

Enhancing FDA's Evaluation of Science to Ensure Chemicals Added to Human Food Are Safe: Workshop Proceedings

Maricel V. Maffini, Heather M. Alger, Erin D. Bongard, and Thomas G. Neltner

Abstract: Science and expert judgment are the foundation for safety assessments of chemicals added to food to ensure their use is safe. Hazard characterization is the first step in a safety assessment. Advances in science and technology pose challenges to the regulatory system and raise questions about whether the current hazard identification and characterization process is able to systematically and transparently encompass such advances while remaining defensible. An April 2011 workshop sponsored by The Pew Charitable Trusts, the Institute of Food Technologists, and the journal *Nature* brought together over 80 experts in science and food policy from government, industry, academia, and public interest organizations to examine the principles underlying the development and use of scientific evidence needed for chemical hazard characterization. Participants discussed challenges of identifying adverse health effects, advances in science, uses of new screening technologies and human biomonitoring data, updating of study designs, and development and review of toxicity test guidelines. Brainstorming sessions allowed participants to propose alternatives to enhance FDA's evaluation of science for safety assessment. Although there was no intention to reach a consensus, several themes emerged including the need for clear procedures to develop validated toxicity tests; importance of regularly updating guidance documents relied upon by regulators and industry; benefits of transparency and public access to information; potential for greater interagency collaboration; opportunities to improve hypothesis-based research to make it more useful to regulatory decision making; and importance of staying abreast of scientific developments to ensure that safety assessments are made using sensitive and relevant methods.

Introduction

Since the passage of the 1958 Food Additives Amendment to the Federal Food, Drug, and Cosmetic Act (Public Law 85-829, 72 Stat. 1784), the Food and Drug Administration (FDA) has developed and implemented a complex regulatory system intended to help ensure that chemicals added directly or indirectly to food are safe. The 1958 law and subsequent legislation established a number of categories of additives (Table 1) with specific requirements for each (reviewed in Neltner and others 2011).

FDA must give premarket approval for all chemical uses defined as food additives (including food contact substances that are reasonably expected to migrate to food) and color additives. The Food Additives Amendment excludes from the definition of "food additive" chemicals already expressly approved by FDA or the U.S. Dept. of Agriculture for use in food before 1958 ("prior-sanctioned substances") and chemicals determined by a food manufacturer, additive supplier, or FDA to be "generally recognized as safe" (GRAS) that are not pesticides, color additives, or animal drugs. There are other exclusions from the statutory

definition of "food additive," such as animal drugs, dietary supplements, and certain pesticides, which were not discussed at the workshop summarized in this article. Table 1 lists the different categories and subcategories of substances added to food. Despite these technical definitions, for the purpose of simplicity the term "food additives" is used in a broader sense of "substances added to food" throughout this article, and we do not restrict the term to the legal definition of food additive.

As the food additive regulatory program developed over time, Toxicology as a discipline grew into a large and sophisticated field of science for assessing the potential impact of chemical exposure on human health. In response to concerns raised by fraudulent practices at some private contract testing laboratories and to improve the transparency and reproducibility of results, FDA adopted a Good Laboratory Practice (GLP) rule in 1978 setting standards for conducting nonclinical laboratory studies such as animal studies and *in vitro* (for example, cell- or bacteria-based) experiments that support or are intended to support applications for research or marketing permits for FDA-regulated products (21 CFR §§58.1–58.219). In 1982, FDA published *Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food*, also known as the "Redbook," to describe how existing information is considered, the criteria used to assess the need for additional studies, and minimum acceptable protocols

MS 20112001. Authors are with The Pew Charitable Trusts, 901 E St., NW Washington, DC 20004, U.S.A. Direct inquiries to author Maffini (E-mail: mmaffini@pewtrusts.org).

Table 1—Categories of substances added to food established by the Food Additives Amendment of 1958 and subsequent legislation (adapted from Neltner and others [2011]).

1. Food additives
A. Direct food additives
B. Indirect food additives
C. Substances covered by food contact substance (FCS) notifications
D. FCSs below threshold of regulation
E. Radiation sources
2. GRAS substances
A. Common food ingredients in use before 1958
B. Manufacturer self-determined
C. Association expert panel-determined
D. FDA-listed
E. FDA-affirmed
F. Substances covered by FDA-reviewed GRAS notification
3. Prior-sanctioned substances
4. Color additives
5. Pesticide chemicals or residues
6. Drugs in animal feed
7. Dietary supplements

for commonly used toxicological studies. A major revision of the Redbook in 1993 reevaluated, updated, and revised the study protocols for assessment of food additive and color additive safety. In 2000, the Redbook became available on FDA's website, allowing chapters to be updated electronically and its title was shortened to *Toxicological Principles for the Safety Assessment of Food Ingredients* (FDA 2000), thus clarifying that the guidance applies to the broad range of substances added to or present in food, not just substances that meet the legal definitions of food additives and color additives. In addition, FDA and industry rely on other forms of guidance, including guidelines developed by international organizations such as the Organisation for Economic Co-Operation and Development (OECD) (OECD 2010a) and the World Health Organization (WHO 2011), in making safety assessments of food additives.

The Current Challenge

The 1983 National Research Council report *Risk assessment in the federal government: Managing the process* (Press 1983) defines *risk assessment* to mean the characterization of the potential adverse health effects of human exposures to environmental hazards. Risk assessment includes several quantitative and qualitative elements: description of the potential adverse health effects based on the evaluation of results of epidemiologic, clinical, toxicologic, and environmental research; extrapolation from those results to predict the type and estimate the extent of health effects in humans under given conditions of exposure; judgments as to the number and characteristics of persons exposed at various intensities and durations; and summary judgments on the existence and overall magnitude of the public-health problem. Risk assessment also includes characterization of the uncertainties inherent in the process of inferring risk.

FDA's premarket approval process consists of a safety assessment that is similar to a risk assessment and a safety management decision designed to judge whether the exposure associated with a proposed use can be deemed safe (but not necessarily risk-free), with an acceptable degree of confidence, according to the governing statutes and regulations (for example, what conditions of exposure are considered). A safety assessment consists of 2 essential parts: (1) hazard identification and characterization; and (2) exposure assessment. Although both parts are equally important, the workshop summarized here focused only on haz-

ard identification. The Pew Health Group plans to organize a similar workshop to discuss exposure assessment in the fall of 2011.

The identification of the potential hazard posed by a substance and the potential health effects associated with it relies on scientific information. Scientific research advances at a rapid pace and new understanding of the meaning of emerging information and its implications for human risk may pose challenges for existing regulatory programs. Historically, as the understanding of new scientific findings has developed, regulatory programs have incorporated them into the process of making regulatory decisions.

Highly sensitive analytical tools now allow investigators to measure very small amounts of chemicals from many sources (for example, air, water, bodily fluids), and to analyze and quantify increasingly subtle events inside single cells, such as gene expression levels. New high-throughput cell-based *in vitro* screening methods and computer-based modeling of chemical-cell interactions are beginning to produce immense quantities of data on the actual and potential biological effects of compounds. Researchers are identifying new biological endpoints related to human health, including gene expression, biochemical pathways, and hormone-receptor interactions. In addition, there is increased awareness of different susceptibilities to harm depending on individuals' life stage (for example, pregnant women, infants, children, and the elderly).

Research shows that many common diseases and disorders are the result of multiple events and emanate from multiple pathways. For instance, new research is beginning to show that obesity may result from the integration of several factors, such as genetic susceptibility and hormone disruption, along with those conventionally thought to cause the disease such as high-calorie diets and inadequate physical activity relative to caloric intake (Koplan 2005; Levi and others 2010). Interindividual or intergroup variability can be important, and recent studies suggest that some chemicals could potentially have negative human health impacts at very low doses. Changes in certain biological activity can now be routinely measured at the level of individual cells or subcellular components exposed to chemicals in *in vitro* test systems. The potential human health impacts of such biological responses cannot be fully understood without also understanding the innate compensatory adaptation and repair mechanisms that operate in the body; that compensation and repair may be diminished or lacking in developing individuals and not fully functional in susceptible populations, and that exposure to chemicals from food and other sources occurs throughout life. Some of these considerations have long been incorporated into regulatory risk assessments when data are available; however, emerging research on toxicity pathways and biological mechanisms holds the promise of considerable refinement to existing safety evaluation and risk assessment approaches in the near future.

Contrasting with the continuous output of research and scientific development, the regulatory system has its own pace. A recent controversy over FDA's assessment of emerging science for a substance it approved in the early 1960s, bisphenol A (BPA), highlighted issues that are likely to continue to arise. In Myers and others (2009), several academic researchers maintained that studies showing that BPA "interferes" with the endocrine system provided sufficient evidence to question its safety and called on FDA to consider relevant research using state-of-the-art techniques in its safety evaluations of chemicals. These scientists said that FDA relied more heavily on GLP studies using standardized protocols consistent with the Redbook than on peer-reviewed studies published in scientific journals. A series of responses to that publication noted

that several weight-of-evidence assessments had not supported the hypothesis that low-dose oral exposure to BPA adversely affects reproductive or developmental health (Becker and others 2009; Tyl 2009). These responses, from a commercial laboratory scientist and representatives of various chemical manufacturers, also noted the importance of GLP regulations in establishing quality standards, assurance, transparency, and reproducibility of studies. A subsequent editorial in *Nature* (Nature 2010) called on scientists who develop cutting-edge biological techniques to put a high priority on validating and standardizing these techniques in ways that make the results usable by regulators. The editorial also called on regulators to find faster ways to get the new techniques incorporated into guideline-based studies (for example, Redbook guidance).

This recent controversy underscores important distinctions between premarket and postmarket safety evaluation and the types of data regulators can access for the former case versus the latter. In an initial premarket safety assessment, only a few guideline-based studies may be available on which to base the initial decision. The Redbook recommends minimum studies based on expected toxicity and exposure. If the toxicity and the exposure are low, a few studies may be required. In addition, FDA does not require human studies. In a postmarket situation where FDA adopted a regulation approving the additive's use, as currently implemented by FDA, the burden that was once on the manufacturer to prove safety is now on the agency when it considers whether new research warrants withdrawing the approval. Nevertheless, although the types of data available to regulators premarket and postmarket may be different, the statutory safety requirement of reasonable certainty of no harm applies to both situations, and the food manufacturer always has the responsibility, under the food law, to produce a product that meets the safety standard.

Goals of the Workshop

Against this backdrop, the Pew Health Group of The Pew Charitable Trusts convened a workshop on April 5 to 6, 2011, entitled "Enhancing FDA's Evaluation of Science to Ensure Chemicals Added to Human Food Are Safe." The Institute of Food Technologists (IFT) and *Nature* agreed to cosponsor the event. FDA and the National Inst. of Environmental Health Sciences (NIEHS) provided essential planning support. The workshop brought together more than 80 experienced scientists and policymakers, including representatives from industry, government, academia, and public interest organizations; the public interest representatives were the smallest contingent. The workshop was not convened to reach a consensus or dwell on controversies involving specific chemicals. Rather, it sought to develop a shared understanding of the current system FDA uses to assess the hazards of chemicals added to human food, and explore opportunities to strengthen that system, while contributing to FDA's *Advancing Regulatory Science Initiative* (FDA 2010b). Subsequent workshops will focus on exposure assessment and risk assessment. To help make the discussions more focused and productive, the workshop centered on the evaluation of potential human health hazards posed by chemicals added directly or indirectly to human food and did not consider pet food or animal feed, animal drugs, pesticide residues, contaminants, or environmental impacts.

This article summarizes the presentations made at the workshop and the discussions that occurred during small group sessions focused on 4 specific questions in the context of FDA's assessment of a chemical's hazards:

- What are the considerations in identifying and validating adverse effects?
- What are the best methods to evaluate study designs and data for regulatory decisions?
- How should validation studies be developed and test guidelines be reviewed?
- What problems have been identified with the current regulatory process and what potential solutions should be considered?

The article does not represent a consensus of the workshop participants or planners; nor does it represent the views of individual participants. Instead, it presents background information on FDA's evaluation of scientific studies on substances added to human food and summarizes the discussions and perspectives of the workshop participants. A draft of the article was reviewed by the workshop participants for accuracy and completeness and to ensure that their views voiced at the meeting had been accurately captured and fairly reported.

Pew Health Group's Framework for Safety Assessments for Human Food

The Pew Health Group developed a framework to illustrate the current system FDA uses to identify and evaluate the hazard information needed to determine whether food additives are safe (Figure 1). As with all frameworks, it simplifies the process and leaves out various nuances, but it captures the essential elements of today's scientific bases for food additive regulation. Note that the definitions of these types of tests used for the purpose of this workshop and article are intended to be shorthand phrases that may oversimplify some concepts and therefore may not conform to the way these terms are understood by the professional toxicology and risk assessment communities. These terms and definitions were developed to provide a common vocabulary among participants from different specialized fields.

The 4 corners of the framework represent 4 types of scientific studies, each of which serves a distinct role in chemical safety assessment:

- *Screening tests* identify potential hazards or, in some cases, actual human exposures to chemicals. Although these tests generally are not sufficient to confirm adverse effects, they may serve as the basis for subsequent hypothesis-based research and guideline-based studies.
- *Hypothesis-based research* means studies that explore whether potential human health hazards exist and determine their significance. This research generally begins with an investigator's hypothesis and consists of experimental protocols designed by the investigator to test the hypothesis. This research is commonly done in an academic setting.
- *Validation studies* focus on the technical aspects of the experimental tests and assess methods and protocols used in toxicology studies to determine whether they can reproducibly measure or predict adverse effects. Once a test is validated, anyone performing testing, including food manufacturers, their suppliers, expert panels, and consultants, can reasonably rely on such tests for their own internal toxicity assessments or for petitions prepared for submission to FDA or other regulatory authorities globally.
- *Guideline-based studies* follow agency-recommended protocols or test guidelines to evaluate and report the potential hazards of a substance in a standardized manner. These studies commonly are carried out in laboratories that specialize in performing the protocols identified in guidelines such as the

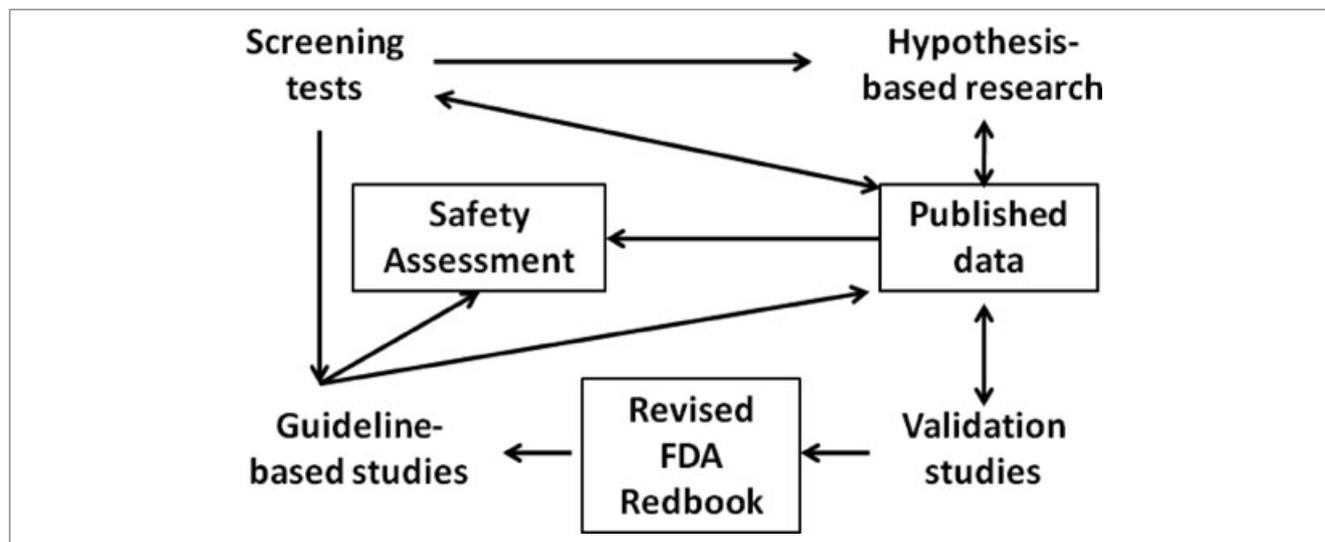


Figure 1—Illustration of the components of hazard-related scientific information used to make safety assessments for human food.

Redbook using GLP for purposes of supporting a regulatory safety determination.

While the results of all 4 types of studies may be published, the majority of the *published data*, especially those in peer-reviewed journals, are from hypothesis-based research by academic investigators, including clinical and epidemiological studies, using protocols and tests usually not included in the *Redbook*. Currently, only methods and protocols that have been validated may be included in the *Redbook*, either as new or revised guidelines. Industry uses the *Redbook* as well as guidance from other organizations, including U.S. Environmental Protection Agency (EPA), Food and Agricultural Organization (FAO)–World Health Organization (WHO), and OECD to design and conduct studies to evaluate safety. This guidance is followed regardless of whether or not the firm will ultimately submit a notification or petition to FDA. Therefore, the *Redbook* plays a pivotal role in guiding industry's safety data development, as well as the determination of safety either by the industry or by FDA, as the case may be.

Safety assessments of any substance added to food rely on available guideline-based studies and published hypothesis-based research. Typically, the results of screening tests and validation studies are not pivotal for the final safety assessment since they are not intended to confirm adverse effects. In the particular case of a GRAS substance, such determination requires general recognition of safety among qualified experts (qualified by training and experience). FDA has interpreted this general recognition standard to mean that any study pivotal to a GRAS determination must be published in some form in the peer-reviewed scientific literature.

Excerpts from the Pre-Workshop Webinar

In a webinar held before the workshop, Antonia Mattia, Director of the Division of Biotechnology and GRAS Notice Review in FDA's Office of Food Additive Safety (OFAS), summarized the major points made in 3 background documents prepared by OFAS personnel and distributed to workshop participants (see FDA 2011b). OFAS regulates food additives, color additives, GRAS substances, and food contact substances under a series of laws passed in 1958 and in subsequent years. The food additive petition process and food contact notification process are both mandatory programs, whereas GRAS notification is a vol-

untary program where FDA encourages firms to submit their internal safety decisions to the agency for review and comment. The same safety standard and standard of review apply in all 3 programs. The safety standard used by FDA requires reasonable certainty in the minds of competent scientists that the substance is not harmful under the intended conditions of use. This standard is commonly known as the "reasonable certainty of no harm" standard. The term "harm" is not defined in food additive law or in the implementing regulations, but FDA views harm, based on the law's legislative history, as an effect that adversely affects human health, not simply an undesirable or unexpected effect that does not adversely affect human health. (As examples, a headache or an episode of diarrhea would not be interpreted by FDA to constitute harm from food.)

Food additive safety assessments require knowledge about what exactly is the food additive (that is, its molecular identity, along with its manufacturing and purity specifications), how much is in food (to estimate probable exposure), whether it is harmful (toxicology data), and other case-specific questions. Specifically, the food additive information submitted to FDA typically includes the identity and composition of the food ingredient, information about its manufacture and use in food, estimated daily intake, analytical methodology used to characterize the chemical (for example, its chemical and physical properties), full reports of any available toxicity data, proposed maximum use limits, and environmental information. Toxicological studies include short-term tests for genetic toxicity, metabolism, and pharmacokinetic studies, subchronic feeding studies, 2-generation reproduction studies, developmental toxicity studies, chronic feeding studies, 2-y carcinogenicity studies, and other studies as needed, such as neurotoxicity or immunotoxicity studies. Based on the *Redbook*, the selection of the appropriate toxicity tests is based on the substance's "level of concern," which is based on a structure–activity relationship and the estimated cumulative dietary human exposure; in this context exposure is synonymous with consumption. There are 3 concern levels from low to high: I, II, and III; the number of required toxicity tests increases with the level of concern (see The Pew Health Group 2011). For instance, substances classified as concern level I undergo genetic toxicity tests and a short-term (28-d) toxicity test with rodents. It may be appropriate, however, given the specific

issues relevant to the use of an additive, to require more elaborate data. Concern levels II and III prescribe more extensive toxicity testing than concern level I and may also require more elaborate data based on intended use.

FDA approaches the safety determination that results from the safety assessment as a consensus decision among the FDA's scientists involved in the review based on an evaluation of all the available data consistent with its Redbook guidance. In the case of a safety assessment based on general recognition of safety, consensus must be shown in the larger scientific community through the weight of evidence in the available literature and can involve, for example, using an expert committee representative of the scientific community. Decisions inevitably entail some level of uncertainty and they are based on FDA's view of the best science available at the time. If the information is inadequate, FDA will continue to raise the remaining safety questions and will require the industry sponsor to fully address such questions including conducting additional studies as part of its premarket assessment. FDA's decisions must be strong enough to withstand scientific, procedural, and legal challenges.

The Redbook is a guidance document that industry and other stakeholders rely upon regarding toxicological information to be submitted to FDA, and it represents the agency's current thinking on toxicological principles for the safety assessment of food ingredients. To develop and issuing guidance documents, such as the Redbook, FDA must follow the Good Guidance Practices (GGP) regulation (21 CFR §10.115). Changes in guidance documents fall into 2 categories: level 1 and level 2. Level 1 guidance involves the initial interpretation of statutory or regulatory requirements, policy changes, complex scientific issues, or controversial issues. For level 1 guidance, FDA announces draft guidance in a *Federal Register* notice, invites public comment, and incorporates suggested changes as appropriate. Level 2 guidance sets forth existing practices or minor changes in the interpretation of policy. The clearance process is less extensive for level 2 guidance and does not include issuing a draft for public comments. Most recent changes to the Redbook have all been at level 2.

Summary of Plenary Session Presentations

This section is an account of the speakers' presentations; each talk was recorded and the transcripts were used as the basis for this summary.

Shelley Hearne, Managing Director of The Pew Health Group, and IFT Fellow Joseph Hotchkiss welcomed the participants to the workshop. Michael Taylor, Deputy Commissioner of Food for FDA, and, in a later plenary session, Mitchell Cheeseman, Acting Director of OFAS, described FDA's regulation of food additives and elaborated on several of the points made in the preworkshop webinar. See Table 2 for the workshop agenda.

Michael Taylor emphasized the need to harness the best science to evaluate the safety of chemicals added to food. The existing framework is based on the principle of prevention, where submitters of petitions and notifications bear the legal burden of proof for safety. Even for GRAS substances, the entity using the substance has the legal obligation to ensure that it is safe.

The legal requirements to demonstrate safety evolve over time as new hazards are identified and better understood, Taylor said. New tests and newly identified endpoints can raise questions about the adequacy of safety evaluation methods used to test substances. Good laboratory practices do not guarantee that the right science is available to answer a question, but they help generate reliable data. Taylor rejected the idea that there is a dichotomy

Table 2—Workshop agenda.

Day 1—April 5, 2011

- Welcome and workshop overview
 - Shelley Hearne, Pew Health Group
 - Joseph Hotchkiss, Institute of Food Technologists
 - Linda Birnbaum, NIEHS
 - Michael Taylor, FDA
- Small-group discussions—Round 1: considerations in identifying and validating endpoints, including adverse effects
 - Endocrine disruption
 - Behavioral impacts
 - Nanomaterial characterization
 - Tox21 & NHANES Screens
- Small-group reports from Round 1
- FDA's safety assessment process and use of computational toxicology
 - Mitchell Cheeseman, U.S. FDA, OFAS
- Small-group discussions—Round 2: evaluating study design and data for regulatory decisions
 - Dose response
 - Transparency
 - Study reproducibility
 - Use of hypothesis-based research
- Small-group reports from round 2
- Beyond FDA: EFSA, JECFA, and OECD
 - Jean-Lou Dorne, European Food Safety Authority
 - Angelika Tritscher, World Health Organization (JECFA and JMPR)

Day 2—April 6, 2011

- Alternative Methods
 - Rodger Curren, Inst. for *In Vitro* Sciences
 - Leon Bruner, Grocery Manufacturers Association
 - Jennifer Sass, Natural Resources Defense Council
 - Raymond Tice, NIEHS
- Small-group discussions—Round 3: developing and reviewing test guidelines
 - Developing test guidelines for review
 - Reviewing and approving test guidelines
- Small-group reports from round 3
- Small-group discussions—Round 4: identifying and evaluating potential solutions
 - Improving hypothesis-based research
 - Improving guideline-based studies
 - Refining the regulatory decision-making process
- Small-group reports from round 4
- Discussion and adjourn

between guideline-based studies and hypothesis-based research because both seek to generate reliable data and conclusions. All science needs to be considered in assessing the safety of substances. At the same time, standard protocols used in properly conducted studies give FDA assurance that the burden of proof has been met. With a newly identified endpoint, established protocols may not be able to establish the safety of a substance. In that case, FDA needs probative evidence to judge the relevance of the endpoint to an evaluation of safety, regardless of the source of that evidence.

Taylor also mentioned several drawbacks of the current regulatory framework. FDA does not have a way of acquiring complete information about the use of a substance in foods after it has been approved. Also, under the GRAS process, substances can be added to food without FDA's awareness and without FDA having access to use or safety data for those substances. Finally, the processes for both promulgating and revoking food additive approvals are legalistic and cumbersome, which reduces the flexibility for public health decision making.

Mitchell Cheeseman stated that, for food additive petitions and food contact notifications, successful submissions result in safety decisions by FDA. For GRAS substances, independent determinations are permitted and manufacturers are not required to inform FDA. A voluntary GRAS notification process encourages manufacturers to make information about the substances in food

available to FDA, but the safety determination remains with the firm.

Cheeseman said that FDA uses the best available data, regardless of GLP status. Cheeseman also noted that FDA's safety decision—or industry's in the case of a GRAS substance—must be based on an evaluation of all relevant data, whether they support or contradict the safety of the proposed use of a substance. When relevant contradictory evidence exists, all data must be considered in a weight-of-evidence approach, taking into account the relative probative value of differing data to reach a conclusion based on a consensus. The safety standard of “reasonable certainty of no harm” does not require proof beyond all possible doubt that no harm will result under any conceivable circumstance. As a practical matter, this means that FDA's safety decisions involve some level of uncertainty, which, he said, is recognized and accounted for in the decision-making process.

The Redbook and other guidance documents provide assistance to submitters and petitioners in developing data to address safety criteria, and they provide a framework for FDA's consideration of information provided to address specific review elements. However, the guidance provided in the Redbook and other documents is a starting point and is not binding, Cheeseman emphasized. He stated that other methods are acceptable to the extent that they address the same probative questions. He emphasized that FDA seeks to use the best science available, whether from guideline-based studies or hypothesis-based research, to make regulatory decisions. This flexibility allows FDA to request and to accept alternative testing methods when those methods address probative questions related to the safety of a substance. Petitioners and FDA staff often engage in a case-by-case dialog on particular testing challenges.

Linda Birnbaum, Director of the National Inst. of Environmental Health Sciences (NIEHS) at the National Inst. of Health, who also oversees the National Toxicology Program (NTP), explained how the institute seeks to support science that can be used in regulatory decision making. NIEHS-supported researchers have explored a very wide range of diseases with a known or suspected environmental component, including lung dysfunctions such as asthma, reproductive dysfunction such as reduced fertility, neurodegenerative diseases such as Alzheimer's and Parkinson's, and neurodevelopmental disorders such as Attention Deficit Hyperactive Disorder (ADHD) and autism. NIEHS considers food and nutrients an essential part of the environment. For example, the NTP's Center for the Evaluation of Risks to Human Reproduction (*Postmeeting note: the center has recently expanded into the Office of Health Assessment and Translation*) has evaluated a number of compounds that are approved food additives, from phthalates to bisphenol A, and has concluded that for some compounds concern about particular reproductive or developmental endpoints is justified. Birnbaum noted that the mission of the Office is now being extended to look beyond reproduction and development. For example, a recent workshop examined the role of environmental chemicals in obesity and diabetes, and other workshops are being planned. The NTP is headquartered at the NIEHS and coordinates nonlegally mandated toxicity testing across the federal government, develops new toxicity testing methods, and has the responsibility for convening the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), of which FDA is a member.

Birnbaum emphasized that NIEHS has a particular interest in the developmental origins of disease. Exposures to substances at

particular ages may have latent effects that are not detectable until later in life. Individuals may have early windows of susceptibility when chemical exposure may result in disease and health problems later in life. Also, depending on the developmental stage, chemical substances may have detrimental effects on the body even at very low doses. Endocrine disruptors are an example of such substances. They can have a variety of effects on the body at very low doses, they are widespread in food and other parts of the environment, and people may be exposed simultaneously to multiple endocrine disruptors that exert their effects through different mechanisms.

Guideline-based or GLP studies can guarantee the observance of protocols, but they cannot guarantee that the study was carried out correctly or that the right question was asked, Birnbaum said. She pointed out that a single study never truly answers all of the regulatory questions. Guideline-based and GLP studies can be valuable as part of a research portfolio, but all of the best available science needs to be used to understand risk and establish a regulatory framework. Researchers doing guideline-based studies and hypothesis-based research need to listen to each other to learn why particular approaches produce particular results.

Sidebar: Safety Evaluations Outside the United States

Two speakers at the workshop—Jean-Lou Dorne of the European Food Safety Authority (EFSA) and Angelika Tritscher of the World Health Organization, who is responsible for the **Joint FAO/WHO Expert Committee on Food Additives (JECFA)**—summarized food safety assessment programs in each organization.

Dorne explained that EFSA is the keystone of the European Union's risk assessment of food and feed safety and nutrition. Formed in response to the food crises of the late 1990s, it was established in 2002. It is governed by a Management Board that has 15 members from government, industry, consumers, scientific community, and other food safety agencies. The Authority does chemical risk assessment and responds to hazard-related questions from stakeholders such as the European Commission, the European Parliament, and EU member states. EFSA focuses on risk assessment and food safety, gathering data for evaluations, and improving risk assessment methodologies by proposing new or alternative ways to assess risks. EFSA can also initiate evaluations on chemicals already in the market when it decides there is an important problem (self-mandates); however, research is not part of EFSA's mission. Its goal is to reach a scientific opinion on risk assessment so that the European Commission or others can take management steps. EFSA's scientific opinions are not necessarily achieved by consensus, and minority opinions are published as well. EFSA communicates its findings to consumers, the media, industry, and other professionals. All scientific evaluations/risk assessments are published on EFSA's website approximately 15 d after adoption by the respective scientific panel (<http://www.efsa.europa.eu>).

Dorne also mentioned that the EFSA Emerging Risks Unit, which was created in 2008, monitors emerging risks through extensive review of the literature and by gathering information from the European Commission, member states, experts, and the media. The unit's projects include identifying and monitoring biological and chemical risks, and investigating new methodologies and data collection techniques.

Tritscher provided background information on the Joint FAO/WHO Expert Committee on Food Additives, JECFA, and its mission (<http://www.who.int/foodsafety/chem/jecfa/en//index.html>). She stated that **JECFA, which was founded in 1956**, is an expert committee of qualified independent international expert scientists run jointly by the United Nations Food and Agriculture Organization (FAO) and the World Health Organization (WHO). JECFA is the international risk assessment body that provides the science base for global standards; it also provides a reliable and independent source of expert advice internationally, contributes to setting standards on a global scale to protect the health of consumers, and develops improved principles and methods for risk assessment of chemicals in food. JECFA receives requests for assessments from the Codex Alimentarius Commission (which develops food standards, guidelines, and related texts such as codes of practice under the Joint FAO/WHO Food Standards Programme) and its subsidiary bodies, such as the Codex Committee on Food Additives, or from member states of FAO and WHO. WHO may also organize *ad hoc* expert consultations to respond to a fast-emerging concern. Countries use information from JECFA in the establishment of their national food safety control programs.

Sidebar: Alternatives to Animal Use and the Validation of Toxicological Tests

In a plenary session beginning the 2nd day, 4 presenters discussed alternative methods to animal testing and the validation of new tests and methodologies. The topic of alternative methods to animal use served as an example of the sometimes complex process of validating new methods and incorporating them into test guidelines and the challenges that the process can encounter. Rodger Curren, President of the Inst. for In Vitro Sciences, pointed out that alternative tests have traditionally been thought of as tests that replace, reduce, or refine the use of animals in toxicity testing. The same issues involved in developing alternatives arise in the development of any new test or methodology, so the same standards should apply. Regulators need to know: Is a test reliable and relevant? What is its applicability? What is its predictive capacity? What are its performance characteristics? Can results from different tests be combined to yield valid results? The process of answering these questions works best when it is carried out as a hypothesis-testing activity, Curren said, where, for example, blinded tests are done to confirm whether a new test meets a particular standard. No test is perfect, but regulators need to know how good a test is if they are going to use it.

Curren noted that in the past, it has taken years or even decades for a new test to be validated, and there were several reasons for this. For instance, the gold standard for comparison has been animal tests, but the animal test itself is often flawed because of poor reproducibility. Also, validation studies sometimes failed because scientists doing the validation did not always adhere to protocols. In addition, funding to support this work has sometimes been difficult to secure. Encouragingly, in recent years the situation has improved, he said, and some tests proposed under the OECD Test Guidelines Program have been accepted within a year.

In the United States, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM; <http://iccvam.niehs.nih.gov>) is the group charged by law with advising on test method development and validation. Originally established as a standing committee by NIEHS in 1997, ICCVAM became, with the ICCVAM Authorization Act of 2000, a permanent committee under the NTP Interagency Center for the Validation of Alternative Toxicological Methods (NICEATM). Composed of representatives from 7 U.S. regulatory and 8 research agencies, its duties, said Raymond Tice, Chief of the NTP Biomolecular Screening Branch at NIEHS, are to evaluate the validation status of new or revised safety testing methods, transmit formal recommendations to federal agencies, promote regulatory acceptance of valid methods, and foster national and international harmonization of test methods. ICCVAM also has been working with comparable national organizations in Canada, the European Union (ECVAM), Japan, and South Korea to expedite the international adoption of valid alternative methods. Since 1999, it has helped 40 alternative safety testing methods gain acceptance or endorsement by U.S. and international agencies, and made recommendations for research and development, translation, and validation activities to further advance alternative methods.

ICCVAM has adopted specific criteria for the validation and acceptance of a toxicological test method. Particularly important criteria, Tice said, are reliability, which measures the extent to which a test method can be performed reproducibly within and among laboratories over time, and relevance, which measures the extent to which a test method will correctly predict or measure the biological effect of interest. Ultimately, an accepted alternative test method should provide for equivalent or improved protection of human and/or animal health or the environment compared with traditional animal testing. Alternative test method nominations and submissions are accepted from any individual or organization (for more information, contact NICEATM at niceatm@niehs.nih.gov).

Leon Bruner, Chief Science Officer for the Grocery Manufacturers Association, observed that human toxicity is generally measured in surrogates for obvious ethical reasons. This means that the most important sources of data for use in human safety assessments come from a broad range of *in vivo*, *in vitro*, or *in silico* tests. Given that data from these tests must be extrapolated to the human situation, and because knowledge of biological processes is incomplete, there is always some level of uncertainty in predictions of human toxicity from such tests. The reason guideline-based tests are used with confidence in the regulatory decision-making process is that they have been validated. Tests are considered valid when there is adequate knowledge of the biological mechanisms behind a test, confidence that data can be reproduced across laboratories and assurance that predictions of toxicity will lead to decisions that protect public health. One of the most significant issues with use of new hypothesis-based tests is that their validity has not been adequately assessed. Positive and negative signals from an unvalidated toxicity test mean little unless a scientifically robust validation process has shown that the test effectively predicts harmful effects in humans. Scientific approaches needed to assess the validity of a new or improved toxicity test have been extensively discussed in the toxicology literature. Effective validation studies require the engagement of qualified participants, a reference set of test substances that cover the

range of toxicity response, and logistical, statistical, and financial support. Researchers working to validate a new or improved hypothesis-based test should build on previous discussions and data while innovating to make the process shorter and less costly. Assessing the validity of a toxicity test is relatively easy when the test is sufficiently developed, Bruner added. He also noted that the real difficulty lies in the development of valid tests.

Jennifer Sass, Senior Scientist at the Natural Resources Defense Council, observed that FDA and other federal agencies will need support to develop alternative testing methods. New technologies and new health concerns are going to place great burdens on these agencies. FDA, for example, will face an immense challenge in dealing with the amounts of data on toxicity testing that new technologies can generate, with the Tox21 Program being a prominent example (discussed later in this article). Sass stated that relevant endpoints should be incorporated into new testing methods to provide a focus for evaluations, and these endpoints will need to include multipathway and multistep disease pathways. She also stated that FDA needs to be systematic and transparent about the data it is using and the process for making decisions, with opportunities for public comment throughout the regulatory process, not just at the end. She concluded that waiting for harmful effects to appear in humans represents tangible harm to individuals and a failure to protect public health, which could damage FDA's credibility.

Summary of Small Group Discussions

There were 4 rounds of small group discussions, each covering an overarching theme. Each discussion group had a moderator and FDA representatives to answer any questions. The participants had the opportunity to indicate the sessions they wanted to attend and were assigned to groups based on their interest. Meeting organizers made an effort to ensure fair representation of all stakeholders in each small group discussion. Plenary sessions were held after each round of small group discussions in which the moderator of each group presented a summary of the discussions that took place.

Identifying and validating endpoints

The first round of small group discussions focused on considerations in identifying and validating relevant endpoints, including adverse effects. Four current and sometimes controversial issues in chemical safety assessment framed the discussions: endocrine disruption, behavioral impacts, nanomaterial characterization, and the use of screening tests to trigger additional toxicology studies. The 2 screening programs used as examples were Tox21 (<http://www.epa.gov/nct/Tox21/>), which uses high-throughput cell-based assays designed to evaluate hazards by assessing for interactions between chemicals and different toxicity pathways, and the National Health and Nutrition Examination Survey (NHANES) (<http://www.cdc.gov/nchs/nhanes.htm>), a program designed to assess the health and nutritional status of adults and children in the United States (in particular, discussions focused on the biomonitoring data on chemicals in blood and urine that can be used as biomarkers of exposures to chemicals in the environment). All of the discussions touched on the issue of how to define harm, a task that is complicated by the lack of a formal definition of harm in the legislation or regulations implementing the food additives provisions. FDA views harm, based on the legislative history of its governing acts, as an effect that affects human health, not simply an undesirable or unexpected effect that does not adversely affect human health. While not designed to produce consensus, the discussions revealed a range of views regarding what constitutes an adverse effect and how to identify and characterize such an effect.

Endocrine disruption. The discussion of endocrine disruption dealt with the contentious issue of whether or not positive results in hypothesis-based research constitute or sufficiently predict adverse effects to be used in FDA safety determinations and to justify incorporation of the endpoints into the Redbook. The materials

distributed before the workshop (see The Pew Health Group 2011) noted that current protocols call for at least 3 doses of a substance to be used for toxicity testing: (1) a dose high enough to induce toxicity, (2) a dose low enough to not induce toxicity, and (3) an intermediate dose high enough to induce effects that eventually may lead to adverse impacts, such as changes in enzyme levels or a slight decrease in body weight. The selection of these doses allows the evaluation and reporting of irreversible, gross adverse effects on study animals and increases the likelihood of identifying the no observed adverse effect level (NOAEL), a parameter often used as a starting point to calculate the chemical's acceptable daily intake. The NOAEL is the highest dose of a substance that does not produce an adverse effect.

Hypothesis-based research suggests that chemicals with hormone-like activity can target organs and induce subtle but important effects at very low levels while higher doses may not produce observable effects. Concerns have been expressed that the lowest doses used in guideline-based toxicology studies are sometimes too high and the evaluation of apical endpoints too insensitive, thus preventing the identification of a substance's potential endocrine-disrupting effects.

Current guidelines generally do not require testing or direct assessment of changes in the functioning of the endocrine system *per se*; rather, they assess the overt manifestations (for example, apical effects such as birth defects or tumors) of any such changes in intact whole organisms. Longer term animal studies do include measures of blood chemistries and tissue pathology that may be early indicators of endocrine-related adverse effects. However, current protocols do not require animals to be exposed to chemicals at doses below the NOAEL, even though there are hypothesis-based studies that suggest that such exposures may be important in influencing the risks of human health effects. Some test guidelines require exposure during specific times of development (for example, such as early gestation and gestation and lactation periods); however, it is not yet standard practice to incorporate *in utero* and/or prepubertal exposure in all chronic and/or carcinogenicity studies, although this option can be exercised. Laboratories conducting guideline-based studies are not prohibited from assessing more sensitive potential endocrine disruption effects in addition to the endpoints included in protocols; however, they have little incentive to do so, especially if additional funding is required. Additionally, the endpoints in the Redbook are not the only ones considered by the FDA in making safety determinations. However, the Redbook is widely used and may constitute the only guidance used, particularly in the case of GRAS determinations. There are ongoing efforts at EPA to update certain test guidelines by incorporating measurements such as thyroid stimulating hormone (TSH), T3, and T4 for the evaluation of thyroid function. Some of these methods are currently undergoing validation; therefore FDA has not yet made a determination on whether or not it will consider incorporating any or all updates into the Redbook.

The moderator of the small-group discussion on endocrine disruption identified 3 themes that ran through the discussion. First, *the participants did not agree on what should be considered an adverse effect in the context of endocrine disruption and whether substances added to food could or could not significantly affect the endocrine system and result in quantifiable adverse effects predictive of human health effects.* Scientists from the regulated community, some from government agencies, and some academics held the view that hormonal function and hormone levels are a continuum and that it is difficult to identify with certainty if or when a perturbation of the endocrine system translates into adverse outcomes. Scientists from the regu-

lated community also noted that the endocrine system is largely an adaptive system in which changes do not necessarily denote harm, and that in some cases changes in the endocrine system confer health benefits. Scientists from the regulated community and some government agencies also mentioned that holistic evaluations accomplished by many guideline-based assays can accomplish the task of evaluating potential effects on multiple endpoints. According to this view, a complex system is only "disrupted" if it is made to operate outside its normal range of variation and adaptive responsiveness. In contrast, many scientists from academia, public interest groups, and some from government agencies referred to "perturbations" of the endocrine system as disruptions, arguing that alteration in hormone levels or function almost always translates into a potentially adverse effect in some populations. These scientists shared the view that small disruptions of the endocrine system have been demonstrated to have great and long-lasting impacts, particularly when this disruption occurs in the young and in developing individuals (that is, tissue specific effects). In addition, they stated that a given hormone or hormonally-active substance can have many effects on the body resulting in an actual increase of diseases such as early puberty, behavioral deficits, asthma, obesity, and diabetes, therefore, a particular endpoint may not be indicative of all the endpoints the substance affects either quantitatively or qualitatively. Accordingly, many scientists from academia, public interest groups, and some government agencies noted that the use of multiple endpoints in the context of a complex system is necessary to effectively assess safety.

Another theme was that *difficulties in selecting health-related endpoints are closely related to the lack of a definition of adverse effects.* For example, some academic scientists proposed that biomarkers of endocrine disruption could be used to predict later overt disease outcomes, noting that some biomarkers have been used for many decades in epidemiological and clinical studies and are highly reliable in predicting adverse effects. According to this view, the use of overt effects as endpoints may not be protective enough. Also, these scientists noted that small alterations in hormonal function may cause adverse outcomes and these outcomes may vary with life stages; thus, a clinical definition of "adverse outcomes" may not always be appropriate or reflective of the existing regulatory framework's principle of prevention. Some government and academic scientists pointed out that EPA has a definition of adverse health effects that includes biological perturbations; they also stressed that endpoints should reflect a continuum of risk instead of answering a yes/no question. Possible endpoints to consider include hormone concentrations in blood, hormone-receptor binding, gene expression, or biological factors important at particular developmental stages such as the correct number and organization of neurons. In contrast, scientists from the regulated community held that biomarkers of endocrine disruption are unlikely to predict adverse outcomes relevant to human disease or dysfunction. These scientists said that most biomarkers resulting from certain *in vitro* and animal test systems have not undergone validation for predictivity or reliability with respect to human adverse effects. From this perspective, much more understanding of chemical modes of action and interactions of biological systems is necessary before such screening results can be useful for regulatory hazard identification and decision making. Scientists from the regulated community are of the opinion that if small alterations in hormonal function are validated to cause adverse outcomes relevant to humans that are different at different life stages, these effects may become important on a case-by-case basis to expand existing regulatory consideration of what constitutes an adverse effect. They stated

that more work is needed on validating biomarkers that are predictive of later health effects.

The 3rd prominent theme was *whether it is important to incorporate multiple endpoints and multiple tests in a weight-of-evidence determination*. Some academic, government, and public interest scientists noted that the effects of an endocrine disruptor could be manifested in multiple organs and in various ways, which may indicate a general and potentially permanent harm. They raised the following issues: How are multiple endpoints from multiple studies best integrated to ensure reasonable certainty of no harm? How much weight is given to guideline-based studies and/or to hypothesis-based studies reporting on different endpoints? How can human exposure data from postmarket exposure assessments and endpoints that emerge from exposed populations be integrated into regulatory decision making? Regulated community scientists questioned giving greater weight to postmarket research than to well-conducted guideline studies, because they do not believe it is possible to find a biologically or clinically relevant NOAEL with any confidence under postmarket circumstances by looking at biomarkers in the consumer population. It is important, they stated, that influences exerted by normal physiological processes such as toxicokinetics (influences of exposure route, absorption, distribution, metabolism, and excretion), adaptation and repair mechanisms, and plasticity not be ignored.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Subpopulations, such as young children and pregnant women, appear to be more sensitive than others to endocrine disrupting chemicals. Therefore, efforts should be made to recommend assessing developmental endpoints routinely in the current guidelines and new assays to evaluate or screen for endocrine disruption.
- Population and subpopulation effects need to be studied rather than focusing primarily on effects occurring in individuals. This issue relates to evaluating population versus individual risk, for which significant literature exists. Epidemiological trends may be better indicators of potential public health effects than identifying clinical abnormalities (that is, apical measures of adverse outcome). For example, recent studies demonstrate that thyroid stimulating hormone (TSH) levels within the population reference range are associated with neurodevelopmental deficits in infants (that is, infants with subtle thyroid suppression) (Haddow and others 1999).
- Studying populations may help understand some effects but is not very useful in identifying cause and effect. Well-designed and conducted epidemiological studies can give indications of health trends.
- Understanding the toxicokinetics of all chemicals, including endocrine disruptors—that is, how the body absorbs, metabolizes, and excretes a substance—is important when assessing toxicity in a holistic manner.
- Life stages, modes of action, and effects early in life should be considered in decision models, and the decision logic should identify relevant endpoints.
- Strategies are needed regarding endpoints found in postmarket studies. For example, what is the biological meaning of the effects observed, when should new findings trigger incorporation of endpoints into guidelines, and how should changes be implemented? Since postmarket data usually come from human studies, a degree of flexibility is needed with respect

to their reproducibility, especially in the presence of coherent and cogent evidence.

- Experiments, both short- and long-term, are needed to shed light on the use of biomarkers to predict overt effects.
- Responses to disruptions of the endocrine system may be apparent months, years, or decades after brief exposures, especially during the neonatal period. Current guidelines should be modified to better identify these effects.
- Guideline-based studies currently in the Redbook are likely to miss adverse outcomes from some endocrine disrupting chemicals' mode of action, such as thyroid action. EPA's approved guideline-based studies for its Endocrine Disruptor Screening Program may be better able to detect some level of interference.

Behavioral impacts. The discussion session on behavioral impacts considered what types of behavioral and/or neurological changes represent adverse effects and whether or not they can and should be measured and reported in nonclinical, guideline-based animal studies. FDA defines neurotoxicity as "any adverse effects on the structure or functional integrity of the developing or adult nervous system," and it considers biochemical, morphological, behavioral, and physiological abnormalities as adverse effects. For nonclinical studies, the Redbook recommends that substances undergo a screen to identify any potential adverse impacts on the nervous system, though the Redbook is not restrictive and does not exclude additional studies. The screen consists of a structure-activity relationship analysis, review of a published literature (if any exists), and experimental data from animal screening tests, which FDA acknowledges is the primary means of obtaining neurotoxicity screening information. Neurological screening tests are conducted as a component of other guideline-based toxicity studies and include assessing the incidence and severity of various clinically obvious endpoints such as seizure, tremor, motor coordination, and alertness.

FDA uses clinical and epidemiological studies, when they are available, as the basis for evaluating specific chemical uses. A recent example involved caffeine in alcoholic beverages (FDA 2010a). FDA's (and EPA's) safety standards for lead as a contaminant also have been based on the learning disorders associated with very low levels of human exposure. But clinical studies on additives are uncommon and, like epidemiological studies, may be carried out only after the substance is in widespread use in human food (that is, postmarket).

The moderator identified 4 broad themes in the small-group discussion of behavioral impacts. First, as with the previous group, *differences arose over the definition of adverse effects or harm in the context of behavioral impacts*. Distinguishing between an undesired and an adverse effect can be particularly difficult with behavioral impacts, where many neurological conditions occur on a spectrum and can be difficult to identify in their milder forms. Also, some subpopulations, such as the fetus or young children, may be more susceptible than others, and long-lasting complex effects can manifest months or years later.

A 2nd theme was that *existing screens may not capture the subtle yet complex human behaviors that may be of concern*. Existing screens capture overt effects such as seizures, paralysis, motor coordination, strength, or obviously abnormal behaviors, but the group appeared to agree that they will not detect more subtle effects on the structural or functional integrity of the nervous system in mature or developing organisms, such as certain forms of learning, memory, anxiety, or hyperactivity. Scientists from FDA and the regulated community noted that the cost and efficiency of some of these

studies have precluded the inclusion of testing for some of the more subtle aspects of behavior that have become more important and challenging to our society in recent times into the existing guidelines. For decades, academic researchers have developed a variety of behavioral test protocols for animals to address specific neurological problems. For the assessment of food additives and pesticides, both FDA and EPA rely on animal studies conducted using a diverse set of protocols and endpoints; however, neither the endpoints nor the research protocols have been validated through ICCVAM or an equivalent validation process. A scientist familiar with EPA's assessment of pesticides noted that these protocols have been subjected to a "validation process," albeit less formal than ICCVAM's, through a collaborative exercise conducted over many years by the relevant expert scientists in academic, government, and industry laboratories, which has produced reliable animal protocols predictive of adverse human health effects. The Redbook recommends several protocols to measure the same endpoint. Although the flexibility of these guidelines provides the opportunity to the manufacturers to select the protocols they are most familiar with, it makes it harder for the regulatory scientists to understand similarities and differences of chemicals from the same class when different protocols are used to measure the same endpoints.

Another theme was that *animal tests need to better reflect complex human behaviors*. An ideal screen would detect endpoints relevant to behavioral impacts, albeit with an acceptable rate of false positives. Screening tests are often very sensitive, with low false negative rates, particularly with uncommon or subtle behavioral adverse effects. Some participants were of the opinion that no existing animal testing method currently listed in the Redbook provides a highly sensitive and reliable screen. It will be difficult to design such a screen in animals, but efforts to identify and validate relevant endpoints predictive of adverse human behavioral changes and to develop tests to measure them reliably should continue. In addition, stand-alone guidelines with relevance to human neurological endpoints would be beneficial.

Finally, *human behaviors thought to be related to chemicals added to food need to be identified which will trigger evaluations or reevaluations of substances already in commerce on a case-by-case basis*. An important task will be to determine the types of animal studies that should be triggered by reliable data on the prevalence of human diseases or disorders and reliable evidence relating such conditions to dietary exposures. It is clear that there is a need to rely more on human data where, as with lead and mercury, large-scale epidemiology studies have provided overwhelming evidence for adverse developmental and behavioral effects.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Guidelines to test for behavioral impacts are needed for food additives, including guidelines for clinical behavioral studies; this will help in the interpretation of data for a particular chemical and across classes of chemicals. Guidelines used by the pharmaceutical industry to test behavioral impacts could serve as models. Also, the group added that daily cage observations are not sufficient to trigger more detailed observations or in-depth studies and should be reevaluated.
- Procedures are needed to validate developments in hypothesis-based research into guidance for toxicity testing.
- Gene-chemical interactions research is necessary to understand observed differences in susceptibility in the human population.

- Some additives were approved for use in food more than 40 y ago, should companies or FDA be required to reassess substances in the light of scientific advances? As a practical matter, some form of prioritization should be applied to decide where to focus limited resources.

Nanomaterials. This discussion considered the need to adequately characterize the physical and chemical properties of nanomaterials in hypothesis-based and guideline-based studies and the related challenges. The manufacture of nanomaterials is an emerging field with the potential to enhance food safety and quality, reduce the environmental impact of processing and packaging, and reduce food losses. However, very little is publicly known about the prevalence or safety of these products, even though some materials are already in the market in food and in food packaging. Recent research indicates that some nanoscale substances—which are defined by the EPA as substances between 1 and 100 nanometers in a single dimension—exhibit unusual physical and chemical properties that can make them especially useful but also may be important toxicologically. Some of these materials may have been specifically developed to have nanoscale dimensions and properties, while others, such as some proteins and enzymes, are naturally formed as nanomaterials.

Postmeeting notes: (1) *EFSA draft guidance on nanomaterials was finalized on May 2011 (EFSA 2011). Considering that the EFSA document will be relied upon by industry and other regulatory agencies when developing their own guidelines, we mention it as a post-meeting note because it was not considered during the workshop discussion. Unlike the OECD guidance (OECD 2010b) and other references noted by the group, the EFSA document is the first practical guidance specific to food-related nanomaterials to appear.* (2) *In June 2011, FDA released a draft guidance for industry, "Considering whether an FDA-regulated product involves the application of nanotechnology" (FDA 2011a). The draft guidance "is intended to help industry and others identify when they should consider potential implications for regulatory status, safety, effectiveness, or public health impact that may arise with the application of nanotechnology in FDA-regulated products." The draft guidance document does not address "the regulatory status of products that contain nanomaterials or otherwise involve the application of nanotechnology."*

In 2006, OECD created the Working Party on Manufactured Nanomaterials to address the safety testing of a reference group of nanomaterials (see The Pew Health Group 2011). In a recent report, OECD identified the following data gaps in the published literature that are necessary for safety determinations:

1. Nanomaterial information/identification;
2. Physical-chemical properties and material characterization;
3. Environmental fate;
4. Environmental toxicity;
5. Mammalian toxicology; and
6. Material safety.

In 2007, FDA's Nanotechnology Task Force released a report (FDA 2007) in which it acknowledged the challenges posed by nanomaterials used in food. The report mentioned the "uncertain nature of the science" and the necessity to improve scientific development to assist FDA in decision making. The situation for nanomaterials is similar in some aspects to biotechnology, where scientific advances led FDA to develop specific guidance regarding the scientific considerations for foods derived from new plant varieties.

While identifying potential adverse effects is critical, the first question to be answered is whether the substance is a nanoscale material, especially under the conditions in which it will be used

(for example, will it clump into larger particles or be bound into a matrix). Failure to accurately characterize the test material used in a study could result in conflicting results that provide little insight regarding its toxicity.

Four major themes arose during the discussion, according to the moderator. First, *material characterization and understanding of the interaction of the material with the matrix in which it is placed are equally important*. Because the nanocharacteristics of a material can be affected—even eliminated—by its surroundings, an understanding of the material's characteristics under conditions of anticipated product use is critical to assessing safety. Characterizing the properties of nanomaterials can be difficult, time consuming, and expensive. Several groups, including OECD and the Minimum Information on Nanoparticle Characterization (MIN-Char) initiative (Maynard 2009), have developed lists that provide physical and chemical parameters to characterize nanomaterials. However, flexibility is needed in choosing the important parameters for specific materials rather than evaluating all properties for all materials. Also, since it is not yet known which properties are critical determinants of toxicity, guiding principles on how to exercise this flexibility in an experimental matrix are needed. These principles could guide the development of a knowledge base so that the critical determinants of toxicity can be identified and understood.

A 2nd theme was that *information is needed regarding absorption (gastrointestinal and dermal), distribution, metabolism, and excretion, and the biological effects of nanomaterials in whole animals*. Little information exists regarding the fate of ingested nanomaterials (for example, absorption rates, interactions with intestinal flora, and whether or how their nanocharacteristics change in body fluids). This characterization is necessary in order to determine what toxicity testing is necessary and what material to test. This level of detailed characterization requires a substantial financial investment from industry and research funding agencies. The group had differing opinions regarding whether existing guideline-based studies are adequate to evaluate the potential adverse effects of nanomaterials. Some felt that some nanomaterials might elicit toxic effects that would not be detected by existing tests (for example, immunotoxicity) while others felt that existing tests were adequate if the materials were adequately characterized. The expertise of an interdisciplinary team of scientists will be required to fully understand the toxic potential of nanomaterials.

Third, *rather than distinguishing between engineered and naturally occurring nanomaterials, the context of their use is critical*. Having access to characterization data for naturally occurring nanomaterials in food as well as new and existing food additives whose particle size distributions might include small amounts of nanosized particles could greatly assist the regulatory and scientific communities in developing an understanding of nanocharacteristics relevant to safety. Based on the characteristics of the materials involved, toxicity studies for nanomaterials may not fit the traditional approaches to safety assessment. Industry and academic researchers that drive innovation in nanomaterials will need to work with regulators to determine a framework for evaluating nanomaterial toxicology and safety.

Finally, *a systematic process is needed to keep up with and distill emerging science and technology*. An expert stakeholder process involving scientists from academia, industry, public interest groups, and regulatory agencies is needed to evaluate emerging science on a regular basis; capture and interpret scientific gains; determine if safety testing guidance needs to be updated; and set priorities for the development, validation, and funding of new methods.

It is worth mentioning that some participants commented on the regulatory status of nanomaterials under the GRAS program, adding that the “general recognition of safety” standard likely cannot be met because the existing toxicological literature on nanomaterials intended for food uses suffers from inadequate characterization of the test materials. As a result, each nanomaterial and its intended use must be evaluated on a case-by-case basis. Participants stated that it would be helpful for FDA to clarify its position on establishing GRAS status for food-related nanomaterials since a manufacturer may self-affirm product as GRAS for use in food without notifying FDA of this determination.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Issues of “contamination” may be much more complex for nanomaterials than for conventional chemicals, because factors such as particle size distribution or surface coatings might affect toxicity. This is a difficult issue to address.
- Balance is necessary between broad characterization that will provide little information about the raw material but can be performed quickly and cheaply, and an exhaustive and very expensive characterization of materials. Guidelines are needed that identify the minimum data set required for nanomaterials characterization.
- Consumers should be informed regarding the safety of any nanomaterials. The FAO/WHO Expert Meeting on the Application of Nanotechnologies in the Food and Agriculture Sectors report suggested that greater participation of scientists in the public debate would assist the public in forming their own conclusions.

Tox21 and NHANES screens. This discussion session considered how the results of screening tests should be used as a trigger for additional studies. In essence, how should positive results from screening tests be used as part of a system to assess the toxicity of a chemical or a mixture of chemicals?

The biomonitoring studies conducted as part of CDC's NHANES program are used to determine the prevalence of major diseases and risk factors for diseases. In the late 1990s, NHANES began measuring some synthetic chemicals in the blood and urine of a nationally representative sample of 5000 Americans. NHANES does not take blood samples from children younger than 1 y or urine samples from children younger than 6. The measured chemicals were selected based on:

- Scientific data suggesting exposure in the U.S. population;
- Serious health effects known or suspected to result from some levels of exposure;
- The need to assess the effectiveness of public health actions to reduce exposure to a chemical;
- The availability of a biomonitoring analytical method with adequate accuracy;
- The availability of adequate blood or urine samples; and
- Incremental analytical costs to perform biomonitoring analysis for the chemical.

The NHANES survey data have the potential to trigger new epidemiological studies and hypothesis-based research. In particular, biomonitoring provides real-life snapshots of chemical levels in urine or blood that can be used to estimate chemical exposure from food and the environment. Biomonitoring data may also be useful in designing toxicological studies based on current human exposure levels if there is sufficient information on the similarities and differences between animal and human metabolism and on the pharmacokinetics of the chemical. However, exposure data

without a solid understanding of the hazard or the route of exposure provide limited insight into human risk or how to implement risk management procedures if required. Regulated community, government scientists, and some academics expressed concern that exposure and risk are not interchangeable and biomonitoring data may result in confusion among many stakeholders.

Tox21 is a collaborative program among the EPA, NIH, and FDA to research, develop, validate, and translate innovative chemical testing methods that characterize toxicity pathways. It will use high-throughput screening tests to evaluate mechanisms of toxicity with the purpose of:

- Identifying mechanisms of chemically induced biological activity;
- Prioritizing chemicals for more extensive toxicological evaluation; and
- Developing more predictive models of *in vivo* biological response.

Tox21 screens, many of which are still in development and have not been validated, use *in vitro* methods to evaluate the potential hazard effects of a wide array of chemicals and mixtures on cells at a wide range of concentrations. The Tox21 program has the capacity to generate large amounts of data regarding chemical interactions with cellular pathways. A positive result or combination of results is unlikely to be considered an adverse effect on its own. Instead, it can serve as a trigger for more detailed studies.

The Tox21 and NHANES discussion focused on the use of these screening tests as a trigger for targeted hypothesis-based research and, perhaps, guideline-based studies. Three important themes emerged. First, *screening tests will not entirely replace animal studies, but they will help to improve toxicity tests.* Tox21 data could inform FDA about additional data that should be requested during premarket review and allow manufacturers to identify needed toxicity tests. Tox21 data may also be used to identify some substances that do not need to undergo guideline-based studies, although some participants commented that this should not happen until it is fairly clear that the Tox21 screens will capture the full range of relevant toxicity pathways, which is not yet the case. An important use for screening data could be prioritization of substances for postmarket reassessment. The aggregated analysis of Tox21 and NHANES biomonitoring data could be used to develop priorities for substances that require further testing and safety reassessment and to guide the selection and design of guideline-based toxicity tests. Participants agreed that Tox21 methodology is not sufficiently validated at this point to be used as the primary basis for a substance's premarket safety assessment. However, some participants mentioned it could be used by FDA as a basis to require additional testing.

Another theme emerged regarding *the best ways to use Tox21 and NHANES data.* The food consumption and biomonitoring data from NHANES are a rich resource for information on representative exposures and the possible contribution of food. However, some academic and regulated community scientists raised concerns that the food consumption survey is based on only a 1-d snapshot and the results may be difficult to translate into an assessment of the source of exposure. FDA also uses NHANES food consumption data in their dietary exposure assessments for chemicals. Tox21 is an attempt to move toxicology into the predictive realm. If successful, the screen can either raise or lessen suspicions about a substance's hazard, both of which are important. Tox21 screen (although in an early stage of development) and NHANES biomonitoring program hold promise particularly for directing and prioritizing future studies. Scientists from the regu-

lated community recommended that continued financial support for maintaining the accuracy and completeness of the NHANES database was important.

A final theme was that *questions still surround some screening methods.* To be considered valid, their performance characteristics should be known and they should be reproducible and reliable. Some scientists commented that Tox21 data have some amount of inherent validation through redundancy, the use of prototype compounds, multiple results, and the use of probabilistic modeling. However, the data Tox21 generates as well as the association with adverse health effects will need to be validated. The stated goal of the Tox21 screens is to identify chemicals showing activity in multiple different assays, the hypothesis being that certain patterns of biological activity can be identified that may be early predictors of potential adverse effects and/or human disease. Chemicals identified in such screening would be flagged and prioritized for more detailed evaluation using hypothesis-based and guideline-based methods. Many scientists agreed that there will be a role for new and validated screening strategies in the foreseeable future. However, scientists from the regulated community also noted that in their view, whole animal toxicology as an integrative look at the whole complex system is currently the best (or only) approach because the requisite detailed knowledge about underlying compensatory responses, mechanisms, and systems interactions is still lacking in many of the screening methods.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Modifications in the procedures used to validate methods may be needed, as recommended by the National Research Council (Krewski 2007), to address the unique nature of pathway-based assays in human cell lines (which cannot be validated against rodent apical assays in the usual manner).
- The validation process for Tox21 assays would benefit from incorporating toxicity data from potential drugs that failed in clinical trials, in addition to using prototype chemicals with known toxicity. Clinical trial libraries of therapeutic drug candidates containing toxicology information regarding chemical interactions with cellular pathways may be useful to test Tox21 assays based on similar pathways.
- Data from Tox21 and NHANES should be added to databases that provide input into computational toxicology and exposure models. Computational toxicology integrates available biological and chemical data and computer sciences in an effort to predict chemical hazard potential, and it is part of a decision support tool box used to assess chemicals for potential risks to humans and the environment.
- Human data from absorption, distribution, metabolism, and excretion studies as well as physiologically based pharmacokinetic models for the chemicals that enable the estimation of internal doses in the relevant organs, tissues, and cells are insufficient; thus NHANES data *per se* may not help in computational toxicology.
- NHANES's serum levels are particularly relevant since they provide information about potential target organs, affected pathways, and relevant doses that may complement already existing animal toxicology data.
- NHANES should be extended to measure chemicals in blood in infants younger than 1 y of age and in urine of children younger than 6 y because of the potential increased risk associated with these populations.

Evaluating study design and data for regulatory decisions

The 2nd round of small-group discussions shifted from a narrow focus on study endpoints to the overall design of both guideline-based studies and hypothesis-based research. The discussions examined:

- Whether the methods to select doses for nonclinical, guidance-based studies need to be modified on the basis of research indicating low-dose effects;
- The challenge of transparency in both hypothesis-based research and guideline-based studies as well as FDA's review of the science;
- How to ensure that studies evaluated by FDA are reproducible across laboratories; and
- How FDA can make better use of hypothesis-based research in its regulatory decisions.

Dose response. Dose–response relationships can have diverse shapes—for example, linear, curvilinear, inverted-U, and so on. Figure 2 illustrates 3 examples of dose–response relationships between the serum concentration of selenium and 3 endpoints: (A) total cholesterol; (B) HDL cholesterol; and (C) triglycerides (Laclustra and others 2010). Dose–response relationships are usually classified as monotonic or nonmonotonic based on the curve's slope. When the slope of the curve does not change directions, the relationship is called monotonic, regardless of whether or not the curve is linear (Figure 2A and 2B). If the slope of the curve does change directions the dose–response relationship is called nonmonotonic (Figure 2C); the shape of nonmonotonic relationships also varies.

The discussion considered how studies that indicate nonlinear, dose–response relationships may be interpreted and incorporated into regulatory decision making. Traditional toxicology assumes that higher doses produce greater effects and that dose–response curves are monotonic. Therefore, if there are no adverse effects at high doses, it is common for toxicologists to deem that there will be no adverse effects at lower doses. For example, the Redbook recommends that chemicals be tested using at least 3 doses. Consistent with the current guidelines, the high dose must produce adverse effects and at least 1 lower dose must be identified that does not produce those effects. The highest dose that produces no adverse effects is then used as the starting point for safety evaluation/risk assessment. Adverse effects are defined in different ways in these studies and there is often inconsistency as to which effects are considered adverse. In addition, recent hypothesis-based research indicates that some biological systems, such as the endocrine system, respond to exposure in a nonmonotonic manner—that is, different effects are observed at low and high doses.

As with the discussions in round 1, the issue of defining an adverse effect came up repeatedly. The moderator identified 3 main themes in the discussion that describe the areas of significant disagreement. First, *hormonally active substances may be to be associated with nonmonotonic dose responses*. Some scientists mentioned that the shape of the dose–response curve is directly related to the substance's biochemical function—that is, chemicals that bind to receptors elicit different responses than those that do not bind to receptors—and is a reflection of the biological function of such receptors. These scientists also pointed out that some drugs are designed to be employed in a manner that takes advantage of their nonmonotonic dose response. However, some government and regulated community scientists and some academics pointed out that even where studies have found that low doses can have an effect where higher doses do not, questions of reproducibility and frequency have not been answered. Government scientists stated

that the current evidence on the relevance of nonmonotonic dose responses is insufficient to be used for regulatory decision making. Some academic scientists said that the evidence is sufficient for some substances.

Second, *doses related to human exposure based on exposure assessment or, if postmarket, on biomonitoring data or epidemiology studies should be incorporated in the dose–response relationship*. Relevant endpoints may differ with the doses, and doses related to human exposure need to be studied. Similarly, the dose response may be different for different effects, such as mutagenic or endocrine effects, and endpoints, such as those shown in Figure 2; additionally, substances may have different effects at different developmental stages. A scientist from the regulated community mentioned that, depending on the type of assay, it is common practice to take into account the human exposure to the test substance and to use the amount of the chemical in the food as a starting point for dose selection. Although there was some agreement that studies should be conducted at doses that reflect estimated human exposures in addition to higher doses, participants also pointed out that study designs using very low doses can be methodologically challenging because they would require very large numbers of animals and sensitive methods in order to yield statistically meaningful results. Study of nonmonotonic dose responses needs to take into account the translation from inbred animal strains to the highly genetically varied human population, the developmental stage of exposure, routes of administration, and the numbers of subjects required to attain sufficient statistical power. A FDA scientist also noted that it is necessary to measure internal doses of a substance due to the current lack of precision in calculating the dose delivered in a feeding study and still-developing understanding of toxicokinetics. Some academic scientists pointed out that statistical power considerations depend significantly upon the methodologies and endpoints employed in the study. They mentioned that the less sensitive apical endpoints recommended in the current guidelines require large number of animals to be able to identify effects occurring at low doses; and added that that using sophisticated methodologies directed at biologically relevant endpoints would require smaller number of animals while reaching high statistical power.

Third, *screening tests may help identify nonmonotonic responses and should help improve the design of dose–response studies to include low doses*. The group discussed, without coming to definite conclusions, whether *in vitro* toxicity screening approaches can be used to assess the potential for nonmonotonic dose response curves and to define relevant doses. Similarly, there also was discussion of whether quantitative structure–activity relationship (QSAR) studies could be used for this purpose, since QSAR quantitatively correlates a substance's chemical structure with a defined process such as biological activity or chemical reactivity.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Animals receiving lower doses may need to be assessed for endpoints other than those measured in high-dose groups because different doses may impact different endpoints.
- Relevant NHANES biomonitoring data should be considered in safety testing, especially for dose selection.
- Measuring the concentration of chemicals in human samples could be technically challenging because of very low levels present in fluids and their chemical conformation (conjugated, metabolized, and so on).

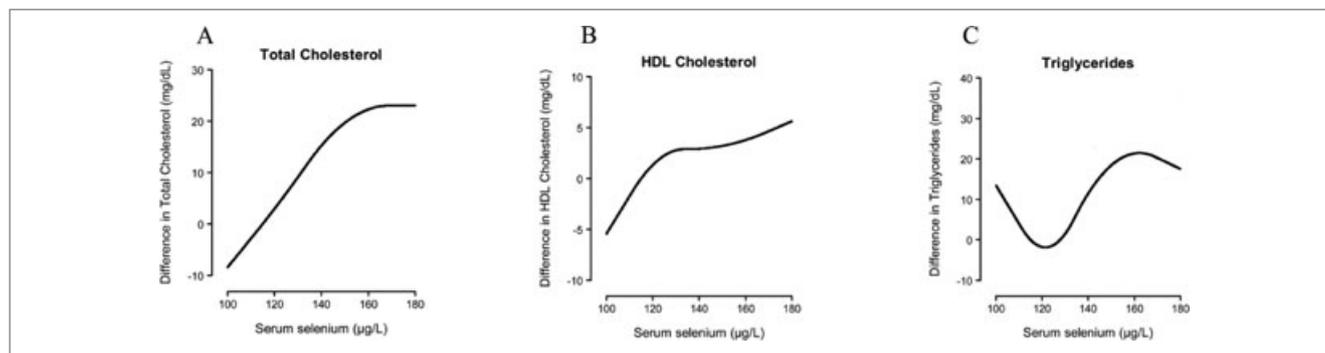


Figure 2—Examples of monotonic (A–B) and nonmonotonic (C) dose–response relationships. The shape of the relationship between serum selenium levels and the serum lipids varies with the measured endpoint. Adapted and reprinted from Laclaustra and others (2010). Copyright (2010). With permission from Elsevier.

- Research on epigenetic effects on the genome may provide clues on the mechanisms for how chemicals at low doses may affect normal development.

Transparency. This discussion considered methods to increase transparency of the data generation and analyses in hypothesis-based research and guideline-based studies. In both cases, independent analysts generally cannot access the raw data without the permission of the researcher, laboratory principal investigator, or study sponsor. While FDA may also request access to the data—and usually receives a favorable response—FDA can demand access to the data only if the study is:

- Conducted by or on behalf of the submitter requesting pre-market authorization; or
- Funded by the federal government and the funding agency requests the information.

When reporting results from guideline-based studies, the laboratory must, according to the GLP rule, provide a description of the transformations, calculations, or operations performed on the data, a summary and analysis of the data, and a statement of the conclusions drawn from the analysis. However, the author of a report from a GLP-compliant study does not have to publish it or submit it to a peer-reviewed journal (and probably could not without the permission of the study sponsor). Thus, because work that was not published and not submitted to FDA is not available for review, the FDA, the scientific community, or a food manufacturer may be unaware of relevant results when, for example, making a GRAS determination or requesting a food additive approval.

Hypothesis-based researchers seldom make their raw data available even when results are published in peer-reviewed journals. Some peer-reviewed journals require that most of the data supporting published articles be publicly available, and other journals provide researchers with the option of posting the raw data on a website to support a published article; but raw data generally remain difficult to access, often because researchers hope to preserve the right to publish future analyses using the same data.

Four themes arose in the small-group discussion. First, *the results of both hypothesis-based research and guideline-based studies would be more credible and useful if FDA and independent analysts had access to raw data, laboratory notes, and detailed data analysis so that these groups could review the analysis.* Academic scientists raised concern that this level of transparency could be disruptive and consume scarce resources and asked whether it was necessary if the article went through peer review and was published in a scientific journal. Public interest scientists also noted that if an agency like FDA needs additional information from a scientific publication, it

should approach the authors and ask for necessary raw data or any other technical information. The authors may or may not provide the requested information based on the funding contract responsibilities. These scientists also underscored that the requirements to access information should be the same for any science being reviewed by regulators, whether from industry or academia, with no double standards for public accessibility and transparency. FDA scientists mentioned that they can and do ask researchers for their data directly. Scientists from the regulated community are of the opinion that the peer-review process does not ensure unbiased review prior to publication.

A 2nd theme was *FDA may not have access to all relevant study results.* For both hypothesis- and guideline-based studies, negative results may be unreported, not accepted for publication in certain journals, or unpublished. Yet these negative data could be very important in understanding the effects of compounds and in developing model. Some FDA scientists noted that this issue is potentially significant from the standpoint of computational modeling. In addition, positive results from a guideline-based study may not be published because the study's sponsor chose not to pursue the product. Some scientists stated that mechanisms are needed to develop a publicly available repository of both positive and negative toxicology data or to change the way many journals view such data.

The 3rd theme was that *FDA should be more transparent in how it makes safety determinations, the data it uses and does not use, and how regulatory decisions are made.* Scientists from the regulated community and public interest organizations noted that they would prefer FDA to have a more transparent process for making scientific and regulatory decisions. Clear guidelines should be available regarding the type and quantity of information that stakeholders should provide to FDA and that FDA uses in its assessment. If FDA accepted the delivery of data in an electronic format, it would make the data easier to use and share with the public. Also, information should be easier to find on FDA's website. One scientist noted that EFSA posts complete safety evaluations on its website, although it acknowledges that the rationale for risk management decisions (made by the European Commission with votes from the European Union member states, Parliament or Council) are not explicitly explained. It was generally observed that FDA makes the GRAS notifications (that is, industry's notification of a GRAS determination submitted voluntarily to FDA for review) available on its website and that other types of submissions such as food contact substance notifications (FCNs) are only available through FOIA requests. FDA is not allowed by statute to make FCNs available until they become effective. Some scientists felt

that there is room for improvement in how FDA communicates with stakeholders and makes information available. Government scientists countered that more outreach to academics would help but that requires resources, which are in short supply. They also added that more information is available that the academic and public interest communities do not appear to seek out.

Fourth, *public interest scientists need access to data, methodologies, and safety determinations*. It is important for them to know which data were used and how they were used. Regulated community scientists mentioned that making information widely available also makes it available to competitors. These scientists said that an important incentive to companies to invest in research and development, including development of safety data, is having some period of competitive advantage due to the exclusive use of those data. Researchers, journals, and FDA all have roles to play in enhancing public access to data from both guideline-based studies and hypothesis-based research. It was noted that FDA publishes some studies done in support of GRAS notifications on its website. In addition, studies reviewed by FDA but not published on its website can become publicly available if a request is filed through the Freedom of Information Act (FOIA); but some scientists from the regulated, academic, and public interest communities raised concerns that the time delay in receiving the information effectively limits transparency. Government scientists discussed a variety of ways that additional transparency could be achieved as well as the resource burdens that greater transparency might require.

Reproducibility. This discussion considered how to ensure that studies evaluated by FDA are reproducible in other laboratories. The workshop materials observed that a common practice in hypothesis-based research is to publish data that have been reproduced in the laboratory several times. In addition, results published in a scientific journal undergo peer-review prior to publication, and articles are intended to be thorough enough to allow peers to reproduce the findings. However, funding for study replication is limited. For guideline-based studies, FDA generally considers studies complying with the GLP rule to be reproducible because of, for example, the large number of animals involved in studies and the strict data reporting required by GLP.

The moderator of the small-group discussion identified 3 themes. First, *reproducibility requires that a methodology is specified, is followed, and is well described*. Reproducibility is an important requirement for a methodology to be validated and included in test guidelines. A new methodology will be treated differently by FDA than a methodology used across many different groups over an extended period. For regulatory decisions, methodologies need to be described in great detail. Additionally, researchers often need to have experience with a methodology, along with good scientific judgment, to produce reproducible results. The reproducibility of a methodology should itself be studied, since data on reproducibility can help determine the certainty or uncertainty of results.

The level of certainty required for regulatory decisions depends, at least in part, on the methodology used. Reproducibility is one of the factors used in judging the weight of evidence for a scientific result. Scientists from the regulated community noted that adherence to GLP is also used by regulators and industry as an indicator of a study's reliability and reproducibility as well as the validity of its results. Some academic, government, and public interest scientists were of the opinion that GLP compliance does not ensure reliability or reproducibility especially when the study does not use internal positive and negative controls. Moreover, these scientists noted that consistency of evidence across studies

done in different laboratories is important. In addition, they asked whether FDA considers reproducibility of studies or reproducibility of endpoints, and how the agency makes safety determinations when only 1 study is available. These scientists pointed out that 1 study is generally not sufficient to be the sole source of toxicology information for decision making, especially if the study contains data gaps. They also observed that FDA does not appear to reach out to academic researchers to discuss the methods used to assess a study. Regulated community and government scientists, however, mentioned that FDA had a number of examples where this had been unsuccessfully attempted in the past. Academic scientists also suggested that when assessing nonclinical studies, the consistency of the evidence across different animal and *in vitro* studies should be favorably compared to the reproducibility of individual studies. The regulated community scientists stated their view that FDA considers consistency of findings across different types of studies as an important factor in the weight of the evidence, noting that repeating the same kind of study and getting the same results is not as important as converging evidence from different types of tests. External scientific expertise may be sought by regulatory agencies in cases where specialized knowledge is needed for data assessment and interpretation and scientific judgment.

A 2nd theme was that *reproducibility is enhanced when methodologies and data are made publicly available*. Academic scientists raised concern that this level of transparency could consume scarce resources and create problems if the data are part of a larger research effort that has not yet been published, since prepublication can undermine chances for getting the results published. Funding agencies could require that investigators allow access to the detailed methodology used and data collection and analysis so results can be reproduced, though intellectual property issues can be an obstacle. These scientists also noted that the same level of detailed methods and data analysis should apply to all studies regardless of their origin. In addition, they were of the opinion that knowing how FDA evaluates the quality and applicability of studies used in both premarket and postmarket decisions would help both industry and the public in better understanding the decision-making process. FDA scientists maintain that the information is available through a FOIA request.

Third, *sharing of tissues and other biological materials among scientists should allow piggybacking of studies to evaluate uncertain results*. For example, the NTP advises other federal agencies on matters of chemical toxicity; it performs its studies using GLP and has a large collection of biological specimens that could be used to validate endpoints, methods, and correlations between chemical exposure and adverse effects. NIEHS has implemented a similar system of sharing tissues among its grantees. Agencies could "advertise" the tissues and other samples available to researchers to test reproducibility of endpoints. Collaboration among agencies would enhance the validation of endpoints and methodologies. Additionally, a mechanism should exist to fund commercial or academic laboratories to repeat a study to examine reproducibility, depending on the level of uncertainty involved. These efforts should be pursued in collaboration with other regulatory agencies and research institutes such as NIH.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Assessments of the weight of the evidence should consider the evidence for harm and no harm across all available studies.
- FDA considers that studies and safety evaluation or risk assessment methodologies relied upon for regulatory decision

making must stand up to potential legal challenge. For example, they must use validated methods, measure endpoints known to be relevant to adverse health effects in humans, and show a dose–response relationship.

Use of hypothesis-based research. This discussion focused on changes both FDA and academic scientists should consider to make better use of hypothesis-based research when the agency conducts a safety assessment. The background materials for the workshop noted that, in general, hypothesis-based research is not conducted with the intention of being used for regulatory purposes but rather to explore new ideas and publish original data. However, some of these studies aim at identifying chemical hazard which could contribute to regulatory decision making. The content of a scientific publication is judged by peers and the format of an article is dictated by the journal. The peer reviewers determine whether or not there are sufficient value and originality in the findings to publish them. Thus, there are no homogeneous criteria for hypothesis-based research, nor is there a “1 size fits all” approach to experimental design and data reporting.

Often hypothesis-based research is conducted after a chemical has been on the market for some time. Therefore, the knowledge and technology available may be different than those available at the time the original safety determination was made, and the new research may question the safety of such chemical use. FDA reported that it reviews hypothesis-based research when it considers the safety of a substance added to food both during premarket and postmarket evaluations. Regardless of the source of the study, FDA noted that it uses 8 criteria derived from a compilation of Redbook, OECD, EPA, and WHO guidelines to determine the adequacy of data for safety assessment:

1. Route of administration;
2. Sample size and statistical analysis;
3. Validity of endpoint measured;
4. Plausibility or relevance to human health;
5. Dose response;
6. Sex of the animals;
7. Repeatability; and
8. Environmental contamination.

In reviewing these criteria, the moderator commented that the group focused on plausibility and repeatability or reproducibility. A plausible biological mechanism derived from hypothesis-based research can be extremely important in serving as the basis for regulatory consideration. Reproducibility may relate to a specific test or to the broader reproducibility of a particular outcome in a range of tests. However, 1 of the 3 themes of the discussion was that *it is important to understand the weight FDA gives to each study analyzed*. A body of evidence is needed to make safety decisions, including any evidence from hypothesis-based research. FDA may interpret the data from a hypothesis-based study in a somewhat different way than do the authors of the study, who are likely to be more interested in exploring the mechanisms of adverse effects than in using the data to make a regulatory decision. FDA maintains that information concerning its safety assessment of food additives is available through FOIA requests. Academic scientists noted that there do not appear to be systematic established communication channels that FDA could use to interact with the academic scientists who conduct basic research and vice versa. However, scientists from the regulated community noted that no codified barriers exist for communication with or by FDA and that numerous avenues for such communication currently exist in FDA’s website.

Second, *education of the academic community regarding the methodologies and endpoints that are valuable to regulators would increase the useful-*

ness of hypothesis-based research for regulatory purposes. Education is a 2-way street: academics would benefit from learning about FDA’s criteria to evaluate science, while FDA scientists would benefit from learning more about new scientific and health results, trends, and methodologies. FDA’s scientists noted that they closely follow scientific developments and participate in professional societies in spite of the constraints on resources.

The 3rd theme was that *collaboration among agencies and better communication would enhance the development of science that is more useful for regulatory purposes*. Federal agencies could hold workshops to focus on key issues. Fact sheets or agency requests could explain to academic researchers the types of studies that are needed. Contracts or requests for funding proposals could encourage investigators to make their research more relevant to public health and the regulatory process. Mechanisms to stimulate translational research also could bridge the gap between these areas of scientific endeavor. Scientists from the regulated community noted that federal funding opportunities for toxicological research have become more difficult to obtain. In some cases, toxicological subspecialty researchers must compete for funding with all researchers in that particular subspecialty of science as the number of NIH study sections has been reduced.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Regulatory agencies should consider support for additional research in areas of biomedical research fields that produce new findings relevant to public health.
- Communication and collaboration between regulatory and research funding agencies should be enhanced to stimulate targeted investigation conducted using some of the principles of GLP, recognizing that academic institutions may not have the capability or funding to perform fully GLP-compliant studies while promoting hypothesis-driven research.
- Hypothesis-based research studies that meet FDA quality criteria can be very useful to FDA in understanding mechanisms of action, dose response, and other factors that enable the agency to determine if further testing or evaluation is needed and, in appropriate cases, whether a substance meets the safety standard of reasonable certainty of no harm.

Developing and reviewing test guidelines

The 3rd round of discussions focused on the development, review, and approval of test guidelines to be included in the Redbook. Participants in 4 small groups were asked to examine 1 of 2 distinct steps in guideline approval: (1) the development, validation, and submission of new or improved draft test guidelines to FDA for consideration, and (2) the review, management, and approval of new or improved draft test guidelines. As with the entire workshop, groups were not expected to reach a consensus but to raise issues and consider and compare current and possible procedures.

Developing test guidelines for review. As a result of the discussions that took place in the 2 sessions, the moderators identified 3 themes. First, *a variety of barriers limit the development and validation of new or improved draft test guidelines*. The most obvious barrier is a lack of financial and human resources, both within FDA and elsewhere. FDA does not currently have the means of undertaking a major program of seeking out and incorporating new testing methods. Academic scientists doing hypothesis-based research face similar constraints since a major source of their funding is the

federal NIH. Industry often supports the development of new and improved tests, but it lacks a clear mandate or incentive to initiate efforts to modify the guidelines and to provide significant financial support, especially since new test methods may require costly modifications to current testing regimes. Scientists from the regulated community stated that industry resources are limited and that a stakeholder process may be advisable to help identify promising possibilities and set priorities for test method development. They indicated that industry has a role to play (this may include funding); however, they must be confident that any new test will improve regulatory decision making. Scientists from the regulated community maintained that new tests must be valid, reliable, relevant, and adequately predictive of effects that are known to be associated with human toxicity or disease.

Additional costs to updating test guidelines include the financial burden of developing and validating new tests and disseminating information about new and improved test methods to the scientific community. Other hindrances include limitations in data availability and usability, and a lack of incentives to develop new tests. Regulated community scientists pointed out that it is important to develop a national strategy to establish priorities for the development and validation of test guidelines. Other regulatory community scientists noted that this already happens with regard to test guidelines developed, validated, and used in the OECD member countries through the national coordinator for each country.

The 2nd theme was that *academic scientists do not believe that FDA does enough to solicit their opinion about its guidance and incorporate their concerns into revisions*. Some academic scientists noted that there may be important information in academic laboratories that could enhance the quality and accuracy of the decision-making process, yet they often do not know whom to contact within FDA to submit their information. During the workshop, FDA explained that it has regulations guiding its development and revision of guidance documents. The rule makes clear that FDA is always open to comments and suggestions. It also publishes its priorities for additions and changes annually in the *Federal Register* to make it easier to know what is coming. However, academic scientists may not closely track the *Federal Register* or these developments. FDA staff expressed their willingness to meet with academic investigators, but they added that forums for doing so are scarce and meetings between regulators and academic researchers rarely occur. One scientist suggested that FDA could hold stand-alone workshops or sponsor them alone or cooperatively at the national or regional chapter meetings of relevant professional organizations such as the Society of Toxicology or the American College of Toxicology, as EPA frequently does. Some academic scientists mentioned that 8 national scientific societies have offered FDA (and EPA) access to leading scientists in diverse fields by means of a public letter published in the journal *Science* (Hunt 2011). In the plenary session, FDA acknowledged the public letter but indicated that it has not yet decided how or when to reach out to the scientific societies.

Clear procedures for developing and validating new test guidelines similar to those by ICCVAM or OECD would be useful. These procedures should consider several questions, including: How should the process of test development and validation be initiated? What are the requirements for this process? What level of validation of tests is needed before submitting a draft guideline? How is a causal association with human disease determined? One approach would be to create a publicly available template detailing the requirements to submit a draft test guideline.

The 3rd theme was that *new funding approaches could enhance the resources for test development*. Multiple federal agencies, such as FDA,

EPA, and NIH, could enhance their coordinated efforts to develop tests that would have benefits for each agency. A concerted effort among stakeholders, including academia, industry, nonprofit organizations, and government, could align efforts and raise the profile of test development. International harmonization through organizations such as Codex (<http://www.codexalimentarius.net>) and OECD could increase the efficiency of test development, though different legal standards and procedures can limit coordination.

Participants also discussed what the appropriate “gold standard” for a new or improved test should be. New and improved tests should provide a level of assurance for consumer safety that is equal or higher than the current level. Should a test be evaluated against a preexisting assay, or should it have a specified predictive value associated with human disease or performance level? Also, a gold standard can change as assays improve, societal values evolve, or biological understanding advances; therefore, the system should be easy to update or adapt.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- FDA and ICCVAM should work together to incorporate new validated methods into the Redbook.
- Guidelines currently used should be systematically reviewed. Several of the currently accepted endpoints, such as organ weight, were grandfathered into the Redbook because of their long history of use and repeatability more than their relevance.
- FDA needs to more effectively make use of all the new knowledge acquired through federal research programs, including Tox21.
- Guidelines incorporating new endpoints into current protocols should be flexible while maintaining the integrity of the regulatory system.

Reviewing and approving test guidelines. The moderators noted that FDA currently has a process to improve guidelines consistent with its good guidance practices rule. Each year, FDA publishes in the *Federal Register* a list of guidelines in development, although it remains unclear how the list is produced and how priorities are identified. FDA regulators also consult with experts in other agencies and outside government and review and consult scientific findings from the United States and the rest of the world. Nevertheless, the process of reviewing and approving new or improved draft test guidelines remains largely *ad hoc* in comparison with procedures used elsewhere (for example, by the OECD). Developing new and improved food additive safety tests is not the highest priority for any 1 office within FDA, nor is there a transparent process to refine and improve tests based on experience or observations made during guideline-based studies.

A major theme of the discussion sessions was that *developing a prioritization system for validating new test guidelines could focus FDA's limited resources on the most pressing public health concerns—for example, those with a known relationship to national rates of morbidity and mortality*. For example, sensitive tests and endpoints for detection of early markers of diabetes, high blood pressure, or obesity could be a top priority. The regulated community may be motivated to support changes that will reduce the cost of the toxicity tests, and it is motivated to support changes that are confirmed to be relevant to the demonstration of safety. Academic scientists were unaware of or poorly informed about the opportunities to introduce proposals for reviewing new or improved guidelines. FDA could be more effective in reaching out to academic scientists, while academic scientists could use existing points of contacts with FDA, such as scientific meetings, more effectively.

A 2nd theme of the small group discussion was that *greater transparency in FDA processes would improve predictability and access to information*. Areas that would benefit from greater transparency include premarket and postmarket assessment, inclusion, and exclusion criteria for scientific studies, data sources, analytical techniques, and handling and communication of uncertainties. However, some FDA scientists pointed out that greater transparency also needs to be weighed against the potential to bog down an already slow approval process, thereby slowing public health decision making. FDA scientists also noted that perhaps more important than bogging down the process is the need for independent review and to shelter reviewers from influences outside and inside the agency. This works against transparency but is absolutely required for a science-based process. FDA scientists stated that they must be able to independently consider the information before them, document their defensible conclusions, those conclusions must be captured in the record, and that no one has the right to look over these scientists' shoulder while they write their review. A participant noted that EPA's Office of Pesticide Programs could serve as an example of a regulatory office having implemented efficiencies that helped mitigate slowing down the process while improving transparency.

Finally, *other agencies and organizations have established procedures for reviewing and accepting test guidelines; coordinating with and adopting those guidelines could save FDA time and resources*. For example, OECD has a rigorous and transparent process for reviewing and accepting guidelines designed to address current public health priorities, and this process could act as a model for FDA. FDA also participates with other federal agencies in the harmonization of U.S. policies and regulations for presentation in global forums, and interactions between FDA and these agencies occur on many levels. However, some participants cautioned that harmonization among agencies can have pitfalls. For example, building a consensus among agencies may result in stifled innovation or create a lowest common denominator effect. Others countered that adopting a process that works for another agency will serve to increase efficiency and is certainly an improvement over having no procedures in place.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- FDA is unlikely to expend limited resources on the review of guidelines that it perceives as not providing utility in safety assessment (for example, guidelines that involve evaluation of effects that FDA currently does not consider adverse).
- Premarket and postmarket scientific developments provide unique challenges for the agency. In the premarket period, FDA maintains that it has control over decisions based on current science. As new science evolves in the postmarket period, FDA does not systematically review new information or reassess its decisions. FDA did undertake systematic reviews in the 1960s and 1970s and found that there were few, if any, concerns for thousands of decisions and substances. Therefore, FDA maintains that any review system would need to be carefully prioritized in light of the agency's limited resources.
- FDA could request that scientific societies, funding agencies, and peer-reviewed journals provide mechanisms for disseminating information to the public about currently accepted regulatory practices, in addition to publishing Redbook guidance in the *Federal Register*.

Identifying and evaluating potential solutions

The final breakout discussions were designed to be brainstorming sessions to identify and evaluate potential solutions to existing problems. Participants were asked to focus on ways to improve hypothesis-based research, improve guideline-based studies, and refine the regulatory decision-making process.

Improving hypothesis-based research. The moderator identified 3 themes that emerged from the discussion of hypothesis-based research. First, *hypothesis-based studies could be modified to be more reflective of the needs and procedures of regulators to enhance their use in regulatory decision making*. Such modifications could include increased emphasis on:

- protocol design;
- characterization of the test substance and matrix;
- dose response;
- statistical methods;
- endpoints relevant to human health; and
- whole-organism evaluations.

Academic scientists pointed out that the strongest incentive would be for funding agencies that support hypothesis-based research, such as NIH, to make relevance to specific regulatory needs part of the criteria for evaluating and making funding decisions on grant applications.

Second, *open communication between all stakeholders and a better understanding of FDA processes among scientists doing hypothesis-based research could enhance the usefulness of this research for regulators*. To improve dialog, academic investigators could meet with FDA staff in formal or informal forums. An increased number of fellowships on the part of both academic institutions and federal agencies could enhance information-sharing. FDA could issue fact sheets and requests for data, through other means than the *Federal Register*, to alert researchers of opportunities. Professional societies, expert scientist meetings, or a common electronic platform could be used for discussions, for submissions to FDA, for data mining, and for requests for information from regulatory scientists. Greater access to data from guideline-based studies could help academic investigators integrate hypothesis-based research with regulatory decision-making processes. More interactions with research scientists in general could help the FDA scientists better appreciate new developments in toxicity science. Equally important is improve to access by FDA to raw data from hypothesis-based research.

Professional societies, funding agencies, and journals could encourage collaboration among laboratories to enhance the availability of resources and support the production of studies relevant to regulators. One scientist suggested that the regulated community could support this work through a third-party, multisource funding mechanisms to reduce conflict-of-interest concerns.

Finally, *early training in academic settings can lead to a greater understanding of the regulatory process and incorporation of hypothesis-based studies into regulatory decision making*. Curricula should include an introduction to the legislative framework and the rules that govern the testing and safety evaluations of chemicals added to food. By incorporating such materials into training, students would gain a greater understanding of the role of regulatory science and how their future work might be used in the regulatory arena.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Journal editors should provide more space for inclusion of detailed information on protocols, standard operating procedures, raw data, statistical methods, and negative

results. Journals could also require the inclusion of raw data and procedures in supplementary materials.

- All studies, including guideline-based studies, must be held to similar publication standards. Publishing detailed methodological information and raw data is considered unnecessary since FDA can always ask for the information. If FDA does not get the requested information, then FDA has a basis to have less confidence in the data. Scientists doing hypothesis-based research should particularly consider how to design experiments to extrapolate from laboratory results to humans and from high doses to low doses.

Improving guideline-based studies. The moderator stated that many participants in the discussion on guideline-based studies observed that such studies have strengths and weaknesses. For example, they can investigate durations and types of exposures that are unlikely to be investigated in hypothesis-based research. Their use of standardized protocols can reduce variability, enhance reproducibility, and provide useful information to regulators, whose decisions must meet applicable statutory and regulatory criteria, and to policymakers. Furthermore, some scientists noted that the Redbook is not a prescriptive list of studies required in every case and can be modified as needed to address specific issues, including emerging public health questions. However, guideline-based studies, which constitute the recommended minimum acceptable standard information package for demonstrating the safety of a chemical's particular intended use, incorporate only endpoints that have been validated as being reliably predictive of adverse health effects. Therefore, they may not ensure that investigators seek to observe particular endpoints that may or may not correlate with emerging research questions.

One theme of the discussion was that *FDA could request the development of new and improved guidelines and other information needed for decision making*. Perhaps even more feasible than developing new guidelines would be to identify additional endpoints or methods that, once validated, would increase the sensitivity of the current guidelines to better identify adverse effects related to public health. For instance, a standing committee could review animal test methods proposed to be incorporated into the guidelines similar to ICCVAM's.

FDA has a range of options for informing the academic community of its needs. FDA could use its own publications, make requests to professional societies, or establish interactions with universities. Academic researchers may not be aware of existing opportunities to work with FDA. Food safety scientists could work with those in academia to help them understand what risk assessment entails and the statutory requirements for food safety.

One idea proposed is to use results or methodologies from hypothesis-based research to develop potential add-ons for guideline-based studies. For example, could an ancillary study of a substance be guided by specific endpoints if those endpoints were sufficiently validated as being relevant to human disease? In many cases, FDA would need to specify, possibly within the Redbook, how the add-ons would relate to the existing guideline-based studies.

A 2nd theme of the discussion was that *many of these steps would require increased resources at FDA*. No-cost or low-cost opportunities may be available to make greater use of hypothesis-based research if this were accorded a higher priority. A combination of private and public funding will be needed to develop and validate new or improved draft test guidelines. An integrated group of agency, industry, and academia representatives could coordinate increased funding. Other federal and international agencies

such as ICCVAM or the Health and Environmental Sciences Inst. of the International Life Sciences Inst. offer potential models for this coordination, though they work in different areas and at a different level of technical detail. In addition, the establishment of priorities may lead to opportunities for increased research funding.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Improving existing guidelines might be more efficient than developing new ones from scratch in some cases. The Redbook would have more flexibility if it allowed for the addition of new endpoints whenever the toxicology data pointed toward new leads worth pursuing.
- Journal editors could provide more space for inclusion of detailed information on protocols, standard operating procedures, raw data, statistical methods, and negative results. Journals also could require the inclusion of raw data and procedures in supplementary materials.

Refining the regulatory process. The moderators observed that *the systematic updating of guidelines would ensure that sensitive and relevant toxicology tests are used in assessing the safety of substances added to food*. The Redbook is published on FDA's website in distinct chapters and sections, which allows the guidance to be more easily and routinely updated. It is a "living document"; FDA can revisit it and update it as new scientific information is presented. FDA's electronic publication of the Redbook demonstrates its commitment to making the document more amenable to changes as new scientific methods relevant to human disease are validated. Although some participants saw this as a positive approach, they also mentioned that FDA could improve their efforts in maintaining an updated Redbook that more clearly reflects scientific advances. This can be especially important as Tox21 methods develop and are better understood and validated. Clear guidance on incorporation of new methods into GRAS determinations, where industry is not required to inform FDA of its actions but uses the Redbook as guidance to demonstrating safety, is important to both industry and FDA.

A 2nd theme of the discussion was that *a prioritized cyclic review of safety decisions can incorporate new science, postmarket surveillance data, and the "human experience"—what actually happens with food ingredients in the body*. The agency could develop a process for reviewing regulations and decisions that is flexible enough to respond to resource limits and changing needs. FDA should articulate and clarify the triggers and processes for postmarket evaluation or reevaluation, since there is no requirement established in regulation. Postmarket decisions also typically involve more and different kinds of data than do premarket approvals. Some participants noted that the review process could be prioritized based on current scientific developments, technological advances, and human exposure data. All stakeholders (including federal agencies, industry, academics, advocacy groups, international bodies, and the public) need to be involved in identifying these priorities and the resources needed to pursue them. Stakeholders could be convened by a neutral third party such as the National Research Council.

FDA noted that there is a balance between transparency and efficiency. Enhancing transparency can slow the decision-making process. Introducing time limits for decision making and reducing transparency can improve efficiency. For example, with food contact substances (under the Food Contact Notification procedures initiated with the passage of the FDA Modernization

Act of 1997), confidential discussions can take place between FDA and manufacturers to remedy problems before a public notice is issued. Approximately 100 food contact notifications are done each year and remain confidential until a decision is made.

Finally, *all stakeholders would benefit if the regulatory decision-making process were more user friendly*. The *Federal Register*, Notices of Proposed Rulemaking, and responses to FOIA requests provide important due process and level the playing field for all potential stakeholders but may not always be clear or intuitive for communicating with stakeholders or the public. A centralized portal for information could be customized to provide information to stakeholders. Communication between FDA and stakeholders requires coordination among regulating agencies where authorities overlap. Communications between FDA and the regulated community, other experts, and the general public requires the use of different vocabularies based on the audience that is reached. Webinars and websites, similar to EPA's new website or FDA's GRAS notification program website, could improve the quantity and quality of communications. In general, the messages from various agencies need increased cohesiveness, not only within the Dept. of Health and Human Services but throughout all federal agencies.

Several other observations and suggestions from some participants arose during the discussion but did not rise to the level of a theme:

- Resource constraints make cyclic reviews challenging and therefore require that resources be targeted toward those regulated materials of greatest potential concern either because of updated exposure assessments or because of new testing data that indicate unanticipated adverse effects that may not have been anticipated when the substance was originally reviewed for safety.
- Enhancing protective capability needs to be done in a way that does not impede the decision-making system.

Postmeeting note: *On May 9, 2011, the National Research Council issued a report entitled "A risk-characterization framework for decision making at the Food and Drug Administration" (Lawrence 2011). The report describes a risk-characterization framework that can be used to evaluate and compare the public health consequences of different decisions concerning a wide variety of products. Recommendations made in the report include adding risk perception and public attitudes about risk (for example, how much control one has over eliminating or reducing risk, and the ability of institutions to detect and/or mitigate adverse effects) to the traditional risk attributes of exposed population, mortality, and morbidity that are used for decision making. It also highlighted the value of multiple points of view and the need for subject-matter experts to identify and evaluate relevant data.*

Conclusion

The workshop "Enhancing FDA's Evaluation of Science to Ensure Chemicals Added to Human Food Are Safe" was held under ground rules that called for constructive engagement, and the participants answered that call. The discussions and informal conversations aimed to provide the participants with a better understanding of the safety assessment system for substances added to food, including its complexity, strengths, and weaknesses. Discussions of specific topics, some highly controversial, successfully allowed the conversation to move beyond specific additives to broader observations. In addition, the workshop contributed to FDA's *Advancing Regulatory Science Initiative* by developing ideas and approaches to adapting science at FDA to meet the challenges of increasingly complex issues and products.

Although there was no intention to reach a consensus during the 2 d of discussions, the authors note that several topics emerged repeatedly, including:

- Importance of communication and outreach between several groups, including between FDA and the scientific community at large, between FDA and stakeholders, and among scientists working on different aspects of research about substances added to food.
- Transparency of the criteria FDA uses to evaluate scientific data submitted to the agency, the decisions made for or against the submission of a substance intended to be added to food, and the strategies used to keep toxicology tests current according to scientific developments and human health relevance.
- New research methods with demonstrated relevance to human toxicity or disease need to be validated and incorporated into the Redbook.
- Importance of postmarket assessment, including strategies and priorities for cyclic reviews of substances added to foods that are already in commerce.
- Lack of a clear definition of harm and adverse health effects results in inconsistencies and confusion among stakeholders as to the risk assessment process.
- Importance of enhancing the Redbook by regularly making updates to stay abreast with scientific developments and to ensure that all safety determinations are made using sensitive and relevant scientific methods within the regulatory framework based on the principle of prevention.
- Importance of increasing funding available for developing and implementing new or revised test guidelines.
- Opportunities and incentives to improve hypothesis-based research to make it more useful to regulatory decision making should be leveraged.

Continuing with our assessment of the food additives regulatory system, the Pew Health Group will hold a similar workshop that will focus on dietary exposure assessment, another major component of a chemical's safety evaluation, in the fall of 2011.

Acknowledgments

The authors thank the workshop participants, sponsors, small group discussion moderators, facilitators, and speakers for their valuable contributions to the workshop discussions and the development of this article. We specially thank the Institute of Food Technologists', U.S. Food and Drug Administration's, National Inst. of Environmental Health Sciences', Grocery Manufacturers Association's, Natural Resources Defense Council's scientists as well as Pew Health Group's expert advisors for their comments during the review process. The authors also thank Steve Olson and Katherine Portnoy for editorial assistance in preparing this summary of the workshop "Enhancing FDA's Evaluation of Science to Ensure Chemicals Added to Human Food are Safe," which was held in Washington, D.C., on April 5–6, 2011. The authors are grateful to Erik Olson and Shelly Hearne for their support and advice. This work was supported solely by The Pew Charitable Trusts.

References

- Becker RA, Janus ER, White RD, Kruszewski FH, Brackett RE. 2009. Good laboratory practices and safety assessments. *Environ Health Perspect* 117(11):A482–3; author reply A483–4.

- [EFSA] European Food Safety Authority. 2011. Guidance on risk assessment concerning potential risks arising from applications of nanoscience and nanotechnologies to food and feed. Available from: http://www.bionanomedcat.org/documentacio/efsa_guidance_nanofood.pdf. Accessed Jul 15, 2011.
- [FDA] Food and Drug Administration (U.S.). 2000. Toxicological principles for the safety assessment of food ingredients. Available from: <http://www.fda.gov/Food/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/FoodIngredientsandPackaging/Redbook/default.htm>. Accessed Jul 15, 2011.
- [FDA] Food and Drug Administration (U.S.). 2007. Nanotechnology task force report 2007. Available from: <http://www.fda.gov/ScienceResearch/SpecialTopics/Nanotechnology/NanotechnologyTaskForceReport2007/default.htm>. Accessed Jul 15, 2011.
- [FDA] Food and Drug Administration (U.S.). 2010a. Caffeinated alcoholic beverages. Available from: <http://www.fda.gov/Food/FoodIngredientsPackaging/ucm190366.htm>. Accessed Jul 15, 2011.
- [FDA] Food and Drug Administration (U.S.). 2010b. Advancing regulatory science for public health. Available from: <http://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/ucm228131.htm>. Accessed Jul 15, 2011.
- [FDA] Food and Drug Administration (U.S.). 2011a. Considering whether an FDA-regulated product involves the application of nanotechnology. Available from: <http://www.fda.gov/RegulatoryInformation/Guidances/ucm257698.htm>. Accessed Jul 15, 2011.
- Food and Drug Administration (U.S.). 2011b. Supplemental materials for "Enhancing FDA's Evaluation of Science to Ensure Chemicals Added to Human Food are Safe" Workshop. Available from: http://www.pewtrusts.org/our_work_detail.aspx?id=1203.
- Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, O'Heir CE, Mitchell ML, Hermos RJ, Waisbren SE, Faix JD, Klein RZ. 1999. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med* 341(8):549–55.
- Hunt PA. 2011. Assessing chemical risk: societies offer expertise science. 331:1136.
- Koplan JP, Catharyn T, Liverman, and Vivica I. Kraak. 2005. Preventing childhood obesity: health in the balance. Washington, D.C.: National Academies Press.
- Krewski D. 2007. Toxicity testing in the 21st century: a vision and a strategy. Washington, D.C.: National Academies Press.
- Laclaustra M, Stranges S, Navas-Acien A, Ordovas JM, Guallar E. 2010. Serum selenium and serum lipids in US adults: National Health and Nutrition Examination Survey (NHANES) 2003–2004. *Atherosclerosis* 210(2):643–48.
- Lawrence RS. 2011. A risk-characterization framework for the decision-making at the food and drug administration. Washington, D.C.: National Academies Press.
- Levi J, Vinter S, St. Laurent R, Segal L. 2010. F as in Fat. Trust for America's Health. Available from: <http://healthyamericans.org/reports/obesity2010/>.
- Maynard A. 2009. MINChar initiative. Characterization Matters. Available from: <http://characterizationmatters.org/>. Accessed Jul 15, 2011.
- Myers JP, vom Saal FS, Akingbemi BT, Arizono K, Belcher S, Colborn T, Chahoud I, Crain DA, Farabollini F, Guillette LJ, Jr., Hassold T, Ho SM, Hunt PA, Iguchi T, Jobling S, Kanno J, Laufer H, Marcus M, McLachlan JA, Nadal A, Oehlmann J, Olea N, Palanza P, Parmigiani S, Rubin BS, Schoenfelder G, Sonnenschein C, Soto AM, Talsness CE, Taylor JA, Vandenberg LN, Vandenberg JG, Vogel S, Watson CS, Welshons WV, Zoeller RT. 2009. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. *Environ Health Perspect* 117(3):309–15.
- Nature. 2010. The weight of the evidence. Better chemical-control legislation is a good start, but scientific reform should parallel legal reform. *Nature* 464:2.
- Neltner TG, Kulkarni NR, Alger HM, Maffini MV, Bongard ED, Fortin ND, Olson ED. 2011. Navigating the U.S. Food Additive Regulatory Program. *Comp Rev Food Sci Food Safety* 10(6):342–370.
- [OECD] Organisation for Economic Co-Operation and Development. 2010a. OECD guidelines for the testing of chemicals. Available from: http://www.oecd.org/document/40/0,3343,en_2649_34377_37051368_1_1_1_1,00.html. Accessed Jul 15, 2011.
- [OECD] Organisation for Economic Co-Operation and Development. 2010b. OECD Programme on the Safety of Manufactured Nanomaterials 2009–2012: Operational plans of the projects. Available from: http://www.oecd.org/findDocument/0,3770,en_2649_37015404_1_119687_1_1_1,00.html. Accessed Jul 15, 2011.
- The Pew Health Group. 2011. Workshop on Enhancing FDA's Evaluation of Science to Ensure Chemicals Added to Human Food are Safe. Available from: http://www.pewtrusts.org/our_work_detail.aspx?id=1203. Accessed Oct 4, 2011.
- Press F. 1983. Risk assessment in the federal government: managing the process. Washington, D.C.: National Research Council. p. 191.
- Tyl RW. 2009. Basic exploratory research versus guideline-compliant studies used for hazard evaluation and risk assessment: bisphenol A as a case study. *Environ Health Perspect* 117(11):1644–51.
- WHO. 2011. Methods for chemical assessment. Available from: <http://www.who.int/ipcs/methods/en/>. Accessed Jul 15, 2011.