health transformative research grant project for mapping the entirety of pathways of toxicity (*2*, *3*), termed the human toxome. With a large consortium, CAAT US started mapping the human toxome for endocrine disruptors. Most importantly, this project will develop the pathways of toxicity concept further by defining how to identify, validate, annotate and share pathways of toxicity via a public database (*4*). CAAT US also works with the regulatory community on these efforts with the aim of bringing the findings to the policy maker community as well. The research on developmental neurotoxic effects forms proof-of-principle work for identifying pathways of toxicity.

The legislative contexts in the United States and the European Union (EU) are summarized below, along with CAAT's activities in both these regions to accelerate change with the goal of accommodating new and alternative approaches for the safety assessment of substances.
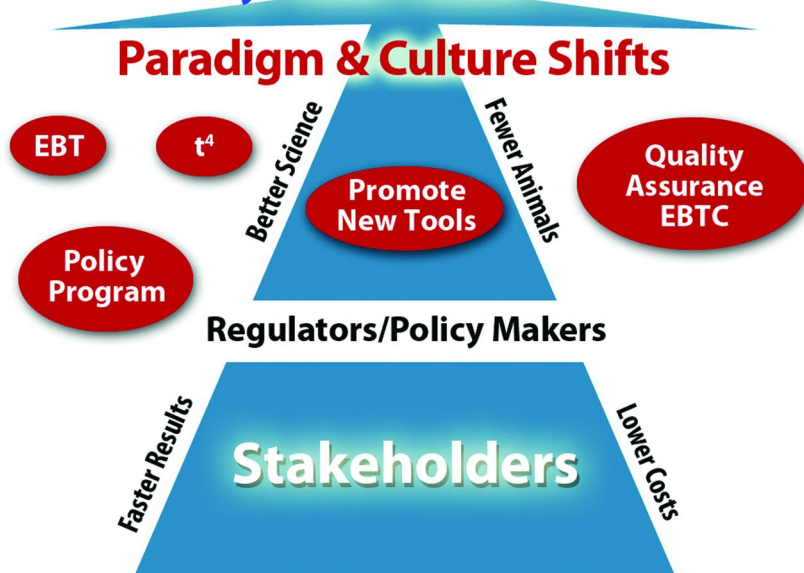
Figure 1. The vision and strategic work components of the Johns Hopkins Center for Alternatives to Animal Testing (CAAT). EBTC, Evidence-based Toxicology Collaboration; t⁴, Transatlantic Think Tank for Toxicology.

United States Legislation and Policy

Food and Drug Law

In the United States, the Food and Drug Administration (FDA), which is part of the Department of Health and Human Services, is the primary federal agency that regulates food and drink (including components and additives) for human and animal consumption (*5*), drugs (including biologicals, such as vaccines and blood), medical devices, cosmetics, tobacco products and radiation-emitting products. The majority of the FDA's legal authority is found in the Federal Food, Drug, and Cosmetic Act (FFDCA), first enacted in 1938 and amended many times since. The United States Department of Agriculture (USDA) and the FDA concurrently regulate some foodstuffs, such as meat and poultry. The USDA has jurisdiction in processing plants and the FDA regulates meat and poultry after they leave the plants (*5*).

Food additives are defined as any substances that are intended for use in or to affect characteristics of food, and will become part of such food (*6*). They must be shown to be safe under the intended conditions of use through testing by the procedures set out in the FDA Redbook (*7*) and according to the FDA's principles of toxicological testing for food (*8*). Some food additives are classified as "generally recognized as safe" (GRAS). The burden of showing that a substance should be classified as toxic in this way is on the registrant of the compound. Compounds added to the GRAS list after 1958 have needed scientific evidence of safety obtained from required tests (*5*, *7*).

The FFDCA does not require that cosmetic products and their ingredients be regulated by the FDA before they are placed on the market, with the exception of color additives. Nevertheless, cosmetics (and their ingredients) must be safe for consumers under labeled or customary conditions of use. Companies and individuals who market cosmetics have a legal responsibility for the safety of their products and ingredients (*9*).

Neither the FFDCA nor FDA's regulations require specific tests to demonstrate the safety of individual products or ingredients. Rather, the FDA has consistently advised manufacturers to use whatever testing is necessary to ensure the safety of their products and ingredients, but to ensure that it be substantiated in a number of ways: "*the safety of a product can be adequately substantiated through (a) reliance on already available toxicological test data on individual ingredients and on product formulations that are similar in composition to the particular cosmetic, and (b) performance of any additional toxicological and other tests that are appropriate in light of such existing data and information*" (*10*, *11*). The cosmetics companies have established a scientific review process, called the Cosmetics Industry Review (*12*), which conducts safety assessments of new cosmetics and ingredients. These assessments rely on published studies, but, if needed, new safety testing can be developed.

The FDA has authority to regulate the drug discovery process and approves all drugs before they can be sold and used in the United States. It has divided drug discovery into two phases: pre-approval, before introduction to the market, and post-approval after introduction to the market. Toxicity testing is carried out during the pre-approval period by the companies seeking approval of the drugs. The FDA reviews manufacturers' applications to market drugs in the United States and continues its oversight of drug safety and effectiveness as long as the drug is on the market (*13*).

Pre-approval drug development is a tightly guarded process at pharmaceutical companies, and, while some information is available publicly about how corporate testing strategies are developed and what tests are used, broader information is generally not widely shared. DeGeorge and colleagues (*14*) provided a detailed discussion about how toxicology testing is used in the development of anti-cancer drugs. Another example of how the pre-approval process works is set out on the website of the United States National Cancer Institute's Developmental Therapeutics Program (*15*). First, cell lines are employed to explore the basic toxicological properties of a compound. Next, animal tests are used to learn about metabolism and basic pharmacology (*16*), as animal data are required for an investigational new drug (IND) application. According to FDA, the IND

application must contain pre-clinical data in three broad areas of study—animal, pharmacology and toxicology—to permit an assessment of reasonable safety for initial testing in humans. The FDA almost always requires data from formally designed, conducted, and analyzed clinical (human) trials to make a decision on a drug's safety and effectiveness. The IND application must be filed by the drug's sponsor (usually its manufacturer) before clinical testing can start, and must include the proposed clinical study design and the principal investigator's qualifications (*13*).

Environmental Law

Environmental law regulates human activity in order to limit ecological impacts that that threaten public health and diversity (*17*). More than 100 laws make up the body of environmental law and are largely organized by category (e.g. endangered species) and/or media (e.g. clean air). The United States Environmental Protection Agency (EPA) is the primary regulatory agency in charge of environmental regulation (*18*). The EPA is organized along media lines (air, water, waste, toxics, etc.).

Two major environmental laws are closely associated with toxicity testing – the Toxic Substances Control Act (TSCA) and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). TSCA governs chemicals in commerce, giving the EPA the authority to call for testing in certain limited circumstances. Under TSCA, chemicals in commerce are divided into two groups: existing and new. Existing chemicals (those in commerce at the time that TSCA's regulations came into effect) do not require testing to remain on the market unless the EPA determines that they are creating a risk of harm. Under section 4 of TSCA, the EPA must, by rule, require the chemical industry to test a chemical for its environmental or health effects if it makes either what is known as a hazard finding or an exposure finding (*19*). EPA must make a hazard finding if

- the chemical poses an unreasonable risk of injury to health or the environment
- there are insufficient data about the chemical to predict its health or environmental effects
- testing is necessary to develop data on these effects.

EPA must make an exposure finding if

- the chemical will be produced in substantial quantities and

  - o it may enter the environment in substantial quantities
  - o there may be substantial human exposure to the chemical

- there are insufficient data about the chemical to predict its health or environmental effects
- testing is necessary to develop data on these effects (*19*).

When EPA makes either of these findings for a chemical, the agency must write regulations, for which testing is required. The test rule will develop health and environmental data if there are gaps. In totality, data presented to the EPA must convince the agency that the chemical does not present an unreasonable risk of injury to health or the environment.

New chemicals and new uses of old chemicals cannot be marketed until EPA approves a pre-manufacture notice or a significant new use regulation (*20*). No new testing need be done; available information, which might include animal toxicity testing, can be submitted. The EPA can, however, ask for additional information to confirm whether the chemical is safe.

Under the FIFRA all pesticides, fungicides, herbicides and rodenticides require testing before being allowed on the market. This testing involves a series of toxicity tests that are outlined in the regulations and guidance developed to ensure data and information requirements of TSCA and FIFRA were satisfied (*19*). A key indicator of EPA policy on toxicity testing is the Series 870 Health Effects Test Guidelines, issued by EPA's Office of Chemical Safety and Pollution Prevention (*21*). The EPA guidelines are harmonized with those published by the Organization for Economic Cooperation and Development (OECD). The testing methodologies set forth in the Series 870 Guidelines primarily reflect traditional mammalian approaches to toxicity testing. Although some of the guidelines do contain *in vitro* methodologies, these appear to be exceptions to the general rule.

In 2008, the EPA Office of Research and Development entered into a Memorandum of Understanding with the National Institute of Environmental Health Sciences/National Toxicology Program and the National Human Genome Research Institute/National Institutes of Health Chemical Genomics Center to launch Tox 21. In 2010, the FDA formally joined this collaboration. Starting from the premise that "[t]he convergence of science, technology, regulatory need, and public opinion has produced an historic opportunity to transform toxicology and risk assessment into more accurate, rapid, and cost-effective sciences," the parties to Tox 21 explain that its purpose is to guide the construction and governance of a detailed research strategy to make the National Research Council Committee's vision a reality (*19*, *22*).

Although TSCA establishes the principal legal framework under which industrial chemicals are regulated (and toxicity testing for those chemicals occurs), pesticides are treated separately and come within the purview of FIFRA. Enacted in its modern form in 1972, FIFRA establishes the framework for pesticide regulation in the United States. The EPA's authority under FIFRA is a balancing standard: the EPA must balance congressional mandate to prevent unreasonable adverse effects on the environment while taking into account the economic, social, and environmental costs and benefits of the use of any pesticide. (Pesticides tolerances are further regulated under the FFDCA, as discussed above.)

Unlike TSCA, FIFRA places the burden to demonstrate a chemical's safety on the manufacturer, not on the EPA. Testing is required but FIFRA does not have provisions on chemical data and testing that approach the level of detail seen in TSCA. The statute places the details of this process almost entirely within the discretion of the EPA Administrator, who must publish and revise guidelines

specifying the kinds of information that will be required to support the registration of a pesticide (*19*).

The EPA also regulates pesticide tolerances on food. It can establish or leave in effect a tolerance for a pesticide residue in or on a food only if it determines that the level is safe. The term safe means that there is a reasonable certainty that no harm will result from aggregate exposure to the chemical residue through dietary and all other exposures. The EPA is required to pay particular attention to information concerning the effects of exposure on infants and children. In setting a tolerance, the EPA is allowed to take into account available data and information on both anticipated and actual (measured) residue levels of a pesticide in or on food. Under certain circumstances, in assessing chronic dietary risk, the EPA can also consider available data and information on the percentage of food actually treated with the pesticide (*23*).

EU Legislation and Policy

From the EU perspective, there are a number of actors to be identified that interact closely with policy makers in order to support and/or shape law based on science and vice versa. These individuals and entities are:

- The European Commission (EC)
- The European Parliament (EP)
- The Council of the EU

The EC

The EC is the only EU institution that has the initiative to propose and draft laws at the EU level. Beginning in 1967, the Parliament-elected Commissioner, supported by a Directorate General (DG), was assigned to the area of Research, Innovation and Science. Other current DGs involved in research include the following: Agriculture and Rural Development, Climate Action, Communications Networks, Content and Technology (Connect), Education and Culture, Energy, Enterprise and Industry, Environment, the Joint Research Centre (JRC), Mobility and Transport, and Regional Policy. Notably, the EU bodies that drive science programs and funding are the European Research Executive Agency and the European Research Council.

The JRC

The JRC has been the DG in charge of science for EU policy support since 1959, although it was originally created to fulfill requirements under the European Atomic Energy Community (Euratom) treaty in Rome in 1957. Since its inception the JRC has extended its expertise to other fields important to policy making, such as life sciences, energy, security and consumer protection. It now comprises seven scientific institutes, each with its own specialty, located in five different

countries across Europe: Ispra (Italy), Geel (Belgium), Petten (Netherlands), Karlsruhe (Germany) and Seville (Spain).

Scientific Committees

When preparing its policy and proposals relating to consumer safety, public health and the environment, the EC relies on independent scientific committees to provide it with sound scientific advice and draw its attention to new and emerging problems.

Since March, 2009, three scientific committees represented by a panel of experts have met regularly in Luxembourg and consulted with the EC on (a) consumer safety, (b) health and environmental risks and (c) emerging and newly identified health risks.

The committee is renewed every 5 years. Whenever it is felt necessary, the scientific committees can call on additional expertise from a pool of scientific advisors and a database of experts.

Chief Scientific Adviser for the EU and President's Science & Technology Advisory Council

In 2012, Anne Glover was appointed the first Chief Scientific Adviser for the EU. The Chief Scientific Adviser may be consulted on any topic linked to science, such as science communication and promotion advising the President of the Commission on specific topics, commenting on topics such as the safety and risk assessment of genetically modified organisms and overseeing debate (e.g. whether to take a threshold or non-threshold approach for testing endocrine disrupting chemicals (*24*)).

The President's Science & Technology Advisory Council was established in January 2013, and is chaired by the Chief Scientific Adviser. It is meant to be an independent and informal group of science and technology experts from academia, business, and civil society. The Council covers a broad range of disciplines and unites expertise from the European Research Area.

The EP

The EP, as stated in the Treaty of Lisbon (*25*), deals with research framework, among other topics. In a nutshell, the Treaty of Lisbon makes the EP a stronger lawmaker by bringing over 40 new fields within the co-decision procedure, under which the EP has equal rights with the EC. (Co-decision is in contrast to the consultation procedure, where the EP only provides an opinion.) The areas covered by co-decision include agriculture, energy security, immigration, justice and home affairs, health, and structural funds.

Science and Technology Option Assessment

Political issues increasingly require expert consultation about scientific progress in order for the Members of the EP to decide legislation (e.g. new

regulations on *in vitro* medical devices, clinical trials, Horizon2020, etc.). The role of the Science and Technology Option Assessment (STOA) is to coordinate requests from the EP Members and, more generally, from the EP committees (e.g. Committee on Industry, Research and Energy or Committee on the Environment, Public Health and Food Safety) for overview and accurate information on ongoing legislative processes. Furthermore, it is the function of STOA to bring experts together on an ad hoc basis, as well as for scientific panels, to reply to the EP needs.

Raising awareness on new trends and/or disrupting technologies is also part of STOA. For example, in 2013, over 17 workshops were held at the EP to discuss issues such as risk and innovation to balance benefits and hazards or how to feed the world in 2050.

Intergroups

Intergroups can be formed by Members of the EP from any political group and any committee. Their aim is to enable informal exchanges of views on particular subjects and promote contact between EP Members and civil society. These groups are not EP bodies and, therefore, might not express the opinion of the EP. During the last parliamentary term (2009–2014) more than 25 intergroups were established.

In this context, the intergroup Welfare and Conservation of Animals works on different aspects of animal welfare and conservation and animal experimentation, including alternatives to animal testing. An intergroup on risk assessment is in the process of getting established for the next parliamentary term.

Council of the EU

The Council of the EU provides and defines the general political directions and priorities for the EC. It does not exercise legislative functions per se, although it sits with the EC and EP to discuss the files. These meetings are known as trilogue. The council consists of the heads of state or government leaders of the EU Member States, together with the Council President and the EC President. Each Member State has a permanent representation in Brussels that always includes a counselor for research and innovation.

EU Legislative Framework and Tools To Promote Alternatives to Animal Testing

In contrast to the United States, the EU has a number of legal mechanisms in place to promote alternatives to animal testing.

Treaty of the Functioning of the EU

Animal welfare is incorporated as a European value in Article 13 of the Treaty of the Functioning of the EU:

"In formulating and implementing the Union's agriculture, fisheries, transport, internal market, research and technological development and space policies, the Union and the Member States shall, since animals are sentient beings, pay full regard to the welfare requirements of animals, while respecting the legislative or administrative provisions and customs of the Member States relating in particular to religious rites, cultural traditions and regional heritage."

However, animal welfare is not an EU policy area. Nevertheless, promotion and use of alternative test methods and the principle of replacement, reduction and refinement (3Rs) are anchored elsewhere within the EU legislation (see later for examples). EU agencies (e.g. the European Chemicals Agency, the European Medicines Agency and the European Food Safety Authority) also contribute to fostering novel technologies such as *in silico* and *in vitro* methods.

Directives and Regulations

It is important to understand the difference between an EU Directive and an EU Regulation. Directives are addressed to national authorities, who must then take action to make them part of national law. If a member state fails to pass the required national legislation, or if the national legislation does not adequately comply with the requirements of the directive, the European Commission may initiate legal action against the member state in the European Court of Justice. Regulations are the most direct form of EU law. As soon as they are passed, they have binding legal force throughout every Member State. National governments do not have to take action to implement EU Regulations.

Directive 2010/63/EU

On January 1, 2013, EU Directive 2010/63/EU on the protection of animals used for scientific purposes (*26*) entered into force for the 28 EU Member States. It repealed the previous Directive 86/609/EEC. Since it is a directive, it allows Member States certain flexibility in the transposition of the Directive into national laws. Among the purposes of this Directive are to give scope; harmonize the current EU understanding of what defines an animal; map the resources, including identifying competent people and authorities; establish a common framework; and promote collaboration of the Member States with the EC to disseminate animal welfare in the EU.

The new Directive applies to live non-human vertebrate animals, including independently feeding larval forms and fetal forms of mammals from the last third of their normal development, and live cephalopods. The directive refers directly to the 3Rs.

Member States must assist the EC in identifying and nominating suitable specialized and qualified laboratories to carry out validation studies of alternative methods.

Cosmetics

The Cosmetics Directive provided the regulatory framework for phasing out animal testing for cosmetics purposes (*27*). It establishes a testing ban on finished cosmetic products and cosmetic ingredients on animals and a marketing ban of finished cosmetic products and ingredients included in cosmetic products that were tested on animals for cosmetics purposes in the EU. The same provisions are contained in the Cosmetics Regulation (EU 1223/2009), which replaced the Cosmetics Directive from July 11, 2013.

REACH

In 2007, REACH legislation (EC 1907/2006) came into force. This Regulation relates to chemicals and their safe use (*28*). The aim of REACH is to improve the protection of human health and the environment through the better and earlier identification of the intrinsic properties of chemical substances. It promotes the use of alternative methods for animal testing but does not oblige the test performer to do so: *"In order to avoid animal testing, testing on vertebrate animals for the purposes of this Regulation shall be undertaken only as a last resort. It is also necessary to take measures limiting duplication of other tests."*

Test Methods Regulation

In parallel to the adoption of REACH, the EC published standardized and accepted methods for testing hazardous properties of chemicals. These were written into the Test Methods Regulation (EC 440/2008), which came into force on May 30, 2008).

> "*The European Union is committed to promoting the development and validation of alternative techniques which can provide the same level of information as current animal tests, but which use fewer animals, cause less suffering or avoid the use of animals completely. Such methods, as they become available, must be considered wherever possible for hazard characterisation and consequent classification and labelling for intrinsic hazards and chemical safety assessment.*"

Regulation for Food Additives, Enzymes and Flavorings

The Regulation on food additives, food enzymes and food flavorings (EC 1331/2008) states that "*It is envisaged, in particular, that food additives, food enzymes and food flavorings, to the extent that the safety of food flavorings must be assessed ... must not be placed on the market or used in foodstuffs for human consumption, in accordance with the conditions laid down in each sectoral food law, unless they are included on a Community list of authorised substances*". The

guidance for submission for food additive evaluations refers to Directive 2010/63/EU and the 3Rs. These two elements must be considered whenever toxicological test methods are necessary. Moreover, the use of a tiered testing approach is developed to encourage the test performers to use *in silico* or *in vitro* tests, as well as validated test methods, under OECD standards, in use for REACH or listed under EC 440/2008.

CAAT Science Strategy and Policy Program

CAAT US is committed to support the paradigm change in regulatory safety assessments enshrined in Tox 21. Lessons can be learned from more than two decades of development and validation of alternative methods (*29–31*). Increasingly, the limitations of animal-based approaches, which we developed over almost a century, have revealed themselves (*32*). We have argued elsewhere that a revolutionary rather than an evolutionary change is required (*33*). Of note, however, is that the new methods also come with many limitations (*34–36*).

Development of Concepts To Enable Implementation of Tox 21

Beside the technological developments, conceptual steering is necessary to enable transition to new approaches and bring together different elements for a new regulatory approach (*37*). With the Transatlantic Think Tank for Toxicology, CAAT has launched a series of workshops and concept papers to promote discussion on this subject. With the creation of CAAT in Europe (CAAT Europe) in March, 2010, this program has a strong and unique transatlantic component (*38*). CAAT Europe complements the strategic initiative as a member of the American Consortium on EU Studies (ACES), an official EU Center of Excellence, and strengthens the two-way communication across the Atlantic. We believe that no approach accepted only on one side of the Atlantic will advance humane science as well as a meeting of minds leading to international harmonization. The costs per year for two workshops are borne by the Doerenkamp-Zbinden Foundation, Switzerland, and additional partners enable further projects on a case-by-case basis. In almost 5 years the program has resulted in more than 25 published workshop reports and commissioned whitepapers.

Assessment of the State of the Art in Toxicology

The doors for a novel approach to safety assessments must be opened by a fair and objective evaluation of current practices. A role model for effecting these changes is evidence-based medicine (EBM), which has been suggested as a template for addressing validation (*29, 39, 40*). The Cochrane Collaboration has engaged 27,000 physicians, scientists and health care providers to produce more than 5,000 guidance documents evaluating clinical practices. Because of the transparency and objectivity of the process, as well as its scientific rigor, when EBM guidance is available, it is considered the best available for a given

clinical question. A similar process, evidence-based toxicology (EBT) (*41*), can and should be developed to guide the evaluation of drugs, chemicals and other entities. It is noteworthy that one of the authors (T. H.) holds the first chair for EBT worldwide. The first conference was held in 2012, hosted by the EPA (*42*). In this context, a consensus paper on the validation of high-throughput assays was prepared (*43*).

Quality Assurance of New Approaches

Emerging technologies and numerous initiatives to promote their use to assess toxicity are being seen worldwide. To assist in the culture change and paradigm shift that we advocate, it is important to establish a mutually beneficial dialog between stakeholders. This dialogue will focus on quality assurance of the novel tools. Traditionally, this was attempted by formal validation; this approach has two principal problems:

- It is costly, takes a long time and is not amenable to change on the basis of new developments in technology, as any change invalidates the validation
- Validation is done using current, imperfect, traditional animal-based methods as the point of reference and thus cannot lead to a paradigm shift

Therefore, a mechanism that assures quality without these limitations is necessary. CAAT's toxicity testing symposia touched on this issue, which was taken up in detail at a CAAT organized conference, *21st Century Validation for 21st Century Tools*, in July, 2010. From that conference, a steering group was formed that includes representatives from CAAT, the EPA, the FDA, the National Institute of Environmental Health Sciences/National Toxicology Program, the American Chemistry Council, CropLife America, the pharmaceutical industry, the Humane Society of the US, the Institute for In-Vitro Sciences, and the International Life Sciences Institute / Health and Environmental Sciences Institute. The group has embraced the concept of EBT as a substitute for traditional validation (*44*) and views the development of this concept as a prime opportunity to collaborate toward change in regulatory toxicology. This group promotes a private–public partnership called the Evidence-based Toxicology Collaboration (EBTC) (*45*) between agencies and industry to promote quality assurance and implementation of new approaches. The EBTC was inaugurated on March 10, 2011, as a satellite activity to the 50th Society of Toxicology conference in Washington, DC, (*46*). A European branch was launched one year later, as a satellite activity to EuroTox in Stockholm, Sweden, 2012. CAAT provides the secretariat for EBTC. While the costs for individual evaluations of new methods must be borne by their developers and promoters, a central steering and publically available repository for guidance and reference documents is necessary (similar to the Cochrane library for EBM).

The secretariat assumes the following responsibilities:

- Central coordination of the steering group, organization of EBTC and the appointment of evaluation committees

- A standing committee for horizontal EBT method development (meta-analysis, quality scoring tools, probabilistic risk assessment etc.)
- An Internet portal for guidance and reference materials.
- Public relations

CAAT US Policy Program

CAAT launched its education, advocacy and outreach program in February 2007. This US policy program is aimed at educating policy makers and legislators about the need for alternatives to the use of animals in toxicity and safety testing and in biomedical research. It advocates for humane sciences in government research and regulations. In the longer term, the CAAT program strives to create a legislative and policy culture that values the lives of animals and promotes the use of alternatives and humane sciences.

Policy makers and regulators represent the best opportunity for a cultural shift and change in regulatory toxicology. CAAT's policy program is recognized as a point of reference and expertise among the policy and decision-making community, especially in the US. During the past 4 years, CAAT has successfully positioned itself as the go-to organization for information on Tox 21 and implementation of the National Academy of Sciences' vision and strategy for toxicity testing. Through its education on Capitol Hill, consensus and constituency building and written materials and presentations, the policy program has been instrumental in advocating the relationship between humane sciences and environmental health protection.

CAAT has developed an effective set of messages regarding humane science and public health protection, which it will continue to bring to policy makers, both at federal agencies and on Capitol Hill (*47*). Our fundamental approach is to find champions for alternatives in toxicity testing and biomedical research. In addition, we reach out to policy makers at US federal agencies that are important to the culture change and paradigm shift we seek.

A key element of CAAT's policy program has been creating and strengthening the relationships with important constituencies, such as the environmental law and policy and animal law communities. One particularly effective tool in constituency building has been the joint implementation of four symposia devoted to new methods in toxicity testing and implementation of the National Academy of Sciences' report. In addition to the benefit of producing intellectual capital, which can be effectively used in education and advocacy, these symposia have helped unite a diverse group of stakeholders to further cement this coalition.

Another goal is strengthening institutional care and use committees (IACUCs) by educating lawyers and religious leaders to serve as public members. These committees are required under US federal laws to oversee animal research and every IACUC is required to have a non-scientist member of the community. Many non-scientist members are ill equipped to understand and meaningfully contribute to the discussions about animal protocols that IACUCs review. CAAT seeks to create a group of appropriately trained and educated non-scientists who, if appointed to IACUCs, can make a difference in these critically important

committees. This will be achieved through research, and (if feasible) a pilot program. CAAT has established a certified program in humane sciences and toxicology policy in Johns Hopkins School of Public Health, and any individual who completes the curriculum can be awarded this certificate. In the past 3 years CAAT has brought 90% of the certificated curriculum online, and the full program will shortly be available worldwide.

CAAT US has established a strategic partnership with the EU delegation in the United States through its selection as part of ACES. This effort is complemented by CAAT EU. For example, ACES funded a symposium held by CAAT on toxicity testing entitled *Implementing the US NAS Toxicity Testing Report: An EU Perspective on the Way Forward*. This symposium allowed CAAT to leverage its policy efforts. To take advantage of the momentum gained from this symposium and of policy developments in Europe (e.g. REACH, the seventh amendment of the Cosmetics Directive and the novel Laboratory Animal Welfare Directive), CAAT will continue to expand its activities through ACES and continue with joint briefings for congressional staff and information days.

CAAT Europe Policy Program

CAAT Europe was established in February 2012, and cemented CAAT's role as a transatlantic bridge for the 3Rs and as a global scientific voice for bringing the 3Rs, humane science, and novel technologies into law, regulations, and guidance.

The program operates along three axes. First, CAAT Europe facilitates cross-sector networking and promotes dialogue. More than 100 face-to-face meetings have been held with EP officials (e.g. Members of the EP, Members' assistants, and policy advisers) since the setup of the EU policy program. All the relevant stakeholders—industry (e.g. cosmetics, chemicals, plant protection, and consumer products), non-governmental organizations, the EC, and ministries or regulatory agencies in Member States—have been contacted. Additionally, cooperation with academia representatives' offices in Brussels facilitates contact between EP Members and the corresponding national 3Rs scientists or regulators.

The second axis is regulatory monitoring, lobbying and/or advocacy for alternatives to animal testing on EU legislative files.2012 and 2013 were busy years for science owing to the preparatory work to launch the next European research-funding scheme, named Horizon 2020, on January 1, 2014. This scheme is aligned with the multiannual financial framework, which also starts in 2014 and ends in 2020. The total worth of the framework €80 billion and was launched with €15 billion assigned to the first 2 years. Although, at first glance, the spending seem impressive, the total corresponds to less than 1% of the total assigned to the multiannual financial framework, where more than 40% goes to the Common Agriculture Policy.

Among other topics, regulations on clinical trials, medical devices and *in vitro* diagnostic devices have been debated in the past 2 years by the EU institutions. Some of these files are still not closed.

Following strong public opinion concerns, the EP has also tackled endocrine disrupters by writing the "own initiative report" *Protection of Public Health on*

Endocrine Disrupters. As mentioned before, the EP has no power to propose new laws. Nevertheless, in order to respond to public pressure, the EP decided to take the lead even if the final report had no more value than a consultation.

The third axis is dissemination and communication. In the past 2 years, to inform EP Members and stakeholders about ongoing legislative works, CAAT Europe suggested and/or participated in workshops held at the European Parliament on multiple topics, such as the following:

- *The Human Toxome*, May, 2012
- *Advancing safety science and health research under Horizon 2020 with innovative non animal tools*, October, 2012
- *The Human Toxome project and endocrine disruption testing*, December, 2012, at the Intergroup on the Welfare and Conservation of Animals.
- *Worldwide Implementation of the 3Rs in Regulatory Toxicology: What are the Leadership Challenges and Opportunities?*, March, 2013
- *New Regulatory Science in Systems Toxicology*, March, 2013
- *Understanding Endocrine Disruptors available methodologies; what can we learn from experience to date?*, United States Mission to the EU, November, 2013
- *Hazard/Risk Assessment from the EU and the US perspectives*, November, 2013

Output and Outreach

Members of the EP or policy advisers may ask CAAT Europe for advice and briefing on topics linked with alternatives to animal testing on an ad hoc basis. Members of the EP have invited representatives of CAAT Europe to participate to Parliamentary events, such as *Risk in innovation: balancing benefits and hazards*, held in January, 2013, and organized by STOA. Likewise, stakeholders have invited CAAT Europe to participate in workshops and explain views on alternatives to animal testing.

In March 2013, CAAT Europe applied for two specific lots out of nine after a call for tenders organized by STOA on behalf of the EP. These were Life Sciences for Human Well-Being and Safety and Security Technologies. In early 2014, CAAT Europe was listed as an official expert contact point for the EP for a period of 4 years.

Conclusions

The state and the dynamics of the political landscape for regulating products in the US and Europe are very different: Europe has taken over from the US as a pacemaker of novel legislation. The accelerating unification process now including 28 member states with 530 million citizens led to enormous efforts in harmonizing and creating legislation, with more than 70% of the national legislations now originating on the EU level. Therefore this chapter included also a description of players on the European side involved in this process. In

contrast, the US has not seen major new legislations for product safety in decades; however, a number of well-established agencies exist, which fill the existing framework with innovative approaches. Notably, the European counterparts are usually more administrative executers of the legislation. The situation varies also for the different industrial sectors with large grade of harmonization for drugs, similar requirements for pesticides and tremendous differences for cosmetics' ingredients and environmental chemicals.

The need to embrace new approaches to product safety is increasingly perceived on both sides of the Atlantic. This requires information for policy-makers and agencies on technical opportunities and in a globalized economy also about the developments in other major economic regions. The example of the policy programs of the Centers for Alternatives to Animal Testing in the US and Europe were given to demonstrate how academia can help shape and accelerate this process. This is in the best interest not only of the animals to be spared, but also of consumers and patients world-wide to benefit from modern safety sciences.

Acknowledgments

References

1. NRC Committee on Toxicity Testing and Assessment of Environmental Agents. *Toxicity testing in the 21st century: A vision and a strategy*; The National Academies Press: New York, 2007, pp 1–196.
2. Hartung, T.; McBride, M. *ALTEX* **2011**, *28*, 83–93.
3. Bouhifd, M.; Hogberg, H. T.; Kleensang, A; Maertens, A.; Zhao, L.; Hartung, T. *Basic Clin. Pharmacol. Toxicol.* **2014**, *115*, 24–31.
4. Kleensang, A.; Maertens, A.; Rosenberg, M.; Fitzpatrick, S.; Lamb, J.; Auerbach, S.; Brennan, R.; Crofton, K. M.; Gordon, B.; Fornace, A. J., Jr.; Gaido, K.; Gerhold, D.; Haw, R.; Henney, A.; Ma'ayan, A.; McBride, M.; Monti, S.; Ochs, M. F.; Pandey, A.; Sharan, R.; Stierum, R.; Tugendreich, S.; Willett, C.; Wittwehr, C.; Xia, J.; Patton, G. W.; Arvidson, K.; Bouhifd, M.; Hogberg, H. T.; Luechtefeld, T.; Smirnova, L.; Zhao, L.; Adeleye, Y.; Kanehisa, M.; Carmichael, P.; Andersen, E. M.; Hartung, T. *ALTEX* **2014**, *31*, 53–61.
5. Piña, K. R.; Pines, W. L. A Practical Guide to FDA's Food and Drug Law and Regulation, 4th ed.; Food and Drug Law Institute: Washington, DC, 2012; pp 1–521
6. United States Food and Drug Administration. Food Ingredients, Packaging and Labeling. http://www.fda.gov/Food/IngredientsPackagingLabeling/ Food AdditivesIngredients/ucm228269.htm (accessed April, 2014).

7. United States Food and Drug Administration. Guidance for industry and other stakeholders: toxicological principles for the safety assessment of food ingredients. Redbook 2000, http://www.fda.gov/downloads/Food/GuidanceRegulation/ucm222779.pdf (accessed June, 2014).

8. United States Food and Drug Administration. Food Regulatory Guidance. http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocuments RegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm2006 826.htm (accessed June, 2014).

9. United States Food and Drug Administration. Cosmetics questions and answers. http://www.fda.gov/Cosmetics/ResourcesForYou/Consumers/ucm 2005206.htm (accessed June, 2014).

10. United States Food and Drug Administration. FDA Authority over cosmetics. http://www.fda.gov/cosmetics/guidanceregulation/lawsregula tions/ucm074162.htm (accessed June, 2014).

11. United States Food and Drug Administration. Cosmetics guidance and regulation. http://www.fda.gov/Cosmetics/GuidanceRegulation (accessed June, 2014).

12. Cosmetic ingredient review. http://www.cir-safety.org (accessed June, 2014).

13. Thaul, S. CRS Report for Congress prepared for members and committees of Congress: how FDA Approves drugs and regulates their safety and effectiveness. http://fas.org/sgp/crs/misc/R41983.pdf (accessed June, 2014).

14. DeGeorge, J. J.; Ahn, C.-H.; Andrews, P. A.; Brower, M. E.; Giorgio á M. Anwar Goheer, D. W.; Lee-Ham, D. Y.; McGuinn, W. D.; Schmidt, W.; Sun, C. J.; Tripathi, S. C. *Cancer Chemother. Pharmacol.* **1998**, *41*, 173–185.

15. Developmental Therapeutics Program NCI/NIH. http://dtp.nci.nih.gov (accessed June, 2014).

16. United States National Cancer Institute. *Drug Discovery at the National Cancer Institute*. http://www.cancer.gov/cancertopics/factsheet/NCI/drug discovery (accessed June, 2014).

17. Lazarus, R. *The Making of Environmental Law*; University of Chicago Press: Chicago, IL, 2004; pp 1–334

18. Bearden, D. M.; Copeland, C.; Luther, L.; McCarthy, J. E.; Tiemann, M.; Esworthy, R.; Yen, J. H. Environmental laws: summaries of major statutes administered by the environmental protection agency. http://www.hsdl.org/?view&did=748067 (accessed June, 2014).

19. Myers, B. D.; Locke, P. A. *NYU Environ. Law J.* **2012**, *20*, 35–95.

20. United States Environmental Protection Agency. New Chemicals. http://www.epa.gov/oppt/newchems/index.htm (accessed June, 2014).

21. United States Environmental Protection Agency Office of Chemical Safety and Pollution Prevention. OCCP harmonized test guidelines: series 870 – health effects test guidelines. http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series870.htm (accessed June, 2014).

22. United States Department of Health and Human Services (HHS) and National Institutes of Health (NIH). Tripartite memorandum of understanding on high throughput screening, toxicity pathway profiling, and biological

interpretation of findings. http://www.niehs.nih.gov/about/highlights/assets/docs/memorandum_of_understanding.pdf (accessed June, 2014).

23. United States Environmental Protection Agency Pesticide Regulation. Implementation of Requirements under the Food Quality Protection Act (FQPA). http://www.epa.gov/pesticides/regulating/laws/fqpa/fqpa_implementation.htm (accessed June, 2014).

24. European Commission Assistant to the Chief Scientific Adviser. Minutes of the expert meeting on endocrine disruptors. http://ec.europa.eu/commission_2010-2014/president/chief-scientific-adviser/documents/minutes_endocrine_disruptors_meeting_241013_final.pdf (accessed June, 2014).

25. Treaty of Lisbon. http://europa.eu/lisbon_treaty/full_text/index_en.htm (accessed June, 2014).

26. Hartung, T. *ALTEX* **2010**, *27*, 285–303.

27. Hartung, T. *ALTEX* **2008**, *25*, 147–162.

28. Hartung, T. *ALTEX* **2010**, *27*, 3–14.

29. Hartung, T. *J. Toxicol. Environ. Health* **2010**, *13*, 277–290.

30. Hartung, T. *Eur. J. Pharmaceutics Biopharmaceutics* **2011**, *77*, 338–349.

31. Hartung, T.; Leist, M. *ALTEX* **2008**, *25*, 91–96.

32. Hartung, T. *ALTEX* **2008**, *25*, 3–9.

33. Hartung, T. *ATLA, Altern. Lab. Anim.* **2008**, *36*, 635–639.

34. Hartung, T. *ALTEX* **2007**, *24*, 143–147.

35. Hartung, T.; Hoffmann, S. *ALTEX* **2009**, *26*, 155–166.

36. Hartung, T. *ALTEX* **2013**, *30*, 275–291.

37. Hartung, T. *Nature* **2009**, *460*, 208–212.

38. Daneshian, M.; Leist, M.; Hartung, T. *ALTEX* **2010**, *27*, 63–69.

39. Griesinger, C.; Hoffmann, S.; Kinsner-Ovaskainen, A.; Coecke, S.; Hartung, T. *Human Exp. Toxicol.* **2009**, *28*, 83–179.

40. Hartung, T. *Toxicol. Sci.* **2009**, *109*, 18–23.

41. Hartung, T. *Human Exp. Toxicol.* **2009**, *28*, 93–94.

42. Stephens, M. L.; Andersen, M.; Becker, R. A.; Betts, K.; Boekelheide, K.; Carney, E.; Chapin, R.; Devlin, D.; Fitzpatrick, S.; Fowle, J. R.; Harlow, P.; Hartung, T.; Hoffmann, S.; Holsapple, M.; Jacobs, A.; Judson, R.; Naidenko, O.; Pastoor, T.; Patlewicz, G.; Rowan, A.; Scherer, R.; Shaikh, R.; Simon, T.; Wolf, D.; Zurlo, J. *ALTEX* **2013**, *30*, 74–104.

43. Judson, R.; Kavlock, R.; Martin, M.; Reif, R.; Houck, K.; Knudsen, T.; Richard, A.; Tice, R.; Whelan, M.; Xia, M.; Huang, R.; Austin, C.; Daston, G.; Hartung, T.; Fowle, J. R., III; Wooge, W.; Tong, W.; Dix, D. *ALTEX* **2013**, *30*, 51–66.

44. Hartung, T. *ALTEX* **2010**, *27*, 241–251.

45. Evidence-based Toxicology Collaboration. http://www.ebtox.com (accessed June, 2014).

46. Zurlo, J. *ALTEX* **2011**, *28*, 152.

47. Locke, P. A.; Myers, D. B., Jr. *ALTEX* **2011**, *28*, 266–272.

Chapter 5

Analytical Procedures and the Regulation of New Drug Development

George Lunn*

Office of New Drug Quality Assessment, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993
E-mail: George.lunn@fda.hhs.gov

Within the United States Food and Drug Administration, the Center for Drug Evaluation and Research is responsible for the regulation of pharmaceuticals that are intended for marketing. In the broadest sense, the type of information that should be submitted to the Food and Drug Administration by the applicant is set by statute and the Code of Federal Regulations, along with various Guidance documents. The recommendations become more specific as we move from the statute to Guidance. The amount of information that should be supplied is determined by the stage of development. Generally, lesser amounts of information are supplied in the early stages and submissions become larger and more detailed as development proceeds. These principles are illustrated by the example of analytical procedures.

Introduction

Within the United States Food and Drug Administration (FDA) the Center for Drug Evaluation and Research (CDER) is responsible for the regulation of pharmaceuticals that are intended for marketing in the United States. New drugs and new variants of existing drugs (e.g. new dosage forms) are evaluated for safety and efficacy by the Office of New Drugs and for quality by the Office of New Drug Quality Assessment. Many other groups within CDER also participate in the approval process. For instance, generic drugs, which are copies of already marketed drugs, are regulated by the Office of Generic Drugs.

In the broadest sense, the type of information that should be submitted to the FDA by applicants is set by statute and the Code of Federal Regulations, along with various Guidance documents. The recommendations become more specific as we move from the statute to Guidance.

The amount of information that should be supplied is determined by the stage of development. Generally, lesser amounts of information are supplied in the early stages and submissions become larger and more detailed as development proceeds.

The Type of Information That Should Be Submitted

The FDA's Authority To Regulate Drugs

The FDA's authority to regulate drugs comes from the Food, Drug and Cosmetics Act, as amended. Drawing from the Act, Title 21 of the Code of Federal Regulations (CFR) (Figure 1) is the FDA's interpretation of how the law should be implemented. As an aid to industry the FDA issues a number of Guidance documents that describe in detail various aspects of the process. For some Guidance additional documents that provide Questions and Answers for specific topics have also been issued. The International Conference on Harmonisation guidelines are also used.
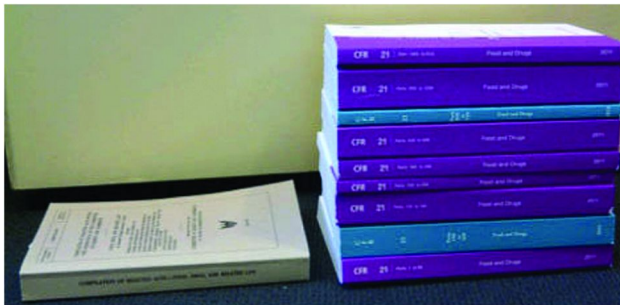
Figure 1. Hard copies of the Food, Drug and Cosmetics Act and Title 21 of the Code of Federal Regulations

The Hierarchy of Regulation

Briefly, the importance of these documents can be ranked as follows:

- Food, Drug and Cosmetics Act
- Title 21 of the Code of Federal Regulations
- Guidance documents
- Questions and Answers for specific topics

As we move down the list the level of detail increases but the mandatory nature of the document decreases. FDA cannot waive the requirements of the law but can waive the requirements of some regulations for sufficient reason (for marketing applications see 21 CFR 314.90 (*1*))

The Guidance spells out FDA policy but need to be grounded in the regulations. How these various requirements mesh together can be illustrated by considering the regulation of analytical procedures. The basic principles, however, apply to all areas of drug development.

Analytical Procedures

Analytical procedures are used to control the quality of a drug. The type of information that should be submitted is set out as follows. From the Food, Drug, and Cosmetic Act (*2*): "…[applicants] shall submit to the Secretary as a part of the application…a full description of the methods used in, and the facilities and controls used for, the manufacture, processing, and packing of such drug." (Section 505 [b] [1]). From Title 21 of the Code of Federal Regulations (*3*): "…the specifications necessary to ensure the identity, strength, quality, purity, potency, and bioavailability of the drug product, including, for example, tests, analytical procedures, and acceptance criteria relating to sterility, dissolution rate, container closure systems…" (from 21 CFR 314.50 [d] [ii] [a], although there are other sections that are relevant).

Various Guidance documents may also be relevant, but principally the following:

- Analytical Procedures and Methods Validation for Drugs and Biologics 2014 (draft) (*4*)
- Reviewer Guidance: Validation of Chromatographic Methods 1994 (*5*)
- Guidelines for Submitting Samples and Analytical Data for Methods Validation 1987 (*6*)
- International Conference on Harmonisation Q2(R1) Validation of Analytical Procedures: Text and Methodology (*7*)

Many other Guidance documents also contain relevant material, such as Changes to an Approved NDA or ANDA 2004 (*8*), which applies to approved new drug applications (NDAs) and abbreviated new drug applications (ANDAs). All FDA Guidance documents can be found at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

Specifically, from the Analytical Procedures and Methods Validation for Drugs and Biologics Draft Guidance (*4*), "…typical validation characteristics are:

- Specificity
- Linearity
- Accuracy
- Precision (repeatability, intermediate precision, and reproducibility)
- Range
- Quantitation Limit
- Detection Limit"

Notice the move from very general to very specific:

- The Food, Drug, and Cosmetic Act provides very specific, but general, standards: "description of the methods … and controls"
- 21 CFR spells it out: "analytical procedures"
- Guidance gets very specific: "specificity, linearity, accuracy…"

Guidance Is Not Mandatory

You can take any approach that you can justify scientifically. Guidance documents typically contain a disclaimer such as: "This guidance…does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible…". Of note, however, if the approach suggested in the guidance is used, it is likely that FDA will accept it without further discussion, whereas if a different approach is chosen, a scientific justification should be submitted and further discussions with the FDA may be required.

The Amount of Information That Should Be Submitted

Applying for Marketing Approval

The process for obtaining approval to market pharmaceuticals is complex and it would take many volumes to describe the entire process. However, the procedure for obtaining approval to market a new drug (i.e. a new molecule that has not previously been used in humans) may be briefly summarized as follows. Please bear in mind that for almost everything discussed there are numerous exceptions and variations, but space only permits the presentation of the barest outline.

Initially, a company submits an investigational new drug (IND) application to cover the process from the first use in humans until the product is ready for market (and even afterwards). When sufficient information has been accumulated a new drug application (NDA) is submitted. This is the marketing application. If approved, the product may be marketed in the United States. For both INDs and NDAs initial submissions are followed by a series of amendments that supply more information. The IND portion of the process typically lasts for a number of years, and the NDA portion can take 6–10 months or longer if major problems are found.

Level of Detail

As the application progresses, the level of detail and the amount of the information that is submitted to the FDA increases. For the initial IND submission the amount of information might be quite limited. With subsequent amendments the amount of information and the level of detail will increase. Generally speaking, as the application progresses larger numbers of people are treated.

When the NDA is submitted it will contain detailed and comprehensive information. Many different types of information are submitted in connection

with any drug application. This information can be divided into three broad categories:

- Quality: will the manufacturing process consistently produce a product that will perform as expected?
- Safety: will the product exhibit untoward toxic effects?
- Efficacy: will the product achieve the desired therapeutic outcome?

For each category one or more groups within the FDA will assess the information, dependent upon the nature of the application. For example, all quality-related information is reviewed by the Office of New Drug Quality Assessment, but a sterile product will also require input from the New Drug Microbiology group of the Office of Pharmaceutical Science. For less complex applications only one person may be assigned from each group, but, for complex applications that need to be reviewed in a short period of time, a multi-person team may be assembled.

Submission of Information about Analytical Procedures

As an illustrative example we will consider the quality section of an application, and within that section we will consider analytical procedures. However, the process is broadly the same for all aspects of the application. First we need some background concerning analytical procedures that are used to make sure that the product will perform as expected.

Attributes of the drug product, such as appearance, identity, assay, impurities, dissolution and sterility, are measured. Notice we may test not only purity (e.g. the levels of various impurities) but also performance (e.g. the speed at which the product dissolves under certain specified conditions). The quality of a drug product is assured by means of the specification, which consists of a series of tests, analytical procedures and acceptance criteria.

A typical (but made up) drug product specification might be analyzed as shown in Table I.

In the above example high performance liquid chromatography (HPLC) is used for assay and impurities. Many different types of analytical procedures are used, but HPLC is regarded as the gold standard for assay and impurities. In its barest essentials, a liquid, such as a mixture of methanol and water, is pumped at high pressure through a column, which is a steel tube filled with specially treated material (generally chemically-modified silica). The sample is introduced through a sample port at one end of the column, and the components of the sample pass through the column to a detector (Figure 2). Different molecules in the sample travel through the column at different rates and emerge at different times. Therefore, the active material can be separated from various impurities, which have different molecular structures.

A graph of the output from the detector (called a chromatogram) would be similar to that in Figure 3.

Table I. Example of a drug product specification[a]

Test	Procedure	Acceptance criterion
Appearance	Visual	Red capsule with white powder fill
Identity	IR	Conforms to reference spectrum
Assay	HPLC	90.0–110.0%
Impurities		
Impurity A	HPLC	NMT 0.7%
Impurity B	HPLC	NMT 0.5%
Each unspecified	HPLC	NMT 0.2%
Total	HPLC	NMT 2.0%
Dissolution	USP <711>	80% in 30 minutes

[a] IR, Infra-Red Spectroscopy; HPLC, high-performance liquid chromatography; NMT, not more than; USP, United States Pharmacopeia.
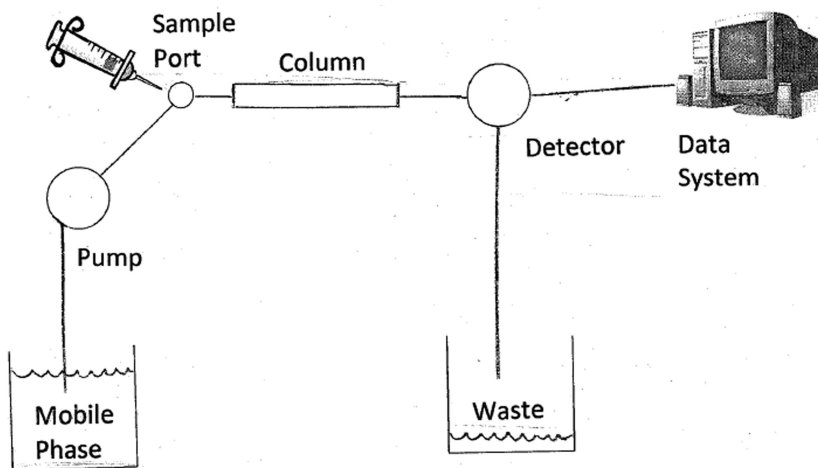
Figure 2. Schematic diagram of a basic high performance liquid chromatography instrument
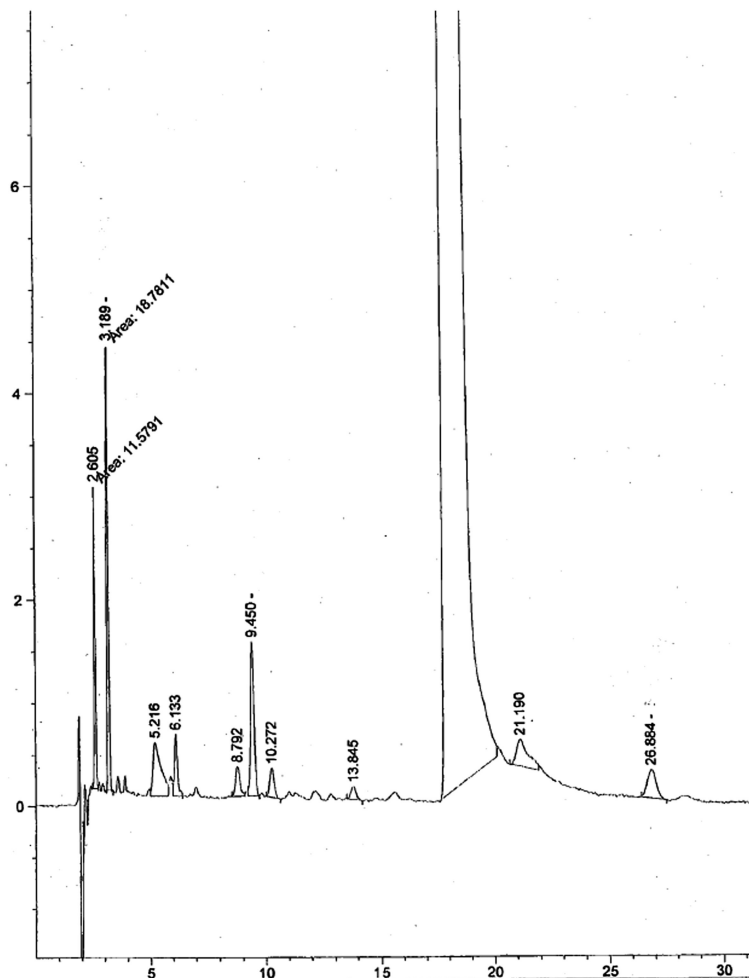
Figure 3. A typical high performance liquid chromatography chromatogram. Graph enlarged to show peaks for impurities. Note that the main peak goes off scale.

The detector response is plotted over time (30 min). Each peak represents a different substance. In the case of Figure 3, the main peak, represents the active pharmaceutical ingredient, and the impurities can be seen as separate smaller peaks. They can be readily quantified by measurement of the peak areas. The maximum acceptable level of each impurity is determined using a variety of considerations, including toxicological testing.

How Might the Submission of Information Be Carried Out in Practice?

The FDA's phase 1 IND Guidance (*9*) suggests that an initial submission should include a brief description of the test methods. Thus, the initial application might just say "a reversed-phase HPLC method". As the IND progresses, a fuller description might be supplied in an amendment. This is illustrated with the following example taken from a published method (*10*):

- Sample preparation: mix 24 mg drug substance with 4 mL acetonitrile and 80 mL mobile phase; sonicate until all solid dissolves, cool, make up to 100 mL with mobile phase; inject a 50 μL aliquot
- Column: 5 μm Supelcosil LC-ABZ (Sigma Aldrich, St. Louis, MO), 150 × 4.6 mm
- Column temperature: 35°C
- Mobile phase: Acetonitrile:25 mM pH 5.0 ammonium dihydrogen phosphate buffer 20:80
- Flow rate: 1 mL/min
- Injection volume: 50 μL
- Detector: Ultraviolet 220 nm
- Retention time: 7.44 min

Thus, enough information is given so that a competent scientist can reproduce the method. As the application moves ahead, the FDA expects to see validation data that increase in detail. When the NDA is submitted it should contain full validation information, together with detailed validation reports.

Using the example cited we might expect to get validation information, such as the following:

- Accuracy: 100.4% (relative standard deviation 0.7%)
- Precision: relative standard deviation 0.36%
- Specificity: no interfering peaks are seen
- Detection limit: 0.001%
- Quantitation limit: 0.003%
- Linearity: 0.99996
- Range: 47–151%

Intermediate precision should also be demonstrated, for instance by showing that different analysts with different equipment on different days can obtain concordant results.

Robustness testing is very important. It consists of making small, deliberate variations in the system parameters and assessing the effect of these changes on the system performance. This provides confidence that small unintentional variations, for example caused by fluctations in the temperature of the room or aging of a pump seal, will not provide unacceptable results. In addition, the system may subsequently need to be changed for operational reasons. If the method stays within the range validated by robustness testing re-validation is

not required. In some cases a parameter, such as buffer pH, may be found to be extremely critical. In that case the method should note that this parameter should be carefully controlled.

An illustration of robustness testing, taken from our example, is hown in Table II. Usually more parameters are varied, but for reasons of space only a few are listed.

Table II. Robustness testing

Variation	Resolution impurity A–nevirapine	Resolution nevirapine–impurity B
Acceptance criteria	> 5.0	> 7.4
Baseline (normal operating conditions)	6.6	10.6
pH 4.9 (lower)	6.8	10.6
pH 5.1 (higher)	6.8	10.6
24.8 mM buffer (lower)	64	10.4
25.2 mM buffer (higher)	6.5	10.5
19% acetonitrile (lower)	7.0	10.8
20.5% acetonitrile (higher)	6.6	10.5

For each change the resolution (separation) of impurity A, impurity B and nevirapine is acceptable and thus we can conclude that the method is robust.

References

1. United States Food and Drug Administration. Title 21 of the Code of Federal Regulations, section 314.90. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.90 (accessed April, 2014).
2. United States Food and Drug Administration. Federal Food, Drug, and Cosmetics Act, section 505(b)(1). http://www.house.gov/legcoun/Comps/FDA_CMD.pdf (accessed April, 2014).
3. United States Food and Drug Administration. Title 21 of the Code of Federal Regulations, section 314.50(d)(ii)(a). http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.50 (accessed April, 2014).
4. United States Department of Health and Human Services; Food and Drug Administration; Center for Drug Evaluation and Research (CDER); Center for Biologics Evaluation and Research (CBER). Guidance for industry: analytical procedures and methods validation (draft). http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM386366.pdf (accessed April, 2014).

5. Center for Drug Evaluation and Research (CDER). Reviewer guidance: validation of chromatographic methods. http://www.fda.gov/downloads/ Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ UCM134409.pdf (accessed April, 2014).
6. United States Food and Drug Administration. Guidelines for Submitting Samples and Analytical Data for Methods Validation. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/ Guidances/ucm123124.htm (accessed April, 2014).
7. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonised tripartite guideline: validation of analytical procedures: text and methodology Q2(R1). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/ Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf (accessed April, 2014).
8. United States Department of Health and Human Services; Food and Drug Administration; Center for Drug Evaluation and Research (CDER). Guidance for industry: changes to an approved NDA or ANDA. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatory Information/Guidances/ucm077097.pdf (accessed April, 2014).
9. Center for Drug Evaluation and Research (CDER); Center for Biologics Evaluation and Research (CBER). Guidance for industry: content and format of investigational new drug applications (INDs) for phase 1 studies of drugs, including well-characterized, therapeutic, biotechnology-derived products. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatory Information/Guidances/UCM071597.pdf (accessed April, 2014).
10. Li, Q. C.; Tougas, T.; Cohen, K.; Lee, R.; Meagan, P.; Corson, M.; Muchnik, T. *J. Chromatogr. Sci.* **2000**, *38*, 246.

Chapter 6

Steps Towards the Analytical Standards Required for Science-Based Tobacco Product Regulation

Christopher Proctor,* Derek Mariner, and Kevin McAdam

British American Tobacco Group Research and Development Centre, Regents Park Road, Southampton SO15 8TL, United Kingdom
*E-mail: christopher_proctor@bat.com

Cigarette smoking is a major cause of a variety of serious diseases due to the frequent and persistent inhalation of a wide range of toxicants found in the smoke. The United States Food and Drug Administration have identified over 90 harmful and potentially harmful constituents in tobacco and tobacco smoke, and scientific studies are seeking both to prioritize these toxicants against the key diseases caused by smoking and to develop robust analytical methods for their measurement in cigarette smoke. Additionally novel tobacco and nicotine products with reduced toxicant levels are being developed, requiring new approaches to toxicant emissions sampling and analysis.

Introduction

Numerous epidemiological studies have shown that cigarette smoking causes a range of serious disease. The relationship between smoking and disease is dose related, with the risks increasing strongly with daily consumption and total duration of smoking. Studies have also found that health risks diminish following the cessation of smoking in a manner dependent on the number of years the subject has smoked and his or her age at cessation Risks either return to levels similar to those in never-smokers, in the case of diseases such as cardiovascular disease and lung cancer, or slow in progression for diseases such as chronic obstructive pulmonary disease (*1*).

This understanding led the United States Institute of Medicine to suggest in a report on the scientific basis for tobacco harm reduction that some of the harm caused by tobacco use might be reduced by the introduction of what it termed potential reduced-exposure products. These were products that (a) result in the substantial reduction in exposure to one or more tobacco toxicants and (b) can reasonably be expected to reduce the risk of one or more specific diseases or other adverse health effects (2). The report was not specific on which toxicants should be reduced nor on the degree of reduction, but rather expected that clinical and other studies would be used to determine whether there was a reasonable expectation that toxicant reductions would result in reductions in health risks.

The Institute of Medicine also noted the importance of regulatory oversight for the assessment and marketing of potential reduced-exposure products. In 2009, the United States Food and Drug Administration (FDA) began regulating tobacco after the introduction of the Family Smoking Prevention and Tobacco Control Act. The Act required the FDA to form a Tobacco Products Scientific Advisory Committee (TPSAC) and to involve this panel in the determination of what harmful or potentially harmful constituents (HPHCs) were present in tobacco and tobacco smoke. This they did, and FDA issued guidance on the identity of these substances and set out further guidance to encourage tobacco product manufacturers in the United States to measure and disclose results for a subset of them. This approach was taken because analytical methods had not been defined for many of the HPHCs rather than through any decision that the subset comprised the most important constituents. The FDA has indicated that the full list of HPHCs will need to be measured in the future. The FDA is also required to publically disclose information on HPHCs levels and may use such information for approval of new tobacco products, including modified-risk tobacco products, and may in the future set standards for tobacco products related to HPHC levels.

Outside of the United States, most countries have ratified the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC). The FCTC is a common framework to guide countries in setting tobacco control regulations. Two of the articles of the framework, Articles 9 and 10, relate to tobacco product regulation, including the measurement and disclosure of tobacco and tobacco smoke constituents and emissions. WHO also has a scientific advisory panel, the Study Group on Tobacco Product Regulation (TobReg), which has issued various reports, including one recommending a possible approach to the mandated lowering of tobacco smoke toxicants (3). This group recommended focus on 18 toxicants in tobacco smoke, nine for potential mandated lowering and nine for monitoring. WHO has also formed the Tobacco Laboratory Network of independent international analytical laboratories (TobLabNet), which is working on establishing standardized methods for assessment of a selection of the toxicants.

Tobacco smoke is a complex mixture of gases, volatile, semi-volatile and involatile compounds. While some of the more than 6,000 constituents of cigarette smoke are present in tobacco, many are formed during combustion (4). Therefore, the combustion conditions within a cigarette, which vary dependent upon the way in which the cigarette is smoked, can affect not only the total yield of an individual compound but also the relative yields of the constituents.

The ultimate purpose of analytical measurement of toxicants is to determine levels of toxicants across different tobacco products in a manner that is relevant to human exposure. There is a considerable variability across any population in the way in which tobacco products are used. In the case of cigarette smokers, this variability includes the number of puffs taken on a cigarette and how large each puff is in terms of the volume of smoke drawn. These factors can have notable effects on the level of exposure to toxicants. Industry researchers, (through the research groups of the Cooperation Centre for Scientific Research Relative to Tobacco, CORESTA), recognized that it was not possible to truly reflect the range of smoking behaviors in a population and standardization bodies (including the International Organization for Standardization [ISO]) set standard parameters for a smoking machine that included the environmental conditions and air flows around the cigarette, the conditioning of the cigarettes, and the mechanics of lighting and puffing the cigarettes (5). Under ISO smoking condition a 35 mL puff is taken of 2 s duration (using a bell-shaped puff profile) every 60 s.

This smoking regime is likely to underestimate the amount of human exposure to toxicants, but it was never intended to be an accurate measure for toxicant exposure. For this purpose various more-intensive regimes have been proposed that take larger and more frequent puffs and that partially or fully block ventilation holes in cigarette filters to prevent the intended dilution of smoke coming from the tobacco rod, as is thought to be done by some smokers with the fingers or mouth. TobLabNet is developing analytical methods based on ISO smoking parameters and what is termed the Health Canada Intense machine smoking regime (after Health Canada, who proposed the method), where the cigarette is prepared with ventilation holes completely sealed and is puffed with a volume of 55 ml over 2 s every 30 s (6). This regime more than doubles the yield of most toxicants relative to yields collected under ISO condition.

That cigarettes are the most risky form of tobacco product is widely accepted, because of the toxicants they emit and the way in which they are used (frequent inhalation of smoke into the lung). Reducing toxicant exposure by changing the route of exposure or the levels or numbers of toxicants present is likely to change the risk profile of the product. For example, epidemiological studies of Swedish snus, a form of oral tobacco product that is both relatively low in toxicant levels and is not smoked, avoiding direct lung exposure, report that use is substantially less risky than cigarette smoking (7). Developing relevant extraction regimes and consequent analytical methods for oral tobacco products poses some different challenges to methods for cigarette smoking. Other tobacco and nicotine products, including electronic cigarettes, are relatively unexplored scientifically, and, although they are likely to have substantially fewer toxicants than cigarette smoke, they might introduce compounds not typically seen in smoke. These gaps in knowledge raise a demand for the introduction of comprehensive non-targeted analysis to supplement targeted analyses of known toxicants.

This chapter considers what science is needed to support the development of regulations on toxicants in a spectrum of tobacco and nicotine products. This includes using computational toxicology and other biological techniques to provide insights into which of the toxicants in most relevant to various tobacco-related diseases, and using this information to focus the development

of robust and reliable analytical methods for the determination of the toxicants. Given that toxicant exposure can vary dependent on how people use these products, analytical approaches need to consider the relevance of the toxicant collection to the purpose of the regulation. Lastly, novel tobacco and nicotine products, including e-cigarettes, will provide new challenges for analysts.

Identification and Prioritization of Toxicants

Identification and characterization of the most important toxicants in cigarette smoke and other tobacco products, in terms of potential to cause disease and dose-response relationships related to various diseases, would be of considerable benefit both to tobacco manufacturers and regulators. Fowles and Dybing, in 2003, described calculations conducted to prioritize the hazards for 158 chemical constituents in tobacco smoke (*8*). On the basis of published cancer potency factors and knowledge of typical yields in smoke, they proposed that 1,3-butadiene, one of the volatile compounds, was the most influential in relation to cancer, and, of the top five cancer-related toxicants, four were aldehydes or small organic compounds. The researchers suggested that around 62.4% of the cancer risk was associated with aldehydes and small organics, a further 18.2% with metals, such as arsenic and cadmium, and only 0.8% with polyaromatic hydrocarbons. From reference exposure levels, the researchers calculated that acrolein and acetaldehyde affected respiratory health, and that hydrogen cyanide and arsenic were connected to cardiovascular disease.

However, Fowles and Dybing noted limitations in their estimates. For example, they estimated the sum of the cancer risk indices that they had calculated and noted that it seemed to be five times lower than would be expected from the cancer mortality attributed to smoking in the USA. Consequently, we have investigated other possible paradigms that might be applicable to these and other tobacco smoke toxicants. Our current risk-assessment paradigm is based on a combination of computer modelling approaches: margin of exposure (MOE) calculations, mode of action (MOA) reviews and physiologically based pharmacokinetic modelling. These models are supplemented with data from *in vitro* models of disease and conventional *in vitro* toxicology assays. The data can be used to generate margins of exposure where *in vivo* data are unavailable, and to provide support for the postulated modes of action for specific chemicals. In addition, this approach can be used to further refine physiologically based pharmacokinetic modelling tools.

MOE and MOA

We propose the application of the MOE model, as described by the European Food Safety Authority (EFSA) guidelines (*9*), which permit the

analysis of genotoxic and carcinogenic compounds. An MOE is the ratio between a benchmark dose (a reference point derived from either experimental or epidemiological dose-response data) and the specific human exposure. The view from the EFSA is that an MOE greater than 10,000 could be considered a low priority for risk management actions. We calculate MOE values from a wide range of different studies with various disease end points to produce a series of values representative of those in the literature. Review of the distribution of the MOE data then allows an assessment to be made on the strength of the relationship by providing a confident representation of the potential risk associated with any specific compound.

A basic criticism of all the risk assessment techniques employed to date is that they have been applied to individual toxicants rather than toxicants within the complex mixture of tobacco smoke. Progress has been made in the field of risk assessment of simple mixtures of chemicals (*10*), but a complex mixture, such as tobacco smoke, presents additional challenges. We have initiated work to investigate the utility of the MOE segregation tool for use in small-scale mixture assessment of three aldehydes (*11*) through careful consideration of their MOA. This MOE model can also be used as part of a quantitative risk assessment paradigm for tobacco smoke toxicants, in conjunction with MOA reviews for the individual toxicants. To generate a combined MOE assessment, two assumptions are made: (a) the compounds involved are similar in structure and (b) they share similar toxicological properties.

The International Programme on Chemical Safety (IPCS) released a proposed framework for evaluating the MOAs of chemicals with carcinogenic and non-cancer effects (*12*). The use of this framework is proposed to lead to a robust evidence-based risk assessment. The first step is to conduct a framework analysis with nine key headings, as set out in the IPCS guidelines. These headings provide a structured and transparent approach to building a network of supporting evidence for the postulated key events and identifying any inconsistencies and data gaps. The postulated MOA can then be evaluated for potential human relevance, taking into account the plausibility of key events and any differences in kinetic and dynamic factors between animals and humans. The assessment will conclude with a statement of confidence, analysis and implications.

The generation of MOA reduces the number of assumptions made in combined MOE assessment by ensuring that similar toxicological properties and lesion types are used. We conducted an MOA review for three aldehydes found in tobacco smoke (acetaldehyde, formaldehyde and propionaldehyde), all of which are saturated aldehydes. On the basis of the available literature and the structural similarities, the following key events were identified and postulated to comprise a common MOA for all three aldehydes: cytotoxicity/genotoxicity, hyperplasia and metaplasia and tumor formation. Cumulative MOEs associated with each of the identified key events could be generated and were postulated to contribute to the induction of respiratory tumors. The MOEs generated were 0.09 for cytotoxicity, 80.55 for genotoxicity, 7.42 for hyperplasia and metaplasia, and 74.15 for tumors. The fact that all four MOEs are considerably below the 10,000 threshold suggests that the three aldehydes in question are of high priority for exposure reduction research (*13*).

Pharmacokinetic Modeling

Predicting the target-tissue concentrations of the toxicologically active chemical species is one of the most technically challenging tasks in assessment of the biological effects of exposure to tobacco smoke. To address this aspect, where possible we are developing physiologically based pharmacokinetic models for individual tobacco smoke toxicants. The output from these models might allow us to determine whether the concentrations that give a positive biological response *in vitro* are relevant to the human target-tissue dose. Although we have made some notable advances in developing experimental methods for *in vitro* toxicant exposure and in the application of biologically based risk assessment, effective collaboration with external laboratories and experts dealing with tobacco and non-tobacco products would greatly aid these endeavors. Only through the combined and concerted effort of the tobacco and chemical industries, government agencies, contract research organizations and academia will significant progress be made in this area. The value in completing such research is a far greater focus for regulators on the toxicants that are most important in tobacco and tobacco smoke and a scientific basis for the potential of reducing health risks by reducing one or more toxicant.

Standardization of Analytical Methods for Toxicants in Tobacco and Cigarette Smoke

Tobacco and tobacco smoke are both complex matrices and present interesting analytical challenges. Many countries require the measurement and disclosure of tar (nicotine-free dry particulate matter), nicotine and carbon monoxide in cigarette smoke, for regulatory reporting, printing on cigarette packs or to ensure that products comply with regulatory limits on the amount of the toxicant allowed. Because of these requirements, a considerable amount of effort has gone into developing and evaluating methods for analysis of these three analytes.

Until recently, very few regulators required data to be collected on an expanded list of toxicants. Health Canada did, requiring measurement and reporting of a list of 44 smoke toxicants (often known as the Hoffmann list, after American Health Foundation scientist Dietrich Hoffmann), the data for which were used to characterize the range of likely toxic constituents in cigarette smoke (*14*). Over the years, these toxicants have been measured in cigarette smoke with many different analytical methods. The methods used in our laboratories are consistent with industry best practices, such as CORESTA Recommended Methods and inter-laboratory ring trials, but most have not been harmonized or internationally validated. Consequently, inter-laboratory differences in measurements range from four to ten times the optimum precision for certain analytes, which could introduce bias in results.

More recently, the FDA has identified 93 HPHCs of interest in tobacco and tobacco smoke. The FDA's Center for Tobacco Products has held workshops to gain information on tobacco analysis, focusing first on tobacco-specific nitrosamines and polyaromatic hydrocarbons. The Center for Tobacco Products

is also seeking to generate new reference materials that might help substantially in the development of analytical methods.

The cigarette smoke toxicants are present in ranges from milligrams (e.g. CO) to attograms (Po-210) per cigarette. Cigarette smoke contains over 6,000 identified but not necessarily quantified constituents (*6*), and probably a much greater number of unidentified constituents (*15*). Given the complexity of the smoke matrix and low yields of many smoke toxicants, it is essential that validated analytical methods are used to measure toxicant yields. Ideally, internationally standardized validated methods would be employed. However, the number of methods that have achieved this status so far is relatively small.

Currently, data from analytical methods that have been developed within a single analytical laboratory must frequently be considered; however, we believe that data should only be accepted from methods developed and validated according to minimum standards of analytical excellence (*16*), developed in laboratories operating under an effective quality management system. The methods should demonstrate robustness, selectivity, and consistency and should be quantitative over the range of yields relevant to cigarette smoke We believe that details of the analytical methodology and the validation data should be made available as part of the reporting documentation, in order that the quality of the method used to determine the yields can be critiqued and that independent verification of smoke yields can take place if necessary.

Choice of Machine Smoking Regime

Three smoking regimes are currently in use for regulatory reporting purposes. The ISO regime is widely used in the European Union and elsewhere; the Commonwealth of Massachusetts and the State of Texas mandate use of a more intense method with 50% filter ventilation blocking, whereas the Canadian Federal government requires smoke yield testing to be conducted with both the ISO3308 and a "maximum emission" Health Canada Intense testing protocol (with 100% ventilation blocking) intended, according to Health Canada, to provide data that reflects the emissions that are actually available to the consumer. In the context of the analytical measurement, ventilation blocking means that the small holes in the cigarette filter that that normally allow the ingress of air to dilute smoke are blocked, completely or partially, by tape or by a shield applied on the smoking engine to reduced dilution.

The ISO regime for machine smoking cigarettes generally underestimates the mean yields of smoke obtained by human smokers (*17*). We and others (*18*) have reported that the Health Canada Intense regime generally overestimates most human behavior or mouth-level exposure (MLE) to smoke constituents. MLE studies measure levels of nicotine in used cigarette filters to estimate the maximum nicotine exposure of the smoker when smoking the cigarette.

We have compared MLE to tar and nicotine in smokers with machine-smoking yields from the same cigarettes for all ISO yields available on the German market. We used the ISO, ISO/TC126 WG9 Option B, Massachusetts regime, Health

Canada Intense and the Kozlowski and O'Connor compensating (*19*) regimes. The results of this comparison showed that the Massachusetts regime data were most similar to mean MLEs for human smokers, for all the products studied, but with considerable smoker to smoker variation.. The Kozlowski and O'Connor compensating and ISO TC126 WG9 Option B regimes provided data that were similar to MLEs for some but not all products, but the Health Canada Intense regime overestimated most mean MLEs (*20*).

MLE studies give some insights into the maximum likely exposure to nicotine. Clinical studies measuring a range of biomarkers of exposure to the toxicants provide complementary information to assess the comparison of smoking machine regimes and the likely toxicant exposures in groups of smokers.

We have presented data from a clinical study (*20*) in which we compared 24 h urinary levels of a number of biomarkers of exposure to toxicants such as acrolein, 1,3-butadiene and tobacco-specific nitrosamines, to machine measured smoke yields from five types of low ISO tar cigarettes tested under ISO, Health Canada Intense, ISO/TC126 Working Group 9 Option B, and a smoking regime with equivalent parameters to those used in Health Canada Intense but with ventilation unblocked. Strong correlations were identified between average 24 h urinary biomarkers of exposure to toxicants and the measured chemistry (determined as the number of cigarettes smoked per day multiplied by the machine yields of toxicants per cigarette), which suggest that machine smoke yields measured under an appropriate regime, together with mean smoking consumption data, can be used successfully to estimate mean smoker exposure to toxicants. Overall, the strongest correlations were obtained with measurements of smoke yields generated with the smoking proposed by ISO/TC126 Working Group 9 Option B regime. Importantly the correlations with these machine yields were stronger than those with other regimes for volatile smoke toxicants—constituents whose yields are strongly influenced by the level of ventilation in a cigarette. Stronger correlations were also obtained between biomarker of exposure levels and machine yields than between biomarker of exposure levels and ratios of yields to nicotine; however, the limited range of nicotine yields in this study weakens this latter analysis.

It appears, therefore, that two alternative approaches can be advocated for the choice of smoking regimes used to generate and measure smoke yield data for cigarettes. The first is use of a bracketing approach wherein two smoking regimes are employed, one such as ISO, which on average underestimates smokers' exposure to cigarette smoke, and one such as Health Canada Intense, which on average overestimates smokers exposure. This approach, which has been proposed by TPSAC, is reasonable in this context as it covers much of the range and extremes of human exposure. The approach is limited, however, in that one tool for the reduction of smoke yields, filter ventilation, is eliminated in the Health Canada Intense regime and, therefore, might produce unrealistic data, particularly for gas-phase constituents. A second approach would be to use a regime that approximates average human exposure for a population. This approach is simpler to interpret but is currently not viable; despite indications that regimes featuring 50% ventilation blocking provide good indications of relative average human exposure to a range of smoke toxicants from different products, this work has yet

to identify an optimum set of parameters relevant to the range of cigarette designs and smoking behavior likely to be encountered on a global basis.

Fit-for-purpose analytical methods need a considerable amount of collaboration. As the list of potentially regulated tobacco smoke toxicants grows, so does the need for inter-laboratory studies. Additionally, the development of further reference materials and products is important, an issue that FDA have recognized as important.

Analytical Strategies for Evaluating Smokeless Tobacco Products

Smokeless tobacco products cover an extremely wide range of tobacco compositions, product configurations, manufacturing techniques, additives and patterns of use (*21*). Many scientists have attempted to group this broad range of products into a collective entity, but such an approach fails to acknowledge that the toxicant levels found in the products can vary widely between the different product styles. For example, the use of some smokeless tobacco products from South Asia is associated with oral cancer, whereas in Sweden substantial epidemiology study has found no increased risk of oral cancer amongst snus users (*7*). These products have quite different chemistries, for example, Swedish style snus typically contains much lower levels of tobacco-specific nitrosamines than some other smokeless tobacco products due to self-imposed toxicant level standards (*21*). However, some epidemiologic studies of smokeless tobacco use conducted in the United States, where some of the smokeless products were likely to have higher nitrosamine levels than found in Swedish snus, showed no increased risk of cancer or cardiovascular disease with smokeless tobacco use (*22*). In the case of smokeless tobacco products, therefore, chemical analysis of their composition may provide an important contribution towards understanding their potential toxicities, but there is no clear understanding of the dose-response relationship between levels of nitrosamines in smokeless tobacco products and disease risks. Several thousand compounds are present in tobacco, and a number are carcinogens or toxicologically active (*21*).

A number of groups have identified constituents of smokeless tobaccos that may be of toxicological concern. TPSAC identified a proposed list of over 40 constituents of smokeless tobacco products covering a range of chemical groups. The International Agency for Research on Cancer reported 28 chemical agents or carcinogens of interest in smokeless tobaccos (*21*). Swedish Match, a manufacturer of Swedish snus, have created a quality standard (GothiaTek), following approaches sometimes taken in foods, that limits the levels of 12 compounds in snus (*23*) and the European Smokeless Tobacco Council have evolved this quality standard into an industry standard for European smokeless tobacco products. TobReg has proposed limits on the levels of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, *N*-Nitrosonornicotine and benzo(a)pyrene in smokeless tobaccos (*24*).

We have investigated the chemical profiles of different contemporary smokeless product types from the United States and Sweden, and have examined

these products for the presence and levels of approximately 100 chemical substances (*25*). Our studies have shown significant differences in these constituents between product types, and a range of constituent levels from milligrams to not detectable for many of those constituents.

A further consideration is the stability of the smokeless tobacco product over the period from manufacture to consumption. Studies (*21*) have shown that the nitrosamine contents of some United States smokeless tobacco styles increases during several weeks' storage at ambient temperatures, and in Sweden many products are refrigerated until they are sold, although a step in the manufacturing process of snus akin to pasteurization is the likely reason that nitrosamine levels remain stable in snus, rather than refrigeration (*23*). We believe that it is useful to collect chemical stability data as part of the product characterization analysis for smokeless tobacco products.

Analysis of Next-Generation Tobacco and Nicotine Products

The development of next-generation products and technologies will require a range of measurement capabilities, including exploratory analyses (e.g. identifying all substances present), rapid semi-quantitative comparisons of samples (to understand differences in chemical profiles quickly) and quantitative analyses of established toxicants (to regulatory standards but preferably in a shorter time). To establish these capabilities, use of techniques, such as high resolution time-of-flight mass-spectrometry, that are capable of resolving and identifying chemicals in a complex mixture, will be required (*26*). We have also been investigating, through collaboration, the potential of nuclear magnetic resonance spectroscopy (*27*). These instruments can simultaneously detect about 2,500 substances in mainstream smoke condensate. Nuclear magnetic resonance spectroscopy can detect and measure 20 of the 44 Hoffmann toxicants in mainstream tobacco smoke condensate with a simple smoke-collection procedure, which is likely to be extended. We have also applied gas chromatography coupled with high-resolution time-of-flight mass spectrometry to the comparative analysis of smoke samples. This method can automatically detect differences in chemical profiles and quantify them for potential risk assessment.

A parallel need is to develop a chemical informatics platform that will conduct rapid processing of multiple sample replicates, enabling correlation of chemical profile data between linked samples. Electronic cigarettes, also known as e-cigarettes, e-cigs or electronic nicotine delivery systems, are typically cigarette-shaped, battery-powered electronic devices that produce an aerosol that users inhale. These products are used by cigarette smokers as a substitute for the experience of smoking tobacco. There is no combustion and, therefore, the user does not inhale smoke and, although the nicotine is derived from tobacco, e-cigarettes contain no actual tobacco. The aerosol is generated by an electric heating element that is activated manually or when the user draws air through the device. This vaporizes a liquid formulation, often referred to as e-juice or e-liquid, and the vapor then condenses to form an aerosol. This formulation is held in a reservoir, often absorbed on to a foam material, and typically contains

nicotine, propylene glycol and/or glycerol and flavorings. Commonly, a colored light-emitting diode at the end of the device is lit during puffing.

Electronic cigarettes are available in many formats and designs, with various lengths and circumferences or shapes differing completely from the cigarette-like format. Most formats are available as one-, two- or three-piece products. Two-piece products consist of a battery and a cartomizer (comprising a cartridge and atomizer); in three-piece products the cartridge and atomizer are separate. The devices may be disposable or reusable, with rechargeable batteries and refillable cartridges or reservoirs.

An electronic cigarette, when operating correctly, should simply transfer the components of the e-liquid formulation into the aerosol for inhalation. Thus, any impurities present might be transferred, and other substances can be formed if the device is not operating correctly (e.g. if the heating coil gets too hot and causes thermal degradation of the e-liquid constituents or components of the e-cigarette or if there are reactions between constituents and/or degradation products). Some impurities and degradation/reaction products could be toxic.

The characteristics of the aerosol depend mainly on the power applied to the heating coil, the physical characteristics of the formulation (viscosity, wettability and so on) and the specific heat capacity of the formulation. In addition, the pressure drop, airflow rate and aerosol density may vary between and within devices (*28*, *29*). In an investigation of numerous cartridges and refills (*30*), formulations were tested in 16 devices with a smoking simulator, with use of puffing parameters derived from user puffing profiles. Mean nicotine yields were 0.3 (±0.2) mg to 8.7 (±1.0) mg from 150 puffs and 0.5 (±0.1) mg to 15.4 (±2.1) mg from 300 puffs.

A further study (*31*) investigated the yields generated from ten puffs of an electronic cigarette under two puffing regimes and three durations. The findings suggested that moving from ISO to intense puffing parameters (puff volume, duration and frequency) showed around 50% increase in particulate matter and nicotine yields, which is much less than the approximately threefold increase seen with conventional cigarettes, but that there were significant differences between devices in terms of sensitivity to puff duration.

Recent analysis has shown that the other alkaloids and degradation products can be present at between 0 and 4.4% of the nicotine content, but for most of the samples tested were present at 1–2% of the nicotine (*32*).

The presence of carbonyl compounds (formaldehyde, acetaldehyde, acrolein, and *o*-methylbenzaldehyde), volatile organic compounds (toluene, *p*-xylene and *m*-xylene), tobacco-specific nitrosamines (4-[methylnitrosamino]-1-[3-pyridyl]-1-butanone and *N*-Nitrosonornicotine) and certain heavy metals (cadmium, nickel, lead) have been found in the aerosols produced by a range of electronic cigarettes, but at levels 9–450 times lower than the typical levels in cigarette tobacco smoke (*33*).

Standardized methods are important to enable comparison of results for electronic cigarettes from different laboratories and products. In the absence of any standardized testing procedures, a number of testing approaches for conventional cigarette have been used. Standard machine smoking parameters for conventional cigarettes, however, require modification for use with electronic

cigarettes, for example, puff shape might need to differ and puff durations may have to be longer to ensure actuation of the devices, and puff volumes may have to be larger to reflect user behavior better. Analytical methodology to characterize these devices specifically needs to be developed and validated to ensure fitness for purpose. Extension of conventional smoke chemistry methodology is not always appropriate.

Conclusions

The analytical challenges faced by those needing to measure toxicants in cigarette smoke have increased, with both a greater number of toxicants to measure and a wider range of tobacco and nicotine products to test. Given the large negative impact of cigarette smoking on public health, and the potential to reduce harm with the next generation of tobacco and nicotine products, scientifically based tobacco product regulation becomes critical. One of the key underpinning sciences to this will be analytical chemistry, and the best way to rapidly develop the methods needed is through collaborative research between the regulators, the regulated industry and other scientific stakeholders.

References

1. Surgeon General. *The health consequences of smoking – 50 years of progress. A report of the Surgeon General*; United States Department of Health and Human Services: Rockville, MD, 2014.
2. Institute of Medicine. *Clearing the Smoke: Assessing the Scientific Base for Tobacco Harm Reduction*; National Acadamy Press: Washington, DC, 2001.
3. Burns, D.; Dybing, E.; Gray, N.; Hecht, S.; Anderson, C.; Sanner, T.; O'Connor, R.; Djordjevic, M.; Dresler, C.; Hainaut, P.; Jarvis, M.; Opperhuizen, A.; Straif, K. *Tob. Control* **2008**, *17*, 132–41.
4. Rodgman, A.; Perfetti, T. *The chemical components of tobacco and tobacco smoke*, 2nd ed.; CRC Press: Boca Raton, FL, 2013.
5. ISO. *Cigarettes–Determination of Total and Nicotine Free Dry Particulate Matter using a Routine Analytical Smoking Machine. ISO 4387:2000*; International Organization for Standardization: Geneva, Switzerland, 2000.
6. Health Canada. *Determination of "tar", nicotine and carbon monoxide in mainstream tobacco smoke official method*; Health Canada: Ottawa, Canada, 1999.
7. UK Tobacco Advisory Group of the Royal College of Physicians, Harm reduction in nicotine addiction. *Harm reduction in nicotine addiction: helping people who can't quit*; Royal College of Physicians: London, 2007.
8. Fowles, J.; Dybing, E. *Tob. Control* **2003**, *12*, 424–430.
9. EFSA. *EFSA meeting summary report. EFSA/WHO international conference with support of ILSI Europe on risk assessment of compounds that are both genotoxic and carcinogenic: 16–18 November, 2005, Brussels, Belgium*; European Food Safety Authority: Parma, Italy, 2006.

10. Meek, M. E.; Boobis, A. R.; Crofton, K. M.; Heinemeyer, G.; Van Raaji, M.; Vickers, C. *Regul. Toxicol. Pharmacol.* **2011**, *60*, S1–S14.

11. Cunningham, F. H.; Fiebelkorn, S.; Johnson, M.; Meredith, C. *Food Chem. Toxicol.* **2011**, *49*, 2921–2933.

12. International Programme on Chemical Safety (IPCS) Harmonization Project. *IPCS Mode of Action Framework*; World Health Organization: Geneva, Switzerland, 2008.

13. Cunningham, F.; Breheny, D.; Fiebelkorn, S.; Dillon, D.; Meredith, C.; Proctor, C. Cumulative risk assessment of three aldehydes present in tobacco smoke: Application of Margin of Exposure (MOE) and Mode of Action (MOA) evaluations. Presented at Society for Research into Nicotine and Tobacco 19th Annual Meeting, 13–16 March, 2013, Boston, MA.

14. Hoffmann, D.; Wynder, E. The chemical composition and tumourgenicity of tobacco smoke. In *The Chemistry of Tobacco and Tobacco Smoke: Proceedings of the Symposium on the Chemical Composition of Tobacco and Tobacco Smoke held during the 162nd National Meeting of the American Chemical Society in Washington, D.C., September 12–17, 1971*; Schmeltz, I., Ed.; Springer Verlag: Dordrecht, Netherlands, 1972.

15. Dallüge, J.; van Stee, L. L. P.; Xu, X.; Williams, J.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. *J. Chromatogr. A* **2002**, *974*, 169–184.

16. International Union of Pure and Applied Chemistry. *Pure Appl. Chem.* **2002**, *74*, 835–855.

17. Mariner, D.; Ashley, M.; Shepperd, J.; Mullard, G.; Dixon, M. *Regul. Toxicol. Pharmacol.* **2010**, *61*, S39–S50.

18. Hammond, D.; Fong, G. T.; Cummings, K. M.; O'Connor, R. J.; Giovino, G. A.; McNeill, A. *Cancer Epidemiol., Biomarkers Prev.* **2006**, *15*, 1495–1501.

19. Kozlowski, L. T.; O'Connor, R. J. *Lancet* **2000**, *355*, 2169–2171.

20. McAdam, K.; Eldridge, A.; Shepperd, C.; Mariner, D.; Proctor, C. Machine smoking regimes as indicators of human exposure to smoke toxicants. http://www.bat-science.com/groupms/sites/BAT_7AWFH3.nsf/vwPagesWebLive/DO8EAMC4/$FILE/Machine%20smoking%20regimes_Final.pdf?openelement (accessed June, 2014).

21. IARC. *Monograph 89: smokeless tobacco and some tobacco-specific nitrosamines*; International Agency for Research into Cancer: Lyon, France, 2007.

22. Accortt, N. A; Waterbor, J. W.; Beall, C.; Howard, G. *Am. J. Epidemiol.* **2002**, *156* (8), 730–737.

23. Rutqvist, L. E.; Curvall, M.; Hassler, T; Ringberger, T.; Wahlberg, I. *Harm Reduction J.* **2011**, *8*, 11.

24. WHO Study Group on Tobacco Product Regulation. *Report on the scientific basis of tobacco product regulation*, WHO technical series 955; World Health Organization: Geneva, Switzerland, 2010.

25. Digard, H.; Gale, N.; Errington, G.; Peters, N.; McAdam, K. *Chem. Cent. J.* **2013**, *7*, 55.

26. Frosina, J.; van Heemst, J.; Carradus, M.; Wright, C. Application of high resolution time-of-flight mass spectrometry for the evaluation of tobacco smoke chemical profiles. http://www.bat-science.com/groupms/

sites/BAT_7AWFH3.nsf/vwPagesWebLive/DO9BMCHQ/$FILE/BMSS_ SC_2013.pdf?openelement (accessed June, 2014).

27. Ticha, J.; van Heemst, J.; Charlton, A. Non-targeted analysis of selected Hoffman toxicants in smoke condensate by Cryoprobe 1H NMR. http://www.bat-science.com/groupms/sites/BAT_7AWFH3.nsf/ vwPagesWebLive/DO9BMDXK/$FILE/TSRC_JT_2013.pdf?openelement (accessed June, 2014).

28. Williams, M; Talbot, P. *Nicotine Tob. Res.* **2011**, *13*, 1276.

29. Trtchounian, A.; Williams, M.; Talbot, P. *Nicotine Tob. Res.* **2011**, *12*, 905.

30. Goniewicz, M. L.; Kuma, T.; Gawron, M.; Knysak, J.; Kosmider, L. *Nicotine Tob. Res.* **2013**, *15*, 158–166.

31. Taylor, M. Effect of puff duration and volume on the yields of e-cigarettes. Presented at CORESTA Congress; 23–27 September, 2012; Sapporo, Japan.

32. Etter, J. F.; Zäther, E.; Svensson, S. Analysis of refill liquids for electronic cigarettes. *Addiction* **2013**, *108*, 1671–1679.

33. Goniewicz, M.; Knysak, J.; Gawron, M.; Kosmider, L.; Sobczak, A.; Kurek, J.; Prokopowicz, A.; Jablonska-Czapla, M.; Rosik-Dulewska, C.; Havel, C.; Jacob, P., 3rd.; Benowitz, N. *Tob. Control* **2014**, *23*, 133–139.

Chapter 7

Regulatory Toxicology: Progress in Science

Helena Hogberg[1] and Thomas Hartung[*,1,2]

[1]Johns Hopkins Bloomberg School of Public Health, Center for Alternatives to Animal Testing, Baltimore, Maryland 21205, United States
[2]University of Konstanz, Center for Alternatives to Animal Testing Europe, Konstanz 78457, Germany
[*]E-mail: thartun1@jhu.edu

Regulatory toxicology is faced by challenges of cost, throughput, human relevance and animal use. Additional issues are unmet challenges, such as mixture toxicology, susceptible subpopulations, new products, hazards and transparency. Stimulated by the 2007 National Research Council report on toxicity testing in the 21st century, a number of activities have started to revamp regulatory toxicology. These include programs by the Federal agencies, human-on-chip programs, the Human Toxome Project, efforts for translating evidence-based medicine to toxicology and so on. The Center for Alternatives to Animal Testing (CAAT) at Johns Hopkins Bloomberg School of Public Health, Baltimore, M.D., is one of the groups involved in these developments. A summary of the CAAT's overall activities in the field is combined with illustrations of proof-of-principle work.

Introduction

Every year, about U.S.$3 billion is spent worldwide on animal tests to ensure the safety of consumer products, including drugs, chemicals, food and cosmetics (*1*, *2*). While pesticides and drugs are extensively tested on animals, food additives mostly are not, and animal testing of cosmetics is banned in some parts of the world. But what are these tests worth when a common, relatively safe drug such as aspirin fails most of them (*3*)? What does it tell us if 23 of 31 tested ingredients in coffee produce cancer in rats (*4*), other than humans are not 70 kg rats? Certainly, animal experiments have contributed to patient and consumer safety, but on the

expense of many false-positive results from high-dose treatments. How many business and regulatory decisions are made on these flawed data? How many potentially useful or even life-saving drugs never made it to the clinics?

Over the past two decades a biotech revolution has taken place that has delivered proof of the principle that things can be done differently. Roughly 50 alternative methods have been internationally validated (*5*) and are increasingly being used to establish product safety. These methods are used mainly to test for acute and local effects (e.g. skin sensitivity and eye irritation), but they serve as door openers for a new era in regulatory science. Of note is that *in vitro* and *in silico* approaches are no less limited than animal tests in predicting human hazard.

Exciting technical advances are enabling the development of the next generation of safety tests. New technologies can visualize and pinpoint at a molecular level what is happening when substances harm a tissue. Major new technologies include miniature 'organs-on-chips' (*6*) to test drugs and 'virtual experiments' created with computer modeling (*7*). Testing strategies that integrate a variety of tools might ultimately replace animal experiments entirely.

The emergence of new scientific tools for creating safer products (including so-called green toxicology for the design of the next generation of non-toxic substances) also allows quick and low-cost development and manufacturing. Of course, it is crucial that the new methods maintain high safety standards. Despite the culture of validation that clings to what many believe is excessive regulation, the acceleration of technologies is forcing the development of new methods of quality assurance, from the establishment of best practices to systematic reviews and meta-analyses, inspired by evidence-based medicine (*3*, *8*, *9*). A new 'regulatory science' (a term increasingly used by regulatory agencies to describe innovative changes for risk assessment) has emerged—one that is evidence-based, humane and predictive for human risk.

This chapter summarizes emerging technologies and approaches, which together are changing regulatory toxicology. The Centers for Alternatives to Animal Testing (CAAT, http://caat.jhsph.edu) on both sides of the Atlantic steer some of these efforts (*10*), which are described here to illustrate the overall change taking place.

Technologies of Toxicity Testing in the 21st Century

Today's mechanistic toxicology, to a large extent, is effectively relying on methodologies that substitute or complement traditional animal tests. The biotechnology and informatics revolution of the past few decades has made such technologies broadly available and useful. Regulatory toxicology has only slowly begun to embrace these new approaches. Major validation efforts, however, have delivered the evidence that new approaches do not lower safety standards and can be integrated into regulatory safety assessments (*11*). In the United States, the National Academy of Sciences has created the Toxicology in the 21st Century (Tox 21) program. The most recent adaptation by the Environmental Protection Agency (EPA) for their toxicity testing strategy has initiated a debate on how to create a novel approach based on human cell cultures,

lower species, high-throughput testing and modeling (*12*). Systematic mapping of the entirety of pathways of toxicity, termed the human toxome, has been started (http://humantoxome.com) (*13*, *14*). This effort should lead to pathway-based tests and, ultimately, to the integration of results in a systems toxicology approach (*15*). The development, validation and acceptance of alternative methods have led to a new approach for regulatory toxicology. A multi-stakeholder process to develop a roadmap for replacing animal-based systemic toxicity testing has been started (*4*). Integrated testing strategies (ITS) and approaches based on pathways of toxicity are most promising when used in accordance with Tox 21. Beside the technical development of new approaches, conceptual steering and an objective assessment of current practices by evidence-based toxicology also seem to be needed (*16*, *17*).

The Most Important of the -omics Is Economics

Animal testing for product safety amounts to $3 billion per year worldwide (*1*, *2*). Although this number is impressive and this market has attracted a number of contract research organizations (*18*), the trade of regulated industries is short of $10 trillion (roughly $3 trillion each for drugs, foods and chemicals and $600 billion for cosmetics), that is, toxicity testing amounts only to 0.03% of turnover. Animal experimentation worldwide directly employs about 73,000 people, and more widely involves about 300,000 researchers. In terms of animals, 12.1 million were used in the European Union (EU) in 2005. This level was similar in 2008 and slightly decreased to 11.5 million in 2011. Of these animals, in 2005, 23.3% were used for regulatory tests plus 31.0% for industrial research and development (R&D); in 2011, 22.7% were still used for regulatory use, but R&D use had declined to 18.8%. This quite dramatic change reflects, as R&D investment is not increasing, a lesser reliance on animal studies by pharmaceutical industry and but also outsourcing of testing to countries outside of the EU.

The contribution to the world market of different industrial sectors varies. For instance the EU markets contribute the following: drugs 28% (86% for vaccines), chemicals 35%, cosmetics 50% and food 29%. The burden of testing, however, is very different dependent on the country and area, ranging for example in the United States from almost nothing for industrial chemicals to more than $10 million for a new pesticide. In addition, indirect testing needs are produced by product safety liabilities. European legislation is typically more prescriptive. With imminent testing on larger scale for Registration, Evaluation, Authorization and restriction of CHemicals (REACH), these numbers are obviously increasing. Of note is that testing costs are only a part of the overall expenses related to complying with regulatory demands. Estimates for costs and animal use for this program are still under debate and very much will depend on the reinforcement of the legislation by the European Chemicals Agency.

International testing requirements for new drugs are consistent owing to the International Conference on Harmonization. In practice, the demands made by the United States Food and Drug Administration (FDA) are strongly dominating, as two-thirds of sales of drugs under patent protection are in the US. Attempts

of the FDA leadership to revamp regulatory sciences are, therefore, of critical importance. With 40,000 animals per new substance entering the market tested for safety and 350,000 for R&D (factoring in the animals used for candidate drugs that do not succeed), pharmaceuticals still has very high animal use.

Food testing demands are relatively small and relate only to a few new food additives per year, with the notable exception of shellfish toxin testing, which involves several hundreds of thousands of mice per year (19). Pesticide licensing requires around 10,000 animals per product, but a low number of new products limits this contribution. Whether new EU legislation, which has effectively taken half of the plant protection products off the market, will lead to strong increases in new submissions remains to be seen.

While the monetary costs of animal testing seem to be bearable, the costs of making the wrong business decisions on the basis of the results might not. Therefore, the development of tools that better predict effects is in the best interest of regulated industries.

Look Back in Anger? What Clinical Trials Tell Us about Pre-Clinical Research

Have animal studies been misleading? A closer look at the outcomes of clinical trials in disorders, such as stroke or septic shock, shows that the preclinical models frequently prove to be of limited value. For all indications, 95% of drugs that enter clinical trials will not make it to the market, despite all the promise of the animal models used to develop them (20): roughly 20% have to be stopped because of side-effects, 40% show no efficacy and most others too little efficacy to allow marketing. Drug development has, however, started to lessen its reliance on animal models, with notable decreases in animal studies seen since 2005.

What does this tell us for areas where there are no or few clinical trials (21)? Toxicology is a prime example: are we selecting the wrong candidate substances? Aspirin would today likely fail the preclinical stage. Rats and mice predict results in each other for only 60% of complex endpoints and together predict only 43% of any clinical toxic effects observed later (22). Of note is that clinical trials do not typically address long-term side-effects of treatments. New approaches that rely on molecular pathways of human toxicity are being put forward in Tox 21 (23).

Doubts about the use of animal models are also increasing for drug efficacy testing. A National Academy of Science panel analyzed the suitability of animal models to assess the human efficacy of countermeasures to bioterrorism. The panel could neither identify any suitable models nor recommend their development. Rather, it called for the establishment of other human-relevant tools. In line with this recommendation, about $200 million have been made available by the National Institutes of Health (NIH), the FDA and United States Department of Defense agencies over the past year to start developing a 'human-on-chip' approach (6).

There is no reason to assume that other preclinical animal research is more predictive than that carried out in drug industry. Begley and Ellis (24) reported that only six of 53 landmark studies in cancer could be reproduced by industry.

Previously, Bayer had reported only about 25% reproducibility (*25*). These finding suggest that publishing fewer studies of better quality that do not rely on the face value of animal studies would be beneficial.

Problems with the Current Approach to Safety Assessments of Agents

A number of problems increasingly urge for changes in the way regulatory toxicology is carried out.

Disparity exists in testing requirements and risk acceptance for different products. The extent of testing required and the methods to be used, as well as risk management decisions are very different for products and across geographical regions. This inconsistency makes scientifically little sense and hampers economic globalization (*26*).

Throughput and costs of testing can compete with testing needs. Only a small proportion of substances in daily use has been evaluated to a large extent (although REACH legislation is making some change here) but the tests have cost several millions of dollars per substance. This gap will be difficult to close (*27*).

Animal testing has limited predictability for humans. Additionally, different animal species predict effects in each other with limited success for complex health endpoints (*22*).

Most animal methods were established for drugs under development, where precaution is advisable before moving into human studies. This approach, however, is not necessarily advisable when testing existing substances that have been in use for decades (*16*). Additionally, ethical concerns about the use of animals are growing worldwide and are increasingly being embraced in legislation (*28*).

Some new products are not suitable for traditional testing, such as biologics, cell therapies, genetically modified and functional foods (nutraceuticals), medical countermeasures to biological and chemical terrorism and warfare agents, medical devices and nanoparticles (*29–31*). In the same vein, some new hazards are not covered, such as endocrine effects, childhood effects (e.g. asthma) and behavioral effects. Obesity and cardiovascular effects are also not adequately tested. Mixtures of toxicants are not adequately addressed, as the many possible combinations, doses and timings of exposure cannot be fully addressed by the costly low throughput animal methods we have. Individual susceptibilities and effects in vulnerable subpopulations cannot be satisfactorily modeled with the traditional tools because they rely on healthy, inbred young rodents.

Finally, poor research and publication standards impair the predictability and usefulness of toxicological studies.

Toxicology is not different in its problems from other fields in the life sciences, and perhaps has even more issues with regards to internationally harmonized methods and quality assurance. Although there is enough reason to promote change, the current approaches must remain in place for a while. The emerging approaches still need to show whether they can more adequately address the challenges of toxicology testing. Change, however, starts with awareness of

the need to tackle such challenges. It requires our willingness to change practices and not to waste time on defending traditional approaches.

Improving the Predictability of Cell Cultures

Efforts to make cell cultures more organotypic and to integrate different organoids by perfusion have been boosted by needs of the United States Department of Defense (*6*). The desire to develop and evaluate drugs as potential countermeasures for biological and chemical threats requires test systems that can substitute for the clinical trials, which are normally crucial for drug development. As animal models have limited predictability for drug efficacy, traditional *in vitro* and *in silico* approaches are not really game-changers here. The substantial investment into novel tools now underway, however, might bring about a second generation of alternative approaches. The avenue pursued focuses primarily on the development of a combination of different human three-dimensional (stem) cell-based organ equivalents combined with microfluidics ('human on chip').

Over the past 2 years, three funding opportunities in the United States provided more than $200 million to advance the 'human-on-chip' concept (*6*). An alliance of United States agencies is tackling the problem of evaluating drugs for which there are no, and hopefully never will be, patients. Three different calls from NIH, FDA and the Department of Defense agencies, the Defense Threat Reduction Agency and the Defense Advanced Research Agency have been made for the development of such models. The obvious problem is the lack of patients for clinical development, which make traditional product registration with the FDA impossible. The original response was the suggestion to use appropriate animal models instead. In May, 2002, the FDA amended its new drug and biological drug product regulation to allow the substituting of evidence of efficacy in humans with that from animal studies if a "reasonably well understood pathophysiological mechanism for the toxicity … and its amelioration or prevention by the product" was given, "effect [was] demonstrated in more than one animal species" or "a single animal species … predicting the response in humans," a "study endpoint [was] … generally the enhancement of survival or prevention of major morbidity," and "pharmacokinetics and pharmacodynamics … in animals and humans [was] sufficiently well understood." Safety evidence from animals cannot substitute for clinical findings (*32*). The US Department of Defense sponsored a National Academy of Sciences report, *Animal Models for Assessing Countermeasures to Bioterrorism Agents*, published in December, 2011 (*33*). One author (TH) had the privilege of being part of the committee. The key findings of the report are that neither animal nor alternative methods are available for this purpose, but the committee discouraged the development of further animal models. Instead it proposed the exploitation of new alternative approaches.

The animal models, which were obviously not fit for purpose, paired with the need to regulate these new products seems to have opened doors for new approaches. The prospects of such approaches, their impact on the field of alternative approaches and the necessary complementary activities need to

be discussed. Similarly, the need to adapt quality assurance measures and experiences from validation has to be stressed.

Integrated Testing Strategies for Safety Assessments

Despite the fact that toxicology uses many stand-alone tests, very often a systematic combination of several information sources is required. For example, all possible outcomes of interest (e.g. modes of action), classes of test substances (applicability domains) or severity classes of effects might not be covered in a single test. Furthermore, sometimes positive test results are rare, for instance because of low prevalence leading to excessive false-positive results, or the gold standard test is too costly or uses too many animals, which creates a need for prioritization by screening. Tests are also combined when one test does not satisfactorily predict effects in humans or when existing data and evidence from various tests are integrated. Increasingly, kinetic information needs to be included to enable *in vivo* extrapolation from *in vitro* data.

The solution to these problems is ITS (*34*). Such strategies have been discussed for more than a decade and some attempts have been made to incorporate guidance for regulatory testing. Despite the obvious potential to revamp regulatory toxicology, however, little guidance is available on the composition, validation and adaptation of ITS for different purposes. Similar to approaches of weight of evidence and evidence-based toxicology, different pieces of evidence and test data need to be weighed and combined. ITS represent the logical way of combining pathway-based tests, as suggested in Tox 21.

Mapping the Human Toxome

In the United States, Tox 21 has created an atmosphere of departure from traditional methods to use of modern technologies based on pathways of toxicity (*35*). Pathways could be modeled in relatively simple cell tests that could be run by robots. The goal is to develop a public database for such pathways, called the human toxome, to enable scientific collaboration and exchange.

Awareness of Tox 21 is growing. It was first embraced by scientists and in the United States. Most importantly, the United States agencies quickly took up the recommendations. After the program was set up in 2008, the EPA made it their chemical testing paradigm in 2009, and the FDA did the same soon afterward. Industry engaged with various organizations, such as the Human Toxicology Project Consortium. By contrast, in Europe, change has been rather delayed, with some adaptation of vocabulary but not necessarily grasping of the new approach. However, interest has now begun to increasing strongly in Europe.

Tox 21 suggests moving to a new resolution with pathways of toxicity (*14, 36*). The problem is that the respective science is only emerging. What is needed is the completed human toxome to provide the comprehensive pathway list, annotated with relevant cell types, species, toxicant classes and hazards. Additionally, information in systems toxicology approaches need to be integrated, reverse dosimetry must be done to enable *in vitro* to *in vivo* extrapolation and the

data need to be properly interpreted, most likely in a probabilistic way. The NIH has provided a transformative research grant for the human toxome project, led by CAAT, since September, 2011. The project involves the EPA's Toxicity Forecaster program (ToxCast™), the Hamner Institute, Agilent and several members of the Tox 21 panel. The approach is shaped around pro-estrogenic endocrine disruption as a test case.

Translating Evidence-Based Medicine to Toxicology

Review of Existing Data

Early on the need for quality assurance for alternative approaches was noted. The Evidence-based Toxicology Collaboration was created in the United States and Europe in 2011 and 2012, respectively (*9*), with the secretariat run by CAAT. This collaboration of representatives from all stakeholder groups aims to develop tools of evidence-based medicine for toxicology (*3*, *8*, *37*). Evidence-based medicine has revolutionized clinical medicine and, driven by the Cochrane Collaboration with tools such as systematic reviews, meta-analyses and quality scoring of studies, has fostered consensus evaluations of controversial areas of medical practice. While the first systematic reviews in toxicology are on their way, tailored tools, such as quality scoring tools (*38*), and new approaches to validation are emerging (*27*, *39*, *40*). Altogether, Tox 21 and its implementation activities, including the human toxome program and the Evidence-based Toxicology Collaboration, promise a credible approach to revamp regulatory toxicology.

ToxCast™

Various research initiatives have been piloted as part of Tox 21. The main research initiative is ToxCast™ (http://www.epa.gov/ncct/toxcast). ToxCast™ was launched in 2007 and uses high-throughput screening assays to assess toxic effects of chemicals. Phase I of the program was completed in 2009 and screened approximately 300 compounds in over 500 biochemical or cell-based assays (*41*). In 2013, phase II was completed for about 1,800 chemicals evaluated in over 800 assays. Chemicals in phase II include drugs that have failed in clinical trials and might provide information on how to build tests that better predict effects in humans. The obtained ToxCast™ data are used to predict hazards and modes of action and/or to prioritize chemicals for further testing. Moreover, the data are used to build virtual prediction models. One of the EPA's most successful models is the virtual embryo project (v-Embryo™, http://epa.gov/ncct/v-Embryo), which uses the ToxCast™ data to simulate how chemicals can cause developmental problems in embryos. This model has been useful to predict disruption of vascular development and to provide valuable mechanistic information on toxicity and adverse outcome pathways (*42*, *43*). The EPA's Endocrine Disruption Program is also making use of ToxCast™ data to prioritize chemicals that need to be tested for potential endocrine-related activity.

Endocrine Disruptors

A different approach intended to implement Tox 21 for testing of endocrine disruption is the NIH funded project, Mapping the Human Toxome by Systems Toxicology, which is led by CAAT (*13*). This project involves Brown University, Hamner Institutes for Health Research, Georgetown University, the EPA's National Center for Computational Toxicology, Agilent Technologies and several members of the Tox 21 panel. The aims is to comprehensively map pathways of endocrine disruption (*44*) as a first step towards mapping the human toxome. The area of endocrine disruption is especially suited to pioneer this approach, as the physiological pathways of the endocrine systems are reasonably well understood. The human toxome project is making use of an immortalized human breast adenocarcinoma cell line (MCF-7) that has been pre-validated by the Interagency Coordinating Committee on the Validation of Alternative Methods for use in screening for substances that induce cell proliferation via estrogen-receptor-mediated pathways (*45*). The cells are exposed to compounds with well-known endocrine disruption potential and the response is evaluated with untargeted mass spectrum-based metabolomics and gene-array-based transcriptomics (*46*, *47*). The obtained -omics data are integrated with biochemical information to identify and annotate pathways of toxicity for a defined set of endocrine disruptors (*36*). The final goal of this project is to use the data to establish the beginnings of a publicly available database of pathways.

Developmental Neurotoxicity

An area of toxicology where Tox 21 and the human toxome project could have notable impact is developmental neurotoxicity. Animal testing strategies for developmental neurotoxicity have limitations: high costs ($1.4 million per substance) and time consumption (*48*, *49*). In addition, there are scientific concerns about the relevance of these studies for human health effects. Consequently, only very few substances have been identified as developmental neurotoxicants (*50*, *51*), for which evidence shows that exposures to environmental chemicals contribute to the increasing incidence of neurodevelopmental disorders in children, such as autism and attention deficit hyperactivity disorder (*52–54*). This increase might be due partly to improved diagnosis of these diseases, but a National Academy of Sciences report suggests that 28% of all major developmental disorders in children are linked entirely or partly to environmental exposures (*55*). Thus, the pressure to develop new, faster and cheaper approaches for developmental neurotoxicity assessments is high, especially as there is need to test large sets of compounds for specific regulatory requirements in Europe and the United States, which is impossible with current *in vivo* test methods.

The main problem in the development of a test strategy with predictive capacity for developmental neurotoxicity is the complexity of the human nervous system. Since chemicals induce toxic effects through various mechanisms, several cell models and endpoints have to be applied. A global effort is needed to identify how the information gained from different models and their respective endpoints can be used to build up an intelligent and reliable developmental neurotoxicity

testing strategy for regulatory purposes. During the past decade new emerging technologies have entered into the toxicology field that could speed up the testing and identification of critical developmental neurotoxicity information (*56*, *57*), either by allowing evaluation of many chemicals (high-throughput screening), or by compiling large amounts of complex information from one test approach for each substance (high-content screening).

Automated testing platforms by robots could allow screening of a high number of chemicals under standardized conditions with a simple cell model and end point (the ToxCast™ approach), which would be especially useful for initial screening. One of the most promising assays for developmental neurotoxicity high-throughput testing is the neurite outgrowth assay, which is based on high content imaging of the morphological changes of the neuronal network, such as the number of cells exhibiting neurites, or the length and branching of axons and dendrites (*58*, *59*). However, as the development of the nervous system involves several critical processes (e.g. cell proliferation, migration, differentiation, glial maturation and apoptosis) it is not likely that one simple test method can capture the complexity of these events. In this case, a pragmatic and realistic approach would be to select endpoints that can cover many different developmental neurotoxicity processes. A useful tool here is -omics technologies, which include genomics, transcriptomics, proteomics and metabolomics. The complete analysis of an organism's response to a perturbation on the genome, transcriptome, proteome and metabolome levels could lead to increased understanding of the mechanisms in complex systems.

The developmental neurotoxicity Tox 21 program, Identification of Pathways of Developmental Neurotoxicity for High Throughput Testing by Metabolomics, is a CAAT project funded by the FDA. It aims to find pathways of developmental neurotoxicity with a metabolomics approach. This project makes use of a rat primary three-dimensional organotypic neural cell model, which closely reproduces the *in vivo* situation of the central nervous system in terms of morphology and biochemical signaling (*60*). This unique model consists of all different types of cells in the central nervous system, such neurons, astrocytes, oligodendrocytes and microglia (*61*). Previous studies that used this neural cell model demonstrated neurotoxic and neuroprotective mechanisms of the central nervous system (*62–66*). Moreover, neurodevelopmental processes have been well characterized, making the model relevant for developmental neurotoxicity studies. The cell model is exposed to developmental neurotoxicity reference compounds (e.g. pesticides, drugs and environmental contaminants) during development, and low-molecular-weight metabolites relevant for neurodevelopment are quantified by mass spectrometry-based metabolomics. Pathway analysis software is used to assess changes in metabolite levels, which could indicate perturbed pathways (*36*). Further use of pathways of toxicity will advance understanding of developmental neurotoxicity and test capabilities for substances' potential risks to human health.

Other novel end points for developmental neurotoxicity include electrical activity measurements assessed by micro-electrode arrays (MEA), a new and promising tool (*67*, *68*). This technique provides a functional and neuronal-specific end point that had previously been used mainly in basic

research. Whole neuronal ensembles can be measured as functional networks, which yields more relevant physiological information than conventional methods for electrophysiology assessment (e.g. patch clamps). Many different *in vitro* systems have been used with this technique, including, primary cells, three-dimensional brain cell cultures and human stem cells (*65, 69, 70*). These cell models recapitulate many functions of neurons *in vivo*, including spontaneous activity (spiking and bursting), plasticity and responsiveness to a wide variety of neurotransmitters and pharmacological agonists and antagonists (*71, 72*).

MEA could serve as a sensitive tool to detect functional changes following exposure to chemicals during critical periods of *in vitro* development and, consequently, might be a useful tool for neurotoxicity and developmental neurotoxicity evaluations. Indeed, *in vitro* studies that have used MEA have established EC_{50} values in the same range as those published in other toxicological studies, and are generally in agreement with those obtained from *in vivo* experiments (*73, 74*). The modeling of learning and behavioral effects *in vitro*, which are considered crucial by regulators for developmental neurotoxicity risk assessment remain challenging. Non-mammalian species, such as zebrafish and Nematoda, have shown to be promising alternatives to these *in vivo* tests as the fundamental principles of key cellular events during brain developmental are remarkably conserved (*75*).

Lately it has become evident that more-complex three-dimensional cell models are necessary to reproduce the *in vivo* situation, especially when modeling the developing brain. The rat primary three-dimensional model described has been one of the most promising models for neurotoxicity and developmental neurotoxicity (*76, 77*), likely due to the increased cell–cell interactions that enhance cell survival and differentiation. To avoid interspecies differences, however, it is crucial to develop a similar human cell model. CAAT and collaborators from the Kennedy Krieger Institute and Johns Hopkins School of Public Health, Baltimore, MD, are funded by the National Center for Advancing Translational Sciences, part of the NIH, to establish and characterize such a model by use of induced pluripotent stem cells (*78*). The use of this type of stem cell allows investigation of gene–environment interactions as well as the potential of chemicals to interfere with epigenetic mechanisms. This project is part of the programmed research initiated by the NIH, FDA and Defense Advanced Research Projects Agency to develop 'human-on-chip' tools to assess the safety and efficacy of countermeasures to biological and chemical terrorism and warfare (*6*). The final goal of the 'human-on-chip' project is to integrate ten or more different human organs and connect them with fluids to mimic whole-body physiology.

Such a complex approach would be an important tool for toxicity testing, but would also be useful for research into in human physiology and pathology.

Conclusions

Regulatory toxicology is currently undergoing changes, which might actually become revolutionary (*28*). The increasing unease with the methodologies that have formed the basis of risk assessment for some decades coincides with

technological opportunities from the life sciences that demand change and make it possible. However, when safety is at stake and large global markets have to be regulated, transition is not easy.

References

1. Bottini, A. A.; Hartung, T. *ALTEX* **2009**, *26*, 3–16.
2. Bottini, A. A.; Hartung, T. *ALTEX* **2010**, *27*, 67–77 (Special Issue).
3. Hartung, T. *ALTEX* **2009**, *26*, 75–82.
4. Basketter, D. A.; Clewell, H.; Kimber, I.; Rossi, A.; Blaauboer, B.; Burrier, R.; Daneshian, M.; Eskes, C.; Goldberg, A.; Hasiwa, N.; Hoffmann, S.; Jaworska, J.; Knudsen, T. B.; Landsiedel, R.; Leist, M.; Locke, P.; Maxwell, G.; McKim, J.; McVey, E. A.; Ouedraogo, G.; Patlewicz, G.; Pelkonen, O.; Roggen, E.; Rovida, C.; Ruhdel, I.; Schwarz, M.; Schepky, A.; Schoeters, G.; Skinner, N.; Trentz, K.; Turner, M.; Vanparys, P.; Yager, J.; Zurlo, J.; Hartung, T. *ALTEX* **2012**, *29*, 3–91.
5. Leist, M.; Hasiwa, M.; Daneshian, M.; Hartung, T. *Toxicol. Res.* **2012**, *1*, 8.
6. Hartung, T.; Zurlo, J. *ALTEX* **2012**, *29*, 251–260.
7. Hartung, T.; Hoffmann, S. *ALTEX* **2009**, *26*, 155–166.
8. Hoffmann, S.; Hartung, T. *Hum. Exp. Toxicol.* **2006**, *25*, 497–513.
9. Stephens, M. L.; Andersen, M.; Becker, R. A.; Betts, K.; Boekelheide, K.; Carney, E.; Chapin, R.; Devlin, D.; Fitzpatrick, S.; Fowle, J. R.; Harlow, P., 3rd; Hartung, T.; Hoffmann, S.; Holsapple, M.; Jacobs, A.; Judson, R.; Naidenko, O.; Pastoor, T.; Patlewicz, G.; Rowan, A.; Scherer, R.; Shaikh, R.; Simon, T.; Wolf, D.; Zurlo, J. *ALTEX* **2013**, *30*, 74–103.
10. Daneshian, M.; Leist, M.; Hartung, T. *ALTEX* **2010**, *27*, 63–69.
11. Hartung, T. *J. Toxicol. Environ. Health, Part B* **2010**, *13*, 277–290.
12. Hartung, T. *Nature* **2009**, *460*, 208–212.
13. Bouhifd, M.; Hogberg, H. T.; Kleensang, A.; Maertens, A.; Zhao, L.; Hartung, T. *Basic Clin. Pharmacol. Toxicol.* **2014**, *115*, 24–31.
14. Hartung, T.; McBride, M. *ALTEX* **2011**, *28*, 83–93.
15. Hartung, T.; van Vliet, E.; Jaworska, J.; Bonilla, L.; Skinner, N.; Thomas, R. *ALTEX* **2012**, *29*, 119–128.
16. Hartung, T. *Toxicol. Sci.* **2009**, *109*, 18–23.
17. Hartung, T.; Daston, G. *Toxicol. Sci.* **2009**, *111*, 233–237.
18. Goldberg, A.; Hartung, T. *Eur. Pharmaceut. Contractor* **2008** (Winter), 42–46.
19. Daneshian, M.; Botana, L. M.; Dechraoui Bottein, M. Y.; Buckland, G.; Campas, M.; Dennison, N.; Dickey, R. W.; Diogene, J.; Fessard, V.; Hartung, T.; Humpage, A.; Leist, M.; Molgo, J.; Quilliam, M. A.; Rovida, C.; Suarez-Isla, B. A.; Tubaro, A.; Wagner, K.; Zoller, O.; Dietrich, D. *ALTEX* **2013**, *30*, 487–545.
20. United States Food and Drug Administration. Innovation or stagnation: challenge and opportunity on the critical path to new medical products.

http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm077262.htm (accessed June, 2013).

21. Hartung, T. *ALTEX* **2013**, *30*, 275–291.
22. Olson, H.; Betton, G.; Robinson, D.; Thomas, K.; Monro, A.; Kolaja, G.; Lilly, P.; Sanders, J.; Sipes, G.; Bracken, W.; Dorato, M.; Van Deun,, K.; Smith, P.; Berger, B.; Heller, A. *Regul. Toxicol. Pharmacol.* **2000**, *32*, 56–67.
23. Kavlock, R.; Dix, D. *J. Toxicol. Environ. Health, Part B* **2010**, *13*, 197–217.
24. Begley, C. G.; Ellis, L. M. *Nature* **2012**, *483*, 531–533.
25. Prinz, F.; Schlange, T.; Asadullah, K. *Nat. Rev. Drug Discovery* **2012**, *10*–712.
26. Bottini, A. A.; Amcoff, P.; Hartung, T. *ALTEX* **2007**, *24*, 255–261.
27. Hartung, T. *ALTEX* **2010**, *27*, 253–263.
28. Hartung, T. *ATLA, Altern. Lab. Anim.* **2008**, *36*, 635–639.
29. Hartung, T. *ALTEX* **2010**, *27*, 87–95.
30. Hartung, T.; Sabbioni, E. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2011**, *3*, 545–573.
31. Silbergeld, E. K.; Contreras, E. Q.; Hartung, T.; Hirsch, C.; Hogberg, H.; Jachak, A. C.; Jordan, W.; Landsiedel, R.; Morris, J.; Patri, A.; Pounds, J. G.; de Vizcaya Ruiz, A.; Shvedova, A.; Tanguay, R.; Tatarazako, N.; van Vliet, E.; Walker, N. J.; Wiesner, M.; Wilcox, N.; Zurlo, J. *ALTEX* **2011**, *28*, 236–241.
32. Food and Drug Administration, HHS. *Fed. Regist.* **2002**, *67*, 37988–37998.
33. National Research Council, NRC. *Animal Models for Assessing Countermeasures to Bioterrorism Agents*; National Academies Press: Washington, DC, 2011; pp 1–153.
34. Hartung, T.; Luechtefeld, T.; Maertens, A.; Kleensang, A. *ALTEX* **2013**, *30*, 3–18.
35. Davis, M.; Boekelheide, K.; Boverhof, D. R.; Eichenbaum, G.; Hartung, T.; Holsapple, M. P.; Jones, T. W.; Richard, A. M.; Watkins, P. B. *Ann. N. Y. Acad. Sci.* **2013**, *1278*, 11–24.
36. Kleensang, A.; Maertens, A.; Rosenberg, M.; Fitzpatrick, S.; Lamb, J.; Auerbach, S.; Brennan, R.; Crofton, K. M.; Gordon, B.; Fornace, A. J.; Gaido, K., Jr.; Gerhold, D.; Haw, R.; Henney, A.; Ma'ayan, A.; McBride, M.; Monti, S.; Ochs, M. F.; Pandey, A.; Sharan, R.; Stierum, R.; Tugendreich, S.; Willett, C.; Wittwehr, C.; Xia, J.; Patton, G. W.; Arvidson, K.; Bouhifd, M.; Hogberg, H. T.; Luechtefeld, T.; Smirnova, L.; Zhao, L.; Adeleye, Y.; Kanehisa, M.; Carmichael, P.; Andersen, M. E.; Hartung, T. *ALTEX* **2014**, *31*, 53–61.
37. Griesinger, C.; Hoffmann, S.; Kinsner, A.; Coecke, S.; Hartung, T. Preface. *Hum. Exp. Toxicol.* **2009**, *28*, 83–86.
38. Schneider, K.; Schwarz, M.; Burkholder, I.; Kopp-Schneider, A.; Edler, L.; Kinsner-Ovaskainen, A.; Hartung, T.; Hoffmann, S. *Toxicol. Lett.* **2009**, *189*, 138–144.
39. Hartung, T.; Hoffmann, S.; Stephens, M. *ALTEX* **2013**, *30*, 119–130.
40. Judson, R.; Kavlock, R.; Martin, M.; Reif, D.; Houck, K.; Knudsen, T.; Richard, A.; Tice, R. R.; Whelan, M.; Xia, M.; Huang, R.; Austin, C.;

Daston, G.; Hartung, T.; Fowle, J. R.; Wooge, W., 3rd; Tong, W.; Dix, D. *ALTEX* **2013**, *30*, 51–56.

41. Sipes, N. S.; Martin, M. T.; Kothiya, P.; Reif, D. M.; Judson, R. S.; Richard, A. M.; Houck, K. A.; Dix, D. J.; Kavlock, R. J.; Knudsen, T . B. *Chem. Res. Toxicol.* **2013**, *26*, 878–895.

42. Kleinstreuer, N. C.; Judson, R. S.; Reif, D. M.; Sipes, N. S.; Singh, A. V.; Chandler, K. J.; Dewoskin, R.; Dix, D. J.; Kavlock, R. J.; Knudsen, T. B. *Environ. Health Perspect.* **2011**, *119*, 1596–1603.

43. Knudsen, T. B.; Kleinstreuer, N. C. *Birth Defects Res., Part C* **2011**, *93*, 312–323.

44. Juberg, D. R.; Borghoff, S. J.; Becker, R. A.; Casey, W.; Hartung, T.; Holsapple, M. P.; Marty, M. S.; Mihaich, E. M.; Van Der Kraak, G.; Wade, M. G.; Willett, C. E.; Andersen, M. E.; Borgert, C. J.; Coady, K. K.; Dourson, M. L.; Fowle, J. R.; Gray, L. E.; Lamb, J. C.; Ortego, L. S.; Schug, T. T.; Toole, C. M.; Zorrilla, L. M.; Kroner, O. L.; Patterson, J.; Rinckel, L. A.; Jones, B. R. *ALTEX* **2014**, *31*, 63–78.

45. NTP Interacgency Center for the Evaluation of Alternative Methods. *Draft validation study report: MCF-7 cell proliferation test method*. http://ntp.niehs.nih.gov/iccvam/methods/endocrine/MCF7/MCF7-ValStudyReport-19Jun12-WCv2-draft.pdf (accessed June 2014)

46. Bouhifd, M.; Hartung, T.; Hogberg, H. T.; Kleensang, A.; Zhao, L. *J. Appl. Toxicol.* **2013**, *33*, 1365–1383.

47. Ramirez, T.; Daneshian, M.; Kamp, H.; Bois, F. Y.; Clench, M. R.; Coen, M.; Donley, B.; Fischer, S. M.; Ekman, D. R.; Fabian, E.; Guillou, C.; Heuer, J.; Hogberg, H. T.; Jungnickel, H.; Keun, H. C.; Krennrich, G.; Krupp, E.; Luch, A.; Noor, F.; Peter, E.; Riefke, B.; Seymour, M.; Skinner, N.; Smirnova, L.; Verheij, E.; Wagner, S.; Hartung, T.; van Ravenzwaay, B.; Leist, M. *ALTEX* **2013**, *30*, 209–225.

48. OECD. *Test no. 426: developmental neurotoxicity study*; OECD guidelines for the testing of chemicals, section 4; OECD Publishing: Paris, France, 2007.

49. United States Environmental Protection Agency. *Health effects guidelines OPPTS 870.6300 developmental neurotoxicity study [EPA 712–C–98–239]*. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0156-0042 (accessed June, 2013).

50. Bal-Price, A. K.; Sunol, C.; Weiss, D. G.; van Vliet, E.; Westerink, R. H.; Costa, L. G. *Neurotoxicology* **2008**, *29*, 520–531.

51. Grandjean, P.; Landrigan, P. J. *Lancet* **2006**, *368*, 2167–2178.

52. Kuehn, B. M. *JAMA* **2010**, *304*, 27–28.

53. Landrigan, P. J. *Curr. Opin. Pediatr.* **2010**, *22*, 219–225.

54. Sagiv, S. K.; Thurston, S. W.; Bellinger, D. C.; Tolbert, P. E.; Altshul, L. M.; Korrick, S. A. *Am. J. Epidemiol.* **2010**, *171*, 593–601.

55. National Academy of Sciences. *Scientific frontiers in developmental toxicology and risk assessment*; National Academy Press: Washington, DC, 2000.

56. MacGregor, J. T. *Toxicol. Sci.* **2003**, *75*, 236–248.

57. Smirnova, L.; Hogberg, H. T.; Leist, M.; Hartung, T. *ALTEX* **2014**, *31*, 129–156.
58. Krug, A. K.; Balmer, N. V.; Matt, F.; Schonenberger, F.; Merhof, D.; Leist, M. *Arch. Toxicol.* **2013**, *87*, 2215–2231.
59. Radio, N. M.; Freudenrich, T. M.; Robinette, B. L.; Crofton, K. M.; Mundy, W. R. *Neurotoxicol. Teratol.* **2010**, *32*, 25–35.
60. Trapp, B. D.; Honegger, P.; Richelson, E.; Webster, H. D. *Brain Res.* **1979**, *160*, 117–130.
61. Honegger, P.; Monnet-Tschudi, F. In *Protocols for Neural Cell Culture*, 1st ed.; Fedoroff, S., Ed.; Humana Press: Totowa, NJ, 2001; pp 199–218.
62. Eskes, C.; Juillerat-Jeanneret, L.; Leuba, G.; Honegger, P.; Monnet-Tschudi, F. *J. Neurosci. Res.* **2003**, *71*, 583–590.
63. Monnet-Tschudi, F.; Zurich, M. G.; Honegger, P. *Brain Res.* **1996**, *741*, 52–59.
64. van Vliet, E.; Morath, S.; Eskes, C.; Linge, J.; Rappsilber, J.; Honegger, P.; Hartung, T.; Coecke, S. *Neurotoxicology* **2008**, *29*, 1–12.
65. van Vliet, E.; Stoppini, L.; Balestrino, M.; Eskes, C.; Griesinger, C.; Sobanski, T.; Whelan, M.; Hartung, T.; Coecke, S. *Neurotoxicology* **2007**, *28*, 1136–1146.
66. Zurich, M. G.; Honegger, P.; Schilter, B.; Costa, L. G.; Monnet-Tschudi, F. *Neurotoxicology* **2000**, *21*, 599–605.
67. Hogberg, H. T.; Sobanski, T.; Novellino, A.; Whelan, M.; Weiss, D. G.; Bal-Price, A. K. *Neurotoxicology* **2011**, *32*, 158–168.
68. Johnstone, A. F.; Gross, G. W.; Weiss, D. G.; Schroeder, O. H.; Gramowski, A.; Shafer, T. J. *Neurotoxicology* **2010**, *31*, 331–350.
69. Makinen, M.; Joki, T.; Yla-Outinen, L.; Skottman, H.; Narkilahti, S.; Aanismaa, R. *J. Neurosci. Methods* **2013**, *215*, 88–96.
70. Yla-Outinen, L.; Heikkila, J.; Skottman, H.; Suuronen, R.; Aanismaa, R.; Narkilahti, S. *Front. Neuroeng.* **2010**, *3*, 00111; doi: 10.3389/fneng.2010.00111.
71. Keefer, E. W.; Gramowski, A.; Gross, G. W. *J. Neurophysiol.* **2001**, *86*, 3030–3042.
72. Streit, J. *J. Neurophysiol.* **1993**, *70*, 871–878.
73. Gross, G. W.; Harsch, A.; Rhoades, B. K.; Gopel, W. *Biosens. Bioelectron.* **1997**, *12*, 373–393.
74. Xia, Y.; Gopal, K. V.; Gross, G. W. *Brain Res.* **2003**, *973*, 151–160.
75. Coecke, S.; Goldberg, A. M.; Allen, S.; Buzanska, L.; Calamandrei, G.; Crofton, K.; Hareng, L.; Hartung, T.; Knaut, H.; Honegger, P.; Jacobs, M.; Lein, P.; Li, A.; Mundy, W.; Owen, D.; Schneider, S.; Silbergeld, E.; Reum, T.; Trnovec, T.; Monnet-Tschudi, F.; Bal-Price, A. *Environ. Health Perspect.* **2007**, *115*, 924–931.
76. Bal-Price, A. K.; Hogberg, H. T.; Buzanska, L.; Lenas, P.; van Vliet, E.; Hartung, T. *Neurotoxicology* **2010**, *31*, 545–554.
77. Forsby, A.; Bal-Price, A. K.; Camins, A.; Coecke, S.; Fabre, N.; Gustafsson, H.; Honegger, P.; Kinsner-Ovaskainen, A.; Pallas, M.; Rimbau, V.; Rodriguez-Farre, E.; Sunol, C.; Vericat, J. A.; Zurich, M. G. *Toxicol. In Vitro* **2009**, *23*, 1564–1569.

78. Hogberg, H. T.; Bressler, J.; Christian, K. M.; Harris, G.; Makri, G.; O'Driscoll, C.; Pamies, D.; Smirnova, L.; Wen, Z.; Hartung, T. *Stem Cell Res. Ther.* **2013**, *4* (suppl 1), S4.

Chapter 8

Cooperation between the United States Environmental Protection Agency and Industry To Develop an *in Vitro* Ocular Hazard Testing Strategy

Rodger D. Curren*

Institute for In Vitro Sciences, 30 W. Watkins Mill Road, Suite 100, Gaithersburg, Maryland 20878
***E-mail: rcurren@iivs.org**

In the United States, cleaning products that claim to have antimicrobial properties must be submitted to the Environmental Protection Agency (EPA) for registration. Traditionally, the EPA has required data from *in vivo* animal tests, but the agency and the industry are both keen to move away from these methods. Therefore, seven cleaning product manufacturers, the Accord Group consulting firm and the Institute for In Vitro Sciences conducted an extensive evaluation of the ability of three *in vitro* assays to predict the EPA hazard category of antimicrobial cleaning products. Subsequently the EPA launched an 18-month pilot study to further investigate whether data from an alternative non-animal *in vitro* testing strategy could appropriately determine level of risk for eye irritation from antimicrobial cleaning products. This chapter describes the three *in vitro* tests selected for the program and how they fit into the risk categorization requirements for registration.

Introduction

Most cleaning products in the United Sates do not have to go through a registration process before they are marketed. Companies who manufacture the product decide how to assure safety. In many cases, the safety assessments can be carried out without testing on animals. However, if a product is claimed to have

antimicrobial effects, then, by United States law, these products must go through a formal registration process with the Environmental Protection Agency (EPA) that requires hazard data. Traditionally, the hazard data are obtained through animal testing, but both the industry and the EPA would prefer to move away from all animal testing. Therefore, various parties agreed that an evaluation of the performance of alternative *in vitro* assays with antimicrobial cleaning products (AMCPs) should be undertaken. This search started with an ocular hazard testing strategy for eye irritation.

The rabbit Draize test has long been the standard animal test for eye irritation but has been heavily criticized by animal welfare advocates, owing to the severe pain and distress caused to the animals, which can last for several days. Additional criticism has arisen from within the scientific community, due to the variability of the test results caused by between-animal differences and the subjective nature of the scoring. The appropriateness of the rabbit model has also been criticized because of the differences from humans in eye structure (e.g. the presence of a nictitating membrane, thinner cornea etc.) and response (e.g. lack of significant tearing). Furthermore, the relevance of the extremely high doses employed has been questioned.

The above concerns led the EPA's Office of Pesticide Programs, a consortium of industry manufacturers (Clorox, Colgate-Palmolive, The Dial Corporation, EcoLabs, JohnsonDiversey Inc. [now SealedAir], S. C. Johnson & Son Inc. and The Procter and Gamble Company), a consulting firm (The Accord Group) and the Institute for In Vitro Sciences to launch an extensive evaluation of the ability of three *in vitro* assays to predict ocular hazard (and the EPA cautionary labeling category) for AMCPs. The assays investigated were: the bovine corneal opacity and permeability (BCOP), the Cytosensor Microphysiometer ([CM] Molecular Devices, Sunnyvale, CA, USA) and the EpiOcular™ ([EO] MatTek Corp, Ashland, MA, USA) three-dimensional tissue model. From the beginning no single test was envisioned to provide sufficient information for all labeling purposes; therefore, use of the three tests in a testing strategy was investigated.

In the development of any alternative testing strategy, the final goal needs to be clearly understood and the components by which to measure success decided. In order to replace animal testing, the following factors must be considered: the alternative method(s) should address known toxicity mechanisms, when possible; data collection and analysis should be as transparent as possible (i.e., testing should be done on coded materials and the relationship between the *in vitro* data and existing animal data should be clearly presented); multiple tests should be considered rather than a single test to ensure that the optimum results can be obtained; and the predictive ability of the animal test for human toxicity must be thoroughly understood (consideration of reproducibility, etc.) so that the quality of the replacement strategy can be correctly assessed. After the evaluation was completed, the consortium proposed an *in vitro* testing strategy based on the three assays. The strategy was positively reviewed by the EPA and this led to the establishment of a voluntary pilot project designed to evaluate the effectiveness of the proposed alternative testing strategy as a replacement for the Draize rabbit eye test in assessment of AMCPs. During the pilot program companies could submit registration claims based on testing with the *in vitro* methods, and their

acceptability was considered on a case-by-case basis. At the end of this volunteer program it was concluded that the *in vitro* tests selected could acceptably classify eye irritation hazards and enable appropriate labeling of products. In 2013 the program was made permanent, from which time data from the *in vitro* testing strategy could replace the *in vivo* requirement for AMCPs. For other classes of pesticides and pesticide products, including conventional, biochemical and other antimicrobial pesticides, data from alternative tests still need to be considered on a case-by-case basis. In this chapter I set out the details of testing with the BCOP, CM and EO assays.

Purpose of Testing

The EPA separates AMCPs (and other pesticides) into four toxicity categories for eye irritation. Category I indicates irreversible corrosive eye damage, category II severe but temporary eye injury, category III moderate eye irritation and category IV very mild effects. Each category has specific labeling requirements aimed to be Purpose of testing

The EPA separates AMCPs (and other pesticides) into four toxicity categories for eye irritation. Category I indicates irreversible corrosive eye damage, category II severe but temporary eye injury, category III moderate eye irritation and category IV very mild effects. Each category has specific labeling requirements intended to be meaningful to consumers. A so-called signal word that summarizes the degree of risk is assigned to each category (*1*). Categories I and II labels (signal words "danger" and "warning", respectively) indicate that the use of protective eyewear (e.g. goggles, a face shield or safety glasses) is compulsory A so-called signal word that summarizes the degree of risk is assigned to each category (*1*). Labels for categories I and II (signal words "danger" and "warning", respectively) indicate that the use of, contact with clothes should be avoided, hands should be washed thoroughly before eating, drinking, smoking or using the toilet and any contaminated clothes should be washed before re-use. For category III (signal word "caution"), the labeling indicates that protective eyewear may be used if appropriate but users should avoid contact with eyes if not. Contact with clothes should also be avoided and hands should be washed thoroughly before eating, drinking, smoking or using the toilet. Category IV substances require no precautionary statements risk, but category III labeling may be used if desired. To avoid confusion with the signal words when highlighting particular statements on labels, the use of "notice" or "attention" is recommended instead. The EPA provides instructions on the location and prominence of labeling that must be followed (*1*).

Understand the Mechanisms

Acceptance of data from *in vitro* assays is more easily obtained if the mechanisms underlying the outcomes of exposure to products and how they have arisen are known. The closer the relationship between the *in vitro* model and the *in vivo* target tissue, the less knowledge of the mechanisms is generally required

for acceptance. Thus the use of *ex vivo* ocular tissues or complete *in vitro* corneal equivalent systems (composed of corneal epithelial, stromal keratinocyte and corneal endothelial cell layers) (*2*) certainly requires less validation than a more 'black box' system.

For an eye irritation validation study to be successful, the critical cellular and molecular changes involved in initial ocular injury, subsequent responses and repair processes should be identified and characterized. These can then be mapped to the activity domain of an *in vitro* model. Fortunately for eye irritation there are several common modes of chemical action. Membrane lysis is associated with surfactants and organic solvents. Protein coagulation or denaturation is associated with acids and organic solvents. Saponification is associated with alkalis, and oxidative damage to macromolecules is associated with reactive chemicals, such as bleaches and peroxides.

Irrespective of mechanism, the extent of the initial injury is suspected to be predictive of the degree and duration of injury and the overall outcome (*2*). Mechanistically-based alternative methods to *in vivo* ocular irritation tests, therefore, need to incorporate microscopic or biochemical measurement of the initial injury.

Assays

The EPA pilot program approved the BCOP, CM and EO assays for eye irritation testing. To aid the decision about which assay or assays to use for which substances, a decision tree was designed (Figure 1). The first step is to evaluate available information on the active ingredients and formulation and existing Draize results or *in vitro* data on related compounds. If the formulation is based on oxidizing (reactive) chemistries or the components fall in a class of chemicals that are severe irritants, it should be tested in the BCOP assay. If no oxidizing chemistries are included, the decision may be based on the type or concentration of formulation ingredients, past registration of similar products and in-use information from similar, non-antimicrobial products about the expected ocular hazard category.

The BCOP assay is suitable for testing products that are expected to have category I or II effects and provides direct evidence of corneal damage. After exposure, measurement of corneal opacity with an opacitometer and permeability with sodium fluorescein and spectrophotometry identify epithelial and stromal changes associated with common modes of eye irritation (membrane lysis, protein coagulation and saponification). The irritation score is calculated as the mean opacity value + (15 × mean permeability OD_{490} value). A cutoff of 75 is used for category I effects (severe irritation or corrosion). If the score is less than 75, the current EPA guidelines classify the material as a category II. Although histological evaluation can be performed on the three major layers of the cornea to provide a direct measure of the depth of injury (*3*) and evidence of a lesser hazard, current EPA guidelines require that the CM or EO assay be used as final evidence for a category III or IV determination.
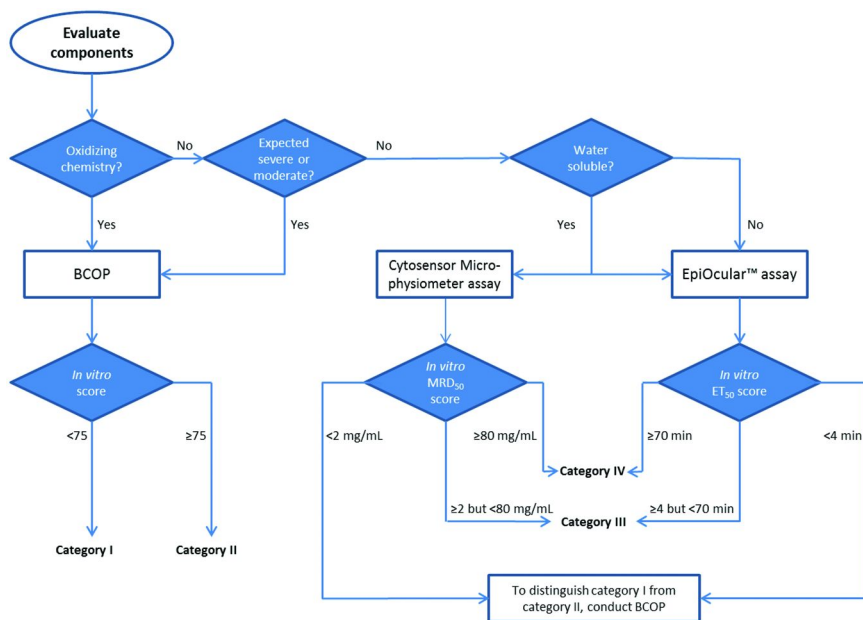
Figure 1. Proposed process for classification of ocular irritation with alternative in vitro tests. MRD_{50}, dose that decreases metabolic rate of cells by 50%; ET_{50}, time of exposure that reduced MTT conversion by 50%. Reproduced from reference (7).

The CM assay, which uses a cell monolayer, can only be used with ingredients or formulations that form fairly complete solutions or slight suspensions. It has primarily been used to assess surfactants and surfactant-containing formulations. The water solubility of a product should be established before testing. Although the assay can detect a wide range of ocular toxicity, including category IV effects, it is not capable of differentiating category II from category I materials. After exposure to the test material, the microphysiometer detects rate of change in extracellular pH, caused by variations in metabolic rate. An increase in release of acidic byproducts reflects alterations (generally decreases) in cell metabolism after exposure to a toxic substance. The endpoint of this assay is the dose of the test material that induces a 50% decrease in metabolic rate relative to a negative control.

The EO assay predicts potential ocular damage after exposure of non-keratinized, three-dimensional, reconstructed cornea-like tissue. The EO tissue construct is grown from normal human-derived epidermal keratinocytes. It can be used to assess very mild (category IV) to moderate (category II) toxic effects and, therefore, is suitable for testing a wide range of products, such as cosmetics, personal hygiene and household products, and some industrial chemicals. It is not suitable for differentiating between category I and category II materials, as most of the tissue is destroyed by category II substances, which means that greater toxicity cannot be measured. In the assay a dose of the ingredient or formulation

being tested is placed on the EO tissue. After exposure the rinsed tissue sample is transferred to MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) solution, which is taken up and reduced in the mitochondria of the living tissue cells. Living cells show a dark purple formazan precipitate, whereas dead cells show no change in color. Thus, the more toxic the material, the less purple is seen in the tissue sample. Assessment of the extracted formazan dye by spectrophotometry reveals the percentage change in cell viability relative to controls. It is important before testing that the substance of interest is screened for direct MTT reduction, otherwise false-negative results may occur.

Generally, one of the three tests described above should be enough to determine the category of a substance, but in some cases the use of a second, more robust or more sensitive test might be appropriate to 'fine tune' the results. For example, if the BCOP was used because a product expected to be category II, but histopathology indicates that it might be a category III, use of the CM or EO assay could help to clarify whether it is actually category III or IV (Figure 2).
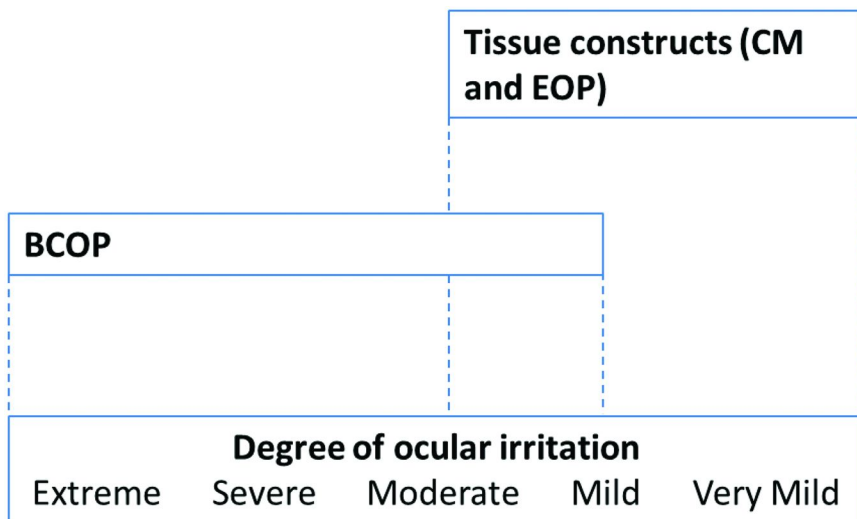
Figure 2. A conceptualized view of effects that can be assessed by the United States Environmental Agency in vitro testing strategy for ocular irritation.

Consideration needs to be given to the active ingredients in AMCPs, as this can affect the suggested testing strategy. Oxidizing materials that contain specific reactive chemicals, such as hypochlorite, peroxide, percarbonate or oxygen bleaches, should be tested with the BCOP assay because the high reactivity means they are likely to be over-predicted by the CM or EO assays. By contrast, high solvent formulations (>5% organic solvent, such as alcohols, glycol or ethers) are likely to be over-predicted by the BCOP assay if the standard exposure time of 10 min is used. Instead, exposure should be limited to 3 min. Finally, the BCOP or EO assays should be used for heavy suspensions or solid materials, since the small tubing size used with the CM assay makes that assay unsuitable.

Data Analysis

The *in vitro* data generated to validate the above assays was analyzed in the context of pre-existing animal data and other published reference values to confirm the relevance and accuracy of the findings. It should not be necessary to perform new animal tests to obtain comparative data, as many robust data in numerous dosing and exposure scenarios are freely available. In addition, the use of prediction models is vital to ensuring that results can be confirmed and validated. Such models convert the results obtained from an alternative method and predict the toxic effects that will be seen *in vivo* (*4*). Prediction models for alternative tests should comprise four major elements: definition of the purpose of the test in terms of end points; definition of all possible results that may be obtained with the method (inputs, e.g. quantitative, censored, qualitative and non-qualified data); an algorithm that translates results into data predictive of *in vivo* effects; and an indication of the accuracy and precision of outputs. The creation of sound prediction models requires knowledge of the conventional *in vivo* test and the alternative *in vitro* method. Relevance of the prediction model should be assessed before validation studies are started. This will ensure that the optimum end points to reflect a valid assay are available and will enable objective comparison of results and help to provide definitive answers on performance of the method. With a prediction model in place that can be interrogated, the validation study can be based on a strong hypothesis (*4*). This process was followed in validating the three assays for the purpose of predicting the ocular hazard category of AMCPs.

Statistical analysis is imperative to make sense of any results. A European study into the validation of alternative methods to the Draize eye irritation test indicated that carefully selected statistical procedures were highly influential in the validation effects for *in vitro* tests (*5*). Recommendations for the application of biostatistical methods during the development and validation of alternative toxicological methods are available from European Union Reference Laboratory for alternatives to animal testing (*6*). Guidance for reporting to meet regulatory requirements is also available (*7*).

Selective use of the assays, owing to their different properties as well as the overlap of their capabilities, means that results are compared not only with controls but also in relation to other factors, such as depth of injury. Thus data should be represented graphically (plotting *in vivo* results versus *in vitro* data) whenever possible to appreciate different distributions of data.

Conclusions

No single *in vitro* assay has yet been proven sufficient for testing substances in all eye irritation categories. Thus, a bottom-up (non-irritant)/top down (severe irritant) strategy is proposed (Figure 1). This approach is conservative and might lead to some category IV materials might be over-predicted as category III (*7*), but this situation was accepted by industry in this case to meet the requirements of the EPA. The testing strategy presented here may eventually be shown to be an improvement on the standard Draize test, which, unlike the *in vitro* assays, is not

strongly predictive of its own results (*4*). To ensure that the tests used match up to the animal gold standard tests – or, more importantly, human studies – and meet regulatory requirements, clear understanding of the mechanisms, capabilities of different models and methods and extensive knowledge of existing toxicological information should be incorporated into every assessment, just as should be done for conventional testing. Intelligent combination and analysis of results should yield well supported results from *in vitro* ocular methods.

References

1. EPA. *Label Review Manual*; United States Environmental Protection Agency: Washington, DC, 2012.
2. Maurer, J. K.; Parker, R. D.; Jester, J. V. *Regul. Toxicol. Pharmacol.* **2002**, *36*, 106–117.
3. Harbell, J. W.; Mun, G.; Curren, R. D. Application of histological evaluation to enhance the bovine opacity and permeability (BCOP) assay. Presented at the 45th Annual Society of Toxicology Meeting, San Diego, CA, March 5–9, 2006.
4. Bruner, L.; Carr, G. J.; Curren, R. D.; Chamberlain, M. *Environ. Health Perspect.* **1998**, *106* (suppl. 2), 477–484.
5. Balls, M.; Botham, P. A.; Bruner, L. H.; Spielmann, H. *Toxicol. In Vitro* **1995**, *9*, 871–929.
6. Holzhutter, H.-G.; Archer, G.; Dami, N.; Lovell, D. P.; Saltelli, A.; Sjostrom, M. *ATLA, Altern. Lab. Anim.* **1996**, *24*, 511–530.
7. Office of Pesticide Programs, U.S. Environmental Protection Agency. Use of an alternate testing framework for classification of eye irritation potential of EPA pesticide products. http://www.epa.gov/pesticides/regulating/eye-policy.pdf (accessed July, 2014).

Chapter 9

Ensuring That Nutrition and Health Claims in the European Union on Foods and Food (Dietary) Supplements Are Justified and Scientifically Substantiated

David P. Richardson*

School of Chemistry, Food and Pharmacy, University of Reading, U.K.
DPR Nutrition Ltd., 34 Grimwade Avenue, Croydon, Surrey CR0 5DG, U.K.
***E-mail: info@drpnutrition.com**

Consumers should be able to make informed choices based on clear and accurate information on the labelling and in the advertising of foods and food (dietary) supplements, and to have confidence in the scientific and regulatory processes used to support nutrition and health claims. The purpose of global and European regulatory developments in the scientific substantiation of nutrition and health claims is to achieve a high degree of consumer protection, to promote fair trade, to stimulate academic research and to encourage innovation in the food industry. The application of chemistry in agriculture, nutrition, food science and food technology has contributed not only to a safe and abundant food supply but also to the growing consumer awareness of the roles that food and food constituents have in maintaining and improving health and reducing the risk of major chronic diseases. This chapter examines the scientific substantiation of health claims, the choice of biomarkers or risk factors used to reflect beneficial physiological effects in humans, the processes for authorization of a health claim and consumer understanding of nutrition and health claims.

Introduction

Diet has beneficial health effects that extend beyond traditionally accepted nutritional effects. The approaches involved in elucidating these beneficial physiological effects are becoming increasingly important, as reflected by the growing research being undertaken and by the development of innovative food and food supplement products.

With consumers' awareness of health benefits of foods and food constituents broadening, the key questions for regulators, food and nutrition scientists, biochemists, analytical chemists and the food industry relate to (a) how consumers can have confidence in the nutrition and health claims on food labels; (b) how the foods and food constituents can be sufficiently characterized; (c) how to develop and validate biomarkers of physiological responses; and (d) what effects are considered to be beneficial to the health of the general public and specific target population subgroups.

The aims of regulatory frameworks in different regions of the world are to ensure that claims are scientifically substantiated, to promote and protect innovation, to improve free movement of goods and ensure fair competition and to achieve a high degree of consumer protection from false and misleading claims.

The application of chemistry involved in nutrition, food science and food technology is substantial, particularly in the areas of food analysis, quality, safety and shelf life and in the implementation of many multidisciplinary applications, such as biotechnology, genomics, microbiology, physical chemistry, engineering, sensory science and toxicology. Science and evidence-based approaches are used around the world to underpin regulatory developments in nutrition and health claims (*1*). The development of functional foods and ingredients is helping to reinvigorate efforts to process and preserve raw materials from agriculture, horticulture, fisheries and aquaculture into a diverse range of foods and dietary supplements.

Health Claim Definitions and Regulatory Frameworks

Definitions and frameworks for nutrition and health claims differ slightly around the world (e.g. the European Commission (*2*), United States Food and Drug Administration (*3*), Codex Alimentarius (*4*)). A nutrition claim typically refers to any representation that states, suggests or implies that a food has particular nutritional properties, including but not limited to energy value, content of protein, fat, carbohydrates, vitamins and minerals. For example, a nutrient content claim can describe the level of a nutrient contained in the food (e.g. "high in fiber", "low in fat", "source of protein", "high in vitamin C") and the criteria for such claims are defined in legislation, such as that in the European Regulation (EC) No 1926/2006 (*2*). Comparative claims compare the nutrient levels and/or energy values of two or more foods (e.g. "reduced sugar", "increased fiber", "lite/light").

In the USA, structure and function claims describe the role of a nutrient or dietary ingredient intended to affect a structure or function in humans. In addition, such claims characterize the means by which a nutrient or dietary ingredient acts

to maintain such a structure or function, or they may describe general well-being from consumption of a nutrient or dietary ingredient.

In the Codex Alimentarius, a health claim refers to any claim that states, suggests or implies that a relationship exists between a food category, a food or one of its constituents and health (*4*). Health claims include:

- Nutrient function claims, which describe the physiological role of a nutrient in growth, development and normal functions of the body
- Other function claims, which describe specific beneficial effects of the consumption of foods and their constituents in the context of the total diet on physiological function or biological activities but do not include nutrient function claims; such claims relate to a positive contribution to health or to the improvement of a function or to the modification or preservation of health
- A reduction of disease risk claim is defined as any health claim that states, suggests or implies that the consumption of a food category, a food or one of its constituents significantly alters a risk factor in the development of a human disease.

The work of the Codex Alimentarius Commission, together with that of the Food and Agriculture Organization of the United Nations and the World Health Organization in their supportive roles, provides a collection of standards, codes or practice guidelines and other recommendations (*5*). The organization has become a global reference point for consumers, food producers and processors, national food control agencies and the international food trade. The main aims are to help nations to join the international community in formulating and harmonizing food standards and ensuring their global implementation. Codex standards have also become the benchmarks against which national food measures and regulations are evaluated within the legal parameters of the World Trade Organization Agreements. From the very beginning in 1963, Codex Alimentarius has been a science-based activity with a focus on food-related scientific research and investigation, particularly in the areas of food chemistry, food technology, undesirable substances and contaminants, hygiene and nutrition. The Codex Committee on Nutrition and Foods for Special Dietary Uses and the Codex Committee on Methods of Analysis and Sampling have played major roles in the development of food and policy framework papers in the field of nutrition, for example, the preparation of Nutrient Reference Values for Labelling Purposes (*6*), the General Principles for the Addition of Essential Nutrients to Foods (*7*), Nutritional Risk Analysis Principles and Guidelines (*8*) and the Codex Guidelines for the Use of Nutrition and Health Claims: Recommendations on the Scientific Substantiation of Health Claims (*4*).

Process of Scientific Substantiation of Health Claims

The Codex guidelines on nutrition and health claims (*4*) focus on the criteria for substantiating a health claim and general principles for the systematic review

of the scientific evidence. Such a process typically addresses the following key areas:

- Identification of the proposed relationship between the food or food constituent and the health effect
- Identification of appropriate, validated measurements for all the food or food constituents and for the health effect
- Identification and categorization of all the relevant scientific data
- Assessment of the quality and interpretation of each relevant study
- Evaluation of the totality of the available relevant scientific data, weighing the evidence across studies and determination of whether and under what circumstances a claimed effect is substantiated

The Codex guidelines state that the totality of the available scientific evidence, including unpublished data where appropriate, should be identified and reviewed. Such evidence includes that to support the claimed effect, evidence that contradicts the claimed effect and evidence that is ambiguous or unclear. Health claims should primarily be based on evidence from well-designed human intervention studies. Generally, human observational studies are not sufficient *per se* to substantiate a health claim, but, where relevant, they may contribute to the totality of evidence. Animal model studies, *ex vivo* and *in vivo* data may be provided as the supporting knowledge base to explain the mechanism(s) of action of a particular relationship between a food or a food constituent and a beneficial health effect, but are insufficient to substantiate any type of health claim.

Overall, the evidence from human studies should demonstrate a consistent association between the food or food constituent and the health effect, with little or no evidence to the contrary. The scientific and analytical data must provide adequate characterization of the food or food constituent considered to be responsible for the beneficial health effect and, where applicable, include a summary of the studies undertaken on the conditions of production, batch-to-batch variability, analytical procedures and results, and state the conclusions of stability studies and studies of storage conditions and shelf-life.

These Codex guidelines were much influenced by the European consensus papers, Scientific Concepts of Functional Foods in Europe (*9*) and the project Process for the Assessment of Scientific Support for Claims on Foods (PASSCLAIM) (*10, 11*).

Choice of Risk Factors or Biomarkers

In the case of a claimed health effect that cannot be measured directly, relevant validated biomarkers may be used. Biomarkers must be both analytically and biologically valid and should reflect a future health outcome at a stage when dietary intervention will be effective (*12*).

A marker or set of markers for a function is a measurable indicator of the state of a particular bodily function and, thus, helps to determine the effect of a food or food constituent on that function and the state of health of an individual.

According to PASSCLAIM (*10*), all markers, whether they are biochemical, physiological or behavioral in nature, should be feasible, valid, reproducible, sensitive and specific. Markers must be rigorously validated and amenable to standard quality control procedures as well as being measurable in easily accessible material, or obtainable by means of methodology that must be both ethical and minimally invasive. In many cases, a battery of markers will be needed for use in a new generation of human intervention studies that will generate readily interpretable, valid and reliable data that can form the basis of future development of functional foods. Currently, the number of validated biomarkers is low (*13*). From the wealth of publications in the area of diet-related cardiovascular disease, low density lipoprotein (LDL) cholesterol and total cholesterol concentrations in serum and blood pressure are well-established markers for risk of cardiovascular disease. Plasma homocysteine, another established marker, is sensitive to dietary factors and is validated methodologically, but it is still not clear to what extent changes in this marker reflect enhanced function and reduction of risk of disease. For hemostatic function and oxidative damage, markers that are sensitive to dietary change need to be developed and validated. In the area of bone health and osteoporosis, bone-mineral density is a measure of the calcium content in bones, and for people older than 50 years with a high risk of fracture, this feature is considered to be a good marker. Thus, changes in bone-mineral density caused by a food or food constituent could provide evidence of a reduction in risks, such as fracture risk. Other examples of surrogate end points of disease risk include glycated hemoglobin as an indicator of long-term hyperglycemia and risk of complications in people with type 2 diabetes and the presence of adenomatous colon polyps is an early indicator of colon cancer (*14*).

Methodological aspects include analytical variability, and efforts have to be made to standardize assessments of all outcome measures and to reduce measurement error as far as possible, for instance by the use of standardized measurement protocols and operating procedures (*14*). Biological variability, such as genetic variation, circadian or seasonal variations, may also introduce systematic bias into results. Hence, it is important to understand the factors underlying this variability and to take samples or adapt the study design appropriately. The thresholds of relevance for nature and size of biological changes or differences in studies should be defined before human studies are initiated to ensure studies are designed with sufficient statistical power to be able to detect effects of such a size if they truly occurred (*15*). A study may show a statistically significant change in a validated biomarker, but the biological, clinical or public health significance must also be considered (*14*).

Process for Authorization of a Health Claim in the European Union

The regulation on nutrition and health claims made on foods (*2*) applies to those made in all commercial communications, whether labelling, presentation or advertising, across all member states of the European Union. The legislation sets out conditions for their use, establishes a system of scientific evaluation

and creates European Union lists of authorized and rejected claims. All claims have to comply with the general principles that they are not false, ambiguous or misleading, and they have to be scientifically substantiated. Health claims based on generally accepted scientific evidence (mostly well-established nutrient function claims) fall under Article 13 (1) of the regulation, whereas those based on newly developed scientific evidence and/or where those claims include a request from the applicant for the protection of proprietary data fall under Article 13 (5). Claims of disease risk reduction and for children's development and health fall under Article 14. Applications for Article 13 (5) and Article 14 claims must follow the procedures set out in the regulation (2) and in the implementing rules (16). Requirements include the submission of a comprehensive dossier of the scientific evidence, name and characterization of the food or food constituent, a proposal for the wording of the claim and specific conditions of use. The characterization of the food or food constituent is of paramount importance, as failure to supply this information is one of the prime reasons for rejection of claims. The source and specification (e.g. physical and chemical properties, composition and, where applicable, microbiological constituents) of the food or food constituent must be provided, along with data on batch variability, analytical methods and details of good laboratory practice, stability information and, if relevant, bioavailability data. The latter should include information on and a rationale for why the constituent being submitted is in the form it is and how it is available to be used by the human body (e.g. absorption studies). The European Food Safety Authority (EFSA) has published scientific and technical guidance for the preparation and presentation of the application for authorization of a health claim (17).

Role of the EFSA in the Scientific Substantiation of Health Claims

The scientific opinions of the EFSA Panel on Dietetic Products, Nutrition and Allergies (EFSA NDA) on the substantiation of health claims are used as the basis for authorization decisions by the European Commission and member states (with scrutiny by the European Parliament). The outcomes are published in the EU Register of Nutrition and Health Claims. In 2011 and 2012, EFSA published a series of guidance documents on the scientific requirements for substantiation of health claims related to:

- Gut and immune function (*18*)
- Antioxidants, oxidative damage and cardiovascular health (*19*)
- Appetite ratings, weight management and blood glucose concentration (*20*)
- Bone, joints, skin and oral health (*21*)
- Nervous system, including psychological function (*22*)
- Physical performance (*23*)

These EFSA guidance documents define a range of claimed physiological effects that are considered to be beneficial, and address the types of human studies, outcome measures and study groups deemed appropriate for their scientific substantiation. EFSA requires conclusive evidence of cause and effect. Most of the successful outcomes of the evaluations have been on extremely well-characterized foods and pure food constituents, for which the physiological effects can be demonstrated by the use of human intervention studies with validated biomarkers. In many cases, this pharmaceutical approach is very difficult to achieve based on state-of-the-art nutrition science, and it poses major challenges to the undertaking of future research that would satisfy EFSA requirements (*13*, *24–26*). Table I highlights the main scientific reasons for rejection of health claim applications to date. The unfavorable responses are, in most cases, totally justified, and the EFSA approach highlights the fact that it is only when a cause and effect relationship has been established between the consumption of the food or food constituent and a health benefit that the outcome is positive. If the evidence is considered to be emerging and/or conflicting and EFSA regards the evidence as not conclusive or too limited, the outcomes are negative (*13*).

Table I. Scientific reasons for failure to achieve EFSA positive opinions and authorization of health claims in the European Union[a]

Reasons for rejection of application
• Foods or food constituents not sufficiently characterized
• Effects of food matrix, processing and stability information, bioavailability and content variability not sufficiently characterized
• A cause and effect relationship was not established between the food or food constituent and the claimed effect
• Lack of systematic literature review and no specific inclusion and/or exclusion criteria
• Criticism of study design, absence of power calculations, insufficient information on background diet and lifestyle, failure to describe target group, intervention trials lacking, no lowered risk factor or measurable effect
• Clinical studies not used as evidence for health effects in the general population

[a] EFSA, European Food Safety Authority.

Physiological Effects Considered To Be Beneficial by the EFSA

According to Regulation (EC) No 1924/2006 (*2*), the use of health claims shall be permitted only if the food or food constituent for which the claim is made has been shown to have a beneficial physiological effect. The EFSA guidance documents state the physiological effects relating to the nervous system, psychological, perceptual (i.e. related to sensory processes), psychomotor and physiological regulatory effects (*18–23*). For function claims, a beneficial effect may relate to maintenance or improvement of a function. For claims of

reduced disease risk, beneficial refers to whether the claimed effect relates to the reduction (or beneficial alteration) of a human disease risk factor. Whether or not the alteration of a factor is considered by EFSA to be beneficial in this context depends on the extent to which it is established that (a) the factor is an independent predictor of disease risk (such a predictor may be established from intervention and/or observational studies), and (b) the relationship of the factor to the development of the disease is biologically plausible.

Each health claim or claim of disease risk reduction is considered by the EFSA NDA on a case-by-case basis, and the population group for which health claims are intended can be the general, healthy population or specific subgroups, such as elderly people, physically active people, women of childbearing age etc. Reference to general non-specific benefits of the nutrient or food for overall good health or health-related well-being may only be made if accompanied by a specific health claim (*18–23*). Table II summarizes the beneficial effects related to antioxidants, oxidative damage and cardiovascular health (*19*).

The EFSA guidance documents and the various scientific opinions published to date provide essential reading for researchers and health claim applicants to identify the key areas of normal metabolism that are considered beneficial physiological effects and appropriate study designs, outcome measures and/or validated biomarkers. Further guidance on the design, conduct and reporting of human intervention studies to evaluate the health benefits of foods, and on the process of scientific assessments can be found in the Scientific Concepts of Functional Foods in Europe (*9*), PASSCLAIM (*10*) and in the papers "A standardized approach towards PROving the efficacy of foods and food constituents for health CLAIMS (PROCLAIM): providing guidance" (*12*) and "Guidelines for the design, conduct and reporting of human intervention studies to evaluate the health benefits of foods" (*14*).

Examples of Authorized Claims in the European Union

Functional Benefits of the Essential Vitamins and Minerals

The well-established functions of essential nutrients are widely documented in the scientific literature. Under Article 13 (1) of Regulation (EC) No 1924/2006 (*2*), health claims can describe or refer to the role of a nutrient or other substance in growth, development and the functions of the body or the impact on psychological and behavioral functions. Details of the authorized health claims, including the nutrient function claims, together with the approved conditions of use, can be found on the European Union register of nutrition and health claims, which is available at http://ec.europa.eu/nuhclaims/.

Key areas of interest for the food and food supplements industries relate to:

- Protection of cell constituents from oxidative damage (e.g. vitamin C, vitamin E, copper, manganese, selenium and zinc)
- Contribution to a normal function of the immune system (e.g. vitamin A [including β-carotene], vitamin D, vitamin B_6, vitamin B_{12}, vitamin C, folic acid [folate], iron, copper, selenium and zinc)

- Contribution to the maintenance of normal bones and teeth (e.g. calcium, vitamin D and phosphorus)
- Contribution to normal energy-yielding metabolism (e.g. thiamin, riboflavin, niacin, vitamin B_{12}, biotin, pantothenic acid, vitamin C, copper, iron and magnesium)
- Contribution to the reduction of tiredness and fatigue (e.g. niacin, vitamin B_6, vitamin B_{12}, pantothenic acid, vitamin C and magnesium)
- Contribution to the normal function of the nervous system; contribution to normal psychological functions (e.g. biotin, vitamin B_6, vitamin C, niacin, vitamin B_{12}, iron and calcium)

Table II. Examples of physiological effects related to antioxidants, oxidative damage and cardiovascular health deemed beneficial by the EFSA[a] (19)

Physiological benefits
• The protection of body cells and molecules such as DNA, proteins and lipids from oxidative damage, including photo-oxidative (UV-induced) damage, may be a beneficial physiological effect.
• Maintenance of normal LDL-cholesterol concentration is a beneficial physiological effect.
• Reduction in LDL-cholesterol concentration within the normal range is considered a beneficial physiological effect in the context of a reduction of disease risk claim for CHD.
• Maintenance of normal HDL-cholesterol concentration is a beneficial physiological effect as long as LDL-cholesterol concentration is not increased.
• Maintenance of normal blood concentration of triglycerides may be a beneficial physiological effect.
• Maintenance of normal blood pressure is a beneficial physiological effect. Reduction in (systolic) blood pressure is considered beneficial in the context of a reduction of disease risk claim for CHD and stroke. An improvement of specific endothelial functions, e.g. endothelium-dependent vasodilation during sustained exposure (e.g. 4 weeks) to the food/constituent may be considered a beneficial physiological effect.
• Decreasing platelet aggregation in subjects with platelet activation during sustained exposure (e.g. 4 weeks) to the food/constituent would be a beneficial physiological effect.
• Maintenance of normal homocysteine metabolism is a beneficial physiological effect.

[a] UV, ultraviolet; LDL, low-density lipoprotein; CHD, coronary heart disease; HDL, high-density lipoprotein.

Authorization of Cardiovascular Health Claims

Oat β-Glucan and Reductions in Blood Cholesterol Levels and Risk of Coronary Heart Disease

The EFSA NDA concluded that oat β-glucan is sufficiently characterized, and that a cause and effect relationship has been established between the consumption of this food constituent and lowering of LDL cholesterol concentrations in blood, pursuant to Article 14 of Regulation (EC) No 1924/2006 (*2*, *27*). The EFSA scientific assessment described β-glucan as a non-digestible, non-starch polysaccharide composed of glucose molecules in long linear polymers with mixed β(1→4) and β(1→3) links whose approximate relative distribution is 70–30%. The molecular weight of oat β-glucan in various commercially available processed food preparations is generally less than the 2000 kDa reported for the source of oats (range about 100 kDa to 2000 kDa). The mixed linkages are important for the physical properties, such as solubility and viscosity. The viscosity is a function of the concentration of dissolved β-glucans and its molecular weight, and further depends on differences in raw materials, processing and methods of determination. Oat β-glucan occurs naturally in the bran of oats and is measurable in foods by established methods.

The substantiation of the claim for reduction in coronary heart disease risk reduction was based on 22 references, which included three meta-analyses and 19 randomized controlled trials pertinent to the health claim. The evidence presented indicates that the cholesterol-lowering effect of oat β-glucan may depend on increased viscosity in the small intestine, which reduces the reabsorption of bile acids, increases the synthesis of bile acids from cholesterol and reduces circulating LDL cholesterol concentrations.

In considering the totality of the available scientific data and weighing the evidence, the EFSA NDA concluded that a cause and effect relationship had been established and determined that, in order to bear the claim, foods should provide at least 3 g of oat β-glucan per day as part of a balanced diet (*27*). The target population was adults who want to lower blood cholesterol concentrations. Following this positive scientific opinion from the EFSA NDA, the health claim was authorized in Commission Regulation EU No 1160/2011 (*28*). The authorized claim is "Oat beta-glucan has been shown to lower/reduce blood cholesterol. High cholesterol is a risk factor in the development of coronary heart disease". The conditions of use of the claim state, "Information shall be given to the consumer that the beneficial effect is obtained with a daily intake of 3 g oat beta-glucan" and "The claim can be used for foods which provide at least 1 g oat beta-glucan per quantified portion".

Effect of Water-Soluble Tomato Concentrate on Platelet Aggregation

Water-soluble tomato concentrate (WSTC) is a lycopene-free and fat-free substance developed in two variants: WSTC I, which is completely water-soluble syrup, and its low-sugar derivative WSTC II, which is supplied in powder format. The two products are prepared from tomato using patented processes, the manufacturing process is clearly described, the chemical specification

of the constituents are provided and batch-to-batch reproducibility has been demonstrated. The WSTCs are standardized on the total quantity of 37 "bioactive" constituents identified and quantified with reversed-phase high performance liquid chromatography mass spectrometry. On the basis of the potentially bioactive compounds, 3 g WSTC I is considered equivalent to 150 mg WSTC II, and correspond approximately to the water-soluble content of 2.5 tomatoes. The EFSA NDA considered that WSTCs I and II were sufficiently characterized for the health claim to be made (29, 30).

This application was the first successful health claim pursuant to Article 13 (5) of Regulation (EC) No 1924/2006, was based on newly developed scientific evidence and included a request for the protection of proprietary data. The scientific substantiation is based on eight human studies, seven of which were human intervention studies on the effects of WSTC on platelet aggregation *ex vivo* in male and female subjects. The human studies consistently showed a reduction in platelet aggregation following consumption of the WSTC under the conditions of use proposed by the applicant and a sustained effect for up to 28 days in subjects that were representative of the target population. The EFSA NDA concluded that a cause and effect relationship was demonstrated between consumption of WSTC and the reduction in platelet aggregation in humans.

The authorized claim and conditions of use were published in Commission Decisions in 2009 (31) and 2010 (32). The claim is that WSTC I and II help maintain normal platelet aggregation, which contributes to healthy blood flow, and the conditions of use are that the beneficial effect is obtained with a daily consumption of 3 g WSTC I or 150 mg WSTC II in up to 250 ml of either fruit juices, flavored drinks or yogurt drinks (unless heavily pasteurized) or with a daily consumption of 3 g WSTC I or 150 mg WSTC II in food supplements, i.e. powdered single-serve sachets, tablets and capsules, when taken with a glass of water or other liquid.

Rejection of Health Claims Related to Dietary Fiber

In 2010, the EFSA NDA concluded that the food constituent dietary fiber was not sufficiently characterized in relation to claimed physiological effects relating to satiety, weight management, normal blood glucose concentrations, normal blood cholesterol concentrations, normal bowel function and regularity, reduction of postprandial glycemic response, decreases in potentially pathogenic gastrointestinal micro-organisms, fecal bulking effects etc. (33) The European Union legal definition of dietary fiber (34) refers to all carbohydrate polymers with three or more monomeric units that are neither digested nor absorbed in the human small intestine, including those occurring naturally, those obtained from raw material by physical, enzymatic or chemical means and edible synthetic fibers. These components include non-starch polysaccharides, resistant starch, resistant oligosaccharides, and other non-digestible but quantitatively minor components when naturally associated with dietary fiber polysaccharides, especially lignin. The terms soluble and insoluble have been used in the literature to classify dietary fiber as viscous soluble in water (e.g. pectins) or as water insoluble (e.g.

cellulose) in an attempt to link different physicochemical properties of fiber components to different physiological effects. The legal definition, therefore, encompasses a large and heterogeneous group of substances for which there is currently no single method of analysis, although several methods have been identified to determine the fiber content of foods. The classification of dietary fiber is method dependent but uses established analytical methods. However, the beneficial physiological effects are fiber specific and depend on the unique physical and chemical characteristics of the fiber components. As a result, the EFSA NDA concluded that, because the generic description dietary fiber could not be sufficiently characterized, a cause and effect relationship could not be established (*33*). Successful health claims have been made subsequently that relate to specific, well-characterized fiber constituents and specific beneficial physiological effect(s) demonstrated by generally accepted scientific evidence.

Consumer Understanding of Nutrition and Health Claims

An important aspect of the European legislation is that it states that the use of nutrition and health claims shall only be permitted if the average consumer can be expected to understand the beneficial effects expressed in the claim (Article 5 [2]) (*2*). The benchmark definition for the average consumer is one "who is reasonably well informed and reasonably observant and circumspect". Recital 16, in the preamble to the regulation, defines further the notion of the average consumer by taking into account different social, cultural and linguistic factors, as interpreted by the European Court of Justice. The key objectives of the legislation are to ensure that nutrition and health claims are truthful, relevant and understood by consumers.

Wills *et al.* (*35*) undertook a comprehensive review of the state of research into how consumers understand and respond to health claims on food and drink products, their attitudes to health claims and their purchasing intentions for foods with health claims on them. Unfortunately, nutrition knowledge is often lacking and consumers are easily confused by details and scientific wording of nutrition and health information. A consistent finding is that consumers prefer simple information on the front of pack, with more detail provided on the back (*36*, *37*).

A good example of the challenges posed by the wording of health claims and consumer understanding is the health relationship for WSTC. The claimed health benefit, as previously described, is a reduction in platelet aggregation, which contributes to overall vascular and cardiovascular health. The applicant's proposed wording was "Helps to maintain a healthy blood flow and benefits circulation". Although the EFSA NDA gave a positive opinion on the supporting scientific evidence, it considered that the findings were not reflected in this wording because only measures of platelet aggregation had been used in the studies presented, whereas "blood flow" and particularly "circulation" depend on many other factors that had not been addressed in the studies provided. After more than 6 months' deliberation with the European Commission and member states and the provision of market research results on consumer understanding,

the compromise wording "WSTC helps maintain normal platelet aggregation which contributes to healthy blood flow" was enshrined in law (*31*, *32*).

This example is just one of many authorized health claims that consumers find difficult to understand. Clearly, the claimed health benefits must not go beyond the scope of the evidence, or confuse or mislead the consumer. However, the area of consumer research is ripe for development of methodologies and approaches to assess consumer understanding of health claims (*38*).

Future Challenges

In addition to more research on consumer understanding to link the totality of the available scientific data and weight of evidence with claims that are truthful and meaningful to consumers, there is a need to assess the strength and limitations of the different sources of scientific evidence. Well-designed randomized controlled intervention trials provide the most persuasive evidence of efficacy in human subjects, and this investigational design permits strong causal interferences. Most other experimental designs, lumped together under the term observational studies, are unable to distinguish whether any observed difference is due to the intervention or to some other unrecognized and often unmeasured factor. Although appropriate study designs and statistical methods can be used to minimize the effects of confounding variables, observational studies can only provide an association and cannot provide definitive proof of cause and effect (*13*). Much of what is known in state-of-the-art human nutrition and health relationships, however, is based on epidemiological evidence, and this source of data underpins most of the national and international dietary recommendations.

The success of randomized controlled trials in evaluating medical treatments and pharmaceuticals does not mean that this method is always the most appropriate for the evaluation of nutritional effects (*26*). Typically, drugs act quickly and their endpoints can be measured over short periods of time. Nutrients and other substances with beneficial physiological effects tend to manifest as small differences over long periods of time. Nutrients work together rather than in isolation. Unlike drugs, there is rarely a nutrient-free state against which nutrient effects can be compared. The dilemmas of focusing on pharmaceutical approaches to assess nutrition and the demonstration of proof of cause and effect are highlighted by Blumberg and Heaney (*26*).

The present predominantly drug-like approach to evidence-based nutrition is based on clinical end points and biomarkers of disease. In several areas of research, no validated physiological biomarkers exist. The identification of further relevant biomarkers to measure food functionality in the human body is one of the most important challenges in nutrition research today. On the basis of nutrition being primarily aimed at maintaining or possibly improving health in normal healthy individuals, new methods and models will need to be developed that better take into account the complexity and balance of homeostatic mechanisms; the term health might be defined more accurately as the ability to adapt to internal and external stimuli. In the case of chronic or slow-developing pathologies, it can be said that there is an adaptation, as individuals can live with

them for a very long time, even without medication. New models are required to illustrate the effects of nutritional interventions on 'normal' biological processes and homeostatic balances in individuals. New biomarkers will be needed to detect early signs of homeostatic disturbance and suboptimum health well before there is any clinical sign of disease (*12*, *24*, *39*).

Clearly, despite the range of nutrition sciences available, they are not necessarily designed to fit the purpose of substantiating health claims, which leaves gaps and uncertainties in the final assessments. Much more attention needs to be paid to developing a suitable scientific framework for weighing the strength and consistency of the evidence in order to embrace state-of-the-art nutrition science and to stimulate future academic research.

There is no doubt that the national and international regulatory developments on nutrition and health claims will impact significantly on existing and new product claims, formulations and recipes, commercial communications to consumers, marketing and research and development strategies and academic research. Future and ongoing regulatory developments in the areas of food and nutrition include mandatory nutrition labelling, nutrient profiles, front-of-pack labelling, such as signposting and traffic lights systems, the setting of maximum safe levels of vitamins and minerals in fortified foods and food (dietary) supplements, various methods of chemical and biochemical analyses and regulatory compliance by enforcement authorities.

The applications of chemical science and technology within the area of food and agriculture have allowed the production of foods in adequate quantities to meet the needs of a growing world population. Today, the production-to-consumption food chain is complex and our food is largely safe, tasty, nutritious, abundant, diverse and convenient, and is less costly and more readily accessible than ever before. The growing awareness of the beneficial interactions between the presence or absence of a food constituent and a specific function or functions in the human body and an improved understanding of the role of food and food constituents in maintaining and improving health and reducing the risk of major chronic diseases will continue to give the impetus for a renaissance in the food biosciences and in analytical chemistry.

References

1. Floros, J. D.; Newsome, R.; Fisher, W.; Barbosa-Canoves, G. V.; Chen, H.; Dunne, C. P.; German, J. B.; Hall, R. L.; Heldman, D. R.; Karwe, M. V.; Knabel, S. J.; Labuza, T. P.; Lund, D. B.; Newell-McGloughline, M.; Robinson, J. L.; Sebranek, J. G.; Shewfelt, R. L.; Tracy, W. F.; Weaver, C. M.; Ziegler, G. R. *Compr. Rev. Food Sci. Food Safety* **2010**, *9*, 572–599.
2. European Commission. Regulation (EC) No 1924/2006. *Off. J. Eur. Union* **2006**, *L404*, 3–18.
3. US Food and Drug Administration. *Fed. Regist.* **1993**, *58*, 2478–2536.
4. Codex Alimentarius. Guidelines for use of nutrition and health claims, CAC/GL 23-1997. http://www.iadsa.org/images/userfiles/Guidelines%20NHC.pdf (accessed June, 2014).

5. WHO; FAO; Codex Alimentarius. *Understanding the Codex Alimentarius*, 3rd ed.; World Health Organization; Food and Agriculture Organization of the United Nations: Rome, Italy, 2008.

6. Codex Alimentarius. Guidelines on nutrition labelling CAC/GL 2-1985. http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.codexalimentarius.net%2Finput%2Fdownload%2Fstandards%2F34%2FCXG_002e.pdf&ei=n_z4Uqn0GI7T7Aa6zIHwCA&usg=AFQjCNFdcdKFltbojzD6K7TVtfPof2Z47g&sig2=TpegbmyccdzjcH0BjrZ9Vw&bvm=bv.60983673,d.ZGU (accessed November, 2013).

7. Codex Alimentarius. General principles for the addition of essential nutrients to foods CAC/GL 9-1987. http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.codexalimentarius.org%2Finput%2Fdownload%2Fstandards%2F299%2FCXG_009e.pdf&ei=8Pz4Uua0BqTX7Aa_tIHoBg&usg=AFQjCNEh81y0Cogovn3zsB6Jn2Fmgk5-GA&sig2=aSfp5mnSDv7D2E2R6pSHbQ&bvm=bv.60983673,d.ZGU (accessed November, 2013).

8. Codex Alimentarius Commission. *Procedural manual: nutritional risk analysis principles and guidelines for application to the work of the Committee on Nutrition and Foods for Special Dietary Uses*, 19th ed.; Section 4; World Health Organization; Food and Agriculture Organization of the United Nations: Rome, Italy, 2010; pp 120–126.

9. Diplock, A. T.; Aggett, P. J.; Ashwell, M.; Bornet, F.; Fern, E. B.; Roberfroid, M. B. *Br. J. Nutr.* **1999**, *81*, S1–S27.

10. Aggett, P. J.; Antoine, J.-M.; Asp, N.-G.; Bellisle, F.; Contor, L.; Cummings, J. H.; Howlett, J.; Müller, D. J. G.; Persin, C.; Pijls, L. T. J.; Rechkemmer, G.; Tuijtelaars, S.; Verhagen, H. *Eur. J. Nutr.* **2003**, *44* (Suppl. 1), 1–30.

11. Richardson, D. P.; Affertsholt, T.; Asp, N-G.; Bruce, A.; Grossklaus, R.; Howlett, J.; Pannemans, S.; Ross, R.; Verhagen, H.; Viechtbauer, V. *Eur. J. Nutr.* **2003**, *42*, 96–111.

12. Gallagher, A. M.; Meijer, G. W.; Richardson, D. P.; Rondeau, V.; Skarp, M.; Stasse-Wolthuis, G. C.; Witkamp, R. *Br. J. Nutr.* **2011**, *106*, S16–S28.

13. Richardson, D. P. *Proc. Nutr. Soc.* **2012**, *71*, 127–140.

14. Welch, R. W.; Antoine, J.-M.; Berta, J.-L.; Bub, A.; de Vries, J.; Guarner, F.; Hasselwander, O.; Hendriks, H.; Jäkel, M.; Koletzko, B. V.; Patterson, C. C.; Richelle, M.; Skarp, M.; Theis, S.; Vidry, S.; Woodside, J. V. *Br. J. Nutr.* **2011**, *106*, S3–S15.

15. European Food Safety Authority. *EFSA J.* **2011**, *9*, 2372.

16. European Commission. Regulation (EC) No 353/2008. *Off. J. Eur. Union* **2008**, *L109*, 11–16.

17. European Food Safety Authority. *EFSA J.* **2011**, *9*, 2170.

18. European Food Safety Authority. *EFSA J.* **2011**, *9*, 1984.

19. European Food Safety Authority. *EFSA J.* **2011**, *9*, 2474.

20. European Food Safety Authority. *EFSA J.* **2012**, *10*, 2604.

21. European Food Safety Authority. *EFSA J.* **2012**, *10*, 2702.

22. European Food Safety Authority. *EFSA J.* **2012**, *10*, 2816.

23. European Food Safety Authority. *EFSA J.* **2012**, *10*, 2817.
24. Mitchell, H. L.; Aggett, P. J.; Richardson, D. P.; Stowell, J. D. *Br. J. Nutr.* **2011**, *105*, 322–328.
25. Biesalski, H. K.; Aggett, P.; Anton, R.; Bernstein, P. S.; Blumberg, J. *Nutrition* **2011**, *62*, 465–473.
26. Blumberg, J.; Heaney, R. P.; Huncharek, M.; Scholl, T.; Stampfer, M.; Vieth, R.; Weaver, C. M.; Zeisel, S. H. *Nutr. Rev.* **2010**, *68*, 478–484.
27. European Food Safety Authority. *EFSA J.* **2010**, *8*, 1885.
28. European Commission. *Off. J. Eur. Union* **2011**, *L296*, 26–28.
29. European Food Safety Authority. *EFSA J.* **2009**, *1105*, 1–15.
30. European Food Safety Authority. *EFSA J.* **2010**, *8*, 1689.
31. European Commission. *Off. J. Eur. Union* **2009**, *L336*, 55–57.
32. European Commission. *Off. J. Eur. Union* **2010**, *L328*, 18–19.
33. European Food Safety Authority. *EFSA J.* **2010**, *8*, 1735.
34. European Commission. Directive 2008/100/EC. *Off. J. Eur. Union* **2008**, *L285*, 9–12.
35. Wills, J. M.; Storcksdieck gennant Bonsmann, S.; Kolka, M.; Grunert, K. G. *Proc. Nutr. Soc.* **2012**, *71*, 229–236.
36. Roe, B. E.; Levy, A. S.; Derby, B. M. *J. Public Policy Mark.* **1999**, *18*, 89–105.
37. Williams, P. G. *Nutr. Rev.* **2005**, *63*, 256–264.
38. Leathwood, P. D.; Richardson, D. P.; Strater, P.; Todd, P. M.; van Trijp, H. C. M. *J. Nutr.* **2007**, *98*, 474–484.
39. van Ommen, B.; Keijer, J.; Heil, S. G.; Kaput, J. *Mol. Nutr. Food Res.* **2009**, *53*, 795–804.

Comprehensive Nuclear Magnetic Resonance Analysis of Honey

István Pelczer,[1,*] Cristine Ha,[1,3] and Joseph G. Lelinho[2]

[1]Department of Chemistry, Frick Chemistry Laboratory, Princeton University, Princeton, New Jersey 08544
[2]Hilltop Honey, North Caldwell, New Jersey 07006
[*]E-mail: ipelczer@princeton.edu
[3]Permanent affiliation: University of Michigan Center of Human Growth and Development, Ann Arbor, Michigan 48104

Honey is a complex natural product with two major classes of components: the dominant sugar derivatives and lower concentrations of other molecules (proteins, vitamins, free amino acids, other organic acids and flavor substances). Molecular-level analysis of honey can be complicated due to, in part, the overwhelming presence of sugars, as well as the sensitivity and fragility of many of the components. Additionally, credible characterization of the composition is essential for quality assurance, as honey can be, and often is, adulterated and falsified as far as origin, composition and preparation are concerned. This issue is very important for human health, as well as a major concern for the market. In this chapter we discuss the opportunities for honey analysis by nuclear magnetic resonance spectroscopy, with focus on assessment of samples in their native condition without any separation.

Introduction

Honey, a product of bees, has been used as source of sugar for thousands of years, and its special medical and general nutritional benefits are widely appreciated. It is a big market commodity with worldwide distribution. This natural material is highly complex. Honey is composed of a large number of

different chemical and biochemical entities, predominantly sugars and, at much lower concentrations, proteins, vitamins, free amino acids, other organic acids, antibiotics, flavor compounds, minerals, etc. Because of high sugar concentrations and the presence of natural antibiotics, honey does not spoil over long periods of time and has been used as a preservation material since ancient times. There are many varieties of honey with different flavors and qualities that result in part from geographical and regional origins and seasonal variations, but also from treatment and preparation processes. Honey has wide ranging uses, but it is primarily used as a natural sweetener on its own or as an ingredient in foods and baked products (Figure 1). It has proven beneficial medical effects when administered orally or topically and has long-term, historic use as a winemaking material (e.g. for mead). More recently, high-end spirits laced with honey have also come to the market. Due to its valuable properties and, therefore, the potentially high market value of good-quality honey, there have been many attempts to manipulate the composition, including replacing bee products with substitutes. A more frequent cause of concern is the presence of 'foreign' materials added intentionally or unintentionally. These include antibiotics used to treat the bees or pesticides sprayed on fields, which can migrate into and/or modulate the products of the hive by adversely affecting the bees' biochemistry/biology. Recent worldwide dwindling of bee colonies for reasons that are still unclear (although role of pesticides is highly suspected) adds to the pressure to manipulate honey and make cheap substitutes and derivatives. Quality assurance and control are big interests for many market players, beekeepers and re-sellers alike, and are definitely of high importance to consumers.

Figure 1. Decorative Hungarian honey breads. (Picture courtesy of Ms. István Pelczer Sr.)

Honey can be analyzed at different levels ranging from bulk methods (*1*, *2*) to detailed component analysis and molecular assessment, including isotope ratio measurements by either nuclear magnetic resonance (NMR) spectroscopy (*3*) or, with higher sensitivity, mass spectrometry (MS) (*4*). Obviously, molecular assessment is more powerful to identify subtle differences by examining contaminants present in tiny quantities and other trace materials in a targeted fashion (*5*, *6*), but it usually requires expensive fine instrumentation. This chapter assesses the capabilities of NMR spectroscopy, with the primary focus being on one-dimensional applications. NMR has special advantages despite its relatively low sensitivity, yet it is only one piece of the toolkit to be considered. The final goal is not only safety and quality assurance but also to learn more about the still enigmatic life of bees and the biology and biochemistry of producing honey and other products in the hive.

Most of the analytical methods available for analysis of honey rely on some kind of separation of the sample or other intervention, but NMR spectroscopy can assess all the organic materials simultaneously and quantitatively without any physical separation. Targeted analysis of small components (*7–9*) and component identification may benefit from separation or concentration of the selected material. Eventually, spiking the original sample can be the ultimate tool to identify suspected ingredients. ^{1}H-, ^{13}C- and ^{31}P-NMR are the primary spectroscopic methods used in such analyses, having different and complementary capabilities that are discussed in more detail below. For the identification of selected components, a variety of two-dimensional correlation methods can also be considered (*10*). The combination of spectroscopic and statistical analyses, which are widely applied to metabolic mixtures (*11–14*), provides a powerful additional avenue to aid classification, mining otherwise hidden information, finding correlations in composition of materials and for component identification.

NMR spectroscopy of honey is quantitative by nature; all components are represented on the spectrum in proportion to their relative concentrations, provided certain simple experimental conditions are met. The technique is robust and highly reproducible, with a high dynamic range (*11*). Samples can often be subjected to NMR analysis in their native condition without any tampering or prior separation. Regardless of the high capital investment required to purchase an NMR instrument, the cost per sample is usually quite acceptable, especially with available automation. The relatively low sensitivity of NMR spectroscopy compared with other major methods (optical spectroscopy and MS) has been greatly compensated by the introduction of cryoprobes (*15–17*).

^{1}H-NMR Spectroscopy

Honey has two distinct classes of components: sugars, which comprise the overwhelming majority of the content, and other ingredients in much smaller concentrations. In Figure 2 ^{1}H-NMR spectra of 15 diverse honey samples are shown overlaid, with very high vertical expansion to highlight the low-concentration components. In this presentation the sugar resonances way exceed the scale and coalesce into a single entity between around 4.2 and 2.9 ppm.

Some of the samples have highly visible, broader amide resonances (around 6.7 and 6.0 ppm, respectively), while the very-low-quantity proteins show with very broad background signals underneath the sharp resonances of the small molecules. ¹H-NMR has much greater sensitivity than ¹³C-NMR; therefore, it is best suited for analysis and identification of the small components (³¹P-NMR is restricted to studying only phosphorus-containing species). In typical circumstances the native honey sample is dissolved in regular water if visualization of exchangeable protons of amides is desired. For lock, either a small amount of ²H₂O is mixed directly into the sample or a small insert capillary filled with deuterated material is used. With high-Q probes, and especially at high magnetic fields, radiation damping may be an issue, leading to significant broadening of the water peak. Even if the sample is prepared entirely with ²H₂O, the natural water content of the honey usually makes it necessary to apply some kind of solvent suppression during the experiment. Unfortunately, some of the sugar resonances fall under or are too close to the water signal's position and will be fully or partially suppressed.
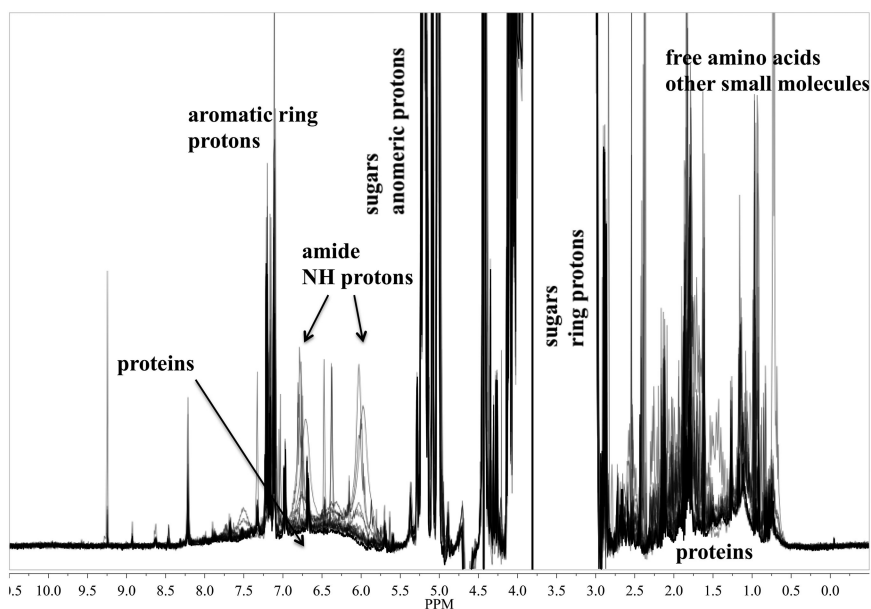
Figure 2. Overlaid ¹H-NMR spectra of 15 honey samples after water suppression. Some components are identified on the plot. The vertical scale is enlarged to highlight low-concentration components to show notable diversity.

The ^{1}H-NMR spectrum usually consists of many overlapping multiplets, which makes analysis, especially direct quantitative analysis, of most components difficult and often impossible. The major sugar components (fructose, glucose, maltose and sucrose) have a seriously overlapping jungle of resonances that obscures the lower-concentration sugar ingredients, which are of potential fingerprint value. This problem may be somewhat reduced by increasing the strength of the magnetic field. Nevertheless, it will not be eliminated, while the cost of the instrumentation will increase exponentially. Two-dimensional *J*-resolved spectroscopy, which can easily be integrated with efficient solvent suppression (I Pelczer, Princeton University, unpublished), can largely separate and resolve many of these multiplets in the mixture (*18*). However, the F2 one-dimensional projection of the two-dimensional spectrum no longer reflects the original relative concentrations (Figure 3). In comparative analyses of many honey samples, a variety of statistical approaches (see later) can be quite useful in identifying components based on their correlated behavior on the spectrum, including the one-dimensional F2 *J*-projections (*18*, *19*) and even in heteronuclear (^{13}C or ^{31}P) experiments.

In order to highlight the low-concentration components, one can consider suppressing the sugar resonances during the experiment by use of band-selective pulses and gradients. This approach, however, comes with a loss of information. Modern instruments have ADC units of high resolution and, therefore, very high dynamic range, which allow all data to be kept. In this case, the sugar region can still be suppressed in data processing when applying statistical analysis in full analogy to removing the regime of the residual water resonance and highlighting the rest of the spectrum (*20*).

^{13}C-NMR Spectroscopy

^{13}C-NMR information has notable advantages when attempting to identify molecules, especially in a complex mixture. As the carbon atoms are situated one layer deeper in the molecular structure than the protons, they are less exposed to environmental conditions (solvent, pH, possible intermolecular effects) and, therefore, provide more-characteristic and more-reliable information about the actual structure. As a consequence, sophisticated spectrum prediction tools can assist identification of individual species. [13]C chemical shifts span over 200 ppm for most organic molecules, and the spectrum consists of singlet resonances after ^{1}H decoupling (apart from possible further heteronuclear interactions with ^{31}P, for example), which significantly reduces the possibility for overlap and is a great benefit for mixture analysis.

Given the high sensitivity of ^{1}H detection, it is a reasonable approach to map the ^{13}C chemical shifts indirectly through proton-detected 'inverse' correlation experiments. Great examples of such an approach have been published, including quantitative analysis of cell extracts (*21*, *22*). However, due to the wide frequency range of the ^{13}C resonances, decoupling of ^{13}C while detecting protons becomes increasingly difficult, especially as the applied magnetic field strength increases.

Possible overheating by ^{13}C decoupling also limits acquisition time, and thus resolution, in the ^{1}H dimension. Additionally, in order to gain sufficient resolution in the ^{13}C domain one a large number of increments is necessary, which may take a very long time and requires long-term stability of the sample. A promising development to largely alleviate this problem is using non-linear sampling for the incremented dimension (*23*).

As honey offers convenient access to high-concentration samples, ^{13}C-NMR analysis of the major carbohydrate ingredients can also be done with room-temperature probe technology within reasonable time (*24*, *25*). With the introduction of cryoprobes, which can be tailored and optimized for ^{13}C-detection, however, direct ^{13}C-detected analysis has become a truly competitive alternative, including for the analysis of the minor components that have fingerprint value (*17*). Such hardware even allows the running of otherwise time consuming quantitative ^{13}C experiments within an acceptable timeframe (typically from 0.5–2.0 h for our honey samples). Solvent suppression is not needed, and the length of the acquisition time can be extended without penalty to provide increased resolution that is limited only by the natural linewidth. Special benefits of direct-detection, one-dimensional ^{13}C-NMR for analysis of complex mixtures at natural abundance were demonstrated early (*16*), including metabolic mixtures for toxicology studies (*26*). An interesting consequence of high-sensitivity ^{13}C-detection over a high dynamic range (possibly as high as five magnitudes) is that many components present at low concentrations but with potential fingerprint value are in the same intensity range as the natural abundance ^{13}C-satellites of the large peaks. As these peaks are not independent and represent nothing new for the composition they have to be identified and sorted out carefully. Fortunately, the systematic behavior of these peaks—0.55% intensity relative to the main peak, doublet splitting by an approximately known coupling constant, slight upfield isotope shift—makes finding them relatively straightforward.

Regardless of the sensitivity previously unheard of when using optimized cryoprobes, ^{13}C-NMR is still best suited for analysis of the sugar components, whereas proteins and other very-low-concentration ingredients remain the subject of ^{1}H-NMR analysis. In Figure 4 the ^{13}C-NMR attached proton test (APT) spectrum (*27*) of a representative sample is presented with vertical expansion of selected downfield and upfield segments. In this spectrum odd and even multiplicity carbons are phased on the opposite sides of the baseline (in this case, signals with even numbers of attached protons are phased positive). Amide and ketone carbonyl signals are easily visible in the downfield segments, and methylene carbons of free proline, for example, can be identified in the upfield region. Yet, such peaks offer limited opportunity for detailed and extensive characterization. Very long acquisition times (often several hours or even overnight) may offer significant improvement of the signal-to-noise ratio, but do not help with comparative analysis of a larger number of samples.
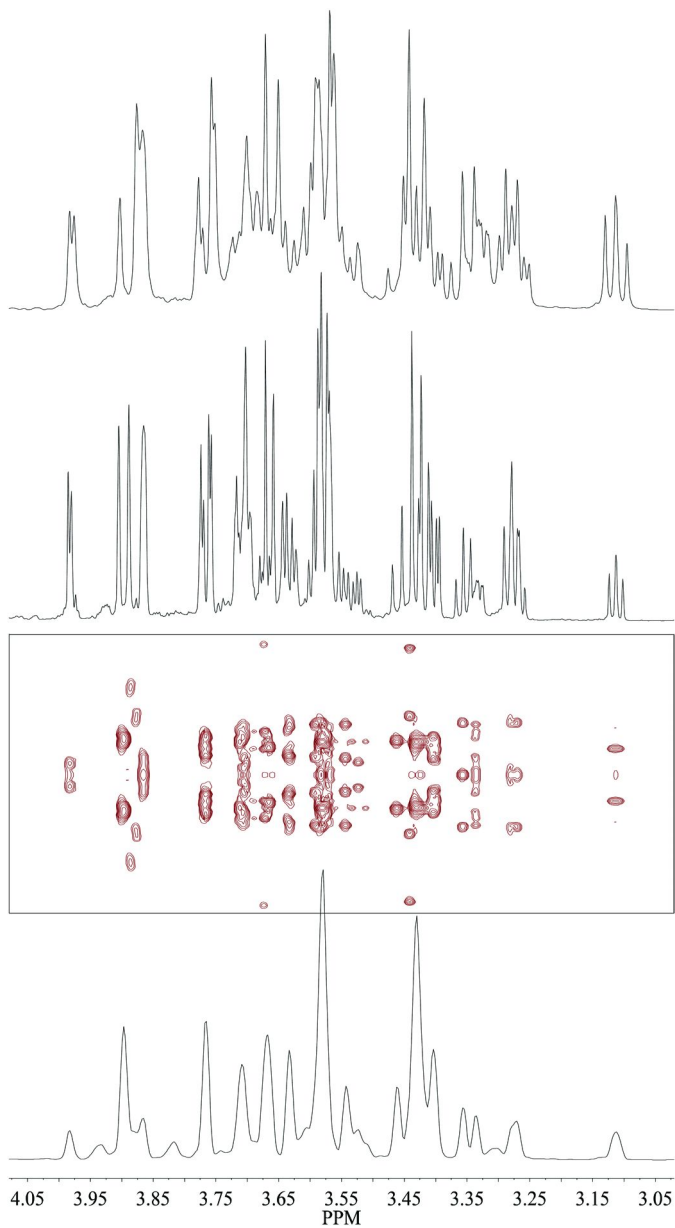
*Figure 3. Comparative plots for the sugar region in the ^{1}H-NMR spectrum, recorded at 500 MHz (top) and 800 MHz (all below). The relevant section of the two-dimensional **J**-spectrum is shown with the F2 projection (bottom).*
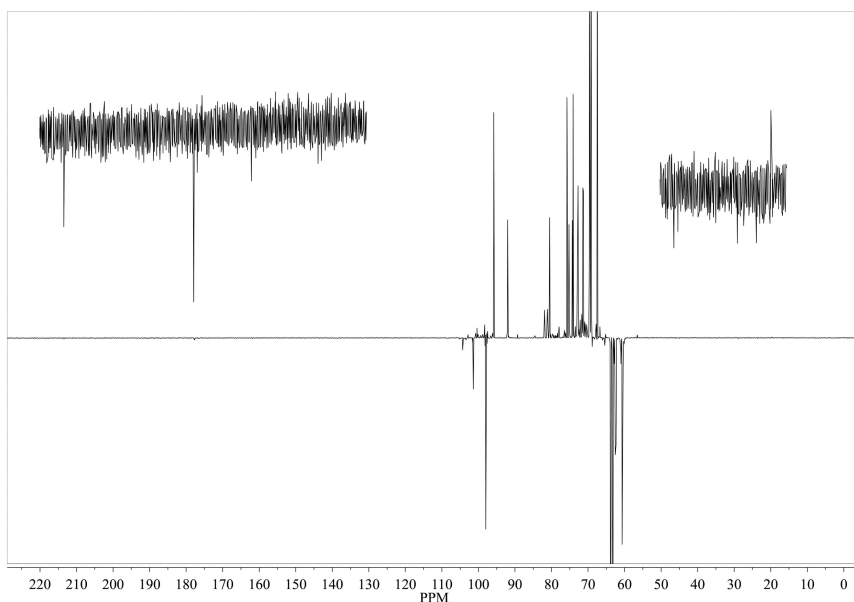
Figure 4. Representative ¹³C-NMR attached proton test (APT) (27) spectrum from the Hilltop Honey collection. Enlarged inserts show amide and ketone carbonyl peaks in the low-field segment and proline CH₂ resonances in the upfield region. Quaternary and CH₂ resonances are phased negative.

An overlaid collection of the expanded sugar region for 18 honey samples is shown in Figure 5a. The quantitative ¹³C-NMR spectra were normalized to total integral and uniformly calibrated, relative to the deuterium shift of the solvent (*28*). Slight yet characteristic variations of the relative concentrations of the major sugar components are quite clear. The many small peaks show just as much, if not more, important information to characterize the brand, origin, possible adulteration and contamination of samples. These ingredients might be possible to extract for additional, separate statistical analysis. Further expansion of a very small segment as a stack-plot (Figure 5b) highlights some of these small components and shows the richness of this kind of information. Analysis of these peaks, for component identification, possibly for use with additional statistical approaches, needs to be handled carefully with a well-designed strategy (see statistics section later).

³¹P-NMR Spectroscopy

Honey contains phosphates and, therefore, we regularly record ³¹P-NMR spectra by use of an optimized cryoprobe (cryo-QNP, Bruker-Biospin, Billerica, MA). Figure 6 shows a set of ³¹P-NMR spectra of selected honey samples. Various phosphate components distinguish the samples, while the differential line broadening relates to differences in pH. Comparative analysis of ³¹P-NMR spectra of the honey samples is in the works.
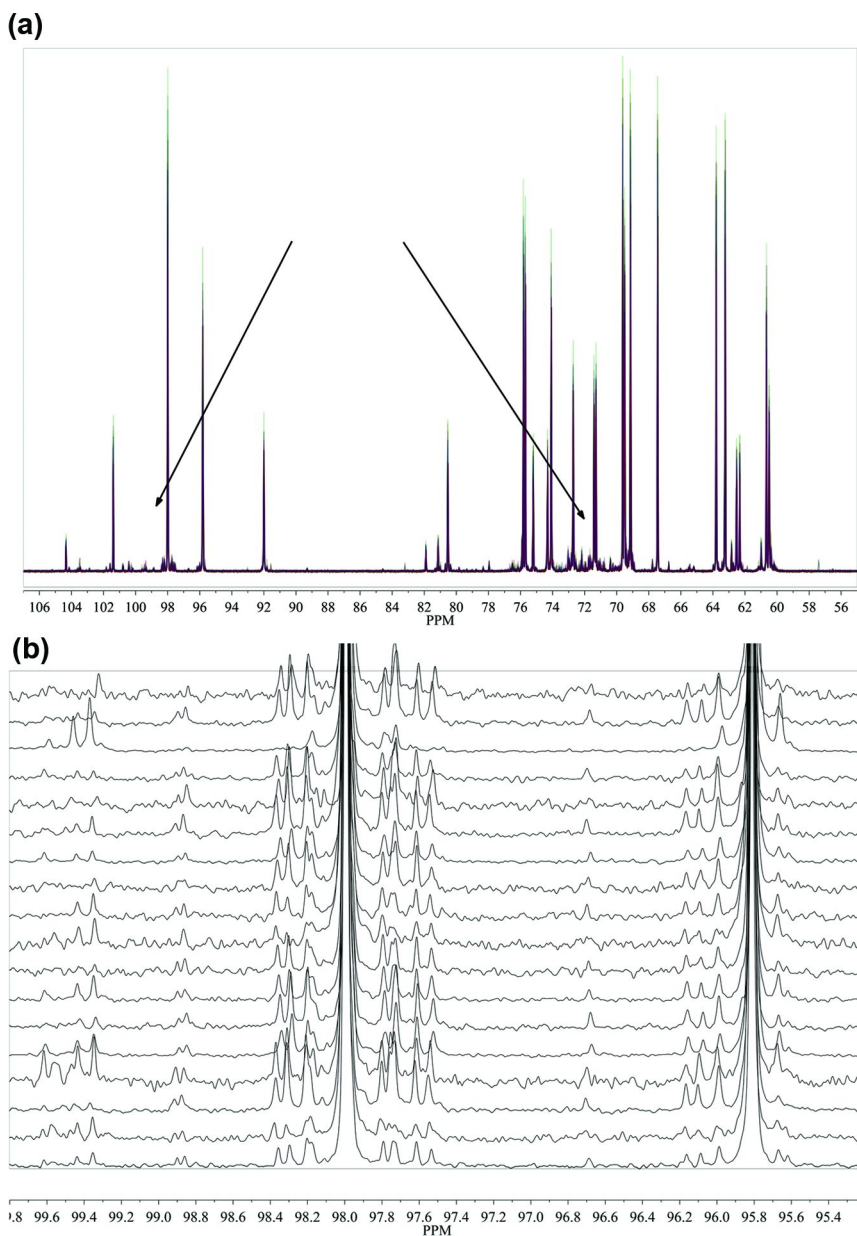
Figure 5. (a) Overlaid ¹³C-NMR spectra (sugar region) of 18 selected honey samples after normalization to total integral, mostly from the Hilltop Honey collection. Small but characteristic variations in concentrations of major sugar components are visible. Arrows indicate low-concentration components of fingerprint value. (b) Stacked expansion of a small region of the same set including the large signals of the anomeric carbons of fructose/β-fructofuranose and glucose/β-glucopyranose carbons at ca. 98.0 and 95.8 ppm, respectively.

70 0.65 0.60 0.55 0.50 0.45 0.40 0.35 0.30 0.25 0.20 0.15 0.10 0.05 0.00 -0.05 -0.10 -0.15 -0.20 -0.25 -0.30 -0.35 -0.40 -0.45 -0.50
PPM
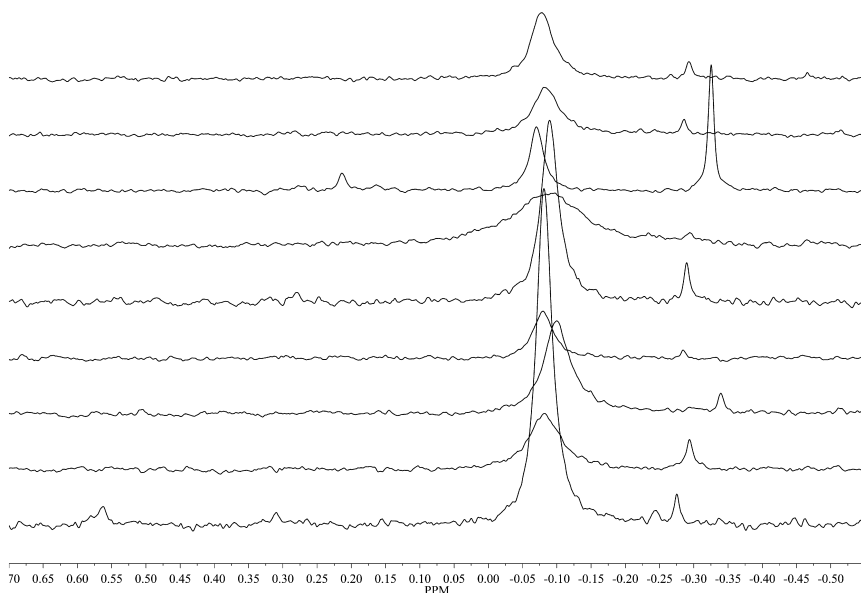
Figure 6. Representative set of ^{31}P-NMR spectra.

Statistical Analyses

Comparative multivariate statistical analysis of complex mixtures (*20*, *29*) can be a very powerful and efficient way to find similarities and differences across groups and to identify components that are responsible for such clustering. Most of the time such analysis is untargeted and identifies patterns, rather than focusing on selected components. The best examples so far have been seen in the area of metabolic mixture analysis (*14*). Appropriate normalization must be applied (*30*) and careful validation of the results is essential (*20*). The multivariate analysis can be complemented with a variety of statistical total correlation spectroscopy (STOCSY) interrogations of the data (*31*), which can help to extract detailed component information and to report about the significance of their contributions (*32*). When analyzing ^{1}H- or ^{13}C-NMR spectra statistically, the dominant sugar resonances can be kept or ignored in order to highlight the information content carried by the low-concentration components in the latter scenario. In the case of ^{1}H-NMR, band-selective pulses can be applied by experimental means to gain extra sensitivity for small components, although with modern instruments capable of very high dynamic range detection this may not be necessary. Various data processing and data management approaches could be used to achieve the same discrimination simply by selecting a subset of the spectrum for analysis through suppression of unwanted data segments, in a manner similar to that of removing residual solvent signals during data processing. This approach might be more complicated when multiple large signals for sugars need to be separated in ^{13}C-NMR, as fine-peak alignments might also be removed. Line-by-line analysis and spreadsheet-based data management, which can rely on frequency domain

curve fitting or Bayesian time domain analysis (*33*), are promising possibilities. Statistical analysis has been successfully applied to honey or ^{1}H-NMR (*34–36*) and, recently, for ^{13}C-NMR spectra at relatively low fields (100 MHz) (*25*). An illustrative scores plot of a multivariate discriminant analysis (O-PLS-DA) applied to the ^{1}H-NMR data of the 15 honey samples previously shown in Figure 2 is presented in Figure 7. It demonstrates the separation of different types of honey, while samples of the same type cluster together. Component-level comparative characterization of the variation between samples can be done best by pairwise analysis, taking into account the one-dimensional coefficient plot (*21*).
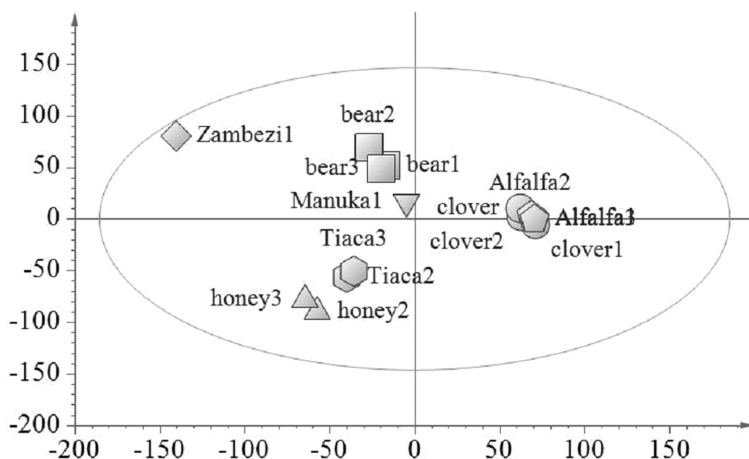
Figure 7. Illustrative untargeted statistical (O-PLS-DA) (20) analysis of selected honey samples from Figure 1 (scores plot). Normalization to total integral and univariate scaling were used. Commercial samples are mixed with those labeled "honey2" and "honey3", which are from the Hilltop Honey collection.

Sample Preparation and Data Acquisition

Honey samples we have tested have been either off-the-shelf commercial products or products produced at the Hilltop Honey apiary. Typically, we prepare honey samples fresh and perform NMR analysis immediately without extra conditioning except mixing them with deionized water at a ratio of about 1:5 and homogenizing them to reduce viscosity. For lock, around 25 μL D_2O is regularly added. Experiments are run on Bruker Avance-III spectrometers (800 or 500 MHz for ^{1}H frequency) at a controlled temperature (295 K) using a TCI (^{1}H/^{13}C/^{15}N//^{2}H) cryoprobe (800 MHz), and ^{13}C-detection optimized dual C/H cryoprobe (DCH, [13]C-^{1}H//^{2}H) or a cryo-QNP probe head (^{1}H/^{31}P,[13]C,[15]N//^{2}H), under the control of TopSpin software (version 3.2; all products Bruker-Biospin). ^{1}H-NMR spectra were acquired without or, most often, with suppressing the water signal by the excitation sculpting method (*37*). Quantitative ^{13}C-NMR spectra

were acquired with a 0.5 min recycle time and typically took up to 2 h acquisition time.

For off-line data processing, including apodization, phase and baseline correction and peak alignment in superimposed spectra, we use MNova, up to version 9.0 (MestreLab Research S.L., Santiago de Compostela, Spain). For line-by-line analysis we have typically used the global spectral deconvolution algorithm in MNova (MestreLab Research S.L.). Component identification and information management has been assisted by Chenomx NMR Suite (Chenomx, Edmonton, CA) and an interface written in house (*38*) to feed ^{13}C data efficiently into a user-defined database for large-scale analysis. We use SIMCA (up to version 13.0, Umetrics, Umeå, Sweden) for statistical analysis after some data preparation (normalization and suppression of selected regions, such as that of the residual water) in Microsoft Excel (Microsoft Corp., Redmond, WA), as well as a variety of in-house written Matlab (MathWorks, Natick, MA) scripts. Data analysis can be highly supported by searching sophisticated databases and spectrum collections, such as that from Bio-Rad Laboratories (Philadelphia, PA), Chenomx's built-in database, and the Madison Metbolomics Consortium Database portal (*39*).

Conclusions

NMR spectroscopy is a powerful, efficient and versatile tool for analysis of complex mixtures, such as honey. The composition of honey can be characterized in detail by ^{1}H-, ^{13}C- and ^{31}P-NMR methods and subsequent statistical analysis to identify origin, seasonal differences, possible contamination and adulteration. It is most effective to characterize low-concentration ingredients with ^{1}H-NMR, given the sensitivity limitations, whereas cryoprobe-assisted ^{13}C-NMR is best suited to assess sugar components, including the large variety of low-concentration derivatives, which carry fingerprint value. Statistical analysis can be tailored and include the dominant sugars or focus on the rest of the spectral data.

Acknowledgments

References

1. Doner, L. W.; Kushnir, I.; White, J. W., Jr. *Anal. Chem.* **1979**, *51*, 224A–228A.

2. Buba, F.; Gidado, A.; Shugaba, A. *Biochem. Anal. Biochem.* **2013**, *2*, 139–146.

3. Lindner, P.; Bermann, E.; Gamarnik, B. *J. Agric. Food Chem.* **1996**, *44*, 139–140.

4. Schellenberg, A.; Chmielus, S.; Schlicht, C.; Camin, F.; Perini, M.; Bontempo, L.; Heinrich, K.; Kelly, S. D.; Rossmann, A.; Thomas, F.; Jamin, E.; Horacek, M. *Food Chem.* **2010**, *121*, 770–777.

5. *Mass Spectrometry Handbook*; Lee, M., Ed.; Wiley: New York, 2012.

6. Cordella, C.; Moussa, I.; Martel, A.-C.; Sbirrazzuoli, N.; Lizzani-Cuvelier, L. *J. Agric. Food Chem.* **2002**, *50*, 1751–1764.

7. Sandusky, P.; Raftery, D. *Anal. Chem.* **2005**, *77*, 2455–2463.

8. Tuchado, P.; Martos, I.; Bortolotti, L.; A. G.Sabatini, A. G.; Ferreres, F.; Tomas-Barberan, F. A. *J. Agric. Food Chem.* **2009**, *57*, 5680–5686.

9. Donarski, J. A.; Stephen, S. A.; Harrison, M.; Driffield, M.; Charlton, A. J. *Food Chem.* **2010**, *118*, 987–994.

10. Richards, S. A., Hollerton, J. C. *Essential Practical NMR for Organic Chemistry*; Wiley: New York, 2010.

11. Lindon, J. C.; Nicholson, J. K. *Annu. Rev. Anal. Chem.* **2008**, *1*, 2.1–2.25.

12. Lenz, E. M.; Wilson, I. D. *J. Proteome Res.* **2007**, *6*, 443–458.

13. Balog, J.; Sasi-Szabó, L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezsö, B.; Damjanovich, L.; Darzi, A.; Nicholson, J. K.; Takáts, Z. *Sci. Transl. Med.* **2013**, *5*, 194ra93.

14. Dumas, M.-E.; Kinross, J.; Nicholson, J. K. *Gastroeneterology* **2014**, *146*, 46–62.

15. Shimba, N.; Kovacs, H.; Stern, A. S.; Nomura, A. M.; Shimada, I.; Hoch, J. C.; Craik, C. S.; Dötsch, V. *J. Biomol. NMR* **2004**, *30*, 175–179.

16. Zhou, Z.; Kümmerle, R.; Stevens, J. C.; Redwine, D.; He, Y.; Qiu, X.; Cong, R.; Klosin, J.; Montañez, N.; Roof, G. *J. Magn. Reson.* **2005**, *200*, 328–333.

17. Kovacs, H.; Moskau, D.; Spraul, M. *Prog. NMR Spectrosc.* **2005**, *46*, 131–155.

18. Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J.; Nicholson, J. K. *Anal. Chem.* **2010**, *82*, 1811–1821.

19. Ayaga, O. Sr. Thesis, Department of Chemistry, Princeton University, Princeton, NJ, 2008.

20. Fonville, J. M.; Richards, S. E.; Barton, R. H.; Boulange, C. L.; Ebbels, T. M. D.; Nicholson, J. K.; Holmes, E.; Dumas, M.-E. *J. Chemometrics* **2010**, *24*, 636–649.

21. Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385–9390.

22. Lewis, I. A.; Karsten, R. H.; Norton, M. E.; Tonelli, M.; Westler, W. M.; Markley, J. L. *Anal. Chem.* **2010**, *82*, 4558–4563.

23. Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luithardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K.; Falchuk, K. H.; Halperin, J. A.; Wagner, G. *J. Am. Chem. Soc.* **2007**, *129*, 5108–5116.

24. Mazzoni, V.; Bradesi, P.; Tomi, F.; Casanova, J. *Magn. Reson. Chem.* **1997**, *35*, S81–S90.
25. Ohmenhauser, M.; Monakhova, Y. B.; Kuballa, T.; Lachenmeier, D. W. *ISRN Anal. Chem.* **2013**, *2013*, 825318; DOI:10.1155/2013/825318.
26. Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskau, D.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2002**, *74*, 4588–4593.
27. Patt, S. L.; Shoolery, J. N. *J. Magn. Reson.* **1982**, *46*, 535–539.
28. Harris, R. K.; Becker, E. D.; Cabral De Menezes, S. M.; Goodfellow, R.; Granger, P. *Pure Appl. Chem.* **2001**, *73*, 1795–1818.
29. Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–479.
30. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal. Chem.* **2006**, *78*, 4281–4290.
31. Robinette, S. L.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2013**, *85*, 1282–1289.
32. Pappalardo, L.; Hoijemberg, P. A.; Pelczer, I.; Bailey, T. A. *Curr. Metabolomics* **2014**, in press.
33. Krishnamurthy, K. *Magn. Reson. Chem.* **2013**, *51*, 821–829.
34. Cotte, J. F.; Casabiance, H.; Chradon, S.; Lheritier, J.; Grenier-Loustalot, M. F. *J. Chromatogr. A* **2003**, *1021*, 145–155.
35. Cordella, C.; Militão, J. S. L. T.; Clément, M.-C.; Drajnudel, P.; Cabrol-Bass, D. *Anal. Chim. Acta* **2005**, *531*, 239–248.
36. Donarski, J. A.; Jones, S. A.; Charlton, A. J. *J. Agric. Food Chem.* **2008**, *56*, 5451–5456.
37. Hwang, T. L.; Shaka, A. J. *J. Magn. Reson., Ser. A* **1995**, *112*, 275–279.
38. Dellapenna, E. Sr. Thesis, Department of Chemistry, Princeton University, Princeton, NJ, 2012.
39. Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghalbina, H. R.; Sussman, M. R.; Markley, J. L. *Nat. Biotechnol.* **2008**, *26*, 162–164.

Rapid Screening Methods for Pharmaceutical Surveillance

Jason D. Rodriguez,* Connie M. Gryniewicz-Ruzicka, Sergey Arzhantsev, John F. Kauffman, and Lucinda F. Buhse

Division of Pharmaceutical Analysis, Center for Drug Evaluation and Research, US Food and Drug Administration, 645 S. Newstead Ave., Saint Louis, Missouri 63110
*E-mail: Jason.Rodriguez@fda.hhs.gov

Over the past few years, the United States Food and Drug Administration has developed rapid spectroscopic screening methods for the surveillance of pharmaceutical materials in the field. The main objective of this program is to enhance public safety and ensure that drug products and ingredients are safe and effective. This chapter presents a broad overview of the technologies that are used in the rapid screening program and highlights their use in improving public safety by increasing the number of regulated products that can be screened before reaching consumers.

Introduction

Pharmaceutical materials entering the United States supply chain from foreign sources have more than doubled in the past 10 years, with 80% of active pharmaceutical ingredients (APIs) and 40% of finished drug products being imported (*1*). These percentages are expected to grow over the next decade, as pharmaceutical materials are imported from more than 200 different countries. Rapid screening of incoming materials is one of the tools that has the potential to dramatically increase the number of products that undergo surveillance before reaching consumers. Such tools are needed since, as of 2010, less than 1% of imported regulated drug products underwent physical inspection (*2*).

The United States Food and Drug Administration (FDA) Division of Pharmaceutical Analysis (DPA) has developed, piloted and supported a rapid screening program (*3*) to enable surveillance of pharmaceutical materials at domestic and foreign sites. Domestically, the program has been in use at mail and import facilities since 2010, and rapid screening instruments were sent to offices in Mexico and India in 2013.

The primary aim of the program is to increase the number of pharmaceutical materials that undergo physical testing. For the purposes of this chapter, we have separated pharmaceutical materials into three classes: finished medicines, APIs and excipients (inactive ingredients). Finished medicines are the drug products dispensed to and consumed by patients and are commonly available as capsules, tablets, suspensions, injectables, etc. These finished medicines typically contain a mixture of APIs and excipients. We also discuss the role that rapid screening has on conducting surveillance on dietary and herbal supplements.

The main incentive for surveillance of finished products is the worldwide problem of counterfeit and sub-standard drugs. The World Health Organization defines counterfeit drugs as those that are deliberately made to have incorrect or insufficient ingredients or fraudulent labels and packaging (*4*). Current estimates indicate that around 10% of medicines worldwide are counterfeit or substandard (*5*). The rise of the Internet and online procurement of drugs by consumers increases the chances of consumers in the United States being exposed to poor quality drugs.

Counterfeit drugs can also include those with adulterated APIs and excipients. Both classes of pharmaceutical materials are especially vulnerable to economically motivated adulteration (EMA), which involves the deliberate substitution of a material with a cheaper alternative. Recent high-profile instances of EMA in the supply chain are the 2008 heparin crisis (*6*) and instances of melamine contamination in pet food, milk and infant formula in 2007–2008 (*7*, *8*). Levels of EMA are frequently in the 15–25% range (*9*) and can be detected with portable spectroscopic instrumentation.

Dietary supplements and herbal remedies are increasing in popularity due to the widespread belief that 'natural' products are safer and healthier than synthesized products. Natural supplements are widely available and claim to benefit consumers by treating various health conditions and promoting general well-being. One of the major concerns related to the safety of dietary supplements is their adulteration with undeclared synthetic pharmaceutical products to enhance the claims stated on the label. Herbal or 'traditional' medicines, which are often purchased via the Internet from foreign countries, have also been reported to contain high levels of toxic metals that may pose additional risks to consumers (*10*).

Compared with prescription and over-the-counter 'conventional' medications, herbal ingredients and finished dietary supplement products are subject to less scrutiny. The United States Federal Food, Drug, and Cosmetic Act was amended in 1994 to establish standards with regard to dietary supplements by the publication of the Dietary Supplement Health and Education Act (*11*). Two key features of this Act are relevant to product safety: 1) manufacturers of dietary supplements do not have to provide the FDA with evidence that they are effective

or safe before marketing; and 2) once a dietary supplement is on the market, the FDA holds the burden of proof to show that the product is not safe in order to restrict its use or to remove it from the marketplace (*11*).

The rapid screening program developed at the DPA uses four types of portable or handheld instrument platforms: near-infrared (NIR) spectroscopy, Raman spectroscopy, ion-mobility spectrometry (IMS) and X-ray fluorescence (XRF) spectroscopy. Each instrument is discussed in detail in the following sections. The selection of a technique for any given analysis depends on the material to be examined and the adulterant of interest. This suite of instruments provides the FDA with a collection of orthogonal methods to test products with the most appropriate technology.

Raman and NIR Spectroscopy

Background

Raman and NIR are complementary spectroscopic techniques used to acquire unique molecular fingerprints of different substances. They can differentiate between chemicals that may physically appear identical to the unaided eye. While both techniques yield vibrational information on the chemical bonds present in a substance, their underlying principles are different. Raman spectroscopy is a scattering technique. It relies on light scattering at a different wavelength from that of the incident light after interacting with a sample (*12*). The differences in energy levels between the incident and the scattered lights are proportional to those of the vibrational modes in a molecule. Thus, different molecules have unique Raman spectra. By contrast, NIR spectroscopy relies on the absorption of light. Upon irradiation with NIR light, typically in the range of 800–2,500 nm, the vibrational modes in a molecule are excited, which gives rise to a unique spectrum for each molecule (*13*). The vibrations observed in the NIR region of the electromagnetic spectrum correspond to O–H, C–H, C–O and N–H overtones and combination bands.

As few as 10 years ago, both techniques were regarded as primarily laboratory methods due to the large footprint required for the two types of spectrometers. Today, miniaturization of the optical and electronic components used to craft these instruments has led to widely available portable or handheld platforms from a variety of commercial vendors. Portable Raman and NIR units have been used extensively in various fields (*14–16*), including food safety, pharmaceutical authentication, law enforcement, forensics and process control. One of the greatest strengths of these instruments is their ability to perform rapid interrogation of the sample under study in its original packaging without any additional preparation that could destroy the sample. This characteristic is well suited to the FDA's rapid screening program, since one of the major goals is to identify substances that need further testing (*3*), which requires the original sample to be intact. Both techniques are an important part of the rapid screening program and may be used interchangeably according to the sample's properties. Samples featuring high levels of fluorescence at baseline on Raman spectroscopy, for example, may be amenable to NIR analysis, whereas hygroscopic samples are

more reliably screened by Raman spectroscopy, since moisture content influences the NIR signatures.

Despite these promising attributes, portable and handheld spectrometers are not without their challenges. Most notably, performance and sensitivity differ substantially between instruments. Each instrument produces a unique spectral response, even when comparing multiple spectrometers of the same model from the same vendor (*17*, *18*). The differences become even more pronounced when instruments from different vendors or data from different platforms are compared (*19*), such as laboratory-based and portable or handheld units. Spectral data acquired and methods built on one instrument, therefore, are unusable on other instruments unless additional standardization and calibration are employed (*15*, *17–19*). The DPA has developed instrument transfer protocols for distribution of methods between instruments.

Methods

The DPA rapid screening program utilizes Raman and NIR instruments to conduct surveillance on pharmaceutical materials vulnerable to EMA (*15*). Screening for EMA is primarily done on raw materials—that is, APIs and excipients. Two main algorithms are used in the development of Raman and NIR methods: spectral library correlation methods and multivariate methods.

Spectral library correlation methods involve comparison between a reference spectrum of a material and a spectrum of the unknown material under study. The metric used for comparison is the spectral correlation (SC) index (also known as the hit quality index) (*20*), and is calculated with the following equation.

$$SC = \frac{(Library \cdot Unknown)^2}{(Library \cdot Library)(Unknown \cdot Unknown)}$$

The SC value is the square of the spectral covariance between the library and unknown spectra, normalized by the squared norms of the two spectral vectors (i.e. the square of the spectral correlation coefficient). SC values range from 0.000 (poorest match possible) to 1.000 (perfect match). The default threshold, based on traditional use of the SC index, is set to 0.95 (*9*). Since the threshold can be affected by transfer of library spectra between different spectrometers (*19*), the optimum pass/fail threshold is best determined on a system-by-system basis to suit the desired application. Preprocessing of spectra greatly influences the SC value (*20*). First-derivative preprocessing enhances the sensitivity of the SC index by reducing the influence of the baseline in spectra and increasing the influence of peaks (which occur as zero crossings in the first derivative spectra). All SC values presented in this chapter have been calculated on first-derivative spectra, which were obtained by applying a second-order, 11-point window Gram polynomial to the original spectrum.

Our initial efforts (*15*) with Raman and NIR screening of systems affected by EMA focused on the adulteration of glycerin with diethylene glycol (DEG), which is a poisonous organic solvent (*17*). Several adverse events have been reported due to adulteration with DEG, including the event that led to the formation of the

FDA (*9*). This system is a prime example of the challenges posed by EMA, since glycerin and DEG are closely related in structure and share many physical and chemical properties, as shown in Figure 1. A high degree of spectral similarities can be seen, most notably in the region between 1,000 cm⁻¹ and 1,600 cm⁻¹. This similarity leads to only a modest decline in the SC value, and samples with as much as 12% of DEG in a mixture would pass this test. To better estimate the exact DEG composition that would be expected to cause the SC test to deliver a failed result, we calculated the SC values for a range of DEG in glycerin compositions and used a polynomial fit to estimate the sensitivity of the SC test (*9*). Figure 2 shows that the SC values for compositions are essentially unchanged until the DEG composition reaches roughly 10%. Thereafter, the SC value begins to drop noticeably, and crosses the 0.95 threshold when DEG is around 18%. Samples containing DEG levels below this would not be flagged for further testing. While the DEG level that the SC method is able to detect is high, EMA is estimated to occur in this range (between 15% and 25% adulterant) in a raw material. EMA may still exist outside this estimated range, and our studies indicate that SC-based tests are useful for screening samples at EMA-relevant ranges or higher (*9*).
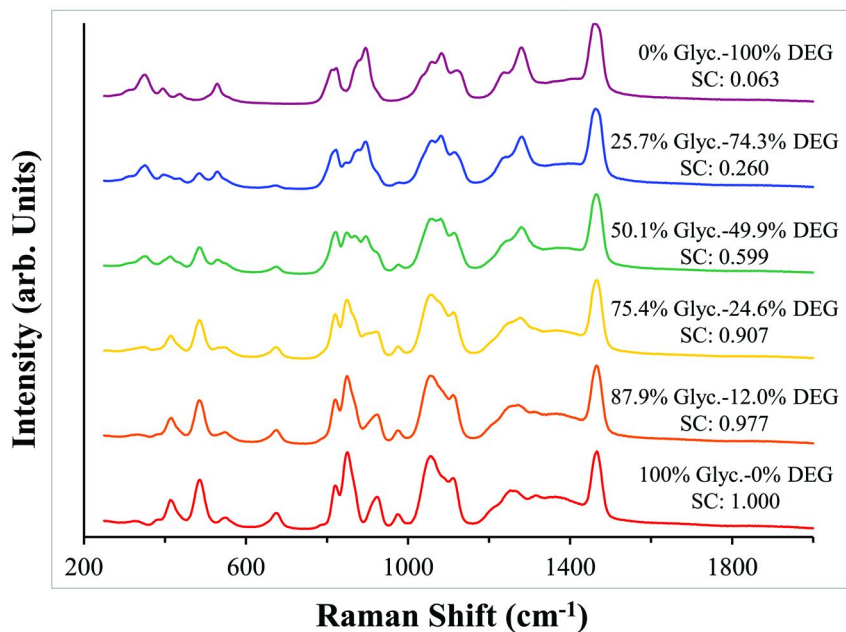
Figure 1. The Raman spectra for a series of pure glycerin samples and spiked samples containing different proportions of DEG. The spectral correlation value decreases as the DEG concentration increases. Glyc, glycerin; DEG, diethylene glycol.
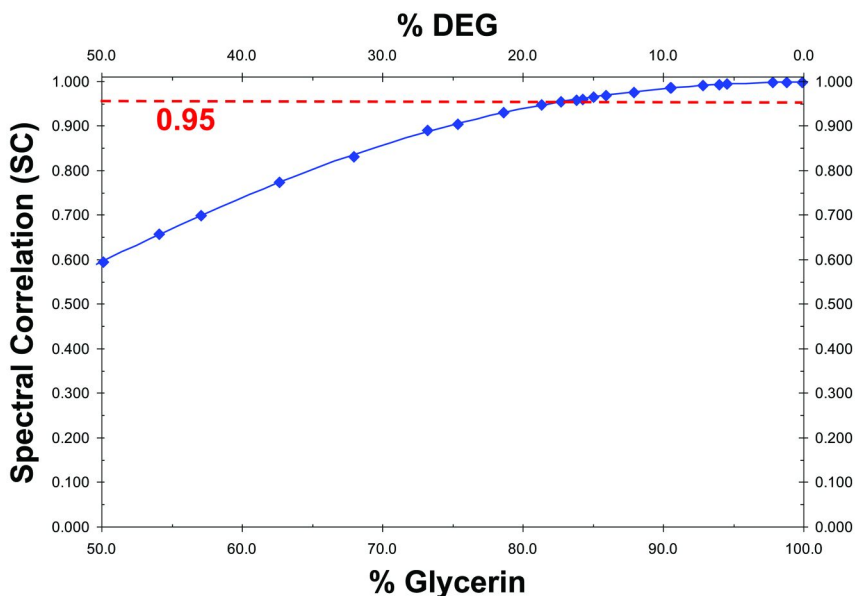
Figure 2. The effect of increasing DEG composition in the Raman spectra for the glycerin–DEG system. The dotted line indicates the 0.95 threshold for a sample to pass the test. The spectral correlation value is expected to drop below the threshold when glycerin samples contain around 18% DEG or more. Adapted with permission from reference (9). Copyright 2011 American Chemical Society. DEG, diethylene glycol.

Raman and NIR techniques may also be implemented for relatively low levels of adulteration with use of multivariate data analysis methods. These types of methods have produced limits of detection less than 2% DEG in propylene glycol by NIR (*15*) and 0.5% DEG in glycerin by Raman spectroscopy (*17*). The multivariate methods used by the DPA involve principal component analysis (PCA) and partial least squares (PLS)-type methods, both of which feature enhanced sensitivities compared with spectral correlation-based methods (*21*).

The Raman spectra of a subset (test set) of glycerin samples containing DEG discussed above were compared with the spectra of certified glycerin, comprising 54 replicate Raman acquisitions from a single source. This comparison is shown in Figure 3. The test set samples, which had DEG compositions of 0–5%, would all be assigned a pass as they lie within the expected variability of the library glycerin spectra. Samples with greater than 5% DEG compositions all lie outside this boundary and would be assigned a fail. The PCA method shows over threefold improvement in sensitivity compared with the SC-based method and, taken collectively with the results of our PLS study (*17*), indicates that Raman and NIR spectroscopies may be used to develop methods with a range of sensitivities.
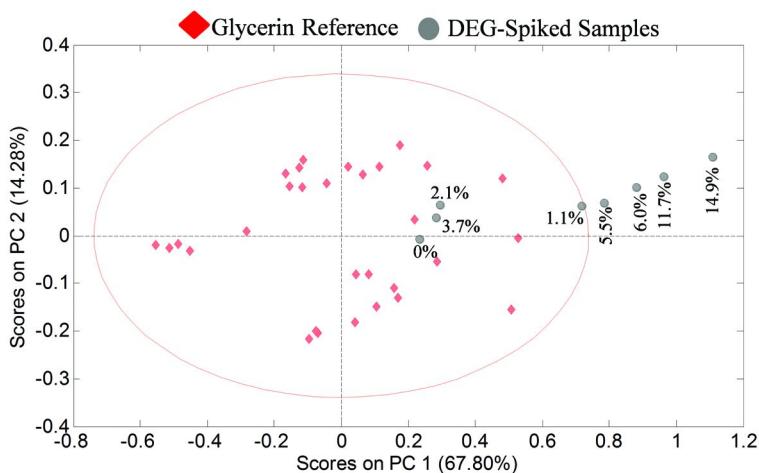
Figure 3. Principal component analysis of the Raman spectra for the glycerin–DEG system. Samples containing more than 5% DEG lie outside the reference area for glycerin values (ellipsis). All spectra were treated by mean centering followed by standard normal variate preprocessing. DEG, diethylene glycol.

Deployment

In the initial field deployment of the Raman instruments in 2012, incoming shipments of glycerin were screened for the presence of DEG. The methods utilized PLS (*17*) and had a limit of detection for DEG of 0.32%. In total, 26 batches of imported glycerin were examined. No DEG was detected in any of the batches tested. Of the 26 total batches screened, nine samples were sent to the lab for confirmatory testing by compendial methods. For the month of July 2012, 15% of large glycerin shipments were physically tested with the Raman methods, a much larger percentage than would have been possible if traditional laboratory testing had been used. None were found to contain adulterants.

Findings and Future Work

The feedback received from the field during the initial Raman deployment was positive, and has been incorporated in current and future plans for Raman and NIR development. The portable instruments allowed the FDA to screen a much larger percentage of containers than would have been possible with only laboratory assessments. To further increase the percentage of material screened, DPA has engaged excipient manufacturers and the International Pharmaceutical Excipients Council in an effort to build a representative library of the most widely used pharmaceutical excipients (*22*). Under this initiative, the spectral library will be used for spectral correlation and multivariate-based analysis. In addition, DPA is building a library of finished medicines used for medical countermeasures against terrorist attacks or pandemic outbreaks (*18*), and another library of

priority APIs. The DPA spectral libraries are being prepared for distribution to field instruments, and will be regularly updated, since libraries are continually evolving. The long-term goal is to use spectral correlation methods as a first screening level to detect any mislabeled incoming finished medicines, APIs and excipients. Multivariate-based methods will then be created for certain high-risk materials and used as a secondary, more targeted screen. Samples failing each of these methods will then be sent to the laboratory for confirmatory analysis.

IMS

Background

IMS is a high-throughput separation method used for detecting and identifying volatile and semi-volatile organic compounds, on the basis of the time required for the ionized species to travel through a drift tube. Upon introduction into the instrument, analyte molecules are vaporized and carried from a heated inlet to the ionization chamber. The volatilized analyte molecules are selectively ionized at ambient pressure by β-radiation from a Ni^{63} source to produce ions. A voltage gradient is applied to the drift region, which causes the swarm of ions to travel down the drift tube towards the detector. A counter current of air keeps the drift region free from neutral analyte molecules to hinder the formation of ion clusters. Molecular ions are detected at the end of the drift region by a Faraday plate detector. Ionic species of different sizes, shapes, charges and collisional cross-sections have different drift velocities and arrive at the detector at different times. The drift velocity of the ions is proportional to the electric field strength, and the proportionality constant is known as the ion mobility (K). Since K is dependent on the buffer gas pressure and temperature, the reduced ion mobility (K_o) can be calculated, correcting for atmospheric pressure and temperature, as follows:

$$K_o = \frac{L}{t_d E} \frac{P}{760} \frac{273.2}{T}$$

where L is the length of the drift tube, t_d is the time required for the ion to drift to the detector, P is the buffer gas pressure in Torr and T=273.3 K. From Equation 2 it is apparent that $K_o \cdot t_d$ is a constant during the measurement of an ion mobility spectrum in a well-controlled instrument. Therefore, an internal calibrant with known ion mobility, $K_{o,C}$, can be used to provide accurate measurement of an analyte's ion mobility, $K_{o,A}$, from its drift time, $t_{d,A}$ and the drift time of the internal calibrant, $t_{d,C}$, with the equation

$$K_{o,A} = K_{o,C} \bullet \frac{t_{d,C}}{t_{d,A}}$$

The internal calibrant is used to correct for variations in instrumental parameters that can occur between measurements made on different days, including variations in P, T and E, and allows reduced ion mobilities to be used for qualitative identification of analytes. The presence of a calibrant in an instrument

also allows methods developed on one portable IMS instrument to be transferred to another without additional method development.

Herbal products have been adulterated with several classes of drugs, including appetite suppressants (sibutramine, rimonabant), diuretics (bumetanide, furosemide), anti-depressants (fluoxetine), laxatives (phenolphthalein) and anti-convulsants (phenytoin). In addition, synthetic analogues of the adulterants (N-desmethyl sibutramine, N-didesmethyl sibutramine and 11-desisobutyl-11-benzyl sibutramine), in which minor modifications were made to the molecular structure of one of the known drug substances, have been observed (23, 25, 26). The consumption of undeclared drugs in dietary supplements can cause adverse health events, including strokes, heart attacks and death.

Sibutramine is one of the most common adulterants found in dietary supplements sold for weight loss. Sibutramine hydrochloride monohydrate was approved by the FDA in 1997 to treat obesity, but was withdrawn from the market in 2010 due to associated cardiovascular risks. While on the market, the product was available by prescription only in dosage strengths of 5 mg, 10 mg or 15 mg sibutramine per capsule. Undeclared sibutramine has been detected in weight loss products at levels of 0.1–40.0 mg per capsule (23), that is, up to around three times more than the formerly approved dose. In the laboratory, dietary supplements are typically analyzed by gas chromatography-mass spectrometry (GC/MS) or high-performance liquid chromatography (HPLC) with detection by ultraviolet, MS or tandem MS (MS/MS). These techniques are not well suited to the rapid screening of adulterants in dietary supplements since the analysis times range from 15–75 min (24, 26–34). Additionally, GC requires an extensive sample preparation procedure in which the sibutramine hydrochloride is converted to the free base form before analysis (26–28).

IMS is an ideal screening tool for the detection of undeclared drugs or adulterants in dietary supplements because of its high speed, selectivity and low detection limits. Portable instruments are commercially available and easy to use. In forensic science and the security industries, portable IMS devices have been used to detect the presence of trace amounts of illicit drugs, explosives and chemical warfare agents (35–41). APIs, such as sibutramine, have complex molecular structures and often contain amine groups. Amines generally provide an intense interference-free response in IMS due to their high proton affinities. Amines protonate readily in the IMS source and the main ionic product in most cases is the protonated molecular ion. The reduced mobility of a protonated molecular ion detected by a portable IMS instrument can, therefore, be used to identify the presence of an undeclared drug or adulterant.

Methods

The DPA rapid screening program has developed an IMS method to screen weight loss products for the presence of undeclared drugs and their analogues (42, 43). Weight loss products can be found in several forms, such as tablets, capsules and powders. A simple extraction procedure is required to separate the API or adulterant from the other ingredients in the product. To prepare individual test

samples, tablets are crushed and the powdered contents of capsules removed, and a quarter of a teaspoon of powder is transferred to a test vial. Isopropyl alcohol or methanol is added to the test vial to dissolve the drug product or adulterant; plant materials found in herbal products are generally not soluble in alcohol. The sample solution is filtered and the supernatant is collected. Because IMS has trace-level sensitivity, the supernatant undergoes a series of dilutions before analysis. The more-dilute sample preparations are analyzed first to prevent saturation of the IMS ionization chamber with high analyte concentrations. A 1 μL sample is deposited onto a substrate and the volatile solvent is allowed to evaporate. The substrate is then introduced into the IMS instrument for analysis. The measurement takes about 30 s and 30–60 s more are required to purge the instrument and prepare it for the next sample analysis. The portable IMS instruments used for field screening are also equipped with a nicotinamide internal calibrant. The calibrant aids in the calculation of the K_o for unknowns analytes.

Alarms can be programmed on the portable instruments using the K_o values for undeclared drugs and adulterants of interest. Analyte reference materials are analyzed on IMS instruments to determine the K_o values. Examples of the ion mobility spectrum for a blank substrate, a solvent blank, and three weight loss adulterant reference standards (sibutramine, desmethyl sibutramine and didesmethyl sibutramine) are shown in Figure 4. Protonated ion peaks for the weight loss reference materials were observed at drift times ranging from 17 ms to 19 ms, corresponding to K_o values ranging from 1.1728 to 1.2170 cm^2/(V•s). The limit of detection for the reference standards was determined as the concentration that resulted in a peak with a signal-to-noise ratio of 3. The limit of detection for sibutramine is 2 ng and for its analogues is 1 ng. The drift time, average K_o, full-width half maximum of the protonated ion peak and the limit of detection for the weight loss reference standards are given in Table I.

Weight loss products analyzed by IMS can be classified as adulterated (alarm) and non-adulterated (pass). Non-adulterated products lack the presence of peaks with mobilities in the range of the undeclared drugs or adulterants and are indicated by a green pass screen. Samples are classified as adulterated when a peak has a K_o value similar to the ion mobilities of the undeclared drugs or adulterants, and a red alarm screen is displayed. If an alarm is not triggered during the analysis of the most dilute sample preparation, then the next concentrated sample is analyzed. This process continues until all dilutions and the supernatant have been analyzed.
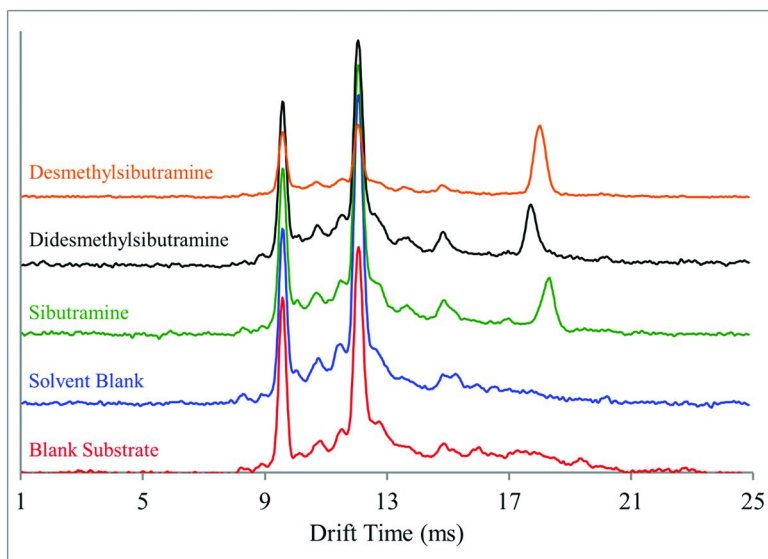
Figure 4. Ion mobility spectra of weight loss reference standards in alcohol. A blank substrate and an alcohol solvent blank are also shown. The spectra have been offset for clarity. The peaks observed at 9.6 ms and 12.1 ms in all ion mobility spectra were attributed to air and the instrument's internal nicotinamide calibrant, respectively.

Table I. Ion-mobility spectrometry characteristics of weight loss reference standards[a]

Reference standard	Drift time (ms)	K_o (cm²/(V•s))	FWHM (μs)	Limit of detection (ng)
Desmethyl sibutramine	18.0	1.1898	440	1
Didesmethyl sibutramine	17.7	1.2100	420	1
Sibutramine	18.3	1.1706	430	2

[a] K_o, reduced ion mobility; FWHM, full-width half maximum of the protonated ion peak.

Deployment

Multiple portable IMS instruments are currently deployed at international mail facilities in the United States (NY, IL, CA, FL and HI) to screen weight loss products for the presence of the undeclared drug sibutramine and its analogues. Samples failing field screening are quarantined by FDA investigators and a portion of the sample is sent to the laboratory for confirmatory analysis. As a control, one of every 10 samples that does not trigger an alarm and is classified as a pass during field screening is sent to the laboratory for additional analysis. If the laboratory results confirm the presence of sibutramine or its analogues in a sample, the products are collected and destroyed.

Table II displays the results of IMS analysis of five products collected and screened at an international mail facility located in New York. The products were tested on site for the presence of sibutramine. The resulting ion mobility spectra for the weight loss products are shown in Figure 5. Samples S-1 and S-2 did not trigger an alarm or exhibit peaks at drift times or reduced ion mobilities similar to those of the weight loss standards, whereas samples S-3, S-4 and S-5 tested positive for the presence of a weight loss adulterant. The five samples were sent to the laboratory for confirmatory analysis and the results from the laboratory were consistent with the IMS field results. Weight loss products S-3, S-4 and S-5 were denied access into the United States and were destroyed.
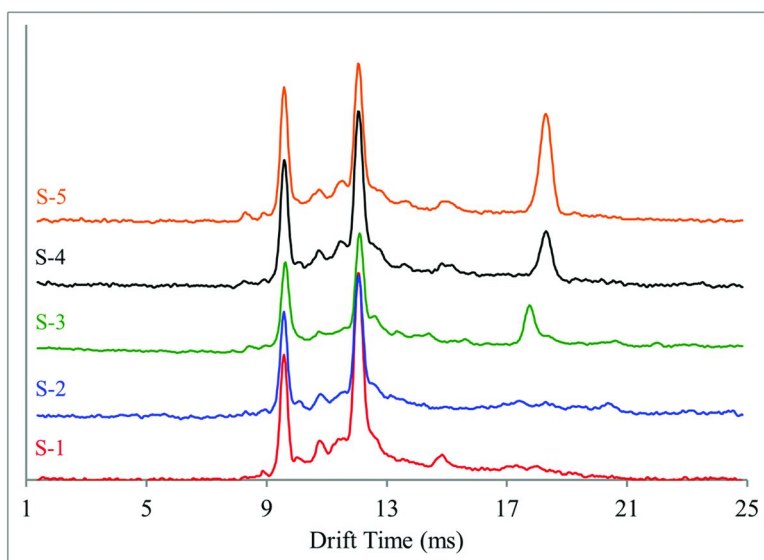
Figure 5. Ion mobility spectra of a dietary supplement marketed for weight loss. Spectra are offset for clarity.

Table II. Ion-mobility spectrometry results for weight loss products screened in a mail facility[a]

Sample	Drift time (ms)	K_o (cm²/(V•s)	Alarm
S-1	–	–	None
S-2	–	–	None
S-3	17.8	1.2123	Didesmethyl sibutramine
S-4	18.3	1.1720	Sibutramine
S-5	18.3	1.1720	Sibutramine

[a] K_o, reduced ion mobility.

Findings and Future Work

In a field investigation, the FDA utilized portable IMS instruments to analyze 225 weight loss products (*3*). Of those, 42 triggered alarms on the IMS instruments and were sent to the laboratory for confirmatory analysis. Twenty three control samples that passed IMS testing were also sent to the laboratory for additional analysis. All 42 samples that had failed IMS field screening were confirmed by the laboratory to contain sibutramine, and five contained other undeclared weight loss adulterants. The 23 passed samples were found to contain no sibutramine, which confirmed the field results. The 42 adulterated weight loss products were destroyed. Additional alarms have been added to the portable instruments for other weight loss adulterants (*43*) and other classes of adulterants. Methods are also being developed to verify the presence of an expected API in a product to provide an additional tool to screen for counterfeit drugs.

XRF

Background

XRF spectroscopy is a well-established analytical technique for elemental analysis in a large variety of matrices that can detect trace to high levels. Portable XRF spectrometers have been widely used to evaluate the composition of alloys, artwork, artifacts and environmental samples (*44*). However, the determination of organic content by XRF is still considered difficult because X-ray cross-sections for light elements (Z<11) are very low. Sensitivity with conventional XRF generally increases with increasing atomic number due to the increase in fluorescence yield with increasing element atomic number (*45*).

XRF spectrometers can measure the elemental content of a sample non-destructively, with little or no sample preparation. Additionally, portable instruments are designed to be safe for field use and several models are available with X-ray tubes that only generate X-rays when the instrument is activated for measurement. The instruments direct gamma X-rays onto the sample and

collect the fluorescent X-rays that are subsequently emitted from the heavy elements. Spectral integration times range from 1 min to 10 min and, therefore, XRF spectrometry is suitable for rapid screening. Much like the techniques discussed in the previous sections, the energy profiles of the fluorescent X-rays are characteristic of the elements from which they emanate and can be used as fingerprints to characterize the elements in a sample. One key attribute of this technique is that XRF spectra are not influenced by the sample matrix and typically have narrow peak widths. Thus, XRF spectroscopy provides a selective method to detect metal impurities in a variety of samples matrices.

One class of adulterants that can be screened for with XRF is toxic metals and metalloids that may be present in pharmaceutical products as a result of insufficient purification of raw materials or contaminated processing equipment (*46*). The toxic metals and metalloids amenable to screening include cadmium, arsenic, lead and mercury, as well as common catalytic metals, such as platinum and palladium. Data on metal impurities in pharmaceutical products are sparse, and the extent of metal contamination in pharmaceuticals is poorly documented (*10, 47–61*) Thus, it is important to conduct screening of elemental impurities in pharmaceutical materials, both to detect adulterated/contaminated products and to obtain a better understanding of baseline levels of toxic metals and metalloids that are present.

Methods

The simplest way to determine the presence of toxic metals and metalloids by XRF is to detect peaks at element-specific energies. Algorithms involving smoothing and baseline subtraction followed by peak detection with peak-intensity threshold criteria can be applied if the background is known and stable (*44, 62, 63*). However, these methods are unsuitable for pharmaceutical surveillance due to matrix dependence effects on the XRF background. To overcome this challenge, we introduced a new method for analysis of XRF spectra based on continuous wavelet transform filters (*64*). The method uses a continuous wavelet transform to filter the signal and noise components of the spectrum. The general idea of the wavelet transform is to represent the signal of interest, s(t), as a linear combination of wavelets $\psi_{a,b}(t)$ whose coefficients are computed according to the following equation:

$$C(a,b) = \int_R s(t)\psi_{a,b}(t)dt$$

The wavelet transformation converts a one-dimensional signal into a two-dimensional wavelet space defined by the scale and the translation variables through the coefficients. The functional form of the mother wavelet defines the wavelet shape, and the scale defines the wavelet width. The limit test compares the wavelet domain signal-to-noise ratios at the energies of the elements of interest to an empirically determined signal-to-noise decision threshold.

To screen drug products and pharmaceutical materials for the presence of toxic and catalytic metals, DPA collected XRF spectra with the manufacturer's

software for analysis with software developed in house. Mexican hat wavelet filters have been applied to the analysis of XRF spectra with emphasis on identification of samples containing toxic elements. The wavelet signal-to-noise algorithm reads the XRF spectrum and then calculates two wavelet coefficients at each spectral energy (i.e. the translation increment is one channel) with scales equal to one and four channels (C[1] and C[4] coefficients) for noise and signal, respectively. The algorithm searches the signal coefficient spectrum for the local maxima within a seven-channel window centered on the known energies of the elements of interest and then calculates the element-specific signal-to-noise ratios as the ratio of C(4) to the average of 11 C(1) values centered on the local C(4) peak. Figure 6 shows the Mexican hat wavelets superimposed on an XRF spectrum for a pharmaceutical tablet containing lead. If the calculated signal-to-noise ratio is larger than a pre-determined threshold value, an undesirable element of interest is present in pharmaceutical materials, and the program alerts the operator of the failed result.

Deployment

The method was tested on spiked tablets in a collaborative study that involved six FDA laboratories in St. Louis, MO, Jamaica, NY, Philadelphia, PA, Atlanta, GA, Detroit, MI and Los Angeles, CA. In total 1,241 tablets and capsules were measured on five different instruments by different analysts, and a prediction was made with the wavelet signal-to-noise method after each measurement. The limit test employed here required only a 90 s accumulation time. The detection limits estimated for arsenic, lead, mercury and chromium were, respectively, 8, 14, 20 and 150 µg/g. Longer accumulation times may lead to reductions in the empirical signal-to-noise thresholds and, therefore, lower detection limits.

Findings and Future Work

The FDA has used XRF to examine the metal content in cosmetics, herbal products and dietary supplements that are suspected of containing high levels of toxic metals. In a recent field investigation, FDA used portable XRF instruments to measure 240 dietary supplements and found 17 suspicious samples. These samples were collected and sent to a laboratory for confirmatory analysis. All 17 suspicious samples were confirmed to contain lead, mercury, arsenic or chromium. In some cases, concentrations of toxic metals as low as 3 ppm were detected. Sixteen additional samples that were not considered suspicious following XRF analysis were also collected to validate the field analysis. Three of these samples were found to have low levels of toxic metals below 3 ppm, and seven were found to have no toxic metals. With these instruments being stationed both domestically and internationally, portable XRF can be used to screen toxic and catalytic metals in pharmaceutical raw materials, cosmetics and herbal products.
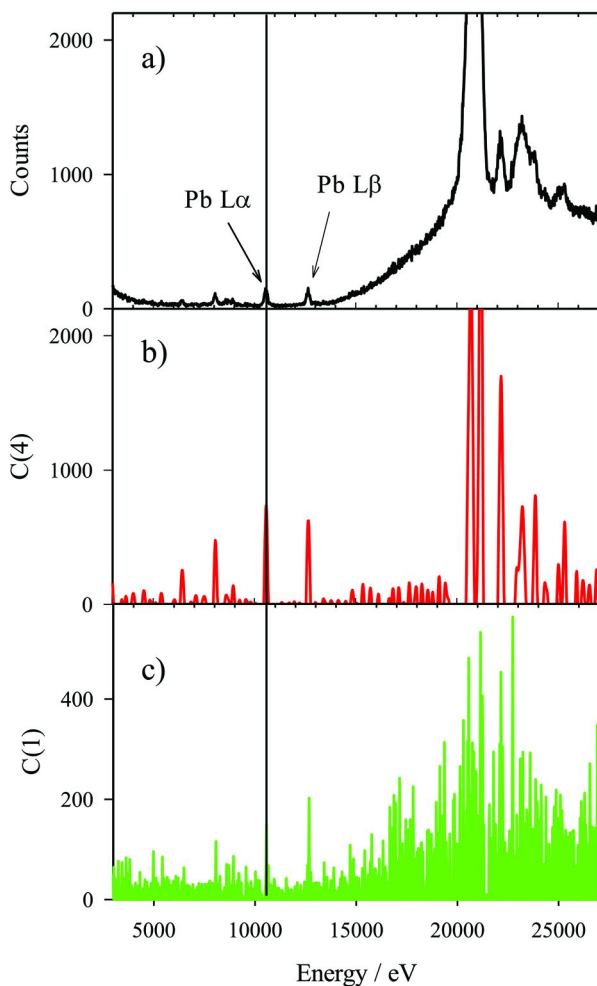
Figure 6. (a) Raw X-ray fluorescence spectrum of a pharmaceutical tablet spiked with 25 ppm lead. (b) The C(4) wavelet coefficient spectrum. (c) The C(1) wavelet coefficient spectrum. The C(4) coefficients are used to characterize the signal, because the width of the a=4 wavelet (8 channels between zero crossings) is close to the half width of the X-ray fluorescence instrument spectral bandpass.

Conclusions

The rapid screening methods described in this chapter are important tools that help the FDA to monitor supply chain integrity and assure the availability of safe and effective drugs. Major ongoing initiatives built around the portable technologies described in this chapter will continue to promote public safety by enhancing the FDA's ability to perform screening on an increasing number of products before they reach consumers. This would not be possible without

the technological advancements that have led to innovation in field-deployable instrumentation. This instrumentation allows the FDA to cast a wider surveillance net on imported pharmaceuticals coming into the United States and direct resources to the products that require the most attention. As technology continues to improve, the rapid screening program can continue to refresh its fleet of field instrumentation with the latest available products and tailor new rapid screening methods to areas of most need. Continued efforts will be dedicated to improving and streamlining method transfer to facilitate distribution of multivariate methods and spectral libraries to instruments from different vintages and vendors. The FDA will continue to develop broadly applicable methods that incorporate algorithms to increase sensitivity and specificity towards adulterants.

Acknowledgments

Disclaimer

The findings and conclusions in this article have not been formally disseminated by the United States Food and Drug Administration and should not be construed to represent any Agency determination or policy.

References

1. United States Food and Drug Administration. Global engagement. http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM298578.pdf (accessed December, 2013).
2. United States Food and Drug Administration Consumer Health Information. Ensuring the safety of imported products. http://www.fda.gov/downloads/ForConsumers/ConsumerUpdates/UCM229925.pdf (accessed December, 2013).
3. Kauffman, J. F.; Rodriguez, J. D.; Gryniewicz-Ruzicka, C. M.; Arzhantsev, S.; D'Sa, A.; Uratani, B.; Wolfgang, S.; Westenberger, B. J.; Buhse, L. F.; Dunn, J. D.; Mecker-Pogue, L. C. *Bio. Pharma. Asia* **2013**, *1*, 28–37.

4. World Health Organization. Medicines: spurious/falsely-labelled/falsified/counterfeit (SFFC) medicines. http://www.who.int/mediacentre/factsheets/fs275/en/ (accessed December, 2013).

5. Dégardin, K.; Roggo, Y.; Margot, P. *J. Pharm. Biomed. Anal.* **2014**, *87*, 167–175.

6. Brustkern, A. M.; Buhse, L. F.; Nasr, M.; Al-Hakim, A.; Keire, D. A. *Anal. Chem.* **2010**, *82*, 9865–9870.

7. Mecker, L. C.; Tyner, K. M.; Kauffman, J. F.; Arzhantsev, S.; Mans, D. J.; Gryniewicz-Ruzicka, C. M. *Anal. Chim. Acta* **2012**, *733*, 48–55.

8. Xiu, C.; Klein, K. K. *Food Policy* **2010**, *35*, 463–470.

9. Rodriguez, J. D.; Westenberger, B. J.; Buhse, L. F.; Kauffman, J. F. *Anal. Chem.* **2011**, *83*, 4061–4067.

10. Saper, R. B.; Kales, S. N.; Paquin, J.; Burns, M. J.; Eisenberg, D. M.; Davis, R. B.; Phillips, R. S. *JAMA J. Am. Med. Assoc.* **2004**, *292*, 2868–2873.

11. *Dietary Supplement Health and Education Act of 1994: Public Law No: 103-417*; Library of Congress: Washington, DC, 1994.

12. McCreery, R. L. *Raman Spectroscopy for Chemical Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, 2001.

13. Cen, H.; He, Y. *Trends Food Sci. Technol.* **2007**, *18*, 72–83.

14. Sorak, D.; Herberholz, L.; Iwascek, S.; Altinpinar, S.; Pfeifer, F.; Siesler, H. W. *Appl. Spectrosc. Rev.* **2011**, *47*, 83–115.

15. Li, X.; Arzhantsev, S.; Kauffman, J. F.; Spencer, J. A. *J. Pharm. Biomed. Anal.* **2011**, *5*, 1001–1006.

16. Zhang, X. F.; Zou, M. Q.; Qi, X. H.; Liu, F.; Zhang, C.; Yin, F. *J. Raman Spectrosc.* **2011**, *42*, 1784–1788.

17. Gryniewicz-Ruzicka, C. M.; Arzhantsev, S.; Pelster, L. N.; Westenberger, B. J.; Buhse, L. F.; Kauffman, J. F. *Appl. Spectrosc.* **2011**, *65*, 334–341.

18. Rodriguez, J. D.; Gryniewicz-Ruzicka, C. M.; Kauffman, J. F.; Arzhantsev, S.; Saettele, A. L.; Berry, K. A.; Westenberger, B. J.; Buhse, L. F. *Am. Pharm. Rev.* **2013**, *16*, 9–18.

19. Rodriguez, J. D.; Westenberger, B. J.; Buhse, L. F.; Kauffman, J. F. *Analyst* **2011**, *136*, 4232–4240.

20. Kauffman, J. F.; Rodriguez, J. D.; Buhse, L. F. *Am. Pharm. Review* **2011**, *14*, 34–40.

21. Gryniewicz-Ruzicka, C. M.; Rodriguez, J. D.; Arzhantsev, S.; Buhse, L. F.; Kauffman, J. F. *J. Pharm. Biomed. Anal.* **2012**, *61*, 191–198.

22. Merrell, P. H.; Buhse, L.; Rodriguez, J. D. *Tablets Capsules* **2013**, *11*, 22–26.

23. United States Food and Drug Administration Consumer Updates. More weight loss products added to consumer alert. http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm103184.htm (accessed August, 2010).

24. *Screening for Sibutramine and Sibutramine Analogs in Weight Loss Supplements by Gas Chromatography-Mass Spectrometry*; U.S. Department of Health and Human Services, FDA.

25. Mans, D. J.; Gucinski, A. C.; Dunn, J. D.; Gryniewicz-Ruzicka, C. M.; Mecker-Pogue, L. C.; Kao, J. L. F.; Ge, X. *J. Pharm. Biomed. Anal.* **2013**, *83*, 122–128.

26. Blachut, D.; Siwinska-Ziolkowska, A.; Kobylecka, A.; Czarnicki, Z. *Z. Zagadnien Nauk Sadowych* **2006**, *66*, 199–211.

27. Blachut, D.; Siwinska-Ziolkowska, A.; Szukalksi, B.; Wojtasiewicz, K.; Czarnicki, Z.; Kobylecka, A.; Bykas-Strekowska, M. *Z Zagadnien Nauk Sadowych* **2007**, *LXX*, 225–235.

28. Blackledge, R. D.; Sorenson, P. D. S. *Microgram* **2000**, *33*, 18–20.

29. Bogusz, M. J.; Hassan, H.; Ai-Enazi, E.; Ibrahim, Z.; Al-Tufail, M. *J. Pharm. Biomed. Anal.* **2006**, *41*, 554–564.

30. Chen, Y.; Zhao, L.; Lu, F.; Yu, Y.; Chai, Y.; Wu, Y. *Food Addit. Contam., Part A* **2009**, *26*, 595–603.

31. Huang, Z. Q.; Xiao, S.; Luo, D.; Chen, B.; Yao, S. Z. *J. Chromatogr. Sci.* **2008**, *46*, 707–711.

32. Jung, J.; Hermanns-Clausen, M.; Weinmann, W. *Forensic Sci. Int.* **2006**, *161*, 221–222.

33. Wang, J.; Chen, B.; Yao, S. *Food Addit. Contam., Part A* **2008**, *25*, 822–830.

34. Zou, P.; Oh, S. S. Y.; Kiang, K. H.; Low, M. Y.; Bloodworth, B. C. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 614–618.

35. Babis, J. S.; Sperline, R. P.; Knight, A. K.; Jones, D. A.; Gresham, C. A.; Denton, M. B. *Anal. Bioanal. Chem.* **2009**, *395*, 411–419.

36. Eiceman, G. A.; Karpas, Z. *Ion Mobility Spectrometry*; CRC Press Taylor & Francis: Boca Raton, FL, 2005.

37. Keller, T.; Keller, A.; Tutsch-Bauer, E.; Monticelli, F. *Forensic Sci. Int.* **2006**, *161*, 130–140.

38. Keller, T.; Miki, A.; Regenscheit, P.; Dirnhofer, R.; Schneider, A.; Tsuchihashi, H. *Forensic Sci. Int.* **1998**, *94*, 55–63.

39. Liu, X. Y.; Nacson, S.; Grigoriev, A.; Lynds, P.; Pawliszyn, J. *Anal. Chim. Acta* **2006**, *559*, 159–165.

40. O'Donnell, R. M.; Sun, X. B.; Harrington, P. D. *TrAC, Trends Anal. Chem.* **2008**, *27*, 44–53.

41. Zimmermann, S.; Barth, S.; Baether, W. K. M.; Ringer, J. *Anal. Chem.* **2008**, *80*, 6671–6676.

42. Dunn, J. D.; Gryniewicz-Ruzicka, C. M.; Kauffman, J. F.; Westenberger, B. J.; Buhse, L. F. *J. Pharm. Biom. Anal.* **2011**, *54*, 469–474.

43. Dunn, J. D.; Gryniewicz-Ruzicka, C. M.; Mans, D. J.; Mecker-Pogue, L. C.; Kauffman, J. F.; Westenberger, B. J.; Buhse, L. F. *J. Pharm. Biomed. Anal.* **2012**, *71*, 18–26.

44. Pessanha, S.; Guilherme, A.; Carvalho, M. L. *Appl. Phys. A* **2009**, *97*, 497–505.

45. Bortoleto, G. G.; Pataca, L. C. M.; Bueno, M. I. M. S. *Anal. Chim. Acta* **2005**, *539*, 283–287.

46. Kauffman, J. F.; Westenberger, B. J.; Robertson, J. D.; Guthrie, J.; Jacobs, A.; Cummins, S. K. *Regul. Toxicol. Pharmacol.* **2007**, *48*, 128–134.

47. Bhatti, H. B. A.; Gogtay, N. J.; Kochar, N. R.; Daivi, S. S.; Kshirsagar, N. A. *Adverse Drug React. Bull.* **2004**, *225*, 1–4.

48. Bourgoin, B. P.; Evans, D. R. *Am. J. Public Health* **1993**, *83*, 1155–1160.

49. Caldas, E. D.; Machado, L. L. *Food Chem. Toxicol.* **2004**, *42*, 599–603.

50. Chuang, I. C.; Chen, K. S.; Huang, Y. L.; Lee, P. N.; Lin, T. H. *Biol. Trace Elem. Res.* **2000**, *76*, 235–244.

51. Ernst, E.; Thompson Coon, J. *Clin. Pharmacol. Ther.* **2001**, *70*, 497–504.

52. Fung, H. T.; Fung, C. W.; Kam, C. W. *Emerg. Med. (Fremantle)* **2003**, *15*, 518–520.

53. Margui, E.; Van Meel, K.; Van Grieken, R.; Buendi, A.; Fonta, C.; Hidalgo, M.; Queralt, I. *Anal. Chem.* **2009**, *81*, 1404–1410.

54. Moore, C.; Adler, R. *Pediatrics* **2000**, *106*, 600–602.

55. Prpic-Majic, D.; Pizent, A.; Jurasovic, J.; Pongracic, J.; Restek-Samarzija, N. *J. Toxicol. Clin. Toxicol.* **1996**, *34*, 417–423.

56. Ross, E. A.; Szabo, N. J.; Tebbett, I. R. *JAMA: J. Am. Med. Assoc.* **2000**, *284*, 1425–1429.

57. Saper, R. B.; Phillips, R. S.; Sehgal, A.; Khouri, N.; Davis, R. B.; Paquin, J.; Thuppil, V.; Kales, S. N. *JAMA: J. Am. Med. Assoc.* **2008**, *300*, 915–923.

58. Scelfo, G. M.; Flegal, A. R. *Environ. Health Perspect.* **2000**, *108*, 309–313.

59. Spriewald, B. M.; Rascu, A.; Schaller, K. H.; Angerer, J.; Kalden, J. R.; Harrer, T. *Occup. Environ. Med.* **1999**, *56*, 282–283.

60. Woolf, A. D.; Hussain, J.; McCullough, L.; Petranonovic, M.; Chomchai, C. *Clin. Toxicol.* **2008**, *46*, 841–844.

61. Palmer, P. T.; Jacobs, R.; Baker, P. E.; Ferguson, K.; Webber, S. *J. Agric. Food Chem.* **2009**, *57*, 2605–2613.

62. Mittermayr, C. R.; Tan, H. W.; Brown, S. D. *Appl. Spectrosc.* **2001**, *55*, 827–833.

63. Morgenstern, P.; Bruggemann, L.; Wennrich, R. *Spectrochim. Acta, Part B* **2005**, *60*, 1373–1379.

64. Arzhantsev, S.; Li, X. A.; Kauffman, J. F. *Anal. Chem.* **2011**, *83*, 1061–1068.

Subject Index