
# Using 21st Century Science to Improve Risk-Related Evaluations



**DETAILS**

200 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-45348-6 | DOI: 10.17226/24635

**CONTRIBUTORS**

Committee on Incorporating 21st Century Science into Risk-Based Evaluations; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine

# USING 21ST CENTURY SCIENCE TO IMPROVE RISK-RELATED EVALUATIONS

Committee on Incorporating 21st Century Science into Risk-Based Evaluations

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

A Report of

*The National Academies of*
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
*Washington, DC*
**www.nap.edu**

**THE NATIONAL ACADEMIES PRESS** **500 Fifth Street, NW** **Washington, DC 20001**

This activity was supported by Contract EP-C-14-005, TO#0002 between the National Academy of Sciences and the US Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

Additional copies of this publication are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

International Standard Book Number-13: 978-0-309-45348-6
International Standard Book Number-10: 0-309-45348-8
Digital Object Identifier: 10.17226/24635
Library of Congress Control Number: 2017930534



Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2017. *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: The National Academies Press. doi: 10.17226/24635.

*The National Academies of*
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C.D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at www.national-academies.org.

*The National Academies of*
SCIENCES • ENGINEERING • MEDICINE

**Reports** document the evidence-based consensus of an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and committee deliberations. Reports are peer reviewed and are approved by the National Academies of Sciences, Engineering, and Medicine.

**Proceedings** chronicle the presentations and discussions at a workshop, symposium, or other convening event. The statements and opinions contained in proceedings are those of the participants and have not been endorsed by other participants, the planning committee, or the National Academies of Sciences, Engineering, and Medicine.

For information about other products and activities of the National Academies, please visit nationalacademies.org/whatwedo.

## COMMITTEE ON INCORPORATING 21ST CENTURY SCIENCE INTO RISK-BASED EVALUATIONS

*Members*

**JONATHAN M. SAMET** (*Chair*), University of Southern California, Los Angeles
**MELVIN E. ANDERSEN**, ScitoVation, Research Triangle Park, NC
**JON A. ARNOT**, ARC Arnot Research & Consulting, Toronto, ON
**ESTEBAN BURCHARD**, University of California, San Francisco
**GEORGE P. DASTON**, Procter & Gamble, Mason, OH
**DAVID B. DUNSON**, Duke University, Durham, NC
**NIGEL GREENE**, AstraZeneca, Waltham, MA
**HEATHER B. PATISAUL**, North Carolina State University, Raleigh
**KRISTI PULLEN FEDINICK**, Natural Resources Defense Council, Washington, DC
**BEATE R. RITZ**, University of California, Los Angeles
**IVAN RUSYN**, Texas A&M University, College Station
**ROBERT L. TANGUAY**, Oregon State University, Corvallis
**JUSTIN G. TEEGUARDEN**, Pacific Northwest National Laboratory, Richland, WA
**JAMES M. TIEDJE**, Michigan State University, East Lansing
**PAOLO VINEIS**, Imperial College London, UK
**MICHELLE WILLIAMS**, Harvard School of Public Health, Boston, MA
**FRED WRIGHT**, North Carolina State University, Raleigh
**LAUREN ZEISE**, California Environmental Protection Agency, Oakland

*Staff*

**ELLEN K. MANTUS**, Project Director
**MARILEE SHELTON-DAVENPORT**, Senior Program Officer
**NORMAN GROSSBLATT**, Senior Editor
**MIRSADA KARALIC-LONCAREVIC**, Manager, Technical Information Center
**RADIAH ROSE-CRAWFORD**, Manager, Editorial Projects
**IVORY CLARKE**, Research Assistant

*Sponsors*

US ENVIRONMENTAL PROTECTION AGENCY
US FOOD AND DRUG ADMINISTRATION
NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES
NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES

## BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY[1]

*Members*

**WILLIAM H. FARLAND** (*Chair*), Colorado State University, Fort Collins
**PRAVEEN AMAR**, Independent Consultant, Lexington, MA
**RICHARD A. BECKER**, American Chemistry Council, Washington, DC
**E. WILLIAM COLGLAZIER**, American Association for the Advancement of Science, Washington, DC
**DOMINIC M. DITORO**, University of Delaware, Newark
**DAVID C. DORMAN**, North Carolina State University, Raleigh
**CHARLES T. DRISCOLL, JR.**, Syracuse University, Syracuse, NY
**ANNE FAIRBROTHER**, Exponent (Retired), Philomath, OR
**LINDA E. GREER**, Natural Resources Defense Council, Washington, DC
**WILLIAM E. HALPERIN**, University of Medicine and Dentistry of New Jersey, Newark
**STEVEN P. HAMBURG**, Environmental Defense Fund, New York, NY
**ROBERT A. HIATT**, University of California, San Francisco
**PHILIP K. HOPKE**, Clarkson University, Potsdam, NY
**SAMEUL KACEW**, University of Ottawa, Ontario
**H. SCOTT MATTHEWS**, Carnegie Mellon University, Pittsburgh, PA
**THOMAS E. McKONE**, University of California, Berkeley
**ROBERT PERCIASEPE**, Center for Climate and Energy Solutions, Arlington, VA
**MARK A. RATNER**, Northwestern University, Evanston, IL
**JOAN B. ROSE**, Michigan State University, East Lansing
**GINA M. SOLOMON**, California Environmental Protection Agency, Sacramento
**ROBERT M. SUSSMAN**, Sussman and Associates, Washington, DC
**DEBORAH L. SWACKHAMMER**, University of Minnesota, St. Paul
**PETER S. THORNE**, University of Iowa, Iowa City
**JOYCE S. TSUJI**, Exponent, Inc., Bellevue, WA

*Senior Staff*

**JAMES J. REISA**, Senior Director
**ELLEN K. MANTUS**, Scholar and Director of Risk Assessment
**RAYMOND A. WASSEL**, Scholar and Director of Environmental Studies
**SUSAN N.J. MARTEL**, Senior Program Officer for Toxicology
**MIRSADA KARALIC-LONCAREVIC**, Manager, Technical Information Center
**RADIAH ROSE-CRAWFORD**, Manager, Editorial Projects

[1]This study was planned, overseen, and supported by the Board on Environmental Studies and Toxicology.

## OTHER REPORTS OF THE BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense (2015)
Review of California's Risk-Assessment Process for Pesticides (2015)
Sustainability Concepts in Decision-Making: Tools and Approaches for the US Environmental Protection Agency (2014)
Rethinking the Components, Coordination, and Management of the US Environmental Protection Agency Laboratories (2014)
Review of the Formaldehyde Assessment in the National Toxicology Program 12th Report on Carcinogens (2014)
Review of the Styrene Assessment in the National Toxicology Program 12th Report on Carcinogens (2014)
Review of EPA's Integrated Risk Information System (IRIS) Process (2014)
Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose-Response Relationships as They Apply to Endocrine Disruptors (2014)
Assessing Risks to Endangered and Threatened Species from Pesticides (2013)
Science for Environmental Protection: The Road Ahead (2012)
Exposure Science in the 21st Century: A Vision and a Strategy (2012)
A Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials (2012)
Macondo Well *Deepwater Horizon* Blowout: Lessons for Improving Offshore Drilling Safety (2012)
Feasibility of Using Mycoherbicides for Controlling Illicit Drug Crops (2011)
Improving Health in the United States: The Role of Health Impact Assessment (2011)
A Risk-Characterization Framework for Decision-Making at the Food and Drug Administration (2011)
Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (2011)
Toxicity-Pathway-Based Risk Assessment: Preparing for Paradigm Change: A Symposium Summary (2010)
The Use of Title 42 Authority at the US Environmental Protection Agency: A Letter Report (2010)
Review of the Environmental Protection Agency's Draft IRIS Assessment of Tetrachloroethylene (2010)
Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use (2009)
Contaminated Water Supplies at Camp Lejeune: Assessing Potential Health Effects (2010)
Review of the Federal Strategy for Nanotechnology-Related Environmental, Health, and Safety Research (2009)
Science and Decisions: Advancing Risk Assessment (2009)
Phthalates and Cumulative Risk Assessment: The Tasks Ahead (2008)
Estimating Mortality Risk Reduction and Economic Benefits from Controlling Ozone Air Pollution (2008)
Respiratory Diseases Research at NIOSH: Reviews of Research Programs of the National Institute for Occupational Safety and Health (2008)
Evaluating Research Efficiency in the US Environmental Protection Agency (2008)
Hydrology, Ecology, and Fishes of the Klamath River Basin (2008)
Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment (2007)
Models in Environmental Regulatory Decision Making (2007)
Toxicity Testing in the 21st Century: A Vision and a Strategy (2007)
Sediment Dredging at Superfund Megasites: Assessing the Effectiveness (2007)
Environmental Impacts of Wind-Energy Projects (2007)
Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget (2007)
Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues (2006)
New Source Review for Stationary Sources of Air Pollution (2006)
Human Biomonitoring for Environmental Chemicals (2006)
Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment (2006)
Fluoride in Drinking Water: A Scientific Review of EPA's Standards (2006)

State and Federal Standards for Mobile-Source Emissions (2006)
Superfund and Mining Megasites: Lessons from the Coeur d'Alene River Basin (2005)
Health Implications of Perchlorate Ingestion (2005)
Air Quality Management in the United States (2004)
Endangered and Threatened Species of the Platte River (2004)
Atlantic Salmon in Maine (2004)
Endangered and Threatened Fishes in the Klamath River Basin: Causes of Decline and Strategies for Recovery (2004)
Cumulative Environmental Effects of Oil and Gas Activities on Alaska's North Slope (2003)
Estimating the Public Health Benefits of Proposed Air Pollution Regulations (2002)
Biosolids Applied to Land: Advancing Standards and Practices (2002)
The Airliner Cabin Environment and the Health of Passengers and Crew (2002)
Arsenic in Drinking Water: 2001 Update (2001)
Evaluating Vehicle Emissions Inspection and Maintenance Programs (2001)
Compensating for Wetland Losses Under the Clean Water Act (2001)
A Risk-Management Strategy for PCB-Contaminated Sediments (2001)
Acute Exposure Guideline Levels for Selected Airborne Chemicals (20 volumes, 2000-2016)
Toxicological Effects of Methylmercury (2000)
Strengthening Science at the US Environmental Protection Agency: Research-Management and Peer-Review Practices (2000)
Scientific Frontiers in Developmental Toxicology and Risk Assessment (2000)
Ecological Indicators for the Nation (2000)
Waste Incineration and Public Health (2000)
Hormonally Active Agents in the Environment (1999)
Research Priorities for Airborne Particulate Matter (four volumes, 1998-2004)
The National Research Council's Committee on Toxicology: The First 50 Years 1947-1997 (1997)
Carcinogens and Anticarcinogens in the Human Diet: A Comparison of Naturally Occurring and Synthetic Substances (1996)
Upstream: Salmon and Society in the Pacific Northwest (1996)
Science and the Endangered Species Act (1995)
Wetlands: Characteristics and Boundaries (1995)
Biologic Markers (five volumes, 1989-1995)
Science and Judgment in Risk Assessment (1994)
Pesticides in the Diets of Infants and Children (1993)
Dolphins and the Tuna Industry (1992)
Science and the National Parks (1992)
Human Exposure Assessment for Airborne Pollutants: Advances and Opportunities (1991)
Rethinking the Ozone Problem in Urban and Regional Air Pollution (1991)
Decline of the Sea Turtles: Causes and Prevention (1990)

*Copies of these publications may be ordered from the National Academies Press*
*(800) 624-6242 or (202) 334-3313*
*www.nap.edu*

# Acknowledgments

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Hugh Barton, Pfizer Inc.
Kim Boekelheide, Brown University
Weihsueh Chiu, Texas A&M University
Graham Colditz, Washington University
Mark Cronin, Liverpool John Moores University
Richard Currie, Syngenta
Margaret Daniele Fallin, Johns Hopkins University
Charles Gallistel, Rutgers University
Kathryn Guyton, International Agency for Research on Cancer
Judy LaKind, LaKind Associates
Thomas McKone, University of California, Berkeley
Giovanni Parmigiani, Harvard University
David Reif, North Carolina State University
Craig Rowlands, Underwriters Laboratories
Maurice Whelan, DG Joint Research Centre

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Joseph Rodricks, Ramboll Environ, and Lynn Goldman, The George Washington University, who were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

The committee gratefully acknowledges the following for their presentations to the committee during open sessions: Christopher Austin, National Center for Advancing Translational Services; David Balshaw, National Institute of Environmental Health Sciences; Linda Birnbaum, National Institute of Environmental Health Sciences; Warren Casey, National Institute of Environmental Health Sciences; Suzanne Fitzpatrick, US Food and Drug Administration; James Jones, US Environmental Protection Agency; Robert Kavlock, US Environmental Protection Agency; Daniel Krewski, University of Ottawa; Michael Landa, US Food and Drug Administration; Chirag Patel, Harvard Medical School; Anton Simeonov, National Center for Advancing Translational Services; Russell Thomas, US Environmental Protection Agency, and David White, US Food and Drug Administration. The committee also thanks Tara Chu, University of Southern California, for compiling information on epidemiological studies that have investigated neurodevelopmental effects of air pollution.

# Acronyms and Abbreviations

| | |
|---|---|
| ACToR | Aggregated Computational Toxicology Resource |
| ADME | absorption, distribution, metabolism, and excretion |
| AHR | aryl-hydrocarbon receptor |
| AOP | adverse outcome pathway |
| B[a]P | benzo[a]pyrene |
| BPA | bisphenol A |
| Cas9 | CRISPR associated protein 9 |
| CC | collaborative cross |
| CCS | collisional cross-section |
| CDC | Centers for Disease Control and Prevention |
| ChEMBL | Chemical European Molecular Biology Laboratory |
| CPT | Continuous Performance Test |
| CRISPR | clustered regularly interspaced short palidromic repeats |
| DNT | developmental neurotoxicity |
| DO | diversity outbred |
| DSSTox | distributed structure-searchable toxicity |
| ECHA | European Chemicals Agency |
| ECVAM | European Centre for the Validation of Alternative Methods |
| EFSA | European Food Safety Authority |
| EPA | US Environmental Protection Agency |
| ES21 | *Exposure Science in the 21st Century: A Vision and a Strategy* |
| ESCAPE | European Study of Cohorts for Air Pollution Effects |
| EURL | European Union Reference Laboratory for Alternatives to Animal Testing |
| EWAS | exposome-wide association study |
| ExpoCast | exposure forecasting |
| FBS | fetal bovine serum |
| FDA | US Food and Drug Administration |
| FIFRA | Federal Insecticide, Fungicide, and Rodenticide Act |
| GC | gas chromatography |
| GPCR | G-protein coupled receptors |
| GWAS | genome-wide association study |
| HELIX | Human Early-Life Exposome Project |
| HERCULES | Health and Exposome Research Center: Understanding Lifetime Exposures |
| HMD | Human Metabolome Database |
| HTS | high-throughput screening |
| IARC | International Agency for Research on Cancer |
| ICCVAM | Interagency Coordinating Committee on Validation of Alternative Methods |
| IMS | ion-mobility spectrometry |
| IOM | Institute of Medicine |
| iPSC | induced pluripotent stem cell |
| IRIS | Integrated Risk Information System |
| IVIVE | in vitro–in vivo extrapolation |
| LC | liquid chromatography |

| | |
|---|---|
| LDL | low-density lipoprotein |
| LUR | land-use regression |
| MCMH | 4-methylcyclohexanemethanol |
| MS/MS | tandem mass spectrometry |
| NASA | National Aeronautics and Space Administration |
| NCATS | National Center for Advancing Translational Sciences |
| NHANES | National Health and Nutrition Examination Survey |
| NHGRI | National Human Genome Research Institute |
| NICEATM | NTP Interagency Center for the Evaluation of Alternative Toxicological Methods |
| NIEHS | National Institute of Environmental Health Sciences |
| NIH | National Institutes of Health |
| NMR | nuclear magnetic resonance |
| NOAEL | no-observed-adverse-effect level |
| NOEL | no-observed-effect level |
| NRC | National Research Council |
| NTP | National Toxicology Program |
| OECD | Organisation for Economic Co-operation and Development |
| OED | oral equivalent dose |
| PAH | polycyclic aromatic hydrocarbon |
| PBPK | physiologically based pharmacokinetics |
| PD | pharmacodynamics |
| PhenX Toolkit | Phenotypes and Exposures ToolKit |
| PM | particulate matter |
| PPARγ | peroxisome proliferator-activated receptor gamma |
| PPRTV | provisional peer reviewed toxicity value |
| PXR | pregnane X receptor |
| QSAR | quantitative structure–activity relationship |
| QSPR | quantitative structure–property relationship |
| RCPM | raven colored progressive matrices |
| REACH | Registration, Evaluation, Authorisation, and Restriction of Chemicals |
| RIX | recombinant inbred intercrosses |
| rTK | reverse toxicokinetics |
| RXR | retinoid X receptor |
| SAP | Science Advisory Panel |
| SAR | structure–activity relationship |
| SEEM | systematic empirical evaluation of models |
| SES | socioeconomic status |
| SEURAT | Safety Evaluation Ultimately Replacing Animal Testing |
| SHEDS-HT | Stochastic Human Exposure and Dose Simulation Model for High-Throughput |
| SHEDS-MM | Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway |
| STROBE | strengthening the reporting of observational studies in epidemiology |
| TCDD | tetrachlorodibenzo-*p*-Dioxin |
| TCE | tricloroethylene |
| Tox21 | *Toxicity Testing in the 21st Century: A Vision and a Strategy* |
| ToxCast | Toxicity Forecaster |
| TTC | threshold of toxicological concern |
| WHO | World Health Organization |
| WPPSI | Wechsler Preschool and Primary Scale of Intelligence |

# Contents

**APPENDIXES**



**BOXES, FIGURES, AND TABLES**

**BOXES**

**FIGURES**

**TABLES**

# USING 21ST CENTURY SCIENCE TO IMPROVE RISK-RELATED EVALUATIONS

# Summary

At the start of the 21st century, several federal agencies and organizations began to recognize the potential of improving chemical risk assessment by using the scientific and technological advances in biology and other related fields that were allowing the biological basis of disease to be better understood. Substantial increases in computational power and advances in analytical and integrative methods made incorporating the emerging evidence into risk assessment a possibility. Strategies were developed to use the advances to improve assessment of the effects of chemicals or other stressors that could potentially affect human health. Building on those efforts, the National Research Council (NRC) report *Toxicity Testing in the 21st Century: A Vision and a Strategy*[1] envisioned a future in which toxicology relied primarily on high-throughput in vitro assays and computational models based on human biology to evaluate potential adverse effects of chemical exposures. Similarly, the NRC report *Exposure Science in the 21st Century: A Vision and a Strategy*[2] articulated a long-term vision for exposure science motivated by the advances in analytical methods, sensor systems, molecular technologies, informatics, and computational modeling. That vision was to inspire a transformational change in the breadth and depth of exposure assessment that would improve integration with and responsiveness to toxicology and epidemiology.

Since release of those two reports, government collaborations have been formed, large-scale US and international programs have been initiated, and data are being generated from government, industry, and academic laboratories at an overwhelming pace. It is anticipated that the data being generated will inform risk assessment and support decision-making to improve public health and the environment. In the meantime, questions have arisen as to whether or how the data now being generated can be used to improve risk-based decision-making. Because several federal agencies recognize the potential value of such data in helping them to address their many challenging tasks, the US Environmental Protection Agency (EPA), US Food and Drug Administration (FDA), National Center for Advancing Translational Sciences (NCATS), National Institute of Environmental Health Sciences (NIEHS) and asked the National Academies of Sciences, Engineering, and Medicine to recommend the best ways to incorporate the emerging science into risk-based evaluations.[3] As a result of the request, the National Academies convened the Committee on Incorporating 21st Century Science into Risk-Based Evaluations, which prepared this report.

## SCIENTIFIC ADVANCES

To approach its task, the committee assessed scientific and technological advances in exposure science and toxicology that could be integrated into and used to improve any of the four elements of risk assessment—hazard identification, dose–response assessment, exposure assessment, and risk characterization. Although the National Academies has not been asked to produce a report on epidemiology comparable with its Tox21 and ES21 reports, epidemiological research is also undergoing a transformation. Because it plays a critical role in risk assessment by providing human evidence on adverse effects of chemical and other exposures, the committee assessed advances in epidemiology as part of its charge. The committee highlights here some of the advances, challenges, and needs in each field in the context of risk assessment. The committee's report provides specific recommendations to address the challenges. Overall, a common theme is the need for a multidisciplinary approach. Exposure scientists, toxicologists, epidemiologists, and scientists in other disciplines need to collaborate closely to ensure that the full potential of 21st century science is realized to help to solve the complex environmental and public-health problems that society faces.

[1]Referred to hereafter as the Tox21 report.

[2]Referred to hereafter as the ES21 report.

[3]The verbatim statement of task is provided in Chapter 1 of the committee's report.

## Exposure Science

A primary objective for improving exposure science is to build confidence in the exposure estimates used to support risk-based decision-making by enhancing quality, expanding coverage, and reducing uncertainty. The many scientific and technological advances that are transforming exposure science should help to meet that objective. Some of the endeavors that the committee considered promising for advancing that objective and in which progress has been made since the ES21 report are highlighted below.

• *Remote sensing, personal sensors, and other sampling techniques*. Remote sensing enhances the capacity to assess human and ecological exposures by helping to fill gaps in time and place left by traditional ground-based monitoring systems. Advances in passive sampling techniques and personal sensors offer unparalleled opportunities to characterize individual exposures, particularly in vulnerable populations. If remote sensing and personal sensors can be combined with global positioning systems, exposure and human-activity data can be linked to provide a more complete understanding of human exposures.

• *Computational exposure tools*. Because exposure-measurement data on many agents are not available, recent advances in computational tools for exposure science are expected to play a crucial role in most aspects of exposure estimation for risk assessments, not just high-throughput applications. However, improving the scope and quality of data that are needed to develop parameters for these tools is critically important because without such data the tools have greater uncertainty and less applicability. Comparisons of calculated and measured exposures are required to characterize uncertainties in the computational tools and their input parameters.

• *Targeted and nontargeted analyses*. Advances in two complementary approaches in analytical chemistry are improving the accuracy and breadth of human and ecological exposure characterizations and are expanding opportunities to investigate exposure–disease relationships. First, targeted analyses focus on identifying selected chemicals for which standards and methods are available. Improved analytical methods and expanded chemical-identification libraries are increasing opportunities for such analyses. Second, nontargeted analyses offer the ability to survey more broadly the presence of all chemicals in the environment and in biofluids regardless of whether standards and methods are available. Nontargeted analyses reveal the presence of numerous substances whose identities can be determined after an initial analysis by using cheminformatic approaches or advanced or novel analytical techniques.

• *-Omics technologies*. -Omics technologies can measure chemical or biological exposures directly or identify biomarkers of exposure or response that allow one to infer exposure on the basis of a mechanistic understanding of biological responses. These emerging technologies and data streams will complement other analyses, such as targeted and nontargeted analyses, and lead to a more comprehensive understanding of the exposure-to-outcome continuum. Identifying biomarkers of exposure to individual chemicals or chemical classes within the complex exposures of human populations remains a considerable challenge for these tools.

• *Exposure matrices for life-span research*. Responding to the need to improve the characterization of fetal exposures to chemicals, researchers have turned to new biological matrices, such as teeth, hair, nails, placental tissue, and meconium. The growth properties (the sequential deposition or addition of tissue with accumulation of chemicals) and availability of the biospecimens offer the opportunity to extract a record of exposure. The question that needs to be addressed now is how concentrations in these matrices are related to and can be integrated with measures of exposure that have been traditionally used to assess chemical toxicity or risk.

• *Physiologically based pharmacokinetic (PBPK) models*. PBPK models are being applied more regularly to support aggregate (multiroute) exposure assessment, to reconstruct exposure from biomonitoring data, to translate exposures between experimental systems, and to understand the relationship between biochemical and physiological variability and variability in population response. An important focus has been on the development of PBPK models for translating exposures between test systems and human-exposure scenarios, development that has been driven by the rapidly expanding use of high-throughput in vitro assays to characterize the bioactivity of chemicals and other materials. That research will remain critical as regulatory agencies, industry, and other organizations increase their dependence on in vitro systems.

The emerging technologies and data streams offer great promise for advancing exposure science and improving and refining exposure measurements and assessment. However, various challenges will need to be addressed. A few are highlighted here.

• *Expanding and coordinating exposure-science infrastructure*. A broad spectrum of disciplines and institutions are participating in advancing exposure methods, measurements, and models. Given the number and diversity of participants in exposure science, the information is mostly fragmented, incompletely organized, and in some cases not readily available or accessible. Thus,

an infrastructure is needed to improve the organization and coordination of the existing and evolving components for exposure science and ultimately to improve exposure assessment. Infrastructure development should include creating or expanding databases that contain information on chemical quantities in and chemical release rates from products and materials, on chemical properties and on processes, and analytical features that can be used in chemical identification.

• *Aligning environmental and test-system exposures*. Aligning information on environmental exposures with information obtained from experimental systems is a critical aspect of risk-based evaluation. Concentrations in test-system components need to be quantified by measurement, which is preferred, or by reliable estimation methods. Knowledge of physical processes, such as binding to plastic and volatilization, and of biological processes, such as metabolism, needs to be improved.

• *Integrating exposure information*. Integration and appropriate application of exposure data on environmental media, biomonitoring samples, conventional samples, and emerging matrices constitute a scientific, engineering, and big-data challenge. The committee emphasizes that integration of measured and modeled data is a key step in developing coherent exposure narratives, in evaluating data concordance, and ultimately in determining confidence in an exposure assessment. New multidisciplinary projects are needed to integrate exposure data and to gain experience that can be used to guide data collection and integration of conventional and emerging data streams.

## Toxicology

The decade since publication of the Tox21 report has seen continued advances in an array of technologies that can be used to understand human biology and disease at the molecular level. Technologies are now available to profile the transcriptome, epigenome, proteome, and metabolome. There are large banks of immortalized cells collected from various populations to use for toxicological research; large compilations of publicly available biological data that can be mined to develop hypotheses about relationships between chemicals, genes, and diseases; and genetically diverse mouse strains and alternative species that can be used for toxicological research. Highlighted below are some assays, models, and approaches for predicting biological responses that have seen rapid advances over the last decade; they are arranged by increasing level of biological organization.

• *Probing interactions with biological molecules*. Chemical interactions with specific receptors, enzymes, or other discrete proteins and nucleic acids have long been known to have adverse effects on biological systems, and development of in vitro assays that probe chemical interactions with cellular components has been rapid, driven partly by the need to reduce high attrition rates in drug development. The assays can provide reliable and valid results with high agreement among laboratories and can be applied in low-, medium-, and high-throughput formats. Computational models have been developed to predict activity of chemical interactions with protein targets, and research to improve the prediction of protein–chemical interactions continues.

• *Detecting cellular response*. Cell cultures can be used to evaluate a number of cellular processes and responses, including receptor binding, gene activation, cell proliferation, mitochondrial dysfunction, morphological changes, cellular stress, genotoxicity, and cytotoxicity. Simultaneous measurements of multiple toxic responses are also possible with high-content imaging and other novel techniques. Furthermore, cell cultures can be scaled to a high-throughput format and can be derived from genetically different populations so that aspects of variability in response to chemical exposure that depend on genetic differences can be studied. In addition to cell-based assays, numerous mathematical models and systems-biology tools have been advanced to describe various aspects of cell function and response.

• *Investigating effects at higher levels of biological organization*. The last decade has seen advances in engineered three-dimensional (3-D) models of tissues. Organotypic or organ-on-a-chip models are types of 3-D models in which two or more cell types are combined in an arrangement intended to mimic an in vivo tissue and, therefore, recapitulate at least some of the physiological responses that the tissue or organ exhibits in vivo. NCATS, for example, has a number of efforts in this field. Although the models are promising, they are not yet ready for inclusion in risk assessment. In addition to cell cultures, computational systems-biology models have been developed to simulate tissue-level response. EPA, for example, has developed virtual-tissue models for the embryo and liver. Virtual-tissue models can potentially help in conceptualizing and integrating current knowledge about the factors that affect key pathways and the degree to which pathways must be perturbed to activate early and intermediate responses in human tissues and, when more fully developed, in supporting risk assessments.

• *Predicting organism and population response*. Animal studies remain an important tool in risk assessment, but scientific advances are providing opportunities to enhance the utility of whole-animal testing. Gene-editing technologies, for example, have led to the creation of transgenic rodents that can be used to investigate specific questions, such as those related to susceptibility or gene–environment interactions. Genetically diverse rodent strains have provided another approach for addressing questions related to interindividual sensitivity to toxi-

cants. Combining transgenic or genetically diverse rodent strains with -omics and other emerging technologies can increase the information gained from whole-animal testing alone. Those targeted studies can help to address knowledge gaps in risk assessment and can link in vitro observations to molecular, cellular, or physiological effects in the whole animal. In addition to the mammalian species, scientific advances have made some alternative species—such as the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and the zebrafish *Danio rerio*—useful animal models for hazard identification and investigation of biological mechanisms.

The assays, models, and tools noted above hold great promise in the evolution of toxicology, but there are important technical and research challenges, a few of which are highlighted below.

- *Accounting for metabolic capacity in assays*. Current in vitro assays generally have little or no metabolic capability, and this aspect potentially constrains their usefulness in evaluating chemical exposures that are representative of human exposures that could lead to toxicity. Research to address the metabolic-capacity issues needs to have high priority, and formalized approaches need to be developed to characterize the metabolic competence of assays, to determine for which assays it is not an essential consideration, and to account for the toxicity of metabolites appropriately.
- *Understanding and addressing other limitations of cell systems*. Cell cultures can be extremely sensitive to environmental conditions, responses can depend on the cell type used, and current assays can evaluate only chemicals that have particular properties. Research is needed to determine the breadth of cell types required to capture toxicity adequately; cell batches need to be characterized sufficiently before, during, and after experimentation; and practical guidance will need to be developed for cell systems regarding their range of applicability and for describing the uncertainty of test results.
- *Addressing biological coverage*. Developing a comprehensive battery of in vitro assays that covers the important biological responses to the chemical exposures that contribute to adverse health effects is a considerable challenge. In addition, most assays used in the federal government high-throughput testing programs were developed by the pharmaceutical industry and were not designed to cover the full array of biological response. As emphasized in the Tox21 report, research is needed to determine the extent of relevant mechanisms that lead to adverse responses in humans and to determine which experimental models are needed to cover these mechanisms adequately. Using -omics technologies and targeted testing approaches with transgenic and genetically diverse rodent species and alternative species will address knowledge gaps more comprehensively.

When one considers the progress in implementing the Tox21 vision and the current challenges, it is important to remember that many assays, models, and tools were not developed with risk-assessment applications as a primary objective. Thus, understanding of how best to apply them and interpret the data is evolving. The usefulness or applicability of various in vitro assays will need to be determined by continued data generation and critical analysis, and some assays that are highly effective for some purposes, such as pharmaceutical development, might not be as useful for risk assessment of commodity chemicals or environmental pollutants. It will most likely be necessary to adapt current assays or develop new assays specifically intended for risk-assessment purposes.

## Epidemiology

The scientific advances that have propelled exposure science and toxicology onto new paths have also substantially influenced the direction of epidemiological studies and research. The factors reshaping epidemiology in the 21st century include expansion of the interdisciplinary nature of the field; the increasing complexity of scientific inquiry; emergence of new data sources and technologies for data generation, such as new medical and environmental data sources and -omics technologies; advances in exposure characterization; and increasing demands to integrate new knowledge from basic, clinical, and population sciences. There is also a movement to register past and present datasets so that on particular issues datasets can be identified and combined.

One of the most important developments has been the emergence of the -omics technologies and their incorporation into epidemiological research. -Omics technologies have substantially transformed epidemiological research and advanced the paradigm of molecular epidemiology, which focuses on underlying biology (pathogenesis) rather than on empirical observations alone. The utility of -omics technologies in epidemiological research is already clear and well exemplified by the many studies that have incorporated genomics. For example, the genetic basis of disease has been explored in genome-wide association studies in which the genomic markers in people who have and do not have a disease or condition of interest are compared. The -omics technologies that have been applied in epidemiological research, however, have now expanded beyond genomics to include epigenomics, proteomics, transcriptomics, and metabolomics. New studies are being designed with the intent of prospectively storing samples that can be used for existing and future -omics technologies. Thus, obtaining data from human population studies that are parallel to data obtained from in vi-

tro and in vivo assays or studies is already possible and potentially can help in harmonizing comparisons of exposure and dose. Furthermore, -omics technologies have the potential for providing a suite of new biomarkers for hazard identification and risk assessment.

Like exposure science and toxicology, epidemiology faces challenges in incorporating 21st century science into its practice. -Omics assays can generate extremely large datasets that need to be managed and curated in ways that facilitate access and analysis. Databases that can accommodate the large datasets, support analyses for multiple purposes, and foster data-sharing need to be developed. Powerful and robust statistical techniques also are required to analyze all the data. And standard ways to describe the data are needed so that data can be harmonized among investigative groups and internationally.

The landscape of epidemiological research is changing rapidly as the focus shifts from fixed, specific cohorts, such as those in the Nurses' Health Study,[4] to large cohorts enrolled from health-care organizations or other resources that incorporate biospecimen banks and use health-care records to characterize participants and to track outcomes. Such studies offer large samples but will need new approaches to estimate exposures that will work in this context. Thus, there will be a need for close collaboration with exposure scientists to ensure that exposure data are generated in the best and most comprehensive way possible. Furthermore, various biospecimens are being collected and stored with the underlying assumption that they will be useful in future studies; researchers involved in such future-looking collections need to seek input from the scientists who are developing new assays so that the biospecimens can be collected and stored in a way that maximizes the potential for their future use. All those concerns emphasize the need to expand the multidisciplinary teams involved in epidemiological research.

## APPLICATIONS OF 21st CENTURY SCIENCE

The scientific and technological advances described above and in further detail in this report offer opportunities to improve the assessment or characterization of risk for the purpose of environmental and public-health decision-making. The committee highlights below several activities—priority-setting, chemical assessment, site-specific assessment, and assessments of new chemistries—that could benefit from the incorporation of 21st century science. Case studies of practical applications are provided in Appendixes B–D.

Priority-setting has been seen as a principal initial application for 21st century science. High-throughput screening programs have produced toxicity data on thousands of chemicals, and high-throughput methods have provided quantitative exposure estimates. Several methods have been proposed for priority-setting, including risk-based approaches that use a combination of the high-throughput exposure and hazard information to calculate margins of exposure (differences between toxicity and exposure metrics). For that approach, chemicals that have a small margin of exposure would be seen as having high priority for further testing and assessment.

Chemical assessment is another activity in which the committee sees great potential for application of 21st century science. Chemical assessments encompass a broad array of analyses. Some cover chemicals that have a substantial database for decision-making, and for these assessments scientific and technical advances can be used to reduce uncertainties around key issues and to address unanswered questions. Many assessments, however, cover chemicals on which there are few data to use in decision-making, and for these assessments the committee finds an especially promising application for 21st century science. One approach for evaluating data-poor chemicals is to use toxicity data on well-tested chemicals (analogues) that are similar to the chemicals of interest in their structure, metabolism, or biological activity in a process known as read-across (see Figure S-1). The assumption is that a chemical of interest and its analogues are metabolized to common or biologically similar metabolites or that they are sufficiently similar in structure to have the same or similar biological activity. The method is facilitated by having a comprehensive database of toxicity data that is searchable by curated and annotated chemical structures and by using a consistent decision process for selecting suitable analogues. The approach illustrated in Figure S-1 can be combined with high-throughput in vitro assays, such as gene-expression analysis, or possibly with a targeted in vivo study to allow better selection of the analogues to ensure that the biological activities of a chemical of interest and its analogues are comparable. The committee notes that computational exposure assessment, which includes predictive fate and transport modeling, is an important complement to the approach described and can provide information on exposure potential, environmental persistence, and likelihood of bioaccumulation.

Site-specific assessment represents another application for which 21st century science can play an important role. Understanding the risks associated with a chemical spill or the extent to which a hazardous-waste site needs to be remediated depends on understanding exposures to various chemicals and their toxicity. The assessment problem contains three elements—identifying and quantifying chemicals present at the site, characterizing their toxicity, and characterizing the toxicity of chemical mixtures—and the advances described in this report can address each element. First, targeted analytical-chemistry

[4]The Nurses' Health Study is a prospective study that has followed a large cohort of women over many decades to identify risk factors for major chronic diseases.

approaches can identify and quantify chemicals for which standards are available, and nontargeted analyses can help to assign provisional identities to previously unidentified chemicals. Second, analogue-based methods coupled with high-throughput or high-content screening methods have the potential to characterize the toxicity of data-poor chemicals. Third, high-throughput screening methods can provide information on mechanisms that can be useful in determining whether mixture components might act via a common mechanism, affect the same organ, or cause the same outcome and thus should be considered as posing a cumulative risk. High-throughput methods can also be used to assess the toxicity of mixtures that are present at specific sites empirically rather than assessing individual chemicals.

Assessment of new chemistries is similar to the chemical assessment described above except that it typically involves new molecules on which there are no toxicity data and that might not have close analogues. Here, modern in vitro toxicology methods could have great utility by providing guidance on which molecular features are associated with greater or less toxicity and by identifying chemicals that do not affect biological pathways that are known to be relevant for toxicity. Modern exposure-science methods might also help to identify chemicals that have the highest potential for widespread environmental or human exposure and for bioaccumulation.

## VALIDATION

Before new assays, models, or test systems can be used in regulatory-decision contexts, it is expected and for some purposes legally required that their relevance, reliability, and fitness for purpose be established and documented. That activity has evolved into elaborate processes that are commonly referred to as validation of alternative methods. One critical issue is that current processes for validation cannot match the pace of development of new assays, models, and test systems, and many have argued that validation processes need to evolve. Important elements of the validation process that need to be addressed include finding appropriate comparators for enabling fit-for-purpose validation of new test methods, clearly defining assay utility and how assay data should be interpreted, establishing performance standards for assays and clear reporting standards for testing methods, and determining how to validate batteries of assays that might be used to replace toxicity tests. The committee discusses those challenges further and offers some recommendations in Chapter 6.

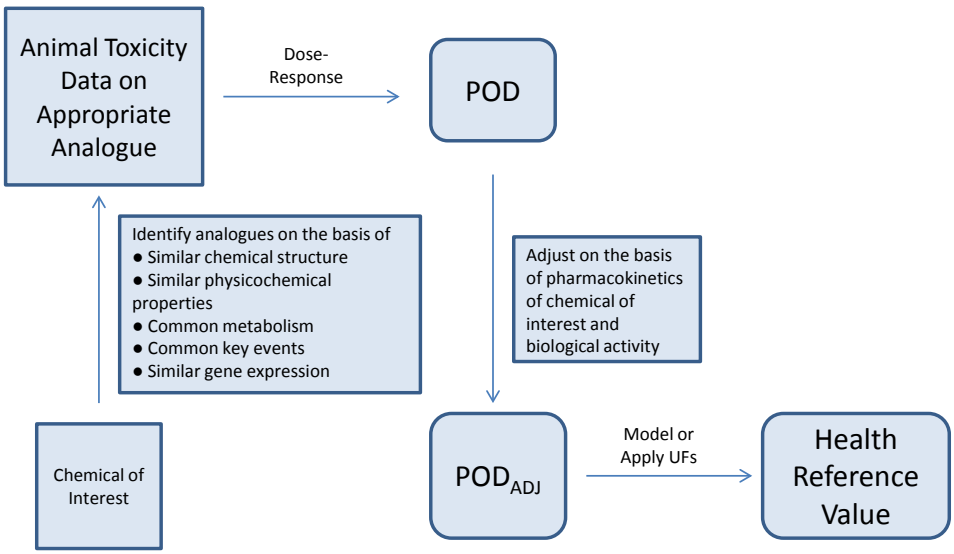



**FIGURE S-1** Approach to deriving health reference values when data on similar chemicals are available. Similarity can be based on such characteristics as chemical structure, physicochemical properties, metabolism, key events in biological pathways, or gene expression; similarity of several characteristics increases confidence in the analogy. The point of departure (POD) of the appropriate analogue would be adjusted on the basis of pharmacokinetic differences between the chemical of interest and the analogue and other important biological factors, such as receptor activation; relevant uncertainty factors would then be applied or models would be used to derive the health reference value. Accounting for uncertainty could include a determination of the degree of confidence in the read-across, including the number of analogues identified, the degree of similarity of the analogues to the chemical of interest, and the extent of the dataset on the analogues.

## A NEW DIRECTION FOR RISK ASSESSMENT AND THE CHALLENGES IT POSES

The advances in exposure science, toxicology, and epidemiology described in this report support a new direction for risk assessment, one based on biological pathways and processes rather than on observation of apical responses and one incorporating the more comprehensive exposure information emerging from new tools and approaches in exposure science. The exposure aspect of the new direction focuses on estimating or predicting internal and external exposures to multiple chemicals and stressors, characterizing human variability in those exposures, providing exposure data that can inform toxicity testing, and translating exposures between test systems and humans. The toxicology and epidemiology elements of the new direction focus on the multifactorial and nonspecific nature of disease causation; that is, stressors from multiple sources can contribute to a single disease, and a single stressor can lead to multiple adverse outcomes. The question shifts from whether A causes B to whether A increases the risk of B. The committee found that the sufficient-component-cause model, which is illustrated in Figure S-2, is a useful tool for conceptualizing the new direction. The same outcome can result from more than one causal complex or mechanism; each mechanism generally involves joint action of multiple components.

Most diseases that are the focus of risk assessment have a multifactorial etiology; some disease components arise from endogenous processes, and some result from the human experience, such as background health conditions, co-occurring chemical exposures, food and nutrition, and psychosocial stressors. Those additional components might be independent of the environmental stressor under study but nonetheless influence and contribute to the overall risk and incidence of disease. As shown in the case studies in this report, one does not need to know all the pathways or components involved in a particular disease to begin to apply the new tools to risk assessment. The 21st century tools provide the mechanistic and exposure data to support dose–response characterizations and human-variability derivations described in the NRC report *Science and Decisions: Advancing Risk Assessment*. They also support the understanding of relationships between disease and components and can be used to probe specific chemicals for their potential to perturb pathways or activate mechanisms and increase risk.

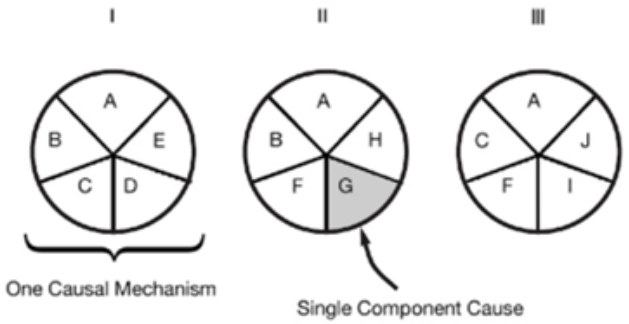



**FIGURE S-2** Multifactorial nature of disease illustrated by using the sufficient-component-cause model in which various overall mechanisms (I, II, and III) for a disease are represented as causal pies of various components (A–J). The committee considers pathways to be components of the mechanism.

The 21st century science with its diverse, complex, and very large datasets, however, poses challenges related to analysis, interpretation, and integration of data and evidence for risk assessment. In fact, the technology has evolved far faster than the approaches for those activities. The committee found that Bradford-Hill causal guidelines could be extended to help to answer such questions as whether specific pathways, components, or mechanisms contribute to a disease or outcome and whether a particular agent is linked to pathway perturbation or mechanism activation. Although the committee considered various methods for data integration, it concluded that guided expert judgment should be used in the near term for integrating diverse data streams for drawing causal conclusions. In the future, pathway-modeling approaches that incorporate uncertainties and integrate multiple data streams might become an adjunct to or perhaps a replacement for guided expert judgment, but research will be needed to advance those approaches. The committee emphasizes that insufficient attention has been given to analysis, interpretation, and integration of various data streams from exposure science, toxicology, and epidemiology. It proposes a research agenda that includes developing case studies that reflect various scenarios of decision-making and data availability; testing case studies with multidisciplinary panels; cataloguing evidence evaluations and decisions that have been made on various agents so that expert judgments can be tracked and evaluated, and expert processes calibrated; and determining how statistically based tools for combining and integrating evidence, such as Bayesian approaches, can be used for incorporating 21st century science into all elements of risk assessment.

## CONCLUDING REMARKS

As highlighted here and detailed in the committee's report, many scientific and technical advances have followed publication of the Tox21 and ES21 reports. The committee concludes that the data that are being generated today can be used to address many of the risk-related tasks that the agencies face, and it provides several case studies in its report to illustrate the potential applications. Although the challenges to achieving the visions of the earlier reports often seem daunting, 21st century science holds great promise for advancing risk assessment and

ultimately for improving public health and the environment. The committee emphasizes, however, that communicating the strengths and limitations of the approaches in a transparent and understandable way will be necessary if the results are to be applied appropriately and will be critical for the ultimate acceptance of the approaches.

# 1

# Introduction

Over the last decade, several large-scale US and international programs have been initiated to incorporate advances in molecular and cellular biology, -omics technologies, analytical methods, bioinformatics, and computational tools and methods into the field of toxicology. The overarching goal of the various programs is to move toxicology from a practice that uses whole-animal testing to one that uses primarily modern in vitro assays and computational approaches to predict toxicity on the basis of an understanding of the biological processes that ultimately lead from the initial chemical exposure to adverse effects. Similar efforts are being pursued in the field of exposure science with the goals of obtaining more accurate and complete exposure data on individuals and populations for thousands of chemicals over the lifespan; predicting exposures from use data and chemical-property information; and translating exposures between test systems and humans. It is hoped that the advances in toxicology and exposure science and better integration of the fields will improve risk assessment and thus better support decision-making to improve public and environmental health. With various efforts under way, diverse data are being generated, and their utility for risk assessment investigated. Although the programs and the data being generated are still evolving and will undoubtedly continue to do so, some data could be used now to help to fill gaps and assess chemical risk better. Several federal agencies recognize the potential value of such data in helping them to address their many challenging tasks. Accordingly, the US Environmental Protection Agency (EPA), the US Food and Drug Administration (FDA), the National Center for Advancing Translational Sciences (NCATS), and the National Institute of Environmental Health Sciences (NIEHS), and the asked the National Academies of Sciences, Engineering, and Medicine to consider the integration of modern and emerging scientific approaches and data into risk-based evaluations and to recommend the best ways to do so. As a result of the request, the National Academies convened the Committee on Incorporating 21st Century Science into Risk-Based Evaluations, which prepared this report.

## TOXICOLOGY IN THE 21st CENTURY

In the early 2000s, several agencies and organizations began to recognize the potential of various scientific advances in biology and related fields and the possibilities provided by increases in computational power to characterize risks of environmental exposures. Roadmaps were developed to incorporate such advances into their strategic plans for assessing chemicals and other agents (EPA 2003; NTP 2004). In 2007, the National Research Council (NRC) released the report *Toxicity Testing in the 21st Century: A Vision and a Strategy*,[1] which envisioned transforming toxicity testing from a system that relies on animal assays to one that relies primarily on high-throughput in vitro assays and computational methods based on human biology. The primary goals behind the vision were "(1) to provide broad coverage of chemicals, chemical mixtures, outcomes, and life stages, (2) to reduce the cost and time of testing, (3) to use fewer animals and cause minimal suffering in the animals used, and (4) to develop a more robust scientific basis for assessing health effects of environmental agents" (NRC 2007). The committee that prepared the 2007 report emphasized that the transformation would require a focused effort over several decades for full implementation. On release of the report, the NIEHS National Toxicology Program, the EPA National Center for Computational Toxicology, and the Chemical Genomics Center[2] of the National Institutes of Health formed a collaboration, known as Tox21, to advance the vision set forth in the 2007 report (Collins et al. 2008). FDA later joined the collaboration.

The goals of the Tox21 collaboration are to identify and characterize specific mechanisms or pathways that lead to adverse effects in humans, to design assays to measure pathway responses, to develop models that can predict toxicity using the assay data, and to set priorities among chemicals for more comprehensive toxicity testing (NCATS 2015a). It is planned that the data generated will ultimately help to inform EPA, FDA, and

[1]Referred to hereafter as the Tox21 report.

[2]The Chemical Genomics Center is now part of NCATS.

other agencies on the hazards posed by the chemicals or products that they regulate and will be used by industry to screen for potential toxicity in product development. A phased approach to the research is being taken. Phase I of Tox21 has been completed and involved testing of about 2,800 chemicals in about 50 assays, including ones to assess cytotoxicity, mitochondrial toxicity, cell signaling, DNA damage, immune response, drug metabolism, nuclear-receptor activation, and inhibition of various molecular targets (Tice et al. 2013; NCATS 2015b). Phase II involves testing of over 10,000 chemicals that occupy a diverse chemical and toxicological space and include "industrial chemicals, sunscreen additives, flame retardants, pesticides and selected metabolites, plasticizers, solvents, food additives, natural product components, drinking water disinfection by-products, preservatives, therapeutic agents, and chemical synthesis by-products" (Tice et al. 2013). Phase III will involve identification of physiologically relevant cells, measurement of gene expression in a large number of molecular pathways, and testing of chemical mixtures and extracts (NCATS 2015b).

In 2007, EPA initiated its Toxicity Forecaster (ToxCast) program, which seeks to develop high-throughput screening (HTS) assays for evaluating biological responses that are relevant to prediction of adverse effects of chemical exposures on humans (EPA 2013). A phased approach to research is also being taken in the ToxCast program. Phase I, which has been completed, involved testing of over 300 well-studied chemicals in several hundred HTS assays (Kavlock and Dix 2010). Phase II has also been completed; it involved testing of over 2,000 chemicals—including industrial and consumer products, food additives, and potentially safer chemical alternatives to existing chemicals—in HTS assays for evaluating various cell responses and over 300 signaling pathways (EPA 2013; Silva et al. 2015). ToxCast data are now being evaluated as a means of setting priorities among chemicals for testing in EPA's Endocrine Disruptor Screening Program and in other programs that require setting priorities for testing.

In addition to US government-led efforts, international efforts are transforming toxicology from an observational to a predictive science. In the European Union, for example, the European Commission and Cosmetics Europe (a trade association for the cosmetics and personal-care industry) have co-funded the research initiative Safety Evaluation Ultimately Replacing Animal Testing (SEURAT 2015). The initiative was started to develop tools to comply with legislation that banned all animal testing for cosmetic ingredients and all marketing of animal-tested cosmetic ingredients and products; a complete ban went into effect in March 2013. Its vision was to eliminate traditional animal testing by adopting a "toxicological mode-of-action framework to describe how any substance may adversely affect human health, and use this knowledge to develop complementary theoretical, computational and experimental (in vitro) models that predict quantitative points of departure needed for safety assessment" (Berggren 2015). The research initiative was a 5-year program (2011–2015) that involved development of in vitro assays that use human pluripotent stem cells, development of a hepatic microfluidic bioreactor, identification and investigation of human biomarkers of chronic toxicity in cellular models, and development of computational tools for predicting chronic toxicity.

Private industry and other organizations are also working to transform the ways in which chemicals are assessed. For example, the pharmaceutical industry has been developing and using in vitro and computational tools as early screens for drug safety for many years (Greene and Song 2011; Bowes et al. 2012). Organizations have developed case studies related to the use of new in vitro assays and computational systems-biology tools for assessment of chemical risk (Daston et al. 2015; Gocht et al. 2015). Cheminformatics research has resulted in the development of rational systems for informing qualitative structure–activity relationship assessments (Wu et al. 2010) and in the development of automated decision trees for identifying toxicity end points, such as developmental and reproductive toxicity (Wu et al. 2013).

Academic institutions are generating a substantial amount of data that could help to inform chemical risk assessment. Academic laboratories tend to focus on end points that are not typically covered in guideline animal studies, such as mammary gland development (Fenton 2006; Soto et al. 2008; Osborne et al. 2015), synaptic morphology and other aspects of nervous system development (Patisaul and Polston 2008), and complex behaviors, including sociality, aggression, cognition, and behavioral hallmarks of psychiatric disorders, such as autism spectrum disorder and attention deficit disorder (Eubig et al. 2010; de Cock et al. 2012; Leon-Olea et al. 2014). Research on genetics, genomics, and epigenetics (including the role of noncoding RNAs) is also abundant and is providing insights on novel biological mechanisms and gene-by-environment interactions (Dolinoy et al. 2007; Rusyn et al. 2010; Tal and Tanguay 2012; Nebert et al. 2013; Yeo et al. 2013). Academic laboratories have been responsible for generating nearly all the data on transgenerational effects (Rissman and Adli 2014); have pioneered the use of nontraditional animal models, including transgenic and population-based models (Churchill et al. 2004; Rusyn et al. 2010; Sullivan et al. 2014); and have conducted most of the epidemiological studies of chemical risk. The enormous volume of data being generated throughout the basic- and clinical-research communities has prompted questions about how the data could best be used for various risk-related activities and decision-making.

## EXPOSURE SCIENCE IN THE 21st CENTURY

Exposure science is undergoing a transformation similar to that affecting toxicology with the advances in molecular technologies, computational tools, bioinformatics, sensor systems, and analytical methods. In 2012, the NRC released the report *Exposure Science in the 21st Century: A Vision and a Strategy*,[3] which articulated a long-term vision for exposure science. The primary long-term goal of the vision was to broaden the reach of exposure science from a traditional focus on discrete exposures to an "integrated approach that considers exposures from source to dose, on multiple levels of integration (including time, space, and biological scale), to multiple stressors, and scaled from molecular systems to individuals, populations, and ecosystems" (NRC 2012). The report described scientific and technological progress that has the potential to transform exposure science, including geographic information technologies that can track sources, exposure concentrations, and receptors; monitoring technologies that can collect data on personal exposure of millions of people; highly sensitive analytical technologies that can identify and measure biomarkers that are indicative of internal exposures; and computational tools that can manage the large amounts of data generated. It also highlighted high-priority research, emphasized the need for interagency collaboration and resources, and elaborated the broad concept of the exposome, defined as "the record of all exposures both internal and external that people receive throughout their lifetime" (Rappaport and Smith 2010). Last, it recognized the interdependence of the fields of toxicology, risk assessment, and exposure science and foresaw the need to evolve the risk-assessment paradigm toward one in which exposure science plays a strong role, specifically, a paradigm that is "influenced by and responsive to human and environmental exposure data." The report described four objectives of exposure science: to set priorities among chemicals for toxicity testing; to provide exposure information to guide toxicity testing; to provide quantitative pharmacokinetic data on absorption, distribution, metabolism, and excretion (ADME) derived from human-exposure studies; and to connect exposure data with biological activity data to identify exposure–response relationships.

In response to the recommendation to improve integration of exposure science throughout the federal government, the Exposure Science in the 21st Century (ES21) Federal Working Group has emerged (EPA 2016a). It consists of representatives of more than 20 federal organizations that have a common interest in exposure-science research and development. The purpose of the working group is to build on the framework recommended in the ES21 report, share information, integrate activities, reduce duplication of efforts among agencies, and promote federal collaboration in the development of exposure science. In addition to the activities of the working group, several research programs are involved in advancing exposure science on paths that are consistent with the vision articulated in the ES21 report. EPA created the Exposure Forecasting (ExpoCast) program, which complements its ToxCast program (EPA 2016b). ExpoCast focuses on developing high-throughput methods for estimating exposure and so far has been used to make exposure predictions related to over 1,900 chemicals. EPA's goal is to combine the exposure estimates from ExpoCast with bioactivity data from ToxCast to predict human health and environmental risks.

NIEHS is also interested in advancing exposure science and has supported research to develop new sensor systems and to identify biomarkers of response to exposure (NIEHS 2015). It has created the Children's Health Exposure Analysis Resource (NIEHS 2016), an infrastructure designed to enable and expand incorporation of environmental exposures into studies of children's health; it includes a data repository, support for statistical analysis, and a network of laboratories to analyze biological samples. The NIEHS strategic plan emphasizes a commitment to supporting research to define and explore the exposome, and the agency has funded the HERCULES center at Emory University to conduct exposome-focused research (NIEHS 2012).

In addition to the efforts in the United States, there are international efforts, such as the Human Early-Life Exposome (HELIX) project and the EXPOsOMICS project. HELIX has the ambitious goal of characterizing early-life exposures and ultimately linking exposures with children's health outcomes (Vrijheid et al. 2014). The project is studying 32,000 mother–child pairs in six European countries. EXPOsOMICS focuses on the external and internal exposome associated with air pollution and water contamination (Vineis et al. 2013, in press). The project will perform personal-exposure monitoring of air pollutants for hundreds of subjects in Europe, and biological samples from thousands of subjects will be analyzed for internal exposure markers by using -omics technologies (CORDIS 2015).

Like the toxicology initiatives, the exposure programs are generating vast amounts of data, but how the data are best used to inform risk-related tasks and decision-making remains to be determined.

## TERMINOLOGY

The recent advances in toxicology and exposure science have given rise to a new vocabulary and a plethora of new terms. Some researchers and practitioners distinguish between terms, but others use the same terms interchangeably and inconsistently. Consequently, there is some confusion as to the specific meanings of various

[3]Referred to hereafter as the ES21 report.

terms. *Mode of action*, *mechanism of action*, and *adverse outcome pathway* are exemplary of the confusion. Each term denotes a progression from some exposure or molecular initiating event to an adverse outcome. *Mechanism of action* is often distinguished from *mode of action* by a greater level of biological detail in the understanding and description of the progression from exposure to outcome (EPA 2005; NRC 2007). *Mode of action* typically describes the progression of key events that result from a chemical exposure whereas *adverse outcome pathway* conceptually describes the sequential chain of causally linked events at various levels of biological organization starting from a molecular initiating event through to the observable adverse outcome (OECD 2013; Berggren et al. 2015). Although all three terms are used to describe the sequence of steps from an initiating event to an adverse outcome, subtle distinctions between the terms have been made. The subtleties are often lost in practice, and the terms are used interchangeably. In the present report, the committee uses primarily *mechanism* and defines the term generally to refer to a detailed description of the process by which an agent causes an effect. It uses *adverse outcome pathway* only in the context of frameworks that have been developed specifically with the phrase. Mechanism is further defined in the context of the new direction of risk assessment in Chapters 5 and 7.

*Exposure* and *dose* are two other terms that are often defined and used inconsistently. The NRC (2012) defined exposure broadly as the contact between a stressor and a receptor at any level of biological organization (organism, organ, tissue, or cell). Given that broad definition, the distinction between *exposure* and *dose* becomes arbitrary, and *dose* becomes unnecessary. Exposure is then characterized by the identity of the stressor and the amount, location, and timing of the stressor that comes into contact with the receptor; timing encompasses both duration and the time at which the contact occurs. The committee uses *exposure* primarily in the present report but acknowledges that it often uses *dose* in conventional phrases, such as dose–response relationship.

Many terms associated with -omics technologies have been coined in recent years. Box 1-1 provides definitions of various terms used throughout this report. Other terms that are specific to topics discussed in various chapters are defined in those chapters. The committee acknowledges that as the science progresses new terms will be needed, but it urges the scientific community to be judicious in inventing new terms. If needed, new terms should be defined clearly and used consistently.

The committee debated how to refer to all the assays, tools, and methods arising from the "21st century visions" for toxicology and exposure science; some are

**BOX 1-1** Definitions of Various -Omics Terms

*Adductomics*: The comprehensive identification of chemicals that bind to DNA or selected proteins, such as albumin.

*Epigenomics*: The analysis of epigenetic changes in DNA, histones, and chromatin that regulate gene expression. Epigenetic changes are changes other than changes in DNA sequence that are involved in gene silencing.

*Exposome*: A term first coined by Wild (2005) to represent the totality of a person's exposure from conception to death; exposome research involves the measurement of multiple exposure indicators by using -omics approaches.

*Genomics*: The analysis of the structure and function of genomes.

*Metabolomics*: The scientific study of small molecules (metabolites) that are created from chemicals that originate inside the body (endogenously) or outside the body (exogenously) (NASEM 2016). For purposes of the present report, metabolomics is assumed to include exogenous chemicals found in biological systems in their unmetabolized forms.

*Proteomics*: The analysis of the proteins produced by cells, tissues, or organisms. Analysis is conducted to understand the location, abundance, and post-translational modification of proteins in a biological sample.

*Transcriptomics*: Qualitative and quantitative analysis of the transcriptome, that is, the set of transcripts (mRNAs, noncoding RNAs, and miRNAs) that is present in a biological sample.

no longer "new," and others are still in development. To simplify the text, the committee often refers to them as Tox21 or ES21 assays, tools, or methods. That notation is meant to be broad and includes all the assays, tools, and methods coming from government, academic, and private laboratories, not only those being developed as part of the Tox21 program previously described.

## THE COMMITTEE AND ITS TASK

The committee that was convened as a result of the agencies' request included experts in toxicology; physiologically based pharmacokinetic modeling; computational methods and bioinformatics; -omics, in vitro models, and alternative methods; epidemiology; exposure assessment; statistics; and risk assessment (see Appendix A for the committee's biographical information). As noted, the committee was asked to consider and recommend the best uses of the various types of emerging data in risk-based evaluations. The committee's verbatim statement of task is provided in Box 1-2.

## THE COMMITTEE'S APPROACH TO ITS TASK

To address its task, the committee held seven meetings, which included three open sessions to hear primarily from various sponsor representatives. Given the potential breadth of its task, the committee devoted substantial time to interpretation of its charge. It used as a basis of its work the risk-assessment framework that was initially proposed in the 1983 report *Risk Assessment in the Federal Government: Managing the Process* (NRC 1983) and updated most recently in the 2009 report *Science and Decisions: Advancing Risk Assessment* (NRC 2009) (see Figure 1-1). The committee considered and describes scientific and technological advances in exposure science, toxicology, and epidemiology that could be integrated into and used to improve any of the four elements of risk assessment (hazard identification, dose–response assessment, exposure assessment, and risk characterization). The report, however, is not a catalog of all scientific and technological advances that have been made since publication of the 2007 and 2012 reports (NRC 2007, 2012), but rather a review of the ones most relevant to risk-based evaluations in EPA and FDA.

The committee identified various agency tasks and decision-making contexts (see Box 1-3)—which require different depths of information—and used the tasks and contexts to frame general and specific examples of applications (case studies) for integrating the new science into various components of risk assessment. The examples provide guidance for communicating to various stakeholders how the new science could be used. The committee then considered how data validation, data integration, and uncertainty analysis might need to be adapted to use the new science. The committee recognizes that there will be challenges in using new tools and concepts in fields that are already heavy with practice standards and set protocols.

## ORGANIZATION OF THIS REPORT

The committee's report is organized into seven chapters and five appendixes. Chapters 2, 3, and 4 describe new or emerging methods and tools in exposure science, toxicology, and epidemiology, respectively. Chapter 5 highlights the new direction of risk assessment and describes practical applications for 21st century science.

**BOX 1-2** Statement of Task

An ad hoc committee under the auspices of the National Research Council (NRC) will provide recommendations on integrating new scientific approaches into risk-based evaluations. Specifically, the committee will first consider the scientific advances that have occurred following the publication of the NRC reports *Toxicity Testing in the 21st Century: A Vision and a Strategy* and *Exposure Science in the 21st Century: A Vision and a Strategy*. Given the various ongoing lines of investigation and new data streams that have emerged, the committee will then propose how best to integrate and use the emerging results in evaluating chemical risk and identify how traditional human-health risk assessment can incorporate the new science. It will consider whether a new paradigm is needed for data validation (or acceptance), how to integrate the divergent data streams, how uncertainty might need to be characterized (or how characterization of uncertainty might need to change), and how best to communicate the new approaches so that they are understandable to various stakeholders. It will focus its recommendations on pragmatic solutions and provide case studies that illustrate its recommendations. Finally, the committee will identify barriers or obstacles to advancing and integrating the various types of science, and ultimately transforming risk assessment.

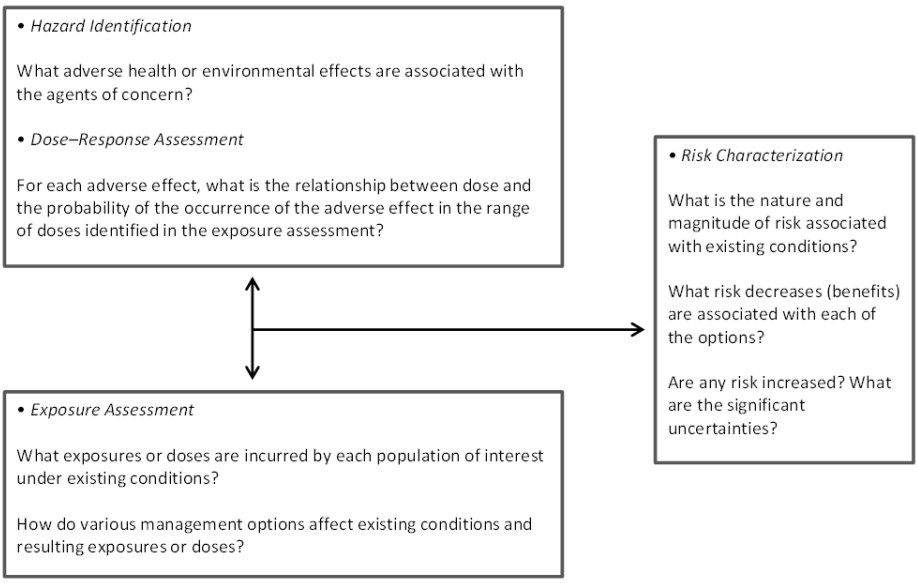



**FIGURE 1-1** The risk-assessment process as defined by its four elements: hazard identification, dose–response assessment, exposure assessment, and risk characterization. Source: Adapted from NRC 2009.

**BOX 1-3** Agency Tasks and Decision-Making Contexts

1. Priority-setting—Can be based on hazard, exposure, or risk.
2. Chemical assessment—Can include Integrated Risk Information System assessments, Provisional Peer Reviewed Toxicity Values, National Toxicology Program Office of Health Assessment and Translation hazard assessments, and assessments of various regulated substances, such as pesticides, drugs, and food additives.
3. Site-specific assessments—Can involve selection of geographic sites or chemicals at a site to evaluate and can involve assessment of data-poor chemicals or mixtures; can also involve assessment of previously unidentified chemicals in the environment.
4. Assessment of new chemistries—Can involve assessment of green chemistry, new-to-the-world technologies, and unexpected environmental degradation products of chemicals in commerce.

Chapter 6 discusses issues surrounding model and assay validation and acceptance. Chapter 7 focuses on interpretation and integration of data and evidence. Appendix A provides biographical information on the committee members, and Appendixes B, C, and D provide case studies that demonstrate practical applications of the committee's recommendations for using new data streams in risk-based evaluations. Appendix E provides a case study in using Bayesian approaches with high-throughput data.

## REFERENCES

Berggren, E. 2015. A Path to Validation-SEURAT-1 Case Studies and the Role of ECVAM. Public Forum-Replacing Animal Testing, October 26-27, 2015, London. ToxBank [online]. Available: http://www.toxbank.net/public-forum/path-validation [accessed January 3, 2017].

Berggren, E., P. Amcoff, R. Benigni, K. Blackburn, E. Carney, M. Cronin, H. Deluyker, F. Gautier, R.S. Judson, G.E. Kass, D. Keller, D. Knight, W. Lilienblum, C. Mahony, I. Rusyn, T. Schultz, M. Schwarz, G. Schüürmann, A. White, J. Burton, A.M. Lostia, S. Munn, and A. Worth. 2015. Chemical safety assessment using read-across: Assessing the use of novel testing methods to strengthen the evidence base for decision making. Environ. Health Perspect. 123(12):1232-1240.

Bowes. J., A.J. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldron, and S. Whitebread. 2012. Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. Nat. Rev. Drug Discov. 11(12):909-922.

Churchill, G.A, D.C. Airey, H. Allayee, J.M. Angel, A.D. Attie, J. Beatty, W.D. Beavis, J.K. Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K.W. Broman, K.J. Buck, E. Buckler, M. Burmeister, E.J. Chesler, J.M. Cheverud, S. Clapcote, M.N. Cook, R.D. Cox, J.C. Crabbe, W.E. Crusio, A. Darvasi, C.F. Deschepper, R.W. Doerge, C.R. Farber,

J. Forejt, D. Gaile, S.J. Garlow, H. Geiger, H. Gershenfeld, T. Gordon, J. Gu, W. Gu, G. de Haan, N.L. Hayes, C. Heller, H. Himmelbauer, R. Hitzemann, K. Hunter, H.C. Hsu, F.A. Iraqi, B. Ivandic, H.J. Jacob, R.C. Jansen, K.J. Jepsen, D.K. Johnson, T.E. Johnson, G. Kempermann, C. Kendziorski, M. Kotb, R.F. Kooy, B. Llamas, F. Lammert, J.M. Lassalle, P.R. Lowenstein, L. Lu, A. Lusis, K.F. Manly, R. Marcucio, D. Matthews, J.F. Medrano, D.R. Miller, G. Mittleman, B.A. Mock, J.S. Mogil, X. Montagutelli, G. Morahan, D.G. Morris, R. Mott, J.H. Nadeau, H. Nagase, R.S. Nowakowski, B.F. O'Hara, A.V. Osadchuk, G.P. Page, B. Paigen, K. Paigen, A.A. Palmer, H.J. Pan, L. Peltonen-Palotie, J. Peirce, D. Pomp, M. Pravenec, D.R. Prows, Z. Qi, R.H. Reeves, J. Roder, G.D. Rosen, E.E. Schadt, L.C. Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M.J. Sillanpaa, L.D. Siracusa, H.W. Snoeck, J.L. Spearow, K. Svenson, L.M. Tarantino, D. Threadgill, L.A. Toth, W. Valdar, F.P. de Villena, C. Warden, S. Whatley, R.W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, and F. Zou. 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36(11):1133-1137.

Collins, F.S., G.M. Gray, and J.R. Bucher. 2008. Transforming environmental health protection. Science 319(5865):906-907.

CORDIS (Community Research and Development Information Service). 2015. EXPOSOMICS Result In Brief: Measuring Environmental Exposure. CORDIS, European Commission [online]. Available: http://cordis.europa.eu/result/rcn/151545_en.html [accessed July 13, 2016].

Daston, G., D.J. Knight, M. Schwarz, T. Gocht, R.S. Thomas, C. Mahony, and M. Whelan. 2015. SEURAT: Safety Evaluation Ultimately Replacing Animal Testing–Recommendations for future research in the field of predictive toxicology. Arch. Toxicol. 89(1):15-23.

de Cock, M., Y.G. Maas, and M. van de Bor. 2012. Does perinatal exposure to endocrine disruptors induce autism spectrum and attention deficit hyperactivity disorders? Review Acta Paediatr. 101(8):811-818.

Dolinoy, D.C., J.R. Weidman, and R.L. Jirtle. 2007. Epigenetic gene regulation: Linking early developmental environment to adult disease. Reprod. Toxicol. 23(3):297-307.

EPA (US Environmental Protection Agency). 2003. Framework for Computational Toxicology Research Program in ORD. EPA/600/R-03/065. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

EPA (US Environmental Protection Agency). 2005. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Risk Assessment Forum, US Environmental Protection Agency, Washington, DC [online]. Available: http://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf [accessed July 13, 2016].

EPA (US Environmental Protection Agency). 2013. Toxicity Forecaster (ToxCast™). Science in Action: Innovative Research for a Sustainable Future. Fact Sheet. Office of Research and Development, US Environmental Protection Agency, Washington, DC [online]. Available: https://www.epa.gov/sites/production/files/2013-12/documents/toxcast-fact-sheet.pdf [accessed July 13, 2016].

EPA (US Environmental Protection Agency). 2016a. The Exposure Science in the 21st Century (ES21) Federal Working Group [online]. Available: https://www.epa.gov/innovation/exposure-science-21st-century-federal-working-group [accessed October 24, 2016].

EPA (US Environmental Protection Agency). 2016b. High-Throughput Exposure Forecasting. Science in Action: Innovative Research for a Sustainable Future. Fact Sheet. Office of Research and Development, US Environmental Protection Agency, Washington, DC. March 2016 [online]. Available: https://www.epa.gov/sites/production/files/2014-12/documents/exposure_forecasting_factsheet.pdf [accessed July 13, 2016].

Eubig, P.A., A. Aguiar, and S.L. Schantz. 2010. Lead and PCBs as risk factors for attention deficit/hyperactivity disorder. Environ. Health Perspect. 118(12):1654-1667.

Fenton, S.E. 2006. Endocrine-disrupting compounds and mammary gland development: Early exposure and later life consequences. Endocrinology 147(6 Suppl.): S18-S24.

Gocht, T., E. Berggren, H.J. Ahr, I. Cotgreave, M.T. Cronin, G. Daston, B. Hardy, E. Heinzle, J. Hescheler, D.J. Knight, C. Mahony, M. Peschanski, M. Schwarz, R.S. Thomas, C. Verfaillie, A. White, and M. Whelan. 2015. The SEURAT-1 approach towards animal free human safety assessment. ALTEX 32(1):9-24.

Greene, N., and M. Song. 2011. Predicting in vivo safety characteristics using physiochemical properties and in vitro assays. Future Med. Chem. 3(12):1503-1511.

Kavlock, R., and D. Dix. 2010. Computational toxicology as implemented by the US EPA: Providing high throughput decision support tools for screening and assessing chemical exposure, hazard, and risk. J. Toxicol. Environ. Health B Crit. Rev. 13(2-4):197-217.

Leon-Olea, M., C.J. Martyniuk, E.F. Orlando, M.A. Ottinger, C.S. Rosenfeld, J.T. Wolstenholme, and V.L. Trudeau. 2014. Current concepts in neuroendocrine disruption. Gen. Comp. Endocrinol. 203:158-173.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2016. Use of Metabolomics to Advance Research on Environmental Exposures and the Human Exposome: Workshop in Brief. Washington, DC: The National Academies Press.

NCATS (National Center for Advancing Translational Sciences). 2015a. Tox21 Program Goals [online]. Available: http://www.ncats.nih.gov/tox21/about/goals [accessed July 13, 2016].

NCATS (National Center for Advancing Translational Sciences). 2015b. Tox21 Operational Model. Available: http://www.ncats.nih.gov/tox21/about/operations [accessed July 13, 2016].

Nebert, D.W., G. Zhang, and E.S. Vesell. 2013. Genetic risk prediction: Individualized variability in susceptibility to toxicants. Annu. Rev. Pharmacol. Toxicol. 53:355-375.

NIEHS (National Institute of Environmental Health Sciences). 2012. Advancing Science, Improving Health: A Plan for Environmental Health Research. 2012-2017 Strategic Plan [online]. Available: http://www.niehs.nih.gov/about/strategicplan/strategicplan2012_508.pdf [accessed July 13, 2016].

NIEHS (National Institute of Environmental Health Sciences). 2015. Exposure Biology and the Exposome [online]. Available: http://www.niehs.nih.gov/research/supported/dert/programs/exposure/ [accessed July 13, 2016].

NIEHS (National Institute of Environmental Health Sciences). 2016. Children's Health Exposure Analysis Resource (CHEAR) [online]. Available: http://www.niehs.nih.gov/research/supported/exposure/chear/ [accessed July 13, 2016].

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.

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

NRC (National Research Council). 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

NRC (National Research Council). 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2004. A National Toxicology Program for the 21st Century: A Roadmap for the Future [online]. Available: https://ntp.niehs.nih.gov/ntp/about_ntp/ntpvision/ntproadmap_508.pdf [accessed July 13, 2016].

OECD (Organisation for Economic Co-operation and Development). 2013. Guidance Document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment. No. 184. ENV/JM/MONO(2013)6. Paris: OECD [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en [accessed July 13, 2016].

Osborne, G., R. Rudel, and M. Schwarzman. 2015. Evaluating chemical effects on mammary gland development: A critical need in disease prevention. Reprod. Toxicol. 54:148-155.

Patisaul, H.B., and E.K. Polston. 2008. Influence of endocrine active compounds on the developing rodent brain. Brain Res. Rev. 57(2):352-362.

Rappaport, S.M., and M.T. Smith. 2010. Environment and disease risks. Science 30(6003):460-461.

Rissman, E.F. and M. Adli. 2014. Transgenerational epigenetic inheritance: Focus on endocrine disrupting compounds. Minireview. Endocrinology 155(8):2770-2780.

Rusyn, I., D.M. Gatti, T. Wiltshire, S.R. Kleeberger, and D.W. Threadgill. 2010. Toxicogenetics: Population-based testing of drug and chemical safety in mouse models. Pharmacogenomics 11(8):1127-1136.

SEURAT (Safety Evaluation Ultimately Replacing Animal Testing). 2015. SEURAT-1 [online]. Available: http://www.seurat-1.eu/ [accessed July 13, 2016].

Silva, M., N. Pham, C. Lewis, S. Iyer, E. Kwok, G. Solomon, and L. Zeise. 2015. A comparison of ToxCast test results with in vivo and other in vitro endpoints for neuro, endocrine, and developmental toxicities: A case study using endosulfan and methidathion. Birth Defects Res. B Dev. Reprod. Toxicol. 104(2):71-89.

Soto, A.M., L.N. Vandenberg, M.V. Maffini, and C. Sonnenschein. 2008. Does breast cancer start in the womb? Basic Clin. Pharmacol. Toxicol. 102(2):125-133.

Sullivan, A.W., E.C. Beach, L.A. Stetzik, A. Perry, A.S. D'Addezio, B.S. Cushing, and H.B. Patisaul. 2014. A novel model for neuroendocrine toxicology: neurobehavioral effects of BPA exposure in a prosocial species, the prairie vole (Microtus ochrogaster). Endocrinology 155(10):3867-3881.

Tal, T.L., and R.L. Tanguay. 2012. Non-coding RNAs–novel targets in neurotoxicity. Neurotoxicology 33(3):530-544.

Tice, R.R., C.P. Austin, R.J. Kavlock, and J.R. Bucher. 2013. Improving the human hazard characterization of chemicals: A Tox21 Update. Environ. Health Perspect. 121(7):756-765.

Vineis, P., K. van Veldhoven, M. Chadeau-Hyam, and T.J. Athersuch. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology. Environ. Mol. Mutagen. 54(7):461-467.

Vineis, P., M. Chadeau-Hyam, H. Gmuender, J. Gulliver, Z. Herceg, J. Kleinjan, M. Kogevinas, S. Kyrtopoulos, M. Nieuwenhuijsen, D. Phillips, N. Probst-Hensch, A. Scalbert, R. Vermeulen, and C.P. Wild. In press. The exposome in practice: Design of the EXPOsOMICS project. EXPOsOMICS Consortium. Int. J. Hyg. Environ. Health.

Vrijheid, M., R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, C. Thomsen, J. Wright, T.J. Athersuch, N. Avellana, X. Basagaña, C. Brochot, L. Bucchini, M. Bustamante, A. Carracedo, M. Casas, X. Estivill, L. Fairley, D. van Gent, J.R. Gonzalez, B. Granum, R. Gražulevičienė, K.B. Gutzkow, J. Julvez, H.C. Keun, M. Kogevinas, R.R.C. McEachan, H.M. Meltzer, E. Sabidó, P.E. Schwarze, V. Siroux, J. Sunyer, E.J. Want, F. Zeman, and M.J. Nieuwenhuijsen1. 2014. The human early-life exposome (HELIX): Project rationale and design. Environ. Health Perspect. 122(6):535-544.

Wild, C.P. 2005. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomarkers Prev. 14(8):1847-1850.

Wu, S., K. Blackburn, J. Amburgey, J. Jaworska, and T. Federle. 2010. A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. Regul. Toxicol. Pharmacol. 56(1):67-81.

Wu, S., J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, and K. Blackburn. 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem. Res. Toxicol. 26(12):1840-1861.

Yeo, M., H. Patisaul, and W. Liedtke. 2013. Decoding the language of epigenetics during neural development is key for understanding development as well as developmental neurotoxicity. Epigenetics 8(11):1128-1132.

# 2

# Advances in Exposure Science

As described in Chapter 1, the National Research Council (NRC) report *Exposure Science in the 21st Century: A Vision and a Strategy* articulated a vision for exposure science that was intended to transform, expand, and invigorate the field (NRC 2012). Recent investments in exposome technologies and programs (CHEAR; NIEHS 2016), in new large-scale longitudinal exposure-epidemiology research programs (HELIX; Vrijheid et al. 2014 and EXPOsOMICS; Vineis et al. 2013), and in the rapidly expanding exposure-science programs headed by the National Exposure Research Laboratory and the National Center for Computational Toxicology of the US Environmental Protection Agency (EPA) are examples of the immediate impact of the ES21 report.[1] Several research fields have seen substantial advances since the ES21 report was published, and these advances create opportunities for providing guidance to EPA, the US Food and Drug Administration, and others on how best to integrate emerging exposure-science data into risk assessments (Egeghy et al. 2016). Accordingly, this chapter describes the major advances in exposure science since the publication of the ES21 report and applications that would be most relevant and useful for risk-based decision-making. It also presents unaddressed opportunities related to decision-making based on exposure or risk and discusses major obstacles to various applications.

The interrelationship among the fields of exposure science, toxicology, and epidemiology is a central theme of this chapter. Figure 2-1 illustrates the series of events from introduction of a stressor into the environment and its movement through the environment via specific pathways to the receptor and the triggering of a biological response of potential regulatory concern. The figure provides a broad conceptual overview of the scope of exposure science and a general organizational framework as envisaged by the ES21 committee and the present committee. The figure also illustrates the points of integration with toxicology and epidemiology and the fundamental distinctions between fields. Although the continuum is depicted as a linear path, the committee recognizes that multiple interconnecting paths are typically involved in the source-to-outcome continuum. In cases where source identification or mitigation rather than toxicology or risk assessment is the goal, one moves from right to left from measured exposures to sources. Box 2-1 provides some definitions of the key terms used in this chapter related to exposure science.

Organizational frameworks for exposure science, such as the one in Figure 2-1, have been used to describe exposure pathways for contaminated sites and are implicit in all models of environmental or biological fate of chemicals (Wania and Mackay 1999; Koelmans et al. 2001; Schenker et al. 2009). The frameworks have been essential in guiding the acquisition of data, the organization of data, and the use of data in modeling to describe or predict exposure quantitatively. Although some frameworks, such as the Conceptual Site Model (Regens et al. 2002; Mayer et al. 2005), are largely qualitative and conceptual and apply to specific exposure settings or specifically to modeling exercises, others, such as the Aggregate Exposure Pathway framework (Teeguarden et al. 2016), attempt to expand on earlier successes by generalizing the approach to support data acquisition, data organization, conceptualization, and modeling in the broader exposure-science community. As the field of exposure science evolves as a result of advances in the tools and approaches described in this chapter, the use of the frameworks will enable the development of infrastructure to support exposure-data acquisition, collection, organization, and access and to improve the accuracy, completeness, efficiency, and transparency of exposure assessment and modeling.

## MAJOR ADVANCES IN EXPOSURE SCIENCE

The committee reviewed advances in the field of exposure science since the publication of the ES21 report with the goal of identifying major advances that have the potential for sustained effects on the important applications described later in this chapter and in the case studies

[1]The present committee refers to *Exposure Science in the 21st Century: A Vision and a Strategy* (NRC 2012) as the ES21 report and to its committee as the ES21 committee.

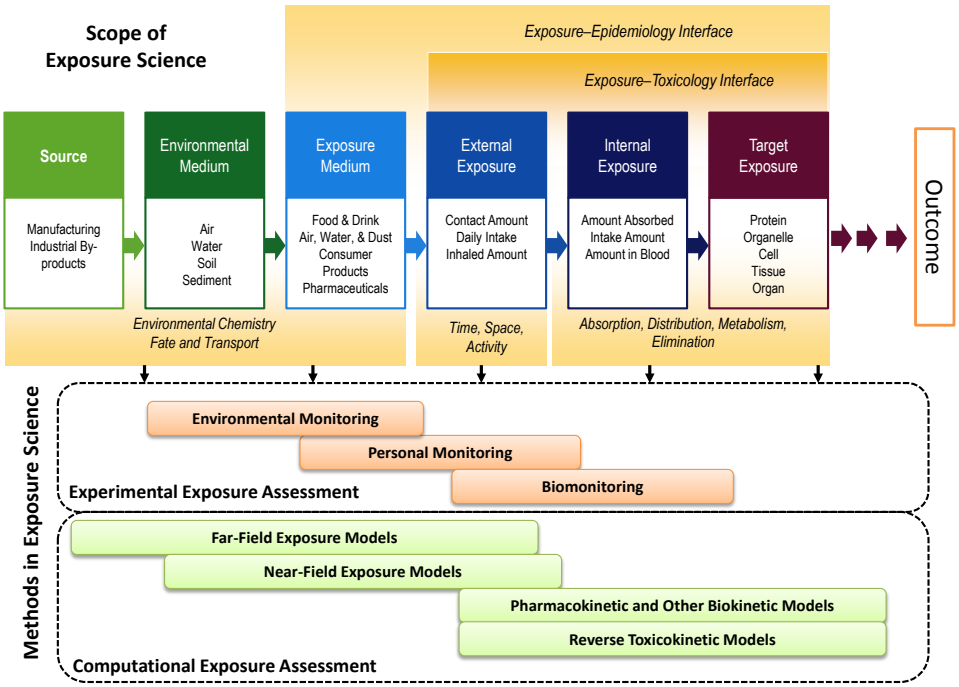



**FIGURE 2-1** Conceptual overview of the scope of and common methods for exposure science. Toxicology and epidemiology have traditionally used both internal-exposure and external-exposure information. The biological interface between exposure and a receptor (such as a human, tissue, or cell) is the test-system or target-site exposure. The main benefit of applying target-site exposures is a reduction in confounding by pharmacokinetic and other factors and has led to increasing use of target-site exposure metrics in toxicology and epidemiology.

described in Appendixes B–D. The advances are summarized in this section.

## Remote Sensing and Geospatial Environmental Exposure Assessment

Several substantial advances in exposure science are the result of innovations in remote sensing, global positioning systems (GPSs), and geographic information systems (GISs). Remote sensing is an important tool for enhancing the capacity to assess human and ecological exposures because it provides information on Earth's surface, water, and atmosphere that cannot be provided by traditional ground-based monitoring systems (Al-Hamdan et al. 2014). Since the ES21 report, remote-sensing data have been used to estimate concentrations of ambient criteria air pollutants ($NO_2$, $O_3$, and $PM_{2.5}$) on a global scale (Brauer et al. 2015; Geddes et al. 2016; van Donkelaar et al. 2015). Models have estimated the changes in global air pollution and have allowed complete global coverage of key pollutants on a relatively fine spatial scale. The application of remote-sensing technologies with ground-based monitoring will continue to improve human exposure assessment. Several recent key advances include the National Aeronautics and Space Administration (NASA) launch of six Earth-observing missions and the addition of three new instruments to the International Space Station (Seltenrich 2014). NASA and the National Oceanic and Atmospheric Administration provide free access to exposure-relevant data, such as $NO_2$ and $PM_{2.5}$ concentrations in the troposphere, and environmental data relevant to exposure assessment and interpretation of monitoring data (Seltenrich 2014).

The studies generated with remote sensing data provide even greater insights into human exposures when coupled with GPS and GIS data on populations of interest. GPS data are used to track people in observational exposure and epidemiological studies (Elgethun et al. 2007), and recent advances have allowed more automated coding of GPS data on activities and microenvironments, such as inside and outside at home and at work (Wu et al. 2011; Breen et al. 2014; Nethery et al. 2014; Andra et al. 2015). Data on microenvironments can be used as input for exposure models to refine exposure estimates based on remote sensing data, ground-based ambient air data, and indoor air monitoring data (Breen et al. 2014). Advances in GPS technologies have also been coupled with sensor technologies that measure basic health data, such as heart and respiratory rates and activity level. Information on such measures can be additional inputs for the exposure models and allow further refinement and improvement of exposure classification (Andersen et al. 2015).

**BOX 2-1** Definitions of Selected Exposure Terms

***Exposure science.*** "The collection and analysis of quantitative and qualitative information needed to understand the nature of contact between receptors (such as people or ecosystems) and physical, chemical, or biologic stressors. Exposure science strives to create a narrative that captures the spatial and temporal dimensions of exposure events with respect to acute and long-term effects on human populations and ecosystems" (NRC 2012).

***Internal and external exposure.*** Internal and external exposures are two commonly used classes of exposure metrics. Blood or tissue concentrations from biomonitoring studies are relatively direct measures of internal exposure; amounts or concentrations in biofluids leaving the body (breath and urine) are less direct measures. Internal measures can be estimated from the less direct measures when supporting pharmacokinetic data and models are available. Air or media concentrations are external measures of exposure from which internal measures of exposure might be derived if necessary. What exposure metric is considered appropriate depends on the decision context, confidence in the measurement, and proximity to the site of action.

***Near-field chemical exposures.*** Near-field human exposures result from chemical release or use near a person. Near-field chemical exposures include direct dermal application (for example, of sunscreen or cosmetics), direct inhalation (for example, of tobacco smoke or pharmaceuticals), and direct ingestion (for example, of pharmaceuticals). Near-field chemical exposures can also result from the intentional use (as in the case of consumer products) and inadvertent release (as in the case of building materials) of chemicals near a person and later near-field transport to a person that results in contact or intake through inhalation, dermal, or ingestion pathways.

***Far-field chemical exposures.*** Far-field human exposures result from release or use distant from a person. They can also result from initial near-field use (indoors) and later fate and transport in the natural environment (outdoors) before the chemical reaches a person. Far-field exposures can result from inhalation of outdoor air and ingestion of drinking water and foods that contain chemicals that have entered the contact media through fate and transport processes in the natural environment.

***Aggregate exposure.*** Aggregate exposure is exposure to a given substance from multiple sources via multiple pathways and routes (that is, combined exposure from multiple sources by dermal, ingestion, and inhalation routes).

### Computational Exposure Assessment

For the vast majority of stressors, there are few exposure measurements (Muir and Howard 2006; Egeghy et al. 2012). Various conceptual, empirical, and predictive exposure models are needed to address those data gaps and to enhance the usefulness and application of measured data to exposure and risk assessment. Since the release of the ES21 report, there has been substantial research activity and advancement in the development of computational exposure tools, particularly for calculating near-field chemical exposures of humans, for quantifying relationships between external and internal exposures and between in vivo and in vitro exposures, and for high-throughput exposure estimation that has been used alone and in combination with bioactivity data to set priorities for chemical assessment.

Egeghy et al. (2011) reviewed tools designed to set priorities rapidly for large numbers of chemicals, and Mitchell et al. (2013) conducted an "exposure model prioritization challenge." A key finding of the challenge was the need to reconcile exposures to chemicals released outdoors (far-field sources) with exposures to chemicals in consumer products applied directly or through indoor-environment exposure pathways (near-field exposures). The recognized absence of tools and exposure information is stimulating research to develop and improve near-field and far-field exposure science. Specifically, the seminal model developed for simulating chemical transport in an indoor environment (Bennett and Furtaw 2004) has been revised to include exposure pathways for which external human exposures (intake fractions) (Shin et al. 2012) and internal exposures (estimates of whole-body concentrations) (Zhang et al. 2014; Webster et al. 2016) can be

estimated. Furthermore, data and models are evolving to improve mechanistic understanding of chemical releases and exposures indoors (Weschler and Nazaroff 2010, 2012; Little et al. 2012). Exposure models for consumer products also are evolving and being evaluated for select chemicals (Young et al. 2012; Gosens et al. 2014; Delmaar et al. 2015; Dudzina et al. 2015). Exposure models and frameworks that combine far-field and near-field pathways for aggregate human exposure assessments are also being developed and applied (Isaacs et al. 2014; Shin et al. 2015; Fantke et al. 2016).

EPA's ExpoCast project conducts research on and uses computational tools for high-throughput exposure estimation or "forecasting" to set testing or assessment priorities. The ExpoCast project combines various models and data sources to estimate exposures, which can then be compared with high-throughput ToxCast data and other sources of toxicity or bioactivity data. As a part of the ExpoCast exposure estimation, the Systematic Empirical Evaluation of Models (SEEM) framework includes calibration and evaluation of exposure-model estimates against chemical concentrations measured in or estimated from blood and urine samples from a group of nonoccupationally exposed US residents over the age of 6 years (Wambaugh et al. 2013, 2014).[2] Exposure-model predictions are compared with available biomonitoring data to estimate the uncertainty in the combined exposure-model calculations (Wambaugh et al. 2013). The Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway chemicals (SHEDS-MM) for exposure-based priority-setting and screening has been revised for high-throughput capacity (SHEDS-HT) (Isaacs et al. 2014) and feeds into the SEEM framework. Other complementary high-throughput aggregate exposure-estimation models that combine existing and emerging tools (see, for example, Shin et al. 2015) can also be incorporated into the SEEM framework, and they are being applied, evaluated, and refined in other contexts.

Improving the amount and quality of the data that are needed to develop parameters for the computational exposure tools is critically important; without such data, the applicability of the tools is limited. Some advances include updated exposure factors (EPA 2011) and the development of the Consumer Product Chemical Profile Database (Goldsmith et al. 2014) and the Chemical/Product Categories Database (Dionisio et al. 2015).[3] Numerous quantitative structure–activity relationship (QSAR) models, quantitative structure–property relationship (QSPR) models, and other computational tools for predicting chemical-property information—such as solubilities, partition coefficients, and degradation rates—continue to evolve. The applicability domains of existing tools for calculating chemical-property information require further examination and more explicit definition to ensure that the models are calibrated and applied within the same chemical space. Integrated testing strategies to obtain more high-quality measurements can then be strategically developed to expand the applicability domains of current QSAR models, QSPR models, and other tools used for property estimation.

Because of the extensive measurement-data gaps, the recent advances in computational tools for exposure science are expected to play a crucial role in most aspects of exposure estimation for risk-based assessments, not only high-throughput applications. Higher-tiered models that link exposure databases and spatial information (see, for example, Georgopoulos et al. 2014) and opportunities to combine and integrate measurements and models to characterize and quantify the source-to-receptor relationship more fully (see, for example, McKone et al. 2007) are being developed and applied. Exposure-model uncertainty and sensitivity analyses are useful computational methods that can be used to set priorities for exposure-science research systematically (Arnot et al. 2012; NRC 2012; Arnold and Ramachandran 2014).

## Personalized Exposure Assessment

Behavior patterns that determine exposure routes, durations, and conditions combined with the variation in environmental concentrations of stressors over space and time result in unique exposure patterns for individuals and populations. Exposure data that are needed to assess personal exposures can now be generated on various spatial and temporal scales with traditional and emerging methods. New opportunities to measure exposures in and outside the body will help to characterize and quantify personal exposures to an array of stressors. Particularly notable are recent advances in the application of passive sampling techniques to determine internal human concentrations (for example, using silicone implants) (Allan et al. 2013a; Gilbert et al. 2015; O'Connell et al. 2015), external exposure concentrations integrated over time and space (for example, using silicone wristbands) (O'Connell et al. 2014a,b), and chemical concentrations and chemical activities[4] in media to which humans are exposed, such as foods (Allan et al. 2013b; Jahnke et al. 2014) and indoor air (Wetzel and Doucette 2015). Por-

[2]Data are from the US National Health and Nutrition Examination Survey.

[3]See http://actor.epa.gov/cpcat.

[4]Chemical activity is related to the energetic state of a chemical, is a measure of the *effective concentration* of a chemical in a given exposure medium (Reichenberg and Mayer 2006; Mackay et al. 2011), and is closely related to the freely dissolved concentration. For example, chemical activity is an improved measure of exposure when interaction with media constituents (such as particles in air and organic matter in water) effectively reduces the amount of chemical free to interact with a biological receptor (such as a human), often referred to as the bioavailable fraction.

table sensors for measuring particles and volatile organic chemicals are being refined and are providing valuable information on personal exposures, particularly in vulnerable populations (McGinn et al. 2016). Mobility-based exposure assessment that uses personal devices, such as cell phones, to provide GPS information, can be used to determine time and location of people relative to exposure levels determined from remote sensing information (Adams et al. 2009; de Nazelle et al. 2013; Su et al. 2015). Consumer product and use databases and market research data can provide population and personal exposure information that can help to inform exposure assessment, for example (Goldsmith et al. 2014). All those emerging technologies and data streams will complement existing tools and techniques in the effort to obtain more comprehensive knowledge of the source-to-outcome continuum.

### Targeted and Nontargeted Exogenous Chemical Exposure Assessment

Important advances in two complementary approaches for characterizing human exposure—targeted and nontargeted analysis—are improving the accuracy and breadth of human and ecological exposure assessment (Fiehn 2002; Park et al. 2012; O'Connell 2014a,b; Go et al. 2015; Mastrangelo et al. 2015; Sud et al. 2016). Both approaches, whether focused on endogenous or exogenous chemicals, are commonly referred to as metabolomics approaches.[5] Targeted analysis focuses on selected chemicals for which standards and methods are available and identifies chemicals on the basis of mass spectrum, elution time, detector signals, or some combination of these measures. Targeted analysis has produced much of the exposure data used in epidemiological studies and risk assessment. The US National Health and Nutrition Examination Survey and the Canadian Health Measures Survey are two extensive biomonitoring programs that use targeted analytical techniques for exposure assessment (Needham et al. 2005; Calafat 2012; Haines and Murray 2012). Although initially limited by throughput and a focus on single chemicals, small groups of chemicals (Casas et al. 2011; Mortensen et al. 2014), or modest-size chemical classes (O'Connell et al. 2014b), targeted methods are emerging for quantitative analysis of hundreds of chemicals (O'Connell et al. 2015). Generally, there is a tradeoff between sensitivity and selectivity that imposes limitations on the number of chemicals that can be analyzed in single runs by using a single instrument or method. Targeted analyses are limited to chemicals for which standards are available. Accepted standards for identification and quantitation have been articulated for most analyte classes (such as metabolites and peptides) (Castle et al. 2006; Fiehn et al. 2006; Goodacre et al. 2007; Sumner et al. 2014), but these standards are inconsistently applied in practice.

Targeted analytical methods for protein and DNA adducts have emerged as an alternative to direct measurement of chemicals in blood. When stable protein or DNA adducts can be easily measured and information on the rates of adduct formation and loss is available, adduct concentrations can be used as proxies for the time-weighted average exposure to the parent chemical. Those approaches are particularly valuable for exposure assessment and exposure reconstruction for short-lived chemicals whose concentrations in blood and other biofluids might be very low and subject to high temporal variability. One example is the use of hemoglobin adducts of acrylamide and its metabolite glycidamide for accurate reconstruction of acrylamide exposure and its concentration in blood over time in humans (Young et al. 2007). Chemical-specific adducts of the carcinogens butadiene, formaldehyde, and acetaldehyde have emerged recently as metrics of exposure to these extremely short-lived chemicals (Swenberg et al. 2007, 2008; Moeller et al. 2013; Yu et al. 2015). The benefits of using stable adducts to measure exposure to short-lived chemicals include the ability to integrate exposure over time (that is, the adducts can serve as integrative measures of exposure because they are more stable) and biological relevance because of the proximity to a target site, such as DNA. Swenberg and co-workers have established highly sensitive methods for specific formaldehyde DNA adducts and pioneered methods for establishing the contribution of endogenous and exogenous formaldehyde to total internal exposure (Edrissi et al. 2013; Moeller et al. 2013; Pottenger et al. 2014; Pontel et al. 2015; Yu et al. 2015). The studies highlight the utility of targeted analysis of adducts for exposure assessment and perhaps a potential for broad assessment of the adductome (Gavina et al. 2014; Pottenger et al. 2014).

Nontargeted analysis has emerged as an approach to provide qualitative information on the large portion of the exposome that is uncharacterized—a portion that includes bioactive endogenous peptides, exogenous chemicals, metabolites, lipids, and other biomolecules. It offers the ability to survey more broadly the presence of all chemicals in the environment and in biofluids regardless of whether standards and methods are available. The nontargeted approach trades selectivity for breadth and produces numerous unidentified analytical features. Comparing unidentified analytical features from large cohorts and correlating them with responses of interest in the cohorts can help to identify analytical features for further investigation (Burgess et al. 2015). Cheminformatics and computational chemistry can be used to identify chemicals with varying levels of confidence; nuclear magnetic resonance spectroscopy can be used to identify chemical structure with high accuracy. Accepted standards for iden-

[5]As defined in Chapter 1 (see Box 1-1), metabolomics is assumed to include exogenous chemicals found in biological systems in their unmetabolized forms.

tification of metabolites (Castle et al. 2006; Fiehn et al. 2006; Sumner et al. 2014) have not been routinely applied to nontargeted approaches, so chemical matches to the analytical features is tentative and association between specific chemicals and disease is uncertain.

Nontargeted approaches are promising, but there are limitations in the use of data produced from nontargeted analyses that should be considered before collecting the data. For example, an unidentified analyte cannot be used to develop a mechanistic argument to support or refute a causal association between the presence of the analyte and a clinical effect, it cannot be quantified in absolute terms, it cannot be subjected to toxicity testing, and it cannot be attributed to sources for purposes of exposure mitigation. Although identifying all analytes is an important objective, reducing the number of analytes—to investigate, for example, on the basis of frequency in samples, membership in an important chemical class, and association with a clinical outcome—will be important until methods for identification of unknown analytes become more efficient.

Initial efforts to understand potential contributions of exogenous and endogenous exposure have led to important insights about the role of each and about potential limitations of analytical technologies. Rappaport and co-workers (2014) reported human blood concentrations of many chemicals, their sources, evidence of chronic-disease risks, and numbers of metabolic pathways. Blood concentrations of endogenous chemicals, food chemicals, and drugs were indistinguishable and spanned 11 orders of magnitude; blood concentrations of pollutants were on the average lower by a factor of about 1,000 (see Figure 2-2). Although the findings cannot be generalized to all chemicals or all exposure scenarios, the blood-concentration ranges highlight the importance of using highly sensitive analytical instrumentation to characterize human exposure (Athersuch 2016; Uppal et al. 2016).

Risk assessment and mitigation of sources and risks all depend on knowing absolute quantities of specific chemicals; therefore, targeted analyses will continue to be the primary source of exposure information. Because the results of nontargeted analyses provide only relative or qualitative exposures, they are not readily applicable to conventional risk assessment. However, when unidentified analytical features can be aggregated according to their toxicity or pharmacokinetic behavior, there will be new opportunities to conduct hazard or risk assessments on the basis of similarity to chemicals whose toxicity is known.

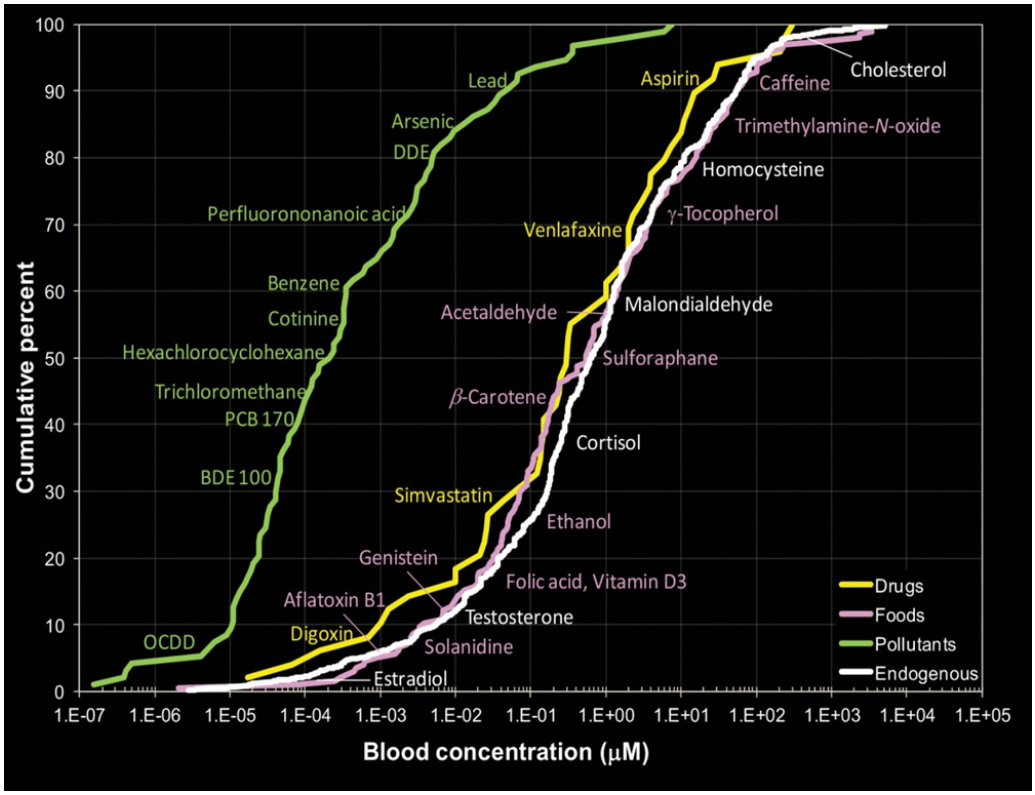



**FIGURE 2-2** A survey of measured blood concentrations shows that for the selected chemicals concentrations of pharmaceuticals and naturally present endogenous chemicals are similar and are generally higher than concentrations of environmental contaminants. The findings highlight the importance of using highly sensitive analytical instrumentation to characterize human exposure. Source: Rappaport et al. 2014.

### Exposure Inference from -Omics Technologies

-Omics technologies that quantify the abundance of biomolecules, such as proteins and transcripts, offer distinct and diverse applications for exposure assessment. In contrast with metabolomic approaches that quantify exposure to specific metabolites of endogenous and exogenous chemicals, proteomic and transcriptomic approaches provide global assessment of biological responses to exposure to multiple stressors. Those -omics approaches can provide biomarkers or biosignatures of response to chemical classes, such as oxidants (Roede et al. 2013; Go and Jones 2014) and potentially genotoxic chemicals (Fenech and Bonassi 2011; Lovreglio et al. 2014; Kalemba-Drozdz 2015; Moro et al. 2015; Tumer et al. 2016). That particular application of -omics technologies, a key element of Wild's original vision of the exposome (Wild 2005, 2012), is used to *infer* exposure to one or more chemicals on the basis of a mechanistic understanding of biological response to them. Some biomarkers of exposure can result from changes in the body that are induced by chemical exposure (for example, changes in metabolite or protein profiles), but these types of biomarkers commonly do not provide quantitative exposure information that can be used for risk estimation. The application of -omics technologies to infer exposure to classes of stressors is expected to grow. Although the initial utility will probably be in qualitative exposure inference and in assembling evidence on biological pathways, application should expand to more confident and more quantitative characterization of exposures to chemical classes or groups of stressors that produce the same biological effect, such as oxidation or inflammation.

### Novel Exposure Matrices for Exposure Reconstruction

Assessment of occupational and environmental exposures will continue to rely on matrices for which there are established methods of collection, analysis, and interpretation. Those matrices include air, water, soil, food, blood, and urine. The expanding computational exposure-science infrastructure (Arnot et al. 2012; Shin et al. 2012, 2015; Wambaugh et al. 2013, 2014; Isaacs et al. 2014), which uses the traditional data streams to construct population-level exposure assessments, will continue to drive the generation of data on the traditional exposure matrices.

Growing emphasis on near-field exposures (Stapleton et al. 2008; Shin et al. 2012; Wambaugh et al. 2014) and on exposures during development, which is the focus of the Children's Health Exposure Resource Centers of the National Institute of Environmental Health Sciences, is poised to drive exposure assessment toward new environmental and biological matrices and new approaches. For example, population-level exposure to hundreds of chemicals was recently shown to be dominated by near-field exposures from consumer-product and household use, not by far-field exposures that take place after chemicals are released into the outdoor environment (Shin et al. 2012; Wambaugh et al. 2014). Increased focus on categorizing chemicals in consumer products and on assembling exposure data for use in exposure assessment is one immediate outcome of the recent studies. Continued efforts to measure and estimate concentrations in multimedia sources—such as indoor air, indoor surfaces, dust, and consumer products—are required to address uncertainty in near-field exposures and pathways.

Characterization of exposures during the toxicologically sensitive period of fetal development has historically been limited to drawing inferences about maternal exposure through periodic maternal blood and urine measurements. Responding to the need to improve the characterization of fetal exposures to chemicals, researchers have turned to novel biological matrices, such as teeth, hair, nails, placental tissue, and meconium. The growth properties (the sequential deposition or addition of tissue) and availability of these biospecimens offer the opportunity to extract a record of exposure. For example, laser-ablation inductively coupled mass spectrometry has been used to reconstruct the timing of shifts in primates' diets that are associated with weaning by measuring calcium:barium ratios in tooth enamel (Austin et al. 2013). The same approach was recently shown to be promising for assessing in utero exposure to manganese. Arora et al. (2012) measured manganese concentrations in tooth dentine specific to the postnatal period and the second and third trimesters and showed a statistically significant relationship between house-dust manganese concentrations and dentine manganese concentrations during the second trimester. Those authors and others (Andra et al. 2015; Palmer et al. 2015) have extended the methods to measure organic chemicals, including phenols and phthalates. Like teeth, hair forms in utero (third trimester), continues to grow, and potentially provides a temporal record of exposure. Initially used widely for forensic analysis of exposure to illicit drugs, hair has emerged as an important matrix for biomonitoring of metals and organic chemicals, such as polybrominated diphenyl ethers (Aleksa et al. 2012; Liu et al. 2015a). Similar methods have been applied to fingernails (Liu et al. 2015a).

Although the new matrices mentioned above have advantages and add valuable information to exposure assessment, they pose challenges in interpretation and application. A common challenge in the use of exposure measures based on the new biological and environmental matrices for quantitative exposure assessment is the need to understand how measured concentrations are related to measures of exposure traditionally used to assess chemical toxicity or risk. Ideally, the new biomonitoring data can be supported by information regarding how measured

concentrations in new matrices are related to conventional measures of internal exposure (serum concentrations, μM) or external exposures (mg/kg-day or mmol/kg-day). New experimental data, such as chemical half-life in the body, and data related to events and processes of exposure, such as time since the exposure, that can inform various relationships and pharmacokinetic models will be useful in interpreting and reconstructing exposures by using the biomonitoring data (see, for example, Lorber and Egeghy 2011; Ritter et al. 2011; Quinn and Wania 2012; Wambaugh et al. 2013; Aylward et al. 2014; Hays et al. 2015). The additional information regarding the exposures provides confidence in using the measured biomonitoring data and supporting the exposure narrative.

## Physiologically Based Pharmacokinetic Models and Models for Translating Exposure Between Systems

Physiologically based pharmacokinetic (PBPK) models have made substantial contributions to exposure assessment for more than 30 years. PBPK models have been applied effectively to characterize target-tissue exposure in test animals and humans, to characterize pharmacokinetic variability, and to extrapolate across species, life stages, exposure routes, and, most recently, ecosystem elements (MacLachlan 2010; Weijs et al. 2012; Sonne et al. 2015). PBPK models now provide a common framework similar to environmental fate and transport models for more integrative exposure assessment and are applied more regularly to support aggregate (multiroute) exposure assessment (Esch et al. 2011; Abaci and Shuler 2015), exposure reconstruction from biomonitoring data, and exposure translation between in vitro and in vivo test systems.

The use of PBPK models for exposure reconstruction, known as reverse dosimetry (Liao et al. 2007; Tan et al. 2007; Bartels et al. 2012; Hays et al. 2012; McNally et al. 2012; Yang et al. 2012; Grulke et al. 2013), has led to important advances in the field of biomonitoring. Internal and external exposures can now be related and predicted on the basis of more limited sets of exposure information—for example, urine biomonitoring data (spot samples)—by applying principles of pharmacokinetics. The tools are used to calculate or estimate margins of exposure to chemicals on the basis of blood or urine spot samples and can be used to inform regulatory levels. New methods offer the ability to evaluate the influence of behavior and physiological variability on exposure distributions (Shankaran and Teeguarden 2014).

The use of PBPK models to characterize the influence of biochemical and physiological variability, particularly the role of polymorphisms of metabolizing enzymes in estimates of metabolism and variability (Beaudouin et al. 2010; Bois et al. 2010; Snoeys et al. 2016), has grown substantially and will continue to contribute to exposure assessment and risk assessment. Those advances help to predict pharmacokinetics of potentially sensitive populations, such as preterm infants (Claassen et al. 2015) and children (Yoon et al. 2012). Recently, PBPK models have been applied to disentangle the role of physiological changes related to disease states from the effects of a chemical on disease and to examine the role of reverse causation in published epidemiological studies (Verner et al. 2015; Wu et al. 2015). Accordingly, PBPK models have emerged as new exposure tools capable of supporting inference in epidemiological studies.

One of the major developments concerning PBPK models has been their use for translating exposures between test systems and human-exposure scenarios. In particular, the rapidly expanding use of high-throughput in vitro cell and cell-free systems to characterize the bioactivity of chemicals and materials, such as nanomaterials, has led to a need to translate in vitro exposure data into corresponding in vivo exposures in test systems and humans. Various terms have emerged to describe the applications to do so—for example, in vitro–in vivo extrapolation (IVIVE), reverse toxicokinetics (rTK), and reverse dosimetry. Each describes a kinetics-based and partitioning-based approach to translating exposures from one system of interest (in vitro) to another (in vivo animal or human), and all strive for mass balance. The use of PBPK models and similar biokinetic models of in vitro test systems has produced important methods that can apply PBPK-modeling principles to a broad set of test systems (Rostami-Hodjegan 2012; Yeo et al. 2013; Campbell et al. 2014; Teeguarden et al. 2014; Martin et al. 2015), including microphysiological organ systems or human-on-a-chip systems (Esch et al. 2011; Abaci and Shuler 2015). However, without clear understanding of how exposures in the systems are related to in vivo exposures or human occupational or environmental exposures, their utility will remain limited, as has been the case for standard in vitro cell-culture and cell-free systems.

IVIVE models can be used to calculate human internal exposure concentrations of chemicals from data obtained in high-throughput in vitro systems (Kesisoglou et al. 2015). That approach uses hepatocyte cultures and other biotransformation systems to measure metabolic rate constants that are used to estimate human intrinsic clearance by the liver, a dominant route of metabolic and total clearance in humans. Clearance values can be obtained for different life stages or for populations that are resistant or vulnerable because of polymorphisms of metabolic enzymes. Renal clearance, another major elimination pathway, is often estimated by using data on glomerular filtration rates and measures of protein binding in serum (Rule et al. 2004; Rotroff et al. 2010; Tonnelier et al. 2012; Wetmore et al. 2012). Other aspects of kidney function, such as tubular processing, can also influence clearance rates (Weaver et al. 2016) and various biomark-

er concentrations. Metabolism in other tissues, which can be important, is not evaluated, and this is a limitation of the current state of these systems.[6] Combining clearance with computational high-throughput methods for estimating average daily contact and intake rates makes it possible to predict internal concentrations expected in humans. Those concentrations can then be compared with effect levels or no-effect levels from toxicity-testing systems. Addressing some limitations—such as not accounting for metabolism by other tissues, for the potential role of transporters, or for human variability—will be important next steps toward higher confidence in the application of the models. New approaches for better understanding of metabolic and genetic determinants of exposure are detailed in the next section.

Key challenges in interpreting and applying IVIVE data include the quantification of relevant concentrations that correspond to observed in vitro bioactivity from assumed nominal (administered) concentrations (see Box 2-2 and Figure 2-3). A consistent approach for comparing and extrapolating results could be the use of the free (dissolved aqueous) concentration in the test system because this metric can be applied to cell-based or cell-free systems. The limitations complicate chemical comparisons for potency and toxicity and reduce confidence in the application of in vitro bioassay data that are based only on nominal concentrations in risk-based assessments. Models to calculate in vitro concentrations that cannot be readily measured with traditional sample extraction and analytical techniques need to be developed, evaluated, and applied. Passive dosing and sampling techniques might prove useful in addressing the current analytical challenges and associated uncertainties in quantifying exposures in smaller in vitro test systems (Kramer et al. 2010).

[6]The committee notes that over-prediction of serum concentrations of parent chemicals and under-prediction of potentially important metabolites is generally a possible outcome of underrepresenting metabolism.

### New Approaches for Assessing Biochemical and Physiological Determinants of Internal Exposure

Metabolism, cellular transport, and other processes that control elimination and distribution of chemicals in organisms are essential considerations and important challenges in exposure science, data interpretation, and risk assessment. Metabolism is a key determinant of chemical residence time in the body and can lead to more or less production of toxic chemicals; thus, it plays an important role in the extent of exposure and chemical toxicity (Leung et al. 2012). Reliable measures of metabolic rates are essential for understanding and characterizing differences in metabolism among species and between in vitro and in vivo test systems and for understanding the extent of variability and its effect on susceptibility or resistance. Computational approaches (PBPK, rTK, and IVIVE) can be used to translate in vitro metabolic rates into estimates of chemical clearance (Wilk-Zasadna et al. 2015) and to quantify differences among species and systems for exposure assessment.

High-throughput in vitro assays can be used to investigate metabolism; they now cover many enzymes and isoforms involved in chemical metabolism, including the phase I cytochrome P450 enzymes and a variety of phase II enzymes (admescope; Tolonen and Pelkonen 2015). Direct measures of activity obtained from the assays complement genomic approaches for characterizing the influence of polymorphisms on metabolism. New proteomic tools that use chemical probes can also be used to measure metabolic activity of specific enzymes directly

**BOX 2-2** Challenges in Estimating In Vitro Test Concentrations

Evidence is accumulating that the prevailing view that stressor concentrations in the in vitro systems can be considered static and can be represented by nominal media concentrations is in many cases not valid (Gulden and Seibert 2003; Gulden et al. 2006; Teeguarden et al. 2007, 2014; Kramer et al. 2012; Armitage et al. 2014; Groothuis et al. 2015). For example, nanomaterials, an emerging class of poorly studied toxicants, undergo transformations (agglomeration and dissolution) in liquid systems and size-dependent and density-dependent diffusion and sedimentation; each process affects delivery of particles to cells in culture. The processes have been shown repeatedly to affect cellular dose and can be expected to affect relative hazard ranking. Chemical concentrations in an in vitro test system can change as a function of the chemical properties, the test system, and time. Measured and estimated dissolved and cell concentrations can be orders of magnitude different from assumed (nominal) in vitro concentrations for various reasons, including chemical volatilization, differential distribution in the test system (Heringa et al. 2004; Kramer et al. 2012; Armitage et al. 2014), metabolism (Coecke et al. 2006; Groothuis et al. 2015; Wilk-Zasadna et al. 2015), and the reasons noted above.

in tissue and cellular preparations (Cravatt et al. 2008; Sadler and Wright 2015). For example, recent publications (Crowell et al. 2013; Sadler et al. 2016) demonstrate that activity-based probes provide better measures of relative enzyme activity for individual enzymes than measures of transcripts or proteins and thus complement conventional metabolism assays. Other in vitro metabolism test systems, such as ones that use hepatocytes and liver spheroids, and computational models to translate metabolic rates and pathways to in vivo clearance continue to evolve (Fitzgerald et al. 2015; Hutzler et al. 2015; Liu et al. 2015b). Higher-throughput systems for measuring and interpreting metabolic rates in hepatocytes have been successful in extending our knowledge from pharmaceuticals to environmental chemicals (Wetmore et al. 2014; Yoon et al. 2014). However, increasing capacity to synthesize chemical standards and test materials will be essential if these approaches are to be successfully applied to the many chemicals in commerce.

As basic hepatic-metabolism data grow, other limitations of the systems to predict chemical kinetics and internal exposures will become important. Extrahepatic metabolism—such as metabolism in the kidney, gastrointestinal tract, and lung—can be important but is not yet addressed in most extrapolations. Similarly, differences in metabolic competence between the cells used in vitro and the in vivo systems can affect the extent of metabolism, the metabolic pathways activated, and the metabolites produced (see, for example, Kolanczyk et al. 2012). Emerging tools that can evaluate potential metabolite production (Tolonen and Pelkonen 2015; Wilk-Zasadna et al. 2015) and the use of multiple in vitro metabolism systems of varied complexity (Zhang et al. 2012) that include more than one tissue or cell type are possible solutions to the challenges. QSAR models that can predict rates of metabolism and clearance in tissues, such as liver and plasma (Berellini et al. 2012; Hsiao et al. 2013), and in the whole body (Obach et al. 2008; Wishart et al. 2008; Arnot et al. 2014) are also promising approaches for obtaining information on metabolism.

Pharmacogenomic profiling has emerged as a valuable approach for characterizing individual and population variabilities in genes that influence absorption, distribution, metabolism, and elimination (ADME) of drugs and environmental chemicals. Variations in ADME processes are important sources of variability in internal exposure. Recent advances in sequencing technologies (De Wit et al. 2015; Heather and Chain 2015; McGinn et al.

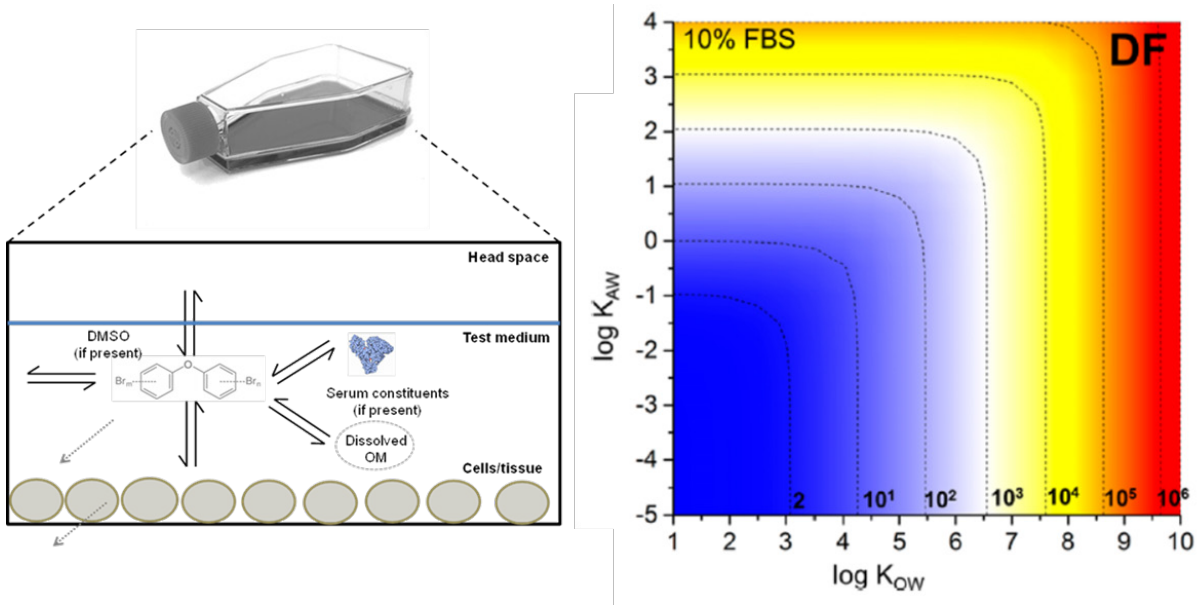



**FIGURE 2-3** (Left) Illustration of chemical distribution in an in vitro test system and (right) illustration of the chemical depletion factor ($DF = C_{nominal}/C_{dissolved}$) in a typical cell-based in vitro test system as a function of chemical partitioning properties. The octanol–water partition coefficient ($K_{OW}$) characterizes chemical partitioning from water to nonaqueous constituents of the test system—such as cell membranes, proteins, plastic, and serum—and the air–water partition coefficient ($K_{AW}$) characterizes chemical partitioning from water into air or head space. In this case, 10% fetal bovine serum (FBS) is assumed present in the test system. The dotted lines (right) are the DFs corresponding to the chemical-property combinations and indicate the order-of-magnitude differences that can occur between assumed (administered or nominal) test concentrations typically used for dose–response calculations and the estimated dissolved (free) concentration in the test system. Source: Armitage et al. 2014.

2016) now offer unprecedented potential for rapid individual and population-level identification of single-nucleotide polymorphisms that affect metabolic, transport, and clearance processes that together influence individual internal-exposure profiles. Recently, the frequencies of polymorphisms in 1,936 proteins that have documented clinical significance for ADME processes were measured and characterized in a Thai population and compared with findings in other ethnicities (Jittikoon et al. 2016). That and other recent analyses that show greater diversity in polymorphisms in American blacks and other ethnicities (Li et al. 2014; Ortega and Meyers 2014) demonstrate the potential for nearly comprehensive assessment of polymorphisms of ADME-related genes in individuals and populations and for internal-exposure predictions on an individual basis. More comprehensive characterization of ADME-related and other polymorphisms in populations and improved understanding of their function and relevance to exposure and toxicity will be valuable in addressing population variability for risk-based decision-making. The committee notes that compartmental and PBPK models for predicting the resulting effects on population distributions of serum concentrations have been used regularly but for only a few metabolic enzymes (EPA 2010).

Another important process to consider is cellular transport; transport proteins influence both tissue and intracellular concentrations. Pharmaceuticals and environmental chemicals are substrates for transporters (Fardell et al. 2011), and the importance of transporters in affecting internal chemical exposure at target sites is recognized (Wambaugh et al. 2014). QSAR models for predicting chemical interactions with transporters (Sedykh et al. 2013) and a variety of in vitro assays (Xie 2008) have been developed to support incorporation of transporters into determinations of internal exposure.

Continued success in using the new tools described here for measuring and calculating biochemical and physiological determinants of internal exposure will improve exposure assessment and ultimately will support the successful integration of in vitro, computational, and in vivo approaches into risk assessment.

## CONFIDENCE LEVELS IN EXPOSURE INFORMATION AND ASSESSMENT

Exposure data from traditional and emerging methods discussed above can be placed in categories spanning the continuum from source to target-site exposure (NRC 2012) (see Figure 2-4). Exposure measures biologically closer to the site of action of the stressor can under some conditions have greater value for linking exposures to effects. For example, the relationship between soil concentrations of a chemical and effects in a population exposed to the soil might be obscured by individual differences in exposure rate, activity patterns, and metabolism. In contrast, individual blood or tissue measures of chemical exposure reflect the combined action of those processes and benefit from being more directly related to the event that initiates adverse effects: interaction of the chemical with a biological receptor (organelle, protein receptor, or DNA). However, soil and air measures of chemicals and biologics can be less confounded sources of information for assessing source contributions to external exposure because fewer processes (absorption, metabolism, and human activity patterns) can obscure relationships between the measured exposure in blood or urine and the source. The committee cautions, however, that internal exposures are not universally better or universally more useful than external exposures for purposes of relating exposures and effects, for example, in epidemiological studies. A long history shows the utility of measures of external exposure for epidemiology. In fact, external exposures might sometimes be superior to internal exposures, for example, when the two are proportional to one another and external measures are easier to acquire. Furthermore, external exposures might be the most biologically relevant when portal-of-entry effects, such as skin sensitization, are the focus. Exposure measures should be carefully selected by considering the strengths and limitations of external and internal measures of exposure and the purpose for which they will be used. Ideally, exposure data are available across the entire spectrum illustrated in Figure 2-4, and approaches for connecting them quantitatively have been developed to enable the use of exposures at any point on the continuum.

There is a spectrum of quality of exposure data from screening-level assessments based on limited information to multiroute, multisource exposure assessments to population-scale longitudinal exposure assessments that use validated exposure biomarkers. Important considerations for the application of exposure data in decision-making are the quality of the data and the context in which the data will be used; data quality can be determined by evaluating accuracy, integrity, suitability, transparency, and concordance of multiple lines of data or evidence (WHO 2016). The degree of confidence that is required for exposure data or exposure assessment is balanced with the cost of data acquisition and determined by the decision context established in problem formulation. In some cases, screening-level exposure data that have greater uncertainty might have sufficient accuracy to support important screening-level decisions made by regulatory agencies and might provide the most cost-effective approach (Wambaugh et al. 2013, 2014; WHO 2016). In those cases, transparency is essential for providing understanding and confidence in decisions that stem from exposure assessment; transparency can be obtained by carefully documenting and reporting data quality, suitability, and integrity (WHO 2016). The use of computationally de-

rived exposure estimates that are based on sparse data is an example of possible applications. That approach might be used to make initial decisions to set priorities among stressors for improved exposure assessment, toxicity assessment, or epidemiological assessment. The same data might also be useful for making initial decisions regarding new applications of a chemical or its inclusion in or removal from new or existing products. In some cases, extensive uncertainty, sensitivity, and variability analyses of exposure-assessment components might indicate that exposures of the magnitude necessary to cause effects fall outside the range of plausibility, in which case such exposure estimates might have sufficient certainty to support decision-making regarding potential risks. As the field moves toward obtaining exposure data on thousands of chemicals in commerce and wider use of cost-effective screening-level analyses, careful reporting of the quality of assessments and associated limitations—for example, through model evaluation and sensitivity analysis—will have high priority. As computational exposure-measurement tools are developed and used, their successful application in risk-based or exposure-based decision-making as described above will involve passing the same quality assessments applied to environmental measures of exposure, for example, by applying EPA or World Health Organization (WHO) guidance to evaluate models (WHO 2005; EPA 2009, 2016a).

Guidance for evaluating exposure data and exposure assessments developed by WHO and EPA and published in the literature focuses more on determining data quality than on establishing confidence in integrating various data streams. For example, integrating emerging data streams (such as computational exposure data) with conventional data (such as those derived from blood and urine biomonitoring and air sampling) is not discussed. Figure 2-5 presents some general considerations for assessing quality of exposure data and for integrating multiple data types. The four attributes for judging the quality of exposure data outlined by WHO—appropriateness, accuracy, integrity and transparency—also apply to Figure 2-5, but there is additional consideration of the strength of agreement between measures and of how each measure is related to the others in the overall exposure narrative. Although computationally derived exposure estimates might be perceived as warranting less confidence than direct measures, consideration of factors related to appropriateness and accuracy might indicate that the computational estimates are of higher quality. For example, direct exposure measures that are made with analytical methods that have not been validated, that are confounded

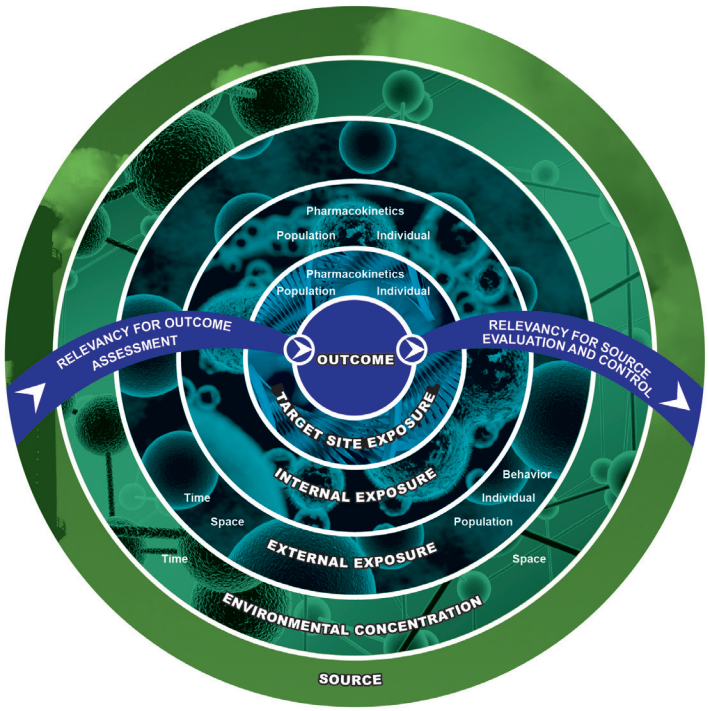



**FIGURE 2-4** Exposure measurements are made along multiple points in the source-to-outcome continuum. The value of exposure data for applications, such as source assessment and mitigation and assessment of public-health effects, might depend on location on the source-to-outcome continuum. Careful consideration should be given to selection of exposure measures by balancing cost, invasiveness, and relevance for the study. For example, although internal exposures might be directly related to the event that initiates adverse effects, external measures of exposure might be more relevant to portal-of-entry effects and have the benefit of being more cost-effective to collect. Source: NRC 2012.

by sample contamination, that are determined without accounting for external-exposure intake rates and half-lives, or that lack temporal resolution necessary for their application in some decision-making contexts might ultimately be less valuable than indirect or proxy measures that are based on a validated exposure metric. Similarly, computationally derived exposure estimates might be useful for some decision-making contexts, particularly when they are based on extensive experimental data—including pharmacokinetics, total external exposure, and patterns of external exposure—and show mass balance throughout the system. Confidence in any exposure assessment is increased when there is concordance, consistency, or agreement between multiple methods of exposure assessment and is greatest when directly measured exposures, indirect measures of exposure, and computationally derived exposure estimates or simulations agree (McKone et al. 2007; Cowan-Ellsberry et al. 2009; Mackay et al. 2011; Ritter et al. 2011; Teeguarden et al. 2013). Agreement between measured and predicted data streams builds confidence in each method of determination. Convergence between exposure measurements (external and internal) and model simulation results (for example, overlap of concentrations or probability distributions of concentrations) indicate higher confidence in an exposure estimate and in resulting risk-based decisions. Although agreement between exposure measures might be a hallmark of quality and of the ideal, multiple concordant measures of exposure are not required to establish levels of quality required for all decision-making contexts.

Consideration of the level of quality and confidence in exposure assessment in the decision-making context will continue to be important, particularly as new exposure data streams emerge from personal sampling data and from use of new exposure matrices, such as bone, teeth, and hair. The potential for using emerging exposure data streams is high, but without careful evaluation, comparison with other types of exposure-assessment data, and a consistent effort to relate measurements to the appropriate level of biological organization (for example, target site or source), confidence in their use or agreement on their best application might be difficult to obtain.

Guidance has been developed to foster confidence, transparency, and reproducibility in calculated data used for exposure and risk assessment. Specific guidance has been developed for QSAR models for predicting chemical properties and toxicity (OECD 2007), for environmental fate and exposure models (EPA 2009; Buser et al. 2012), and for pharmacokinetic models (McLanahan et al. 2012). As new exposure metrics emerge, it will be important to develop guidance for integrating the various exposure measures and to understand their value and relationships with each other.

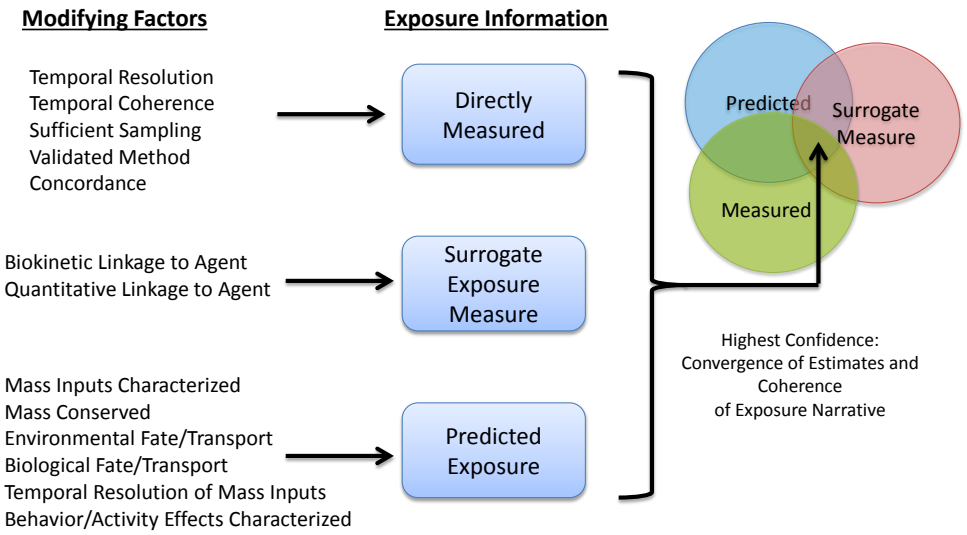



**FIGURE 2-5** Confidence increases with more complete characterization of the exposure pathway and associated exposure determinants. Confidence might be higher for direct measures of the stressor—for example, at the site of action—but if such measures do not consider important modifying factors, confidence might be higher for surrogate exposure measures or predicted exposure measures that do consider such factors. The greatest confidence occurs when there is concordance between multiple exposure-estimation approaches or between multiple exposure measures, especially when divergent exposure metrics are considered. The confidence that is required for exposure data and assessments should be determined by data-acquision costs and the decision context; the highest levels of confidence are not required for many decision contexts.

## APPLICATIONS FOR EXPOSURE SCIENCE

To provide practical guidance on the use of emerging exposure-science data streams for decision-making, the following sections describe applications expected to have near-term and lasting influence on exposure assessment and on risk-based decision-making (see Box 2-3). Each application uses one or more of the advances presented earlier in this chapter to provide a new basis for decision-making, to refine exposure data, or to provide new forms of exposure data.

### Aligning Exposures Between Test Systems and Humans

Comparison of biological responses across diverse experimental systems is nearly always an essential step in risk assessment. For example, risk assessors are faced with aligning toxicity data that are based on disparate measures of exposure: nominal liquid concentrations or cell concentrations in in vitro systems and air concentrations, inhaled amounts, or administered doses in rodent studies and human biomonitoring studies. Specificity, sensitivity, and concordance of observed effects across the test systems underlie the value and strength of evidence supporting conclusions about hazard and risk associated with exposure. To compare the responses from different test systems adequately, the exposures (concentrations) need to be expressed in consistent (comparable) units and with due consideration for the matrix in which the chemical is present. For example, a chemical concentration in whole blood that corresponds to an in vivo response can differ from the total concentration in an in vitro test system that corresponds to a related response, although the free (dissolved) concentrations in the aqueous phases in each system might be equal. Thus, the alignment of exposures in the systems is one important step in comparing exposure–response relationships across systems and evaluating concordance and consistency. As in vitro systems, organotypic, or co-culture systems augment or replace traditional animal studies, biological effects are compared over a more diverse array of assay systems and, from an exposure standpoint, over more types of exposure. For example, the most biologically sound comparison of biological effects shown in a cell-free assay, a cell-based assay, and an inhalation-exposure rodent study would involve comparisons of target-site exposures across all three systems: free-liquid concentrations in the cell-free assay, free cell concentrations in the cell-based assay, and free cell concentrations in the target cells of the rodent. As a practical matter, measured free-liquid concentrations in the in vitro assays and serum concentrations in rodent assays or from human studies would typically be considered appropriate measures of exposure-based alignment of the biological effects. However, there are circumstances in which serum concentrations are not good surrogates for tissue dose—for example, when transport proteins facilitate the uptake to and efflux from the tissue (Koch and Brouwer 2012; Wambaugh et al. 2014). The committee emphasizes that for any metric used to align exposure concentrations between systems, one should consider system conditions that might influence the value or interpretation of the data. For example, is the chemical concentration determined under steady-state or dynamic conditions or is the chemical ionic, in which case pH must be considered?

Each experimental system and human exposure situation has a unique set of processes that control or influence the timing, duration, and extent of exposure at the site of action (see Figure 2-6). Many of the processes are biokinetic and measurable with conventional approaches. Characterizing the processes in each test system allows the measurement, calculation, or simulation of chemical exposure at a common site of action. Consistent metrics of exposure, such as free or cell concentration, represent a possible ideal for comparison across systems and do not have the limitations associated with nominal concentrations. The chemical-activity approach has been proposed for ecological risk assessment (Mackay et al. 2011; Gobas et al. 2015) because it can integrate various multimedia exposure data streams (measured and predicted) and tox-

**BOX 2-3** High-Value Applications for Exposure Sciences

- Aligning exposures between test systems and humans
- Improving exposure assessment for epidemiological studies
- Exposure-based screening and priority-setting
- Identifying new chemical exposures for toxicity testing
- Predicting exposure to support registration and use of new chemicals
- Identifying, evaluating, and mitigating sources of exposure
- Assessing cumulative exposure and exposure to mixtures

icity data streams (in vitro and in vivo) into a framework with consistent units and might be useful for human health evaluations. Other exposure metrics might be suitable for some decision contexts if they are adequately justified on the basis of pharmacokinetics, physical chemistry, and biology of the end point of interest.

Alignment of exposures between systems can be completed under data-poor and data-rich conditions. High-throughput methods for estimating hepatic and renal clearance can provide data needed for estimating human serum concentrations of chemicals that can be compared with cell-culture concentrations. That approach reflects one extreme—the data-poor case—for which data limitations can be overcome by focused, efficient in vitro and computational methods. Recently, an example of alignment of exposures under data-rich conditions—those with data from in vitro assays, whole-animal studies, and human biomonitoring—was published for systemic effects. Human urine and serum time-course concentration data from multiple studies provided empirical pharmacokinetic data that showed a relationship between serum bisphenol A (BPA) concentrations and urine BPA concentrations (Teeguarden et al. 2011, 2015; Thayer et al. 2015). The empirical relationships were used to calculate the range of human serum concentrations expected in a population of more than 28,000 people on whom there were published biomonitoring urine data. The resulting range of serum concentrations was compared directly with liquid concentrations in low-dose BPA cell-culture and aquatic studies (Teeguarden et al. 2013, 2015). Conclusions concerning the probability of biological effects in humans were drawn by aligning exposures across human biomonitoring and two divergent test systems—vertebrates and cell-culture systems—that used a measure of exposure proximal to target-tissue exposure. Although the role of protein binding was not addressed in that example, the data and tools to do so for BPA and other estrogens have been developed for rodent test systems and humans (Plowchalk and Teeguarden 2002; Teeguarden et al. 2005) and in vitro test systems (Teeguarden and Barton 2004).

A separate set of challenges has prevented widespread alignment of particle and nanoparticle exposures between in vitro and in vivo systems. The deposition of particles in the upper and lower airways of rodents and nonhuman primate toxicity-testing systems and of humans is governed by physical processes (gravity, diffusion, and impaction), breathing patterns, airway structure (size, branching pattern, and geometry), and particle characteristics (size, shape, and density). Similar processes affect gravitational and diffusional transport and eventual particle deposition

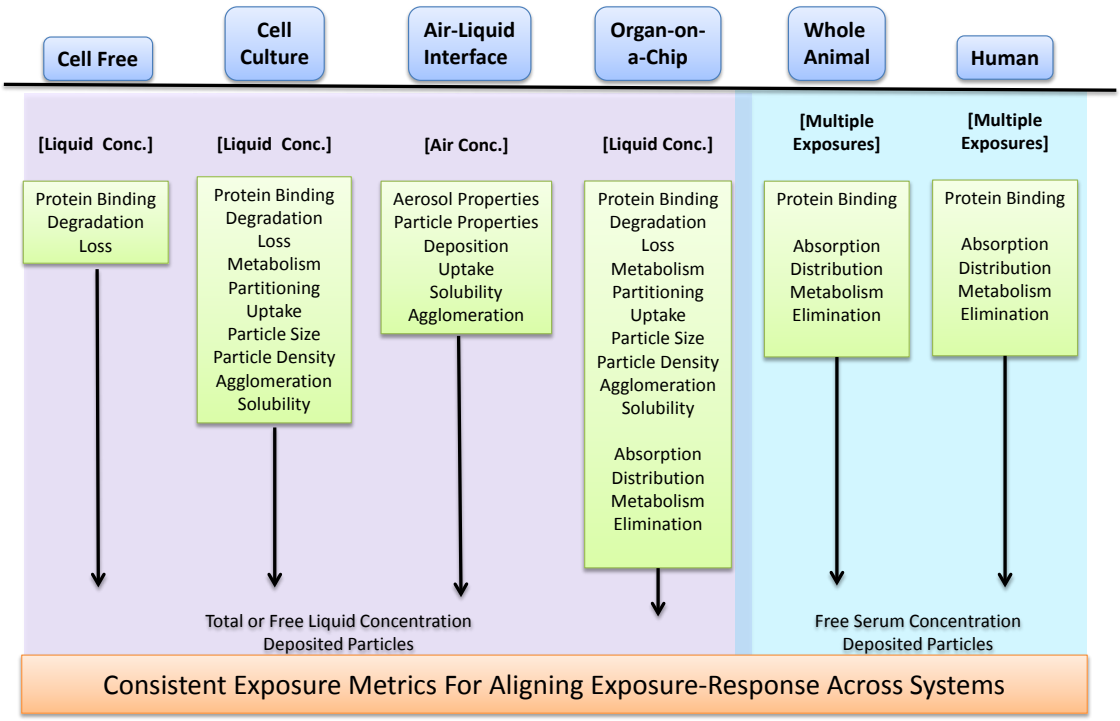



**FIGURE 2-6** Alignment of exposures across experimental toxicity-testing systems can be achieved by understanding, measuring, and applying this information on the processes that control the time course of concentrations and delivery of chemicals and particles to target cells in each system. Common target-cell exposure metrics could be total or free concentrations, peak concentrations, or area under the concentration–time curve.

on target cells in liquid cell-culture systems and include agglomeration capacity; particle size, shape, density, and agglomeration size and density; media height; and diffusion (Teeguarden et al. 2007; Hinderliter et al. 2010; Cohen et al. 2014; DeLoid et al. 2014). Until recently, toxicity data on particles from in vivo and in vitro systems were compared on different exposure scales—for example, air concentrations and liquid cell concentrations (Sayes et al. 2007)—and this potentially obscured relationships between biological effects in the systems. More recently, direct measurement of target-cell doses has become more common. In addition, with the advent of computational tools that can capture the unique kinetics of particles in solution (Hinderliter et al. 2010) and of supportive experimental methods (Davis et al. 2011; Cohen et al. 2014), computational estimation of cellular doses in in vitro systems is becoming more common. With similar tools for measuring or calculating lung-tissue doses of particles after inhalation exposure (Anjilvel and Asgharian 1995; Asgharian and Anjilvel 1998; Asgharian et al. 1999, 2001, 2006, 2012; Asgharian 2004; Asgharian and Price 2007), approaches that allow comparison of in vitro and in vivo models of experimental particle toxicity have emerged (Teeguarden et al. 2014). The consistency of observed effects between the in vitro and in vivo systems might be revealed by making comparisons with a consistent, biologically relevant measure of exposure. For example, iron oxide nanoparticles were shown to cause expression of the same cytokines in macrophages in vitro and in mouse lungs in vivo when exposures were compared on a particle mass or cell basis.

Research in and development of new methods and more frequent application of existing methods to produce consistent measures of biologically appropriate exposure for toxicity across various test and receptor systems is a potentially high-value application for exposure science.

## Improving Exposure Assessment for Epidemiological Studies

Causal inference based on epidemiological evidence can be strengthened when information on health outcomes is combined with clear measures of exposure at the biological site of action or a surrogate for the site of action (such as serum) that is temporally related to the causative biological events. Although that assertion is based on fundamental principles of pharmacology, it is not true that internal exposures are universally better than external exposure for purposes of assessing associations or inferring causation. External-exposure measures have been and will continue to be sufficient, and in some cases superior to internal-exposure measures, for example, where portal-of-entry effects are involved or large population-scale exposure assessments are necessary and internal-exposure assessments are impractical. Reducing or eliminating exposure misclassification and broadening exposure assessment to identify new chemicals that might be causative agents or confounders of existing associations would substantially strengthen the interpretation of epidemiological studies and improve their value for public-health decision-making.

Several advances in the field of exposure science are particularly well suited for improving exposure assessment for epidemiological studies. High-throughput targeted and nontargeted analytical-chemistry tools and new matrices for exposure assessment (such as hair, teeth, and nails) are together expected to offer more temporally relevant exposure assessment of many more chemicals and expand exposure assessment over the full life span. Emerging high-throughput computational-exposure models of external exposure will provide exposure estimates that complement those made through expanded biomonitoring programs. Personal biomonitors and sensor wristbands (O'Connell et al. 2014a,b) offer an unparalleled opportunity to characterize individual exposures and provide temporally and spatially resolved data for understanding patterns of exposure, variability, and the role of behavior and activity levels on exposure. PBPK models could improve exposure assessment by

- Reconstructing exposures from limited biomonitoring samples on the basis of pharmacokinetic understanding (Tan et al. 2006, 2012; Yang et al. 2012).
- Translating external exposures or biomonitoring data into more biologically relevant internal exposures (Teeguarden et al. 2013).
- Reducing the likelihood of reverse causation in epidemiological studies by more clearly delineating the sequences of chemical-induced physiological changes that lead to disease states (Verner et al. 2015; Wu et al. 2015).
- Accounting for population variability that is characterized directly or through the application of pharmacogenomics approaches (Teeguarden et al. 2008; EPA 2010; Ginsberg et al. 2010).

The greater availability of internal-exposure information obtained from direct biomonitoring of human populations or from a combination of computational tools would be of particular value by providing human exposure concentrations at the site of action (tissue or blood). Such information could be compared with measurements in animal and cell-culture studies and might enhance causal inferences derived from epidemiological studies.

## Exposure-Based Screening and Priority-Setting

Several exposure-based priority-setting approaches that benefit from the emerging exposure-science tools and data streams have been developed. In an exposure-based

approach, chemicals in the top exposure category are assigned a higher priority for additional tiered toxicological, hazard, or risk assessment than those in the low exposure category; this provides a reproducible, transparent, and knowledge-based framework to inform decisions for testing priorities (Egeghy et al. 2011; Wambaugh et al. 2013, 2014). The European Food Safety Authority and WHO have reviewed the threshold-of-toxicological-concern (TTC) approach as a screening and priority-setting tool that can be used for chemical assessments in cases where hazard data are insufficient and human exposure can be estimated (EFSA 2016). The TTC approach is used principally as a screening tool to assess low-dose chemical exposures and to identify those on which further data are necessary for assessing human health risk.[7] In some cases following certain requirements, "exposure-based waiving" for toxicity testing or "exposure-based adaptation of information requirements" approaches can be considered under the European Registration, Evaluation, Authorisation and Restriction of Chemicals legislation (Vermeire et al. 2010; Rowbotham and Gibson 2011). Exposure-based waiving has also been used to propose acceptable exposure levels determined on the basis of generalized chemical-toxicity data and without chemical-specific toxicity data. Such approaches might be useful in making initial decisions about the public-health importance of chemical exposures in lieu of complete exposure and hazard data. Within the bounds of uncertainty and variability of the data, some immediate decisions could be made about the low potential for risk posed by exposures below preselected "critical levels" (Vermeire et al. 2010; Rowbotham and Gibson 2011). Cumulative exposures to chemicals in specific classes might move some chemicals up in priority—an outcome of improved exposure data. Structure-based alerts and TTCs can be applied in such screening contexts to complement the exposure-based decision-making process. EPA recently demonstrated integration of nontargeted and targeted chemical analysis of house-dust samples for exposure-based and bioactivity-based ranking of chemicals for further biomonitoring or toxicity testing as shown in Figure 2-7 (Rager et al. 2016).

[7]The committee notes that TTC approach depends on the set of chemicals used to establish the toxicity distribution that is used to derive the TTC value. The ability of the TTC approach to screen chemicals properly will depend on whether the toxicities of the chemicals of interest are well represented by the toxicities of the chemicals used to establish the distribution.

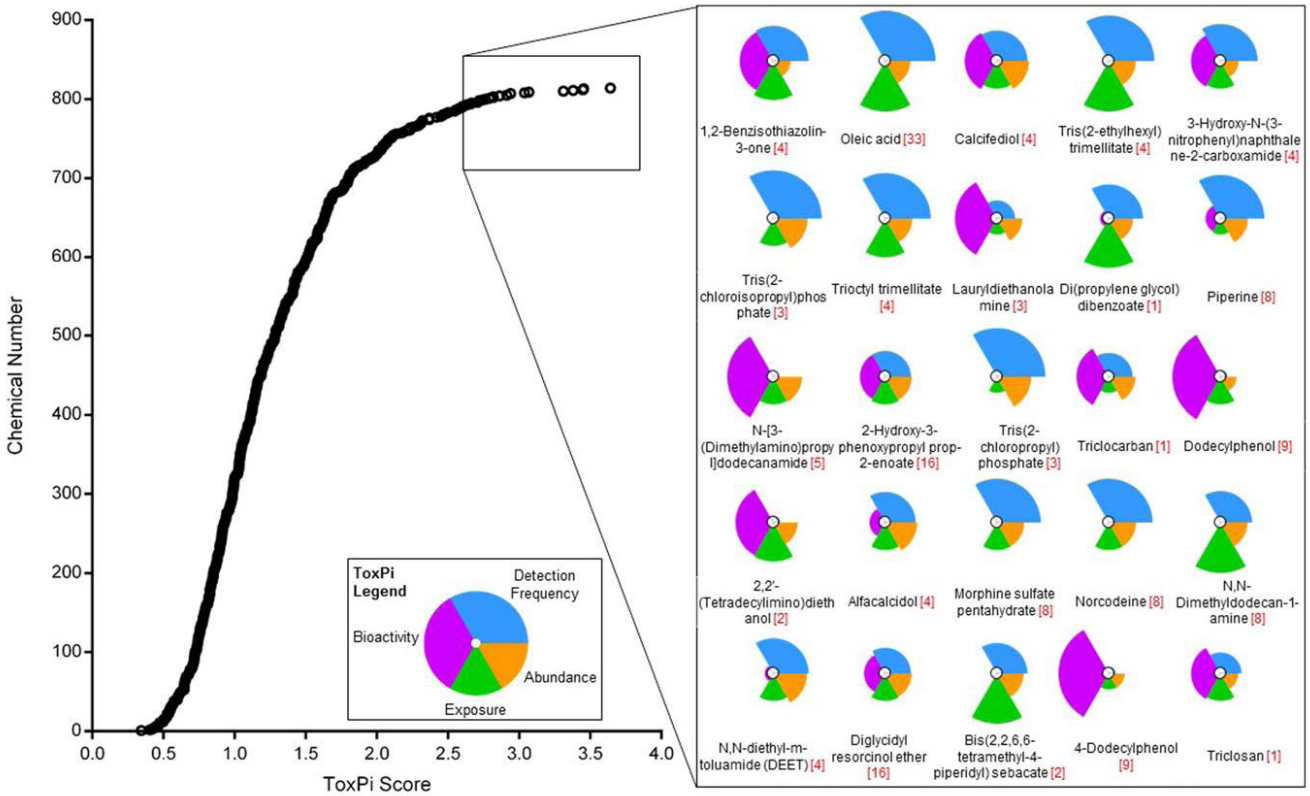



**FIGURE 2-7** Data from nontargeted and targeted analysis of dust samples were used with toxicity data to rank chemicals for further analysis and testing. Source: Rager et al. 2016. Reprinted with permission; copyright 2016, *Environment International*.

Biomonitoring data and environmental-monitoring data on most chemicals in commerce are missing or insufficient for exposure-based decision-making. Application of advanced biomonitoring, personal monitoring, and computational exposure-science tools described in this chapter can support high-throughput screening-level exposure assessment and exposure-based priority-setting for later toxicity testing. Exposure models can be applied to screen large numbers of chemicals in commerce and set priorities among specific chemicals or chemical classes on which there are no or few toxicity-testing data (McLachlan et al. 2014). Chemicals that have predicted high concentrations in humans and environmental media can then be used to identify toxicity-data gaps and set priorities for toxicity-testing for risk-based applications. The committee notes that priority-setting based only on exposure might assign a lower priority to chemicals that might be given a higher priority on the basis of toxicity or risk.

Translation of high-throughput data into risk-based rankings is an important application of exposure data for chemical priority-setting. Recent advances in high-throughput toxicity assessment, notably the ToxCast and Tox21 programs (see Chapter 1), and in high-throughput computational exposure assessment (Wambaugh et al. 2013, 2014) have enabled first-tier risk-based rankings of chemicals on the basis of margins of exposure—the ratio of exposures that cause effects (or bioactivity) to measured or estimated human exposures (Wambaugh et al. 2013, 2014; Wetmore et al. 2013, 2014; Shin et al. 2015). Building on work by Wetmore et al. (2012) and Rotroff et al. (2010), Shin et al. (2015) demonstrated a high-throughput method for screening and setting priorities among chemicals on the basis of quantitative comparisons of exposure data with in vitro bioactivity data (bioactivity quotients); this is similar to the margin-of-exposure approach used in risk priority-setting. They used human intake rates estimated with computational exposure models and toxicokinetic models for the in vitro–in vivo extrapolation of ToxCast toxicity data and identified 38 of 180 chemicals for which total estimated exposures equaled or exceeded the estimated oral dose expected to result in blood concentrations that cause a 50% response in an in vitro toxicity-testing system. Population variability due to differences in metabolic capacity was incorporated into the process (Wetmore et al. 2014). Screening-level exposure assessment was used to establish margins of exposure for that group of chemicals for purposes of priority-setting. The committee notes, however, that limitations of such analyses (see section "New Approaches for Assessing Biochemical and Physiological Determinants of Internal Exposure" above) need to be taken into account. Although exposure estimates that exceed in vitro effect estimates might not be conclusive evidence of risk and exposures that fall below in vitro activities might not be conclusive evidence of no risk, the committee sees the potential for the application of computational exposure science to be highly valuable and credible for comparison and priority-setting among chemicals in a risk-based context.

Human-exposure data on a much larger suite of chemicals than is now available would provide important new data for guiding selection of chemicals and exposure concentrations for hazard testing and mechanistic toxicology. The rapid expansion and use of high-throughput in vitro methods for hazard assessment and mechanistic studies presents a growing opportunity to test chemicals for bioactivity at human-exposure levels—levels lower than those typically used in traditional toxicity-testing studies. In vitro test systems—which are less subject to statistical-power limitations, are less expensive, and have fewer ethical considerations than whole-animal studies—might be better suited for testing exposures lower than those in traditional animal studies. Recent animal studies, however, provide useful examples of applying human exposure information to in vivo test systems. For example, recent studies have included exposures at or near those experienced by humans in animal-testing protocols for genistein and synthetic estrogens (NTP 2008; Delclos et al. 2009, 2014; Rebuli et al. 2014; Hicks et al. 2016). For those animal studies, exposures were selected on the basis of measured serum concentrations obtained in pilot animal studies, values estimated with pharmacokinetic models, and measured or estimated serum concentrations in humans. The use of target-tissue exposures or biologically relevant accessible proxies, such as serum, for selecting can in some cases be of greater relevance than the use of external exposure measures. Thus, there is an opportunity to apply many of the new tools described in this chapter—expanded biomonitoring, new biological matrices, and high-throughput computational exposure models—as a guide for the selection of exposures for use in toxicity testing (Gilbert et al. 2015).

## Identifying New Chemical Exposures for Toxicity Testing

The totality of exposure that makes up the exposome includes registered chemicals that are used in commerce, their environmental and metabolic degradation products, and endogenously produced chemicals. Traditionally, hazard-testing paradigms focus on satisfying regulatory needs for supporting product registration and contain guidelines for testing commercial chemicals, not their degradation products, metabolites, or similar chemicals produced endogenously. Identification of chemicals that make up the latter groups of untested chemicals has become a key goal of federally funded exposure-science programs, such as the Children's Health Exposure Analysis Resource. Owing to advances in high-throughput nontargeted analysis (Fiehn 2002; Park et al. 2012; Go et al. 2015; Mastrangelo et al. 2015; Sud et al. 2016), exposure science is in a more

effective position for discovery-based exposure assessment. Combined with environmental-degradation studies to identify novel chemicals, higher-throughput targeted analytical methods also contribute to overall exposure discovery for toxicity testing. For example, researchers in the Oregon State University Superfund Research Program recently discovered novel oxygenated and nitrogenated polycyclic aromatic hydrocarbons produced by conventional remediation methods and have subjected these environmental degradation products to toxicity testing (Knecht et al. 2013; Chibwe et al. 2015; Motorykin et al. 2015). In collaboration with academic scientists, EPA (Rager et al. 2016) recently demonstrated a workflow for nontargeted analysis of house dust with a transition to targeted analysis (measurement of specific target analytes) for ToxCast chemicals and use of frequency of detection information on chemicals as exposure data for priority-setting shown in Figure 2-8. The committee sees the use of nontargeted and targeted analysis as one innovative approach for identifying and setting priorities among chemicals for additional exposure assessment, hazard testing, and risk assessment that complements the current hazard-oriented paradigm.

### Predicting Exposure to Support Registration and Use of New Chemicals

About 1,000–2,000 chemicals are introduced into commerce each year (EPA 2004). For newly introduced chemicals, exposure assessment means forecasting likely environmental concentrations or total daily human exposures resulting from expected uses and is not a regular part of the decision-making process. The case of methyl tertiary-butyl ether, a gas additive introduced without fate and transport calculations and later found to be widely distributed in the environment, is a poignant example of the value of predictive exposure modeling (Davis and Farland 2001). A recent NRC report, *A Framework to Guide Selection of Chemical Alternatives*, found that despite the known importance of exposure, many frameworks for selecting chemical alternatives downplay its importance and focus on inherent hazards posed by chemicals (NRC 2014). The committee that prepared the report recommended an increased emphasis on comparative exposure assessment and stated that inherent hazard should be the focus only in cases where the exposure routes and concentrations of the chemical of concern and its alternatives

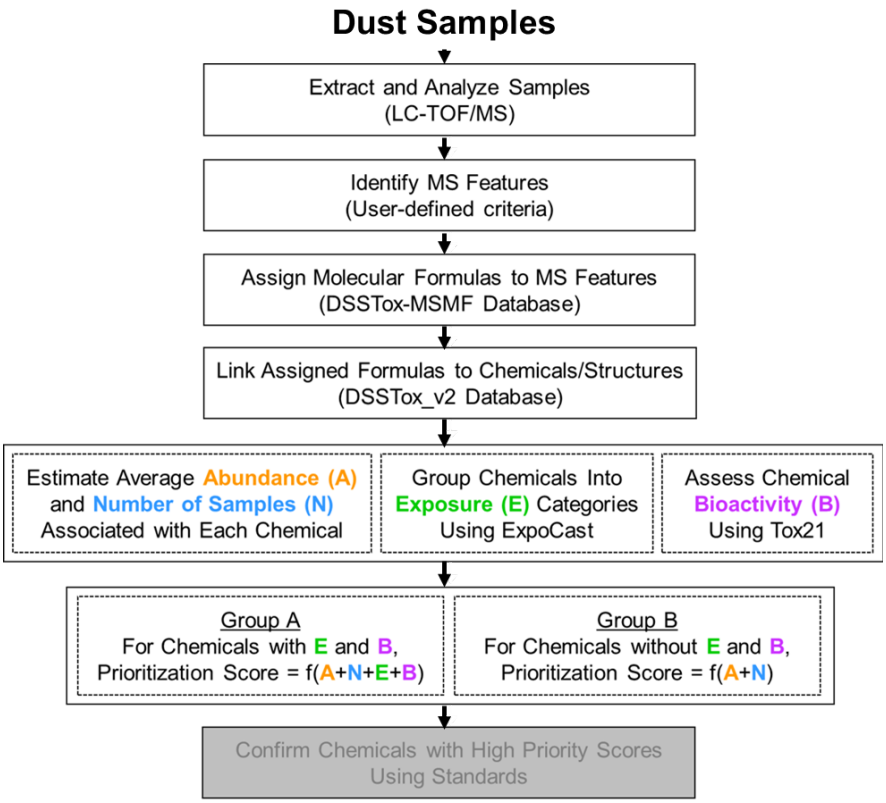



**FIGURE 2-8** Workflow for nontargeted and targeted analysis of the house-dust exposome for chemical priority-setting and testing. Abbreviations: DSSTox-MSMF, Distributed Structure-Searchable Toxicity Database-Mass Spectroscopy Molecular Formula; LC-TOF/MS, liquid chromatography time-of-flight mass spectroscopy; and MS, mass spectrometry. Source: Rager et al. 2016. Reprinted with permission; copyright 2016, *Environment International.*

are not expected to differ substantially; that is, equivalent exposures should not be automatically assumed. And, it recommended greater reliance on physicochemical data and modeling tools, when high-quality analytical data on exposure are unavailable, to aid in predicting the partitioning of contaminants in the environment and the potential for their persistence, bioaccumulation, and toxicity. Although approaches that are based on both hazard and exposure data are preferred, approaches that are based principally on exposure or hazard data will continue to be valuable depending on the decision context.

Tools to predict chemical properties (environmental or tissue-partitioning properties), stability (degradation and metabolism half-lives), and proposed use scenarios can be used to set parameter values for exposure models that are used to predict concentrations in environmental media and humans, over life spans, and on local and national scales. The estimated concentrations can guide selection of toxicity-testing exposures and can be compared with emerging toxicity data for risk-based assessments. Green-chemistry modeling initiatives can be applied to prescreen candidate chemicals according to the likelihood of biodegradation (Boethling 2011). Candidate chemicals can also be screened by applying more comprehensive methods that consider environmental fate and transport and various chemical use scenarios (release pattern and quantities) (see, for example, Gama et al. 2012). Confidence in the prescreening methods will be greatest when the models and tools cover the applicability domain of the chemicals that are being evaluated and when the tools have already been shown to be effective in predicting fate and transport of chemicals that have similar properties (for example, structural similarity or similar use categories). Hence there is a need to test and evaluate exposure modeling tools and data streams systematically with existing commercial chemicals to foster confidence in applying the same and emerging tools for new premarket chemicals.

## Identifying, Evaluating, and Mitigating Sources of Exposure

For chemicals that have multiple relevant exposure pathways, it can be challenging to identify and rank exposure sources for mitigation. Exposure models can be used to reconstruct and identify the sources, behaviors, and pathways that are driving exposures to a particular stressor. Good examples of emerging computational exposure tools that can be used to trace exposures to sources are exposure models for consumer products (Gosens et al. 2014; Delmaar et al. 2015; Dudzina et al. 2015) and exposure models and frameworks that combine far-field and near-field pathways for aggregate human exposure assessments (Isaacs et al. 2014; Shin et al. 2015). For example, Shin et al. (2014) combined exposure models and human-biomonitoring data for nine chemicals to estimate the proportions of total production volumes that are used in selected use categories that correspond to exposure pathways. The models can be used to develop targeted strategies to reduce or virtually eliminate exposures to a particular stressor. For some chemicals, such as those used in pharmaceuticals and personal-care products, the dominant exposure pathways and chemical use rates are relatively obvious, and source mitigation, if necessary, might be relatively straightforward.

The combination of sensor technologies, including personal sensors, with GIS data systems offers new capabilities to identify sources of exposure. Personal sensors—for example, cell-phone–based sulfur oxide and nitrogen oxide sensors—use native GIS systems to collect real-time exposure data, which can be used to identify locations with high exposures and the source locations that contribute to the exposures. Remote sensing can identify high-exposure locations and source locations on a regional or population scale by mapping pollutant concentrations and identifying exposure patterns that might be related to sources.

Some chemicals and materials are poorly degraded and persist in the environment long after production and use are stopped. Some of the highly persistent chemicals also have long residence times in the human body. It can take years or decades for exposures to decline substantially after regulatory action is initiated. Accordingly, highly persistent chemicals that show unacceptable risk should have high priority for mitigation. Models and supporting experimental studies that screen for rates of chemical degradation in environmental media and overall persistence in the environment and in humans can be used to identify persistent chemicals before commercial use and prevent or mitigate potential exposure by finding alternatives.

Emerging exposure-assessment tools can also be used to mitigate sources of exposure to chemicals that cannot be identified confidently. Specifically, nontargeted analysis of environmental samples—air, dust, water, and soil—can be combined with analysis of ecological or human biomonitoring samples to select analytical features that represent internal exposures of potential concern. Geographical mapping of relative concentrations or detection frequency in environmental and human samples can lead to source identification that might in turn help to identify the chemical classes.

## Assessing Cumulative Exposure and Exposure to Mixtures

Humans, animals, plants, and other organisms are exposed to numerous stressors that vary in composition and concentration over space and time. For the most part, traditional toxicity testing has been conducted largely on single chemicals, so there are important uncertainties in assessing potential short-term and long-term effects of exposures to a mixture. That issue is a well-recognized

concern for chemical assessment. With advances in exposure data streams and the potential for high-throughput toxicity screening, there are opportunities to address the uncertainty related to potential effects of mixture exposures better. Measurements obtained from human tissue and from environmental media to which humans are exposed can be used directly or indirectly to formulate environmentally relevant concentrations of mixtures for toxicity screening and testing. For example, internal concentrations of persistent organic pollutants from in vivo exposure of humans (silicone implants) were used to determine and test mixture toxicity in in vitro assays (Gilbert et al. 2015). It is also possible to use environmental-monitoring data (sampled water concentrations) to formulate exposure mixtures for toxicity testing (Allan et al. 2012), including approaches that consider population variability in responses to environmentally relevant chemical-mixture concentrations (Abdo et al. 2015). The substantial advances in analytical chemistry noted in this report are producing more complete data on the extent of cumulative exposure to chemicals. Personal sampling devices, such as wristbands and air-sampling devices, provide data on complex cumulative exposures of individuals. -Omics tools appropriate for measuring the aggregate biological response to cumulative exposures to chemical classes that act through similar mechanisms can be combined with measures of real-world cumulative exposures to assess the effects of cumulative exposures more comprehensively. Aggregate-exposure model calculations for individual chemicals could be combined to obtain estimates of cumulative exposures to mixtures, for example, by using models of exposure to consumer products that are supported by databases of chemical concentrations in the product and product-use rates. The exposure-model calculations could be used to address mixture exposures and potential toxicity; this approach would require mixture-toxicity data or mixture-toxicity models for risk-based assessment. For that case, estimating exposure to a mixture of chemical stressors for risk-based assessments is theoretically possible. The reliability of and confidence in the exposure calculations require further evaluation, and methods for including metabolites and nonchemical stressors in cumulative risk-based evaluations are also required.

## CHALLENGES AND RECOMMENDATIONS FOR ADVANCING EXPOSURE SCIENCE

A principal objective of improving exposure science is to build confidence in exposure estimates by addressing or reducing uncertainty in the estimates used to support risk-based decision-making. That objective is best met by developing and further integrating monitoring, measurement, and modeling efforts and by harmonizing exposures among test systems, the multimedia environment, and humans. Incrementally increasing the number of chemicals included in monitoring programs can help in evaluating and refining exposure models and in developing new approaches to integrate exposure data and constitutes an initial and pragmatic path. However, increased environmental monitoring alone will not be sufficient to improve exposure science. Interpreting the monitoring data and appropriately applying exposure data in risk-based evaluations will require continued complementary development and evaluation of exposure-assessment tools and information, such as fate and transport models, PBPK models, and data on chemical quantity and use, partitioning properties, reaction rates, and human behavior.

In this section, challenges and recommendations to advance exposure science are discussed further. The points include some guidance initially presented in the ES21 report and some new, more pragmatic points, specifically related to the application of exposure science to risk-based evaluations. The points build on the advances and applications detailed in this chapter, which present key development opportunities for the field recommended by the committee. Generally, the recommendations and challenges cover a continuum: preparation of infrastructure, collection of data, alignment of exposures between systems, integration of exposure data, and use of data for priority-setting. The ES21 Federal Working Group (EPA 2016b) is particularly well-positioned to coordinate and support the recommendations outlined below by further strengthening federal partnerships for the efficient development of exposure-science research and by engaging with other stakeholders to address the challenges that face the development and application of exposure information for risk-based evaluations. The committee notes that several recommendations below call for developing or expanding databases. In all cases, data curation and quality evaluation should be a routine part of database development and maintenance.

### Expand and Coordinate Exposure Science Infrastructure to Support Decision-Making

*Challenge*: A broad spectrum of disciplines and institutions are participating in advancing exposure methods, measurements, and models. Given the many participants in exposure science, most information is fragmented, incompletely organized, and not readily available or accessible in some cases. Thus, the full potential of the existing and emerging information for exposure-based and risk-based evaluations cannot be realized. The committee emphasizes that the rapid growth in exposure science presents unprecedented opportunities for more efficient, complete, and holistic use of exposure information, especially if the information can be well organized into a readily accessible format.

*Recommendation*: An infrastructure for exposure information should be developed to organize and coordinate better the existing and rapidly evolving components of exposure science and ultimately to improve exposure assessment. The infrastructure should be organized by using conceptual and systems-based frameworks that are commonly used in exposure assessment and should facilitate the generation, acquisition, organization, access, evaluation, integration, and transparent application and communication of exposure information. The infrastructure might best be comprised of an Internet-based network of databases and tools rather than one database and could expand on existing infrastructure and databases. Guidance for generating, evaluating, and applying exposure information (WHO 2005; EPA 2009) should be expanded to enable inclusion of data in the databases.

*Recommendation*: Coordination and cooperation should be encouraged among the large network of agencies, institutions, and organizations that produce and use exposure information for different but ultimately connected and complementary objectives. Cooperation should increase the efficiency with which the infrastructure described above is developed, and a common ontology of exposure science (Zartarian et al. 2005; Mattingly et al. 2012; EPA 2016b) should continue to evolve to facilitate interdisciplinary communication in the development and application of exposure information.

### Identify Chemicals or Other Stressors and Quantify Sources and Exposures

*Challenge*: Nontargeted analysis in environmental and human media indicates that there are many unknown chemicals in complex uncharacterized mixtures to which humans are exposed. Analytical methods and standards are not available for most chemicals and degradation products, and this hinders the capacity to identify and quantify chemical exposures. Furthermore, uncertainty in source information—product composition, chemical quantity, use, and release rate—is a major obstacle to exposure estimation for most chemicals.

*Recommendation*: Current efforts to obtain and organize information on chemical quantities in and rates of release from products and materials, particularly consumer products and materials in the indoor environment, should be expanded substantially. Curated databases that contain analytical features that can be used in chemical identification should be expanded, and increasing the availability of analytical standards for chemicals and their degradation products should have high priority. Ultimately, the capacity to conduct targeted and nontargeted analyses to identify and quantify new and existing chemicals and mixtures in environmental media and humans should be increased.

### Improve Knowledge of Processes That Determine Chemical Fate in Systems

*Challenge*: Understanding the influence of processes that control the fate, transport, and ultimately concentration of chemicals in environmental compartments and in animal and cell-based test systems is essential for characterizing and predicting exposures. Information on system properties, processes, and transformation pathways that contribute to chemical exposure is nonexistent, incomplete, and inconsistent, and this limits the capacity for more comprehensive, quantitative exposure-based and risk-based evaluations.

*Recommendation*: Databases of chemical properties and information on rates and processes that control chemical fate in vitro, in vivo, and in environmental systems should be developed. Information is needed, for example, on partitioning (distribution) coefficients, degradation and transfer rates, and metabolic and environmental transformation pathways. Information might be obtained through experiments or modeling.

*Recommendation*: Methods for measuring and predicting chemical transformation pathways and rates in environmental media, biological media, and biological test systems should be developed and applied. The methods should be used to quantify human exposures to chemical mixtures (parent chemicals and metabolites) over time and to interpret results from test systems in the context of actual human exposures. In particular, knowledge of environmental, human, and test-system properties and conditions that influence exposures should be improved. Human pharmacokinetic data on metabolism, chemical transporters, and protein binding should be generated for chemicals in consumer products and food-related applications to improve the interpretation of human biomonitoring data from urine, blood, and emerging matrices.

### Align Environmental and Test-System Exposures

*Challenge*: Aligning environmental exposures and information obtained from experimental systems is a critical aspect of risk-based evaluation and is required for improving environmental epidemiology. Various units of quantification, such as administered or unmeasured dose, are often applied, and assumptions, such as steady-state or equilibrium conditions, are made. However, pharmacokinetic and fate processes and other factors often confound the interpretation and translation of exposure information between humans and the environment and experimental systems.

*Recommendation*: Concentrations in the test-system components should be quantified over time by measurement, which is preferred, or with reliable estimation methods. Methods and models that explicitly translate quantitative information between actual exposures and test-system exposures should be developed and evaluated.

*Recommendation*: Chemical concentrations that reflect human exposures as derived from biomonitoring measurements or from predictive exposure models should be considered when designing testing protocols for biological assays. Improving knowledge of processes that determine chemical fate in biological and test systems will be necessary to meet this recommendation.

### Integrate Exposure Information

*Challenge*: Integration and appropriate application of exposure data from environmental media, biomonitoring samples, conventional samples (blood and urine), and emerging matrices (hair, nails, teeth, and meconium) is a scientific, engineering, and big-data challenge. The committee emphasizes that integration of measured and modeled data is a key step in developing coherent exposure narratives, in evaluating data concordance, and ultimately in determining confidence in an exposure assessment.

*Recommendation*: New interdisciplinary projects should be initiated to integrate exposure data and to gain experience that can be used to guide data collection and integration of conventional and emerging data streams. The projects might start as an extension of existing cooperative projects among federal and state agencies, nongovernment organizations, academe, and industry that focus on integrating measurements and models for improved quantitative exposure assessment. High priority should be placed on extending existing (EPA, Centers for Disease Control and Prevention, and WHO) guidance on quality of individual exposure data and assessments to include weighing and evaluating the quality of integrated experimental and modeled information from multiple matrices and data streams.

### Determine Exposure-Assessment Priorities

*Challenge*: All the many uses of exposure data—from selection of chemicals for use in new products to risk-based decision-making to exposure ranking—require exposure data, often for thousands of chemicals, over time and space. Whether or not analytical methods are available for the chemicals, the resources and time that are required for direct measures of exposure are not available, and resource-intensive, high-confidence exposure measurements might not be necessary in some cases. A key challenge for exposure science is how best to focus resources on the highest-priority chemicals, chemical classes, mixtures, and exposure scenarios.

*Recommendation*: Continued development of computational and experimental tools that maximize the value of existing knowledge for estimating exposure should have high priority. Those approaches might initially focus on selected near-field exposures that are known to be important, on chemical classes that are of high interest because of data on biological effects, or on other objectives, such as exposure ranking of members of a chemical class that are being investigated for use in new products.

*Recommendation*: Continued development of approaches for exposure-based priority-setting that use uncertainty analysis to establish and communicate levels of confidence to support decision-making should be encouraged. The need to improve models or data that are used for priority-setting should be evaluated on the basis of the level of uncertainty and the tolerance for uncertainty in the decision-making context. Uncertainty and sensitivity analyses should guide selection and priority-setting among data gaps to be filled.

## REFERENCES

Abaci, H.E., and M.L. Shuler. 2015. Human-on-a-chip design strategies and principles for physiologically based pharmacokinetics/pharmacodynamics modeling. Integr. Biol. (Camb) 7(4):383-391.

Abdo, N., B.A. Wetmore, G.A. Chappell, D. Shea, F.A. Wright, and I. Rusyn. 2015. In vitro screening for population variability in toxicity of pesticide-containing mixtures. Environ. Int. 85:147-155.

Adams, C., P. Riggs, and J. Volckens. 2009. Development of a method for personal, spatiotemporal exposure assessment. J.Environ. Monit. 11(7):1331-1339.

Al-Hamdan, M.Z., W.L. Crosson, S.A. Economou, M.G. Estes, Jr., S.M. Estes, S.N. Hemmings, S.T. Kent, M. Puckett, D.A. Quattrochi, D.L. Rickman, G.M. Wade, and L.A. McClure. 2014. Environmental public health applications using remotely sensed data. Geocarto. Int. 29(1):85-98.

Aleksa, K., J. Liesivuori, and G. Koren. 2012. Hair as a biomarker of polybrominated diethyl ethers' exposure in infants, children and adults. Toxicol. Lett. 210(2):198-202.

Allan, I.J., K. Baek, A. Kringstad, H.E. Roald, and K.V. Thomas. 2013a. Should silicone prostheses be considered for specimen banking? A pilot study into their use for human biomonitoring. Environ. Int. 59:462-468.

Allan, I.J., K. Baek, T.O. Haugen, K.L. Hawley, A.S. Hogfeldt, and A.D. Lillicrap. 2013b. In vivo passive sampling of nonpolar contaminants in brown trout (*Salmo trutta*). Environ. Sci. Technol. 47(20):11660-11667.

Allan, S.E., B.W. Smith, R.L. Tanguay, and K.A. Anderson. 2012. Bridging environmental mixtures and toxic effects. Environ. Toxicol. Chem. 31(12):2877-2887.

Andersen, Z.J., A. de Nazelle, M.A. Mendez, J. Garcia-Aymerich, O. Hertel, A. Tjonneland, K. Overvad, O. Raaschou-Nielsen, and M.J. Nieuwenhuijsen. 2015. A study of the combined effects of physical activity and air pollution on mortality in elderly urban residents: The Danish Diet, Cancer, and Health Cohort. Environ. Health Perspect. 123(6):557-563.

Andra, S.S., C. Austin, R.O. Wright, and M. Arora. 2015. Reconstructing pre-natal and early childhood exposure to multi-class organic chemicals using teeth: Towards a retrospective temporal exposome. Environ. Int. 83:137-145.

Anjilvel, S., and B. Asgharian. 1995. A multiple-path model of particle deposition in the rat lung. Fundam. Appl. Toxicol. 28(1):41-50.

Armitage, J.M.., F. Wania, and J.A. Arnot. 2014. Application of mass balance models and the chemical activity concept to facilitate the use of in vitro toxicity data for risk assessment. Environ. Sci. Technol. 48(16):9770-9779.

Arnold, S.F., and G. Ramachandran. 2014. Influence of parameter values and variances and algorithm architecture in ConsExpo model on modeled exposures. J. Occup. Environ. Hyg. 11(1):54-66.

Arnot, J.A., T.N. Brown, F. Wania, K. Breivik, and M.S. McLachlan. 2012. Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. Environ. Health Perspect. 120(11):1565-1570.

Arnot, J.A., T.N. Brown, and F. Wania. 2014. Estimating screening-level organic chemical half-lives in humans. Environ. Sci. Technol. 48(1):723-730.

Arora, M., A. Bradman, C. Austin, M. Vedar, N. Holland, B. Eskenazi, and D.R. Smith. 2012. Determining fetal manganese exposure from mantle dentine of deciduous teeth. Environ. Sci. Technol. 46(9):5118-5125.

Asgharian, B. 2004. A model of deposition of hygroscopic particles in the human lung. Aerosol Sci. Technol.38(9):938-947.

Asgharian, B., and S. Anjilvel. 1998. A multiple-path model of fiber deposition in the rat lung. Toxicol. Sci. 44(1):80-86.

Asgharian, B., and O. T. Price. 2007. Deposition of ultrafine (nano) particles in the human lung. Inhal. Toxicol. 19(13):1045-1054.

Asgharian, B., F.J. Miller, and R. P. Subramaniam. 1999. Dosimetry software to predict particle deposition in humans and rats. CIIT Activities 19(3):1-6.

Asgharian, B., W. Hofman, and R. Bergmann. 2001. Particle deposition in a multiple-path model of the human lung. Aerosol Sci. Technol. 34(4):332-339.

Asgharian, B., O. Price, and G. Oberdorster. 2006. A modeling study of the effect of gravity on airflow distribution and particle deposition in the lung. Inhal. Toxicol. 18(7):473-481.

Asgharian, B., O. Price, G. McClellan, R. Corley, D.R. Einstein, R.E. Jacob, J. Harkema, S.A. Carey, E. Schelegle, D. Hyde, J.S. Kimbell, and F.J. Miller. 2012. Development of a rhesus monkey lung geometry model and application to particle deposition in comparison to humans. Inhal. Toxicol. 24(13):869-899.

Athersuch, T. 2016. Metabolome analyses in exposome studies: Profiling methods for a vast chemical space. Arch. Biochem. Biophys. 589:177-186.

Austin, C., T.M. Smith, A. Bradman, K. Hinde, R. Joannes-Boyau, D. Bishop, D.J. Hare, P. Doble, B. Eskenazi, and M. Arora. 2013. Barium distributions in teeth reveal early-life dietary transitions in primates. Nature 498(7453):216-219.

Aylward, L.L., J.J. Collins, K.M. Bodner, M. Wilken, and C.M. Bodnar. 2014. Intrinsic elimination rate and dietary intake estimates for selected indicator PCBs: Toxicokinetic modeling using serial sampling data in US subjects, 2005-2010. Chemosphere 110:48-52.

Bartels, M., D. Rick, E. Lowe, G. Loizou, P. Price, M. Spendiff, S. Arnold, J. Cocker, and N. Ball. 2012. Development of PK- and PBPK-based modeling tools for derivation of biomonitoring guidance values. Comput. Methods Programs Biomed. 108(2):773-788.

Beaudouin, R., S. Micallef, and C. Brochot. 2010. A stochastic whole-body physiologically based pharmacokinetic model to assess the impact of inter-individual variability on tissue dosimetry over the human lifespan. Regul. Toxicol. Pharmacol. 57(1):103-116.

Bennett, D.H., and E.J. Furtaw. 2004. Fugacity-based indoor residential pesticide fate model. Environ. Sci. Technol. 38(7):2142-2152.

Berellini, G., N.J. Waters, and F. Lombardo. 2012. In silico prediction of total human plasma clearance. J. Chem. Inf. and Model. 52(8):2069-2078.

Boethling, R.S. 2011. Incorporating environmental attributes into musk design. Green Chem. 13(12):3386-3396.

Bois, F.Y., M. Jamei, and H.J. Clewell. 2010. PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology 278(3):256-267.

Brauer, M., G. Freedman, J. Frostad, A. van Donkelaar, R.V. Martin, F. Dentener, R. Van Dingenen, K. Estep, H. Amini, J.S. Apte, K. Balakrishnan, L. Barregard, D.M. Broday, V. Feigin, S. Ghosh, P.K. Hopke, L.D. Knibbs, Y. Kokubo, Y. Liu, S. Ma, L. Morawska, J.L. Texcalac Sangrador, G. Shaddick, H.R. Anderson, T. Vos, M.H. Forouzanfar, R.T. Burnett, and A. Cohen. 2015. Ambient air pollution exposure estimation for the global burden of disease 2013. Environ. Sci. Technol. 50(1):79-88.

Breen, M.S., T.C. Long, B.D. Schultz, J. Crooks, M. Breen, J.E. Langstaff, K.K. Isaacs, Y.M. Tan, R.W. Williams, Y. Cao, A.M. Geller, R.B. Devlin, S.A. Batterman, and T.J. Buckley. 2014. GPS-based microenvironment tracker (MicroTrac) model to estimate time-location of individuals for air pollution exposure assessments: Model evaluation in central North Carolina. J. Expo. Sci. Environ. Epidemiol. 24(4):412-420.

Burgess, L.G., K. Uppal, D.I. Walker, R.M. Roberson, V. Tran, M.B. Parks, E.A. Wade, A.T. May, A.C. Umfress, K.L. Jarrell, B.O. Stanley, J. Kuchtey, R.W. Kuchtey, D.P. Jones, and M.A. Brantley, Jr. 2015. Metabolome-wide association study of primary open angle glaucoma. Invest. Ophthalmol. Vis. Sci. 56(8):5020-5028.

Buser, A.M., M. MacLeod, M. Scheringer, D. Mackay, M. Bonnell, M.H. Russell, J.V. DePinto, and K. Hungerbühler. 2012. Good modeling practice guidelines for applying multimedia models in chemical assess-ments. Integr. Environ. Assess. Manage. (4):703-708.

Calafat, A.M. 2012. The US National Health and Nutrition Examination Survey and human exposure to environmental chemicals. Int. J. Hyg. Environ. Health 215(2):99-101.

Campbell, J.L., M.E. Andersen, and H.J. Clewell. 2014. A hybrid CFD-PBPK model for naphthalene in rat and human with IVIVE for nasal tissue metabolism and cross-species dosimetry. Inhal. Toxicol. 26(6):333-344.

Casas, L., M.F. Fernandez, S. Llop, M. Guxens, F. Ballester, N. Olea, M.B. Irurzun, L.S. Rodriguez, I. Riano, A. Tardon, M. Vrijheid, A.M. Calafat, J. Sunyer. and I. Project. 2011. Urinary concentrations of phthalates and phenols in a population of Spanish pregnant women and children. Environ. Int. 37(5):858-866.

Castle, A.L., O. Fiehn, R. Kaddurah-Daouk, and J.C. Lindon. 2006. Metabolomics standards workshop and the development of international standards for reporting metabolomics experimental results. Brief Bioinform. 7(2):159-165.

Chibwe, L., M.C. Geier, J. Nakamura, R.L. Tanguay, M.D. Aitken, and S.L. Simonich. 2015. Aerobic bioremediation of PAH contaminated soil results in increased genotoxicity and developmental toxicity. Environ. Sci. Technol. 49(23):13889-13898.

Claassen, K., K. Thelen, K. Coboeken, T. Gaub, J. Lippert, K. Allegaert, and S. Willmann. 2015. Development of a physiologically-based pharmacokinetic model for preterm neonates: Evaluation with in vivo data. Curr. Pharm. Des. 21(39):5688-5698.

Coecke, S., H. Ahr, B.J. Blaauboer, S. Bremer, S. Casati, J. Castell, R. Combes, R. Corvi, C.L. Crespi, M.L. Cunningham, G. Elaut, B. Eletti, A. Freidig, A. Gennari, J.F. Ghersi-Egea, A. Guillouzo, T. Hartung, P. Hoet, M. Ingelman-Sundberg, S. Munn, W. Janssens, B. Ladstetter, D. Leahy, A. Long, A. Meneguz, M. Monshouwer, S. Morath, F. Nagelkerke, O. Pelkonen, J. Ponti, P. Prieto, L. Richert, E. Sabbioni, B. Schaack, W. Steiling, E. Testai, J.A. Vericat, and A. Worth. 2006. Metabolism: A bottleneck in in vitro toxicological test development – The report and recommendations of ECVAM workshop 54. Altern. Lab. Anim. 34(1):49-84.

Cohen, J.M., J.G. Teeguarden, and P. Demokritou. 2014. An integrated approach for the in vitro dosimetry of engineered nanomaterials. Part. Fibre Toxicol. 11:20.

Cowan-Ellsberry, C.E., M.S. McLachlan, J.A. Arnot, M. Macleod, T.E. McKone, and F. Wania. 2009. Modeling exposure to persistent chemicals in hazard and risk assessment. Integr. Environ. Assess. Manag. 5(4):662-679.

Cravatt, B.F., A.T. Wright, and J.W. Kozarich. 2008. Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. Annu. Rev. Biochem. 77:383-414.

Crowell, S.R., A.K. Sharma, S. Amin, J.J. Soelberg, N.C. Sadler, A.T. Wright, W.M. Baird, D.E. Williams, and R.A. Corley. 2013. Impact of pregnancy on the pharmacokinetics of dibenzo[def,p]chrysene in mice. Toxicol. Sci. 135(1):48-62.

Davis, J.A., J.S. Gift, and Q.J. Zhao. 2011. Introduction to benchmark dose methods and US EPA's benchmark dose software (BMDS) version 2.1.1. Toxicol. Appl. Pharmacol. 254(2):181-191.

Davis, J.M., and W.H. Farland. 2001. The paradoxes of MTBE. Toxicol. Sci. 61(2):211-217.

de Nazelle, A., E. Seto, D. Donaire-Gonzalez, M. Mendez, J. Matamala, M. J. Nieuwenhuijsen, and M. Jerrett. 2013. Improving estimates of air pollution exposure through ubiquitous sensing technologies. Environ. Poll. 176:92-99.

De Wit, P., M.H. Pespeni, and S.R. Palumbi. 2015. SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. Mol. Ecol. 24(10):2310-2323.

Delclos, K.B., C.C. Weis, T.J. Bucci, G. Olson, P. Mellick, N. Sadovova, J.R. Latendresse, B. Thorn, and R.R. Newbold. 2009. Overlapping but distinct effects of genistein and ethinyl estradiol (EE(2)) in female Sprague-Dawley rats in multigenerational reproductive and chronic toxicity studies. Reprod. Toxicol. 27(2):117-132.

Delclos, K.B., L. Camacho, S.M. Lewis, M.M. Vanlandingham, J.R. Latendresse, G.R. Olson, K.J. Davis, R.E. Patton, G. Gamboa da Costa, K.A. Woodling, M.S. Bryant, M. Chidambaram, R. Trbojevich, B.E. Juliar, R.P. Felton, and B.T. Thorn. 2014. Toxicity evaluation of bisphenol A administered by gavage to Sprague Dawley rats from gestation day 6 through postnatal day 90. Toxicol. Sci. 139(1):174-197.

Delmaar, C., B. Bokkers, W. ter Burg, and G. Schuur. 2015. Validation of an aggregate exposure model for substances in consumer products: A case study of diethyl phthalate in personal care products. J. Expo. Sci. Environ. Epidemiol. 25(3):317-323.

DeLoid, G., J.M. Cohen, T. Darrah, R. Derk, L. Rojanasakul, G. Pyrgiotakis, W. Wohlleben, and P. Demokritou. 2014. Estimating the effective density of engineered nanomaterials for in vitro dosimetry. Nat. Commun. 5:3514.

Dionisio, K.L., A.M. Frame, M.R. Goldsmith, J.F. Wambaugh, A. Liddell, T. Cathey, D. Smith, J. Vail, A.S. Ernstoff, P. Fantke, O. Jolliet, and R.S. Judson. 2015. Exploring consumer exposure pathways and patterns of use for chemicals in the environment. Toxicol. Rep. 2:228-237.

Dudzina, T., C.J. Delmaar, J.W. Biesterbos, M.I. Bakker, B.G. Bokkers, P.T. Scheepers, J.G. van Engelen, K. Hungerbuehler, and N. von Goetz. 2015. The probabilistic aggregate consumer exposure model (PACEM): Validation and comparison to a lower-tier assessment for the cyclic siloxane D5. Environ. Int. 79:8-16.

Edrissi, B., K. Taghizadeh, B.C. Moeller, D. Kracko, M. Doyle-Eisele, J.A. Swenberg, and P.C. Dedon. 2013. Dosimetry of N(6)-formyllysine adducts following [(1)(3) C(2)H(2)]-formaldehyde exposures in rats. Chem. Res. Toxicol. 26(10):1421-1423.

EFSA (European Food Safety Authority). 2016. Review of the Threshold of Toxicological Concern (TTC) Approach and Development of New TTC Decision Tree [online]. Available: http://www.efsa.europa.eu/en/supporting/pub/1006e [accessed July 15, 2016].

Egeghy, P.P., D.A. Vallero, and E.A. Cohen Hubal. 2011. Exposure-based prioritization of chemicals for risk assessment. Environ. Sci. Pol. 14(8):950-964.

Egeghy, P.P., R. Judson, S. Gangwal, S. Mosher, D. Smith, J. Vail, and E.A. Cohen Hubal. 2012. The exposure data landscape for manufactured chemicals. Sci. Total Environ. 414:159-166.

Egeghy, P.P., L.S. Sheldon, K.K. Isaacs, H. Özkaynak, M.R. Goldsmith, J.F. Wambaugh, R.S. Judson and T.J. Buckley. 2016. Computational exposure science: An emerging discipline to support 21st-century risk assessment. Environ. Health Perspect. 124(6):697-702.

Elgethun, K., M.G. Yost, C.T. Fitzpatrick, T.L. Nyerges, and R.A. Fenske. 2007. Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns. J. Expo. Sci. Environ. Epidemiol. 17(2):196-206.

EPA (US Environmental Protection Agency). 2004. Observational Batteries and Motor Activity. Office of Research and Development, US Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryID=36922 [accessed April 12, 2016].

EPA (US Environmental Protection Agency). 2009. Guidance on the Development, Evaluation and Application of Environmental Models. EPA/100/K-09/003. Office of the Science Advisor, Council for Regulatory Environmental Modeling, US Environmental Protection Agency, Washington, DC [online]. Available: https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1003E4R.PDF [accessed October 24, 2016].

EPA (US Environmental Protection Agency). 2010. Potential for Incorporation of Genetic Polymorphism Data in Human Health Risk Assessment. US Environmental Protection Agency, Washington, DC.

EPA (US Environmental Protection Agency). 2011. Exposure Factors Handbook: 2011 Edition. EPA/600/R-09/052F. National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency, Washington, DC [online]. Available: http://www.nrc.gov/docs/ML1400/ML14007A666.pdf [accessed July 15, 2016].

EPA (US Environmental Protection Agency). 2016a. Modeling EPA Guidance and Publications Developed by Other EPA Organizations [online]. Available: https://www.epa.gov/modeling/modeling-epa-guidance-and-publications-developed-other-epa-organizations [accessed October 24, 2016].

EPA (US Environmental Protection Agency). 2016b. The Exposure Science in the 21st Century (ES21) Federal Working Group [online]. Available: https://www.epa.gov/innovation/exposure-science-21st-century-federal-working-group [accessed October 24, 2016].

Esch, M.B., T.L. King, and M.L. Shuler. 2011. The role of body-on-a-chip devices in drug and toxicity studies. Annu. Rev. Biomed. Eng. 13:55-72.

Fantke, P., A.S. Ernstoff, L. Huang, S.A. Csiszar, and O. Jolliet. 2016. Coupled near-field and far-field exposure assessment framework for chemicals in consumer products. Environ. Int. 94:508-518.

Fardell, J.E., J. Vardy, I.N. Johnston, and G. Winocur. 2011. Chemotherapy and cognitive impairment: Treatment options. Clin. Pharmacol. Ther. 90(3):366-376.

Fenech, M., and S. Bonassi. 2011. The effect of age, gender, diet and lifestyle on DNA damage measured using micronucleus frequency in human peripheral blood lymphocytes. Mutagenesis 26(1):43-49.

Fiehn, O. 2002. Metabolomics – The link between genotypes and phenotypes. Plant Mol. Biol. 48(1-2):155-171.

Fiehn, O., B. Kristal, B. van Ommen, L.W. Sumner, S.A. Sansone, C. Taylor, N. Hardy, and R. Kaddurah-Daouk. 2006. Establishing reporting standards for metabolomic and metabonomic studies: A call for participation. OMICS. 10(2):158-163.

Fitzgerald, K.A., M. Malhotra, C.M. Curtin, F.J. Brien, and C.M. O'Driscoll. 2015. Life in 3D is never flat: 3D models to optimise drug delivery. J. Control. Release. 215:39-54.

Gama, S., D. Mackay, and J.A. Arnot. 2012. Selecting and designing chemicals: Application of a mass balance model of chemical fate, exposure and effects in the environment. Green Chem. 14:1094-1102.

Gavina, J.M., C. Yao, and Y.L. Feng. 2014. Recent developments in DNA adduct analysis by mass spectrometry: A tool for exposure biomonitoring and identification of hazard for environmental pollutants. Talanta 130:475-494.

Geddes, J.A., R.V. Martin, B.L. Boys, and A. van Donkelaar. 2016. Long-term trends worldwide in ambient NO concentrations inferred from satellite observations. Environ. Health Perspect. 124(3):281-289.

Georgopoulos, P.G., C.J. Brinkerhoff, S. Isukapalli, M. Dellarco, P.J. Landrigan, and P.J. Lioy. 2014. A tiered framework for risk-relevant characterization and ranking of chemical exposures: Applications to the national children's study (NCS). Risk Anal. 34(7):1299-1316.

Gilbert, D., P. Mayer, M. Pedersen, and A.M. Vinggaard. 2015. Endocrine activity of persistent organic pollutants accumulated in human silicone implants – Dosing in vitro assays by partitioning from silicone. Environ. Int. 84:107-114.

Ginsberg, G., K. Guyton, D. Johns, J. Schimek, K. Angle, and B. Sonawane. 2010. Genetic polymorphism in metabolism and host defense enzymes: Implications for human health risk assessment. Crit. Rev. Toxicol. 40(7):575-619.

Go, Y.M., and D.P. Jones. 2014. Redox biology: Interface of the exposome with the proteome, epigenome and genome. Redox Biol. 2:358-360.

Go, Y.M., D.I. Walker, Y. Liang, K. Uppal, Q.A. Soltow, V. Tran, F. Strobel, A.A. Quyyumi, T.R. Ziegler, K.D. Pennell, G.W. Miller, and D.P. Jones. 2015. Reference standardization for mass spectrometry and high-resolution metabolomics applications to exposome research. Toxicol. Sci. 148(2):531-543.

Gobas, F.A., S. Xu, G. Kozerski, D.E. Powell, K.B. Woodburn, D. Mackay, and A. Fairbrother. 2015. Fugacity and activity analysis of the bioaccumulation and environmental risks of decamethylcyclopentasiloxane (D5). Environ. Toxicol. Chem. 34(12):2723-2731.

Goldsmith, M.R., C.M. Grulke, R.D. Brooks, T.R. Transue, Y.M. Tan, A. Frame, P.P. Egeghy, R. Edwards, D.T. Chang, R. Tornero-Velez, K. Isaacs, A. Wang, J. Johnson, K. Holm, M. Reich, J. Mitchell, D.A. Vallero, L. Phillips, M. Phillips, J.F. Wambaugh, R.S. Judson, T.J. Buckley, and C.C. Dary. 2014. Development of a consumer product ingredient database for chemical exposure screening and prioritization. Food Chem. Toxicol. 65(5):269-279.

Goodacre, R., D. Broadhurst, A. Smilde, B. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg, and F. Wulfert. 2007. Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 3(3):231-241.

Gosens, I., C.J. Delmaar, W. ter Burg, C. de Heer, and A.G. Schuur. 2014. Aggregate exposure approaches for parabens in personal care products: A case assessment for children between 0 and 3 years old. J. Expo.. Sci. Environ. Epidemiol. 24(2):208-214.

Groothuis, F.A., M.B. Heringa, B. Nicol, J.L. Hermens, B.J. Blaauboer, and N.I. Kramer. 2015. Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. Toxicology 332:30-40.

Grulke, C.M., K. Holm, M.R. Goldsmith, and Y.M. Tan. 2013. PROcEED: Probabilistic reverse dosimetry approaches for estimating exposure distributions. Bioinformation 9(13):707-709.

Gulden, M., and H. Seibert. 2003. In vitro-in vivo extrapolation: estimation of human serum concentrations of chemicals equivalent to cytotoxic concentrations in vitro. Toxicology 189(3):211-222.

Gulden, M., P. Dierickx, and H. Seibert. 2006. Validation of a prediction model for estimating serum concentrations of chemicals which are equivalent to toxic concentrations in vitro. Toxicol In Vitro 20(7):1114-1124.

Haines, D.A., and J. Murray. 2012. Human biomonitoring of environmental chemicals – early results of the 2007-2009 Canadian Health Measures Survey for males and females. Int. J. Hyg. Environ. Health 215(2):133-137.

Hays, S.M., D.W. Pyatt, C.R. Kirman, and L.L. Aylward. 2012. Biomonitoring equivalents for benzene. Regul. Toxicol. Pharmacol. 62(1):62-73.

Hays, S.M., L.L. Aylward, and B.C. Blount. 2015. Variation in urinary flow rates according to demographic characteristics and body mass index in NHANES: Potential confounding of associations between health outcomes and urinary biomarker concentrations. Environ. Health Perspect. 123(4):293-300.

Heather, J.M., and B. Chain. 2015. The sequence of sequencers: The history of sequencing DNA. Genomics 107(1):1-8.

Heringa, M.B., R. Schreurs, F. Busser, P.T. Van Der Saag, B. Van Der Burg, and J.L. Hermens. 2004. Toward more useful in vitro toxicity data with measured free concentrations. Environ. Sci. Technol. 38(23):6263-6270.

Hicks, K.D., A.W. Sullivan, J. Cao, E. Sluzas, M. Rebuli, and H.B. Patisaul. 2016. Interaction of bisphenol A (BPA) and soy phytoestrogens on sexually dimorphic sociosexual behaviors in male and female rats. Horm. Behav. 84:121-126.

Hinderliter, P.M., K.R. Minard, G. Orr, W.B. Chrisler, B.D. Thrall, J.G. Pounds, and J.G. Teeguarden. 2010. ISDD: A computational model of particle sedimentation, diffusion and target cell dosimetry for in vitro toxicity studies. Part Fibre Toxicol. 7(1):36.

Hsiao, Y.W., U. Fagerholm, and U. Norinder. 2013. In silico categorization of in vivo intrinsic clearance using machine learning. Mol. Pharm. 10(4):1318-1321.

Hutzler, J.M., B.J. Ring, and S.R. Anderson. 2015. Low-turnover drug molecules: A current challenge for drug metabolism scientists. Drug Metab. Dispos. 43(12):1917-1928.

Isaacs, K.K., W.G. Glen, P. Egeghy, M.R. Goldsmith, L. Smith, D. Vallero, R. Brooks, C.M. Grulke, and H. Özkaynak. 2014. SHEDS-HT: An integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. Environ. Sci. Technol. 48(21):12750-12759.

Jahnke, A., P. Mayer, M.S. McLachlan, H. Wickstrom, D. Gilbert, and M. MacLeod. 2014. Silicone passive equilibrium samplers as 'chemometers' in eels and sediments of a Swedish lake. Environ. Sci. Process. Impact. 16(3):464-472.

Jittikoon, J., S. Mahasirimongkol, A. Charoenyingwattana, U. Chaikledkaew, P. Tragulpiankit, S. Mangmool, W. Inunchot, C. Somboonyosdes, N. Wichukchinda, P. Sawanpanyalert, Y. He, H.L. McLeod, and W. Chantratita. 2016. Comparison of genetic variation in drug ADME-related genes in Thais with Caucasian, African and Asian HapMap populations. J. Hum. Genet. 61(2):119-127.

Kalemba-Drozdz, M. 2015. The interaction between air pollution and diet does not influence the DNA damage in lymphocytes of pregnant women. Environ. Res. 136:295-299.

Kesisoglou, F., B. Xia, and N.G. Agrawal. 2015. Comparison of deconvolution-based and absorption modeling IVIVC for extended release formulations of a BCS III drug development candidate. AAPS. J. 17(6):1492-1500.

Knecht, A.L., B.C. Goodale, L. Truong, M.T. Simonich, A.J. Swanson, M.M. Matzke, K.A. Anderson, K.M. Waters, and R.L. Tanguay. 2013. Comparative developmental toxicity of environmentally relevant oxygenated PAHs. Toxicol. Appl. Pharmacol. 271(2):266-275.

Koch, K., and K.L. Brouwer. 2012. A perspective on efflux transport proteins in the liver. Clin. Pharmacol. Ther. 92(5):599-612.

Koelmans, A.A., A. van der Heijde, L.M. Knijff, and R.H. Aalderink. 2001. Integrated modelling of eutrophication and organic contaminant fate & effects in aquatic ecosystems. A review. Water Res. 35(15):3517-3536.

Kolanczyk, R.C., P. Schmieder, W.J. Jones, O.G. Mekenyan, A. Chapkanov, S. Temelkov, S. Kotov, M. Velikova, V. Kamenska, K. Vasilev, and G.D. Veith. 2012. MetaPath: An electronic knowledge base for collating, exchanging and analyzing case studies of xenobiotic metabolism. Regul. Toxicol. Pharmacol. 63(1):84-96.

Kramer, N.I., F.J. Busser, M.T. Oosterwijk, K. Schirmer, B.I. Escher, and J.L. Hermens. 2010. Development of a partition-controlled dosing system for cell assays. Chem. Res. Toxicol. 23(11):1806-1814.

Kramer, N.I., M. Krismartina, A. Rico-Rico, B.J. Blaauboer, and J.L. Hermens. 2012. Quantifying processes determining the free concentration of phenanthrene in basal cytotoxicity assays. Chem. Res. Toxicol. 25(2):436-445.

Leung, L., A.S. Kalgutkar, and R.S. Obach. 2012. Metabolic activation in drug-induced liver injury. Drug Metab. Rev. 44(1):18-33.

Li, J., X. Lao, C. Zhang, L. Tian, D. Lu, and S. Xu. 2014. Increased genetic diversity of ADME genes in African Americans compared with their putative ancestral source populations and implications for pharmacogenomics. BMC Genet 15:52.

Liao, K.H., Y.M. Tan, and H.J. Clewell, III. 2007. Development of a screening approach to interpret human biomonitoring data on volatile organic compounds: Reverse dosimetry on biomonitoring data for trichloroethylene. Risk Anal. 27(5):1223-1236.

Little, J.C., C.J. Weschler, W.W. Nazaroff, Z. Liu, and E.A. Cohen Hubal. 2012. Rapid methods to estimate potential exposure to semivolatile organic compounds in the indoor environment. Environ. Sci. Technol. 46(20):11171-11178.

Liu, L.Y., A. Salamova, K. He, and R.A. Hites. 2015a. Analysis of polybrominated diphenyl ethers and emerging halogenated and organophosphate flame retardants in human hair and nails. J. Chromatogr A. 1406:251-257.

Liu, R., P. Schyman, and A. Wallqvist. 2015b. Critically assessing the predictive power of QSAR models for human liver microsomal stability. J. Chem. Inf. Model. 55(8):1566-1575.

Lorber, M., and P.P. Egeghy. 2011. Simple intake and pharmacokinetic modeling to characterize exposure of Americans to perfluorooctanoic acid, PFOA. Environ. Sci. Technol. 45(19):8006-8014.

Lovreglio, P., F. Maffei, M. Carrieri, M.N. D'Errico, I. Drago, P. Hrelia, G.B. Bartolucci and L. Soleo. 2014. Evaluation of chromosome aberration and micronucleus frequencies in blood lymphocytes of workers exposed to low concentrations of benzene. Mutat. Res. Genet. Toxicol. Environ. Mutagen. 770:55-60.

Mackay, D., J.A. Arnot, F. Wania, and R.E. Bailey. 2011. Chemical activity as an integrating concept in environmental assessment and management of contaminants. Integr. Environ. Assess. Manag. 7(2):248-255.

MacLachlan, D.J. 2010. Physiologically based pharmacokinetic (PBPK) model for residues of lipophilic pesticides in poultry. Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess. 27(3):302-314.

Martin, S.A., E.D. McLanahan, P.J. Bushnell, E.S. Hunter, III, and H. El-Masri. 2015. Species extrapolation of life-stage physiologically-based pharmacokinetic (PBPK) models to investigate the developmental toxicology of ethanol using in vitro to in vivo (IVIVE) methods. Toxicol. Sci. 143(2):512-535.

Mastrangelo, A., A. Ferrarini, F. Rey-Stolle, A. Garcia, and C. Barbas. 2015. From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. Anal. Chim. Acta 900:21-35.

Mattingly, C.J., T.E. McKone, M.A. Callahan, J.A. Blake, and E.A. Cohen Hubal. 2012. Providing the missing link: The exposure science ontology ExO. Environ. Sci. Technol. 46(6 ):3046-3053.

Mayer, H.J., M.R. Greenberg, J. Burger, M. Gochfield, C. Powers, D. Kosson, R. Keren, C. Danis, and V. Vyas. 2005. Using integrated geospatial mapping and conceptual site models to guide risk-based environmental cleanup decisions. Risk Anal. 25(2):429-446.

McGinn, S., D. Bauer, T. Brefort, L. Dong, A. El-Sagheer, A. Elsharawy, G. Evans, E. Falk-Sorqvist, M. Forster, S. Fredriksson, P. Freeman, C. Freitag, J. Fritzsche, S. Gibson, M. Gullberg, M. Gut, S. Heath, I. Heath-Brun, A.J. Heron, J. Hohlbein, R. Ke, O. Lancaster, L. Le Reste, G. Maglia, R. Marie, F. Mauger, F. Mertes, M. Mignardi, L. Moens, J. Oostmeijer, R. Out, J.N. Pedersen, F. Persson, V. Picaud, D. Rotem, N. Schracke, J. Sengenes, P.F. Stahler, B. Stade, D. Stoddart, X. Teng, C.D. Veal, N. Zahra, H. Bayley, M. Beier, T. Brown, C. Dekker, B. Ekstrom, H. Flyvbjerg, A. Franke, S. Guenther, A.N. Kapanidis, J. Kaye, A. Kristensen, H. Lehrach, J. Mangion, S. Sauer, E. Schyns, J. Tost, J.M. van Helvoort, P.J. van der Zaag, J.O. Tegenfeldt, A.J. Brookes, K. Mir, M. Nilsson, S. Will-

cocks and, and I.G. Gut. 2016. New technologies for DNA analysis — a review of the READNA Project. N. Biotechnol. 33(3):311-330.

McKone, T.E., R. Castorina, M.E. Harnly, Y. Kuwabara, B. Eskenazi, and A. Bradmanm. 2007. Merging models and biomonitoring data to characterize sources and pathways of human exposure to organophosphorus pesticides in the Salinas Valley of California. Environ. Sci. Technol. 41(9):3233-3240.

McLachlan, M.S., A. Kierkegaard, M. Radke, A. Sobek, A. Malmvärn, T. Alsberg, J.A. Arnot, T.N. Brown, F. Wania, K. Breivik, and S. Xu. 2014. Using model-based screening to help discover unknown environmental contaminants. Environ. Sci. Technol. 48(13):7264-7271.

McLanahan, E.D., H.A. El-Masri, L.M. Sweeney, L.Y. Kopylev, H.J. Clewell, J.F. Wambaugh, and P.M. Schlosser. 2012. Physiologically based pharmacokinetic model use in risk assessment-why being published is not enough. Toxicol. Sci. 126(1):5-15.

McNally, K., R. Cotton, J. Cocker, K. Jones, M. Bartels, D. Rick, P. Price, and G. Loizou. 2012. Reconstruction of exposure to m-xylene from human biomonitoring data using PBPK modelling, Bayesian inference, and Markov Chain Monte Carlo simulation. J. Toxicol. 2012:760281.

Mitchell, J., J.A. Arnot, O. Jolliet, P.G. Georgopoulos, S. Isukapalli, S. Dasgupta, M. Pandian, J. Wambaugh, P. Egeghy, E.A. Cohen Hubal, and D.A. Vallero. 2013. Comparison of modeling approaches to prioritize chemicals based on estimates of exposure and exposure potential. Sci. Total Environ. 458-460:555-567.

Moeller, B.C., L. Recio, A. Green, W. Sun, F.A. Wright, W.M. Bodnar, and J.A. Swenberg. 2013. Biomarkers of exposure and effect in human lymphoblastoid TK6 cells following [13C2]-acetaldehyde exposure. Toxicol. Sci. 133(1):1-12.

Moro, A.M., N. Brucker, M.F. Charao, E. Sauer, F. Freitas, J. Durgante, G. Bubols, S. Campanharo, R. Linden, A.P. Souza, C. Bonorino, R. Moresco, D. Pilger, A. Gioda, S. Farsky, A. Duschl, and S.C. Garcia. 2015. Early hematological and immunological alterations in gasoline station attendants exposed to benzene. Environ. Res. 137:349-356.

Mortensen, M.E., A.M. Calafat, X. Ye, L.Y. Wong, D.J. Wright, J.L. Pirkle, L.S. Merrill, and J. Moye. 2014. Urinary concentrations of environmental phenols in pregnant women in a pilot study of the National Children's Study. Environ. Res. 129:32-38.

Motorykin, O., J. Schrlau, Y. Jia, B. Harper, S. Harris, A. Harding, D. Stone, M. Kile, D. Sudakin, and S.L. Massey Simonich. 2015. Determination of parent and hydroxy PAHs in personal $PM_{2.5}$ and urine samples collected during Native American fish smoking activities. Sci. Total Environ. 505:694-703.

Muir, D.C., and P.H. Howard. 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. Environ. Sci. Technol. 40(23):7157-7166.

Needham, L.L., D.B. Barr, and A.M. Calafat. 2005. Characterizing children's exposures: Beyond NHANES. Neurotoxicology 26(4):547-553.

Nethery, E., G. Mallach, D. Rainham, M. Goldberg, and A. Wheeler. 2014. Using global positioning systems (GPS) and temperature data to generate time-activity classifications for estimating personal exposure in air monitoring studies: an automated method. Environ. Health 13(1):33.

NIEHS (National Institute of Environmental Health Sciences). 2016. Children's Health Exposure Analysis Resource (CHEAR) [online]. Available: http://www.niehs.nih.gov/research/supported/exposure/chear/ [accessed July 15, 2016].

NRC (National Research Council). 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

NRC (National Research Council). 2014. A Framework to Guide Selection of Chemical Alternatives. Washington, DC: The National Academies Press.

NTP (National Toxicoloy Program). 2008. NTP-CERHR Monograph on the Potential Human Reproductive and Developmental Effects of Bisphenol A. NIH Publication No. 08-5994 [online]. Available: https://ntp.niehs.nih.gov/ntp/ohat/bisphenol/bisphenol.pdf [accessed July 15, 2016].

Obach, R.S., F. Lombardo, and N.J. Waters. 2008. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab. Dispos. 36(7):1385-1405.

O'Connell, S.G., L.D. Kind, and K.A. Anderson. 2014a. Silicone wristbands as personal passive samplers. Environ. Sci. Technol. 48(6):3327-3335.

O'Connell, S.G., M.A. McCartney, L.B. Paulik, S.E. Allan, L.G. Tidwell, G. Wilson, and K.A. Anderson. 2014b. Improvements in pollutant monitoring: Optimizing silicone for co-deployment with polyethylene passive sampling devices. Environ. Poll. 193:71-78.

O'Connell, S.G., N.I. Kerkvliet, S. Carozza, D. Rohlman, J. Pennington, and K.A. Anderson. 2015. In vivo contaminant partitioning to silicone implants: Implications for use in biomonitoring and body burden. Environ. Int. 85:182-188.

OECD (Organisation for Economic Co-operation and Development). 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. Series on Testing and Assessment No. 69. Paris: OECD [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en [accessed July 15, 2016].

Ortega, V.E., and D.A. Meyers. 2014. Pharmacogenetics: Implications of race and ethnicity on defining genetic profiles for personalized medicine. J. Allergy Clin. Immunol. 133(1):16-26.

Palmer, R.F., L. Heilbrun, D. Camann, A. Yau, S. Schultz, V. Elisco, B. Tapia, N. Garza, and C. Miller. 2015. Organic compounds detected in deciduous teeth: A replication study from children with autism in two samples. J. Environ. Public Health 2015:862414.

Park, Y.H., K. Lee, Q.A. Soltow, F.H. Strobel, K.L. Brigham, R.E. Parker, M.E. Wilson, R.L. Sutliff, K.G. Mansfield, L.M. Wachtman, T.R. Ziegler, and D.P. Jones. 2012. High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. Toxicology 295(1-3):47-55.

Plowchalk, D.R., and J. Teeguarden. 2002. Development of a physiologically based pharmacokinetic model for estradiol in rats and humans: A biologically motivated quantitative framework for evaluating responses to estradiol and other endocrine-active compounds. Toxicol. Sci. 69(1):60-78.

Pontel, L.B., I.V. Rosado, G. Burgos-Barragan, J.I. Garaycoechea, R. Yu, M.J. Arends, G. Chandrasekaran, V. Broecker, W. Wei, L. Liu, J.A. Swenberg, G.P. Crossan, and K.J. Patel 2015. Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen. Mol. Cell 60(1):177-188.

Pottenger, L.H., L.S. Andrews, A.N. Bachman, P.J. Boogaard, J. Cadet, M.R. Embry, P.B. Farmer, M.W. Himmelstein, A.M. Jarabek, E.A. Martin, R.J. Mauthe, R. Persaud, R.J. Preston, R. Schoeny, J. Skare, J.A. Swenberg, G.M. Williams, E. Zeiger, F. Zhang, and J.H. Kim. 2014. An organizational approach for the assessment of DNA adduct data in risk assessment: Case studies for aflatoxin B1, tamoxifen and vinyl chloride. Crit. Rev. Toxicol. 44(4):348-391.

Quinn, C.L., and F. Wania. 2012. Understanding differences in the body burden-age relationships of bioaccumulating contaminants based on population cross sections versus individuals. Environ. Health Perspect. 120(4):554-559.

Rager, J.E., M.J. Strynar, S. Liang, R.L. McMahen, A.M. Richard, C.M. Grulke, J.F. Wambaugh, K.K. Isaacs, R. Judson, A.J. Williams, and J.R. Sobus. 2016. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environ. Int. 88:269-280.

Rappaport, S.M., D.K. Barupal, D. Wishart, P. Vineis, and A. Scalbert. 2014. The blood exposome and its role in discovering causes of disease. Environ. Health Perspect. 122(8):769-774.

Rebuli, M.E., J. Cao, E. Sluzas, K.B. Delclos, L. Camacho, S.M. Lewis, M.M. Vanlandingham, and H.B. Patisaul. 2014. Investigation of the effects of subchronic low dose oral exposure to bisphenol A (BPA) and ethinyl estradiol (EE) on estrogen receptor expression in the juvenile and adult female rat hypothalamus. Toxicol. Sci. 140(1):190-203.

Reichenberg, F., and P. Mayer. 2006. Two complementary sides of bioavailability: Accessibility and chemical activity of organic contaminants in sediments and soils. Environ. Toxicol. Chem. 25(5):1239-1245.

Regens, J.L., K.R. Obenshain, C.T. Quest, and C. Whipple. 2002. Conceptual site models and multimedia modeling: Comparing MEPAS, MMSOILS, and RESRAD. Hum. Ecol. Risk Asses. 8(2):391-403.

Ritter, R., M. Scheringer, M. MacLeod, C. Moeckel, K.C. Jones, and K. Hungerbühler. 2011. Intrinsic human elimination half-lives of polychlorinated biphenyls derived from the temporal evolution of cross-sectional biomonitoring data from the United Kingdom. Environ. Health Perspect. 119(2):225-231.

Roede, J.R., K. Uppal, Y. Liang, D.E. Promislow, L.M. Wachtman, and D.P. Jones. 2013. Characterization of plasma thiol redox potential in a common marmoset model of aging. Redox Biol. 1:387-393.

Rostami-Hodjegan, A. 2012. Physiologically based pharmacokinetics joined with in vitro-in vivo extrapolation of ADME: A marriage under the arch of systems pharmacology. Clin. Pharmacol. Ther. 92(1):50-61.

Rotroff, D.M., B.A. Wetmore, D.J. Dix, S.S. Ferguson, H.J. Clewell, K.A. Houck, E.L. Lecluyse, M.E. Andersen, R.S. Judson, C.M. Smith, M.A. Sochaski, R.J. Kavlock, F. Boellmann, M.T. Martin, D.M. Reif, J.F. Wambaugh, and R.S. Thomas. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol. Sci. 117(2):348-358.

Rowbotham, A.L., and R.M. Gibson. 2011. Exposure-driven risk assessment: Applying exposure-based waiving of toxicity tests under REACH. Food Chem. Toxicol. 49(8):1661-1673.

Rule, A.D., H.M. Gussak, G.R. Pond, E.J. Bergstralh, M.D. Stegall, F.G. Cosio, and T.S. Larson. 2004. Measured and estimated GFR in healthy potential kidney donors. Am. J. Kidney Dis. 43(1):112-119.

Sadler, N.C., and A.T. Wright. 2015. Activity-based protein profiling of microbes. Curr. Opin. Chem. Biol. 24:139-144.

Sadler, N.C., P. Nandhikonda, B.J. Webb-Robertson, C. Ansong, L.N. Anderson, J.N. Smith, R.A. Corley and A.T. Wright. 2016. Hepatic cytochrome P450 activity, abundance, and expression throughout human development. Drug Metabol. Dispos. 44(7):984-991.

Sayes, C.M., K.L. Reed, and D.B. Warheit. 2007. Assessing toxicity of fine and nanoparticles: Comparing in vitro measurements to in vivo pulmonary toxicity profiles. Toxicol. Sci. 97(1):163-180.

Schenker, U., M. Scheringer, M.D. Sohn, R.L. Maddalena, T.E. McKone, and K. Hungerbühler. 2009. Using information on uncertainty to improve environmental fate modeling: A case study on DDT. Environ. Sci. Technol. 43(1):128-134.

Sedykh, A., D. Fourches, J. Duan, O. Hucke, M. Garneau, H. Zhu, P. Bonneau, and A. Tropsha. 2013. Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions. Pharm. Res. 30(4):996-1007.

Seltenrich, N. 2014. Remote-sensing applications for environmental health research. Environ. Health Perspect. 122(10):A268-A275.

Shankaran, H., and J. Teeguarden. 2014. Improving urine-based human exposure assessment of short-lived chemicals using reverse dosimetry and nhanes physiological and behavior data: A value-of-information approach for bisphenol A. Toxicologist CD138(1).

Shin, H.M.., T.E. McKone, and D.H. Bennett. 2012. Intake fraction for the indoor environment: A tool for prioritizing indoor chemical sources. Environ. Sci. Technol. 46(18):10063-10072.

Shin, H.M., T.E. McKone, and D.H. Bennett. 2014. Attributing population-scale human exposure to various source categories: Merging exposure models and biomonitoring data. Environ. Int. 70:183-191.

Shin, H.M., A. Ernstoff, J.A. Arnot, B.A. Wetmore, S.A. Csiszar, P. Fantke, X. Zhang, T.E. McKone, O. Jolliet, and D.H. Bennett. 2015. Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays. Environ. Sci. Technol. 49(11):6760-6771.

Snoeys, J., M. Beumont, M. Monshouwer, and S. Ouwerkerk-Mahadevan. 2016. A mechanistic understanding of the non-linear pharmacokinetics and inter-subject variability of simeprevir: A PBPK-guided drug development approach. Clin. Pharmacol. Ther. 99(2):224-234.

Sonne, C., K. Gustavson, R.J. Letcher, and R. Dietz. 2015. Physiologically-based pharmacokinetic modelling of distribution, bioaccumulation and excretion of POPs in Greenland sledge dogs (Canis familiaris). Environ. Res. 142:380-386.

Stapleton, H.M., J.G. Allen, S.M. Kelly, A. Konstantinov, S. Klosterhaus, D. Watkins, M.D. McClean and T.F. Webster. 2008. Alternate and new brominated flame retardants detected in US house dust. Environ. Sci. Technol. 42(18):6910-6916.

Su, J.G., M. Jerrett, Y.Y. Meng, M. Pickett, and B. Ritz. 2015. Integrating smart-phone based momentary location tracking with fixed site air quality monitoring for personal exposure assessment. Sci. Total Environ. 506:518-526.

Sud, M., E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K.S. Nair, S. Sumner, and S. Subramaniam. 2016. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res. 44(D):D463-D470.

Sumner, L.W., Z.T. Lei, B.J. Nikolau, K. Saito, U. Roessner, and R. Trengove. 2014. Proposed quantitative and alpha-numeric metabolite identification metrics. Metabolomics 10(6):1047-1049.

Swenberg, J.A., G. Boysen, N. Georgieva, M.G. Bird, and R.J. Lewis. 2007. Future directions in butadiene risk assessment and the role of cross-species internal dosimetry. Chem. Biol. Interact. 166(1-3):78-83.

Swenberg, J.A., E. Fryar-Tita, Y.C. Jeong, G. Boysen, T. Starr, V.E. Walker, and R.J. Albertini. 2008. Biomarkers in toxicology and risk assessment: Informing critical dose-response relationships. Chem. Res. Toxicol. 21(1):253-265.

Tan, Y.M., K.H. Liao, R.B. Conolly, B.C. Blount, A.M. Mason, and H.J. Clewell. 2006. Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform. J. Toxicol. Environ. Health A 69(18):1727-1756.

Tan, Y.M., K.H. Liao, and H.J. Clewell, III. 2007. Reverse dosimetry: Interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. J. Expo. Sci. Environ. Epidemiol. 17(7):591-603.

Tan, Y.M., J. Sobus, D. Chang, R. Tornero-Velez, M. Goldsmith, J. Pleil, and C. Dary. 2012. Reconstructing human exposures using biomarkers and other clues. J. Toxicol. Environ. Health B Crit. Rev. 15(1):22-38.

Teeguarden, J.G., and H.A. Barton. 2004. Computational modeling of serum-binding proteins and clearance in extrapolations across life stages and species for endocrine active compounds. Risk Anal. 24(3):751-770.

Teeguarden, J.G., J.M. Waechter, H.J. Clewell, T.R. Covington, and H.A. Barton. 2005. Evaluation of oral and intravenous route pharmacokinetics, plasma protein binding, and uterine tissue dose metrics of bisphenol A: A physiologically based pharmacokinetic approach. Toxicol. Sci. 85(2):823-838.

Teeguarden, J.G., P.M. Hinderliter, G. Orr, B.D. Thrall, and J.G. Pounds. 2007. Particokinetics in vitro: Dosimetry considerations for in vitro nanoparticle toxicity assessments. Toxicol. Sci. 95(2):300-312.

Teeguarden, J.G., M.S. Bogdanffy, T.R. Covington, C. Tan, and A.M. Jarabek. 2008. A PBPK model for evaluating the impact of aldehyde dehydrogenase polymorphisms on comparative rat and human nasal tissue acetaldehyde dosimetry. Inhal. Toxicol. 20(4):375-390.

Teeguarden, J.G., A.M. Calafat, X. Ye, D.R. Doerge, M.I. Churchwell, R. Gunawan, and M.K. Graham. 2011. Twenty-four hour human urine and serum profiles of Bisphenol A during high-dietary exposure. Toxicol. Sci. 123(1):48-57.

Teeguarden, J.G., S. Hanson-Drury, J.W. Fisher, and D.R. Doerge. 2013. Are typical human serum BPA concentrations measurable and sufficient to be estrogenic in the general population? Food Chem. Toxicol. 62:949-963.

Teeguarden, J.G., V.B. Mikheev, K.R. Minard, W.C. Forsythe, W. Wang, G. Sharma, N. Karin, S.C. Tilton, K.M.

Waters, B. Asgharian, O.R. Price, J.G. Pounds, and B.D. Thrall. 2014. Comparative iron oxide nanoparticle cellular dosimetry and response in mice by the inhalation and liquid cell culture exposure routes. Part. Fibre Toxicol. 11:46.

Teeguarden, J.G., N. Twaddle, M.I. Churchwell, X. Yang, J.W. Fisher, L.M. Seryak, and D.R. Doerge. 2015. 24-hour human urine and serum profiles of bisphenol A: Evidence against sublingual absorption following ingestion in soup. Toxicol. Appl. Pharmacol. 288(2):131-142.

Teeguarden, J.G., C. Tan, S. Edwards, J.A. Leonard, K.A. Anderson, R.A. Corley, M.L. Kile, S.M Simonich, D. Stone, K.M. Waters, S. Harper, and D.E. Williams. 2016. Completing the link between exposure science and toxicology for improved environmental health decision making: The aggregate exposure pathway framework. Environ. Sci. Technol. 50(9):4579-4586.

Thayer, K.A., D.R. Doerge, D. Hunt, S.H. Schurman, N.C. Twaddle, M.I. Churchwell, S. Garantziotis, G.E. Kissling, M.R. Easterling, J.R. Bucher, and L.S. Birnbaum. 2015. Pharmacokinetics of bisphenol A in humans following a single oral administration. Environ. Int. 83:107-115.

Tolonen, A., and O. Pelkonen. 2015. Analytical challenges for conducting rapid metabolism characterization for QIVIVE. Toxicology 332:20-29.

Tonnelier, A., S. Coecke, and J.M. Zaldíva. 2012. Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model. Arch. Toxicol. 86(3):393-403.

Tumer, T.B., S. Savranoglu, P. Atmaca, G. Terzioglu, A. Sen, and S. Arslan. 2016. Modulatory role of GSTM1 null genotype on the frequency of micronuclei in pesticide-exposed agricultural workers. Toxicol. Ind. Health. 32(12):1942-1951.

Uppal, K., D.I. Walker, K. Liu, S. Li, Y.M. Go, and D.P. Jones. 2016. Computational metabolomics: A framework for the million metabolome. Chem. Res. Toxicol. 29(12):1956-1975.

van Donkelaar, A., R.V. Martin, M. Brauer, and B.L. Boys. 2015. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. Environ. Health Perspect. 123(2):135-143.

Vermeire, T., M. van de Bovenkamp, Y.B. de Bruin, C. Delmaar, J. van Engelen, S. Escher, H. Marquart and T. Meijster. 2010. Exposure-based waiving under REACH. Regul. Toxicol. Pharmacol. 58(3):408-420.

Verner, M.A., A.E. Loccisano, N.H. Morken, M. Yoon, H. Wu, R. McDougall, M. Maisonet, M. Marcus, R. Kishi, C. Miyashita, M.H. Chen, W.S. Hsieh, M.E. Andersen, H.J. Clewell, III, and M.P. Longnecker. 2015. Associations of perfluoroalkyl substances (PFASs) with lower birth weight: An evaluation of potential confounding by glomerular filtration rate using a physiologically based pharmacokinetic model (PBPK). Environ. Health Perspect. 123(12):1317-1324.

Vineis, P., K. van Veldhoven, M. Chadeau-Hyam, and T.J. Athersuch. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology. Environ. Mol. Mutagen. 54(7):461-467.

Vrijheid, M., R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, C. Thomsen, J. Wright, T.J. Athersuch, N. Avellana, X. Basagaña, C. Brochot, L. Bucchini, M. Bustamante, A. Carracedo, M. Casas, X. Estivill, L. Fairley, D. van Gent, J.R. Gonzalez, B. Granum, R. Gražulevičienė, K.B. Gutzkow, J. Julvez, H.C. Keun, M. Kogevinas, R.R.C. McEachan, H.M. Meltzer, E. Sabidó, P.E. Schwarze, V. Siroux, J. Sunyer, E.J. Want, F. Zeman, and M.J. Nieuwenhuijsen. 2014. The human early-life exposome (HELIX): Project rationale and design. Environ. Health Perspect. 122(6):535-544.

Wambaugh, J.F., R.W. Setzer, D.M. Reif, S. Gangwal, J. Mitchell-Blackwood, J.A. Arnot, O. Joliet, A. Frame, J. Rabinowitz, T.B. Knudsen, R.S. Judson, P. Egeghy, D. Vallero, and E.A. Cohen Hubal. 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast Project. Environ. Sci. Technol. 47(15):8479-8488.

Wambaugh, J.F., A. Wang, K.L. Dionisio, A. Frame, P. Egeghy, R. Judson, and R.W. Setzer. 2014. High throughput heuristics for prioritizing human exposure to environmental chemicals. Environ. Sci. Technol. 48(21):12760-12767.

Wania, F., and D. Mackay. 1999. The evolution of mass balance models of persistent organic pollutant fate in the environment. Environ. Pollut. 100(1-3):223-240.

Weaver, V.M., D.J. Kotchmar, J.J. Fadrowski, and E.K. Silbergeld. 2016. Challenges for environmental epidemiology research: Are biomarker concentrations altered by kidney function or urine concentration adjustment? J. Expo. Sci. Environ. Epidemiol. 26(1):1-8.

Webster, E.M., H. Oian, D. Mackay, R.D. Christensen, B. Tietjen, and R. Zaleski. 2016. Modeling human exposure to indoor contaminants: External source to body tissues. Environ. Sci. Technol. 50(16):8697-8704.

Weijs, L., A. Covaci, R.S. Yang, K. Das, and R. Blust. 2012. Computational toxicology: Physiologically based pharmacokinetic models (PBPK) for lifetime exposure and bioaccumulation of polybrominated diphenyl ethers (PBDEs) in marine mammals. Environ. Pollut. 163:134-141.

Weschler, C.J., and W.W. Nazaroff. 2010. SVOC partitioning between the gas phase and settled dust indoors. Atmos. Environ. 44(30):3609-3620.

Weschler, C.J., and W.W. Nazaroff. 2012. SVOC exposure indoors: Fresh look at dermal pathways. Indoor Air 22(5):356-377.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell, III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thom-

as. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125(1):157-174.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, L. Li, H.J. Clewell, III, R.S. Judson, K. Freeman, W. Bao, M.A. Sochaski, T.M. Chu, M.B. Black, E. Healy, B. Allen, M.E. Andersen, R.D. Wolfinger, and R.S. Thomas. 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol. Sci. 132(2):327-346.

Wetmore, B.A., B. Allen, H.J. Clewell, III, T. Parker, J.F. Wambaugh, L.M. Almond, M.A. Sochaski, and R.S. Thomas. 2014. Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicol. Sci. 142(1):210-224.

Wetzel, T.A., and W.J. Doucette. 2015. Plant leaves as indoor air passive samplers for volatile organic compounds (VOCs). Chemosphere 122:32-37.

WHO (World Health Organization). 2005. Principles of Characterizing and Applying Human Exposure Models. Harmonization Project Document No. 3. Geneva: WHO [online]. Available: http://apps.who.int/iris/bitstream/10665/43370/1/9241563117_eng.pdf [accessed October 24, 2016].

WHO (World Health Organization). 2016. Exposure Assessment. International Programme on Chemical Safety [online]. Available: http://www.who.int/ipcs/methods/harmonization/areas/exposure/en/ [accessed October 24, 2016].

Wild, C.P. 2005. Complementing the genome with an exposome: The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomarkers Prev. 14(8):1847-1850.

Wild, C.P. 2012. The exposome: From concept to utility. Int. J. Epidemiol. 41(1):24-32.

Wilk-Zasadna, I., C. Bernasconi, O. Pelkonen, and S. Coecke. 2015. Biotransformation in vitro: An essential consideration in the quantitative in vitro-to-in vivo extrapolation (QIVIVE) of toxicity data. Toxicology 332:8-19.

Wishart, D.S., C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, and M. Hassanali. 2008. DrugBank: A knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 36:D901-D906.

Wu, H., M. Yoon, M.A. Verner, J. Xue, M. Luo, M.E. Andersen, M.P. Longnecker, and H.J. Clewell, III. 2015. Can the observed association between serum perfluoroalkyl substances and delayed menarche be explained on the basis of puberty-related changes in physiology and pharmacokinetics? Environ. Int. 82:61-68.

Wu, J., C. Jiang, D. Houston, D. Baker, and R. Delfino. 2011. Automated time activity classification based on global positioning system (GPS) tracking data. Environ. Health 10:101.

Xie, H. 2008.Activity assay of membrane transport proteins. Acta. Biochim. Biophys. Sin. 40(4):269-77.

Yang, Y., Y.M. Tan, B. Blount, C. Murray, S. Egan, M. Bolger, and H. Clewell. 2012. Using a physiologically based pharmacokinetic model to link urinary biomarker concentrations to dietary exposure of perchlorate. Chemosphere 88(8):1019-1027.

Yeo, K.R., J.R. Kenny, and A. Rostami-Hodjegan. 2013. Application of in vitro-in vivo extrapolation (IVIVE) and physiologically based pharmacokinetic (PBPK) modelling to investigate the impact of the CYP2C8 polymorphism on rosiglitazone exposure. Eur. J. Clin. Pharmacol. 69(6):1311-1320.

Yoon, M., J.L. Campbell, M.E. Andersen, and H.J. Clewell. 2012. Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. Crit. Rev. Toxicol. 42(8):633-652.

Yoon, M., A. Efremenko, B.J. Blaauboer, and H.J. Clewell. 2014. Evaluation of simple in vitro to in vivo extrapolation approaches for environmental compounds. Toxicol. In Vitro 28(2):164-170.

Young, B.M., N.S. Tulve, P.P. Egeghy, J.H. Driver, V.G. Zartarian, J.E. Johnston, C.J. Delmaar, J.J. Evans, L.A. Smith, G. Glen, C. Lunchick, J.H. Ross, J. Xue, and D.E. Barnekow. 2012. Comparison of four probabilistic models (CARES(®),Calendex™, ConsExpo, and SHEDS) to estimate aggregate residential exposures to pesticides. J. Expo. Sci. Environ. Epidemiol. 22(5):522-532.

Young, J.F., R.H. Luecke, and D.R. Doerge. 2007. Physiologically based pharmacokinetic/pharmacodynamic model for acrylamide and its metabolites in mice, rats, and humans. Chem. Res. Toxicol. 20(3):388-399.

Yu, R., Y. Lai, H.J. Hartwell, B.C. Moeller, M. Doyle-Eisele, D. Kracko, W.M. Bodnar, T.B. Starr, and J.A. Swenberg. 2015. Formation, accumulation, and hydrolysis of endogenous and exogenous formaldehyde-induced DNA damage. Toxicol. Sci. 146(1):170-182.

Zartarian, V., T. Bahadori, and T. McKone. 2005. Adoption of an official ISEA glossary. J. Expo. Anal. Environ. Epidemiol. 15(1):1-5.

Zhang, D., G. Luo, X. Ding, and C. Lu. 2012. Preclinical experimental models of drug metabolism and disposition in drug discovery and development. Acta Pharm. Sinic. B. 2(6):549-561.

Zhang, X., J.A. Arnot, and F. Wania. 2014. Model for screening-level assessment of near-field human exposure to neutral organic chemicals released indoors. Environ. Sci. Technol. 48(20):12312-12319.

# 3

# Advances in Toxicology

The decade since the publication of *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC 2007) has seen continued advances in an array of technological and biological tools used to understand human function and disease at the molecular level. Some advances were initially catalyzed by the Human Genome Project, which of necessity required technological innovations and large-scale collaborations to reach the ultimate goal of mapping the sequence of DNA. Other developments came from advances made by the pharmaceutical industry to screen for chemicals that have specific biological functionality but minimal off-target effects. As a result of those advances, an era of big-data development and of public access and data-sharing has arrived with ever-increasing data-storage capacity, computational speed, and open-access software. Research has also become more multidisciplinary; project teams today often include geneticists, toxicologists, computer scientists, engineers, and statisticians.

A number of advanced tools can now be used in toxicological and epidemiological research; some examples are listed below.

- Large banks of immortalized cells that are derived from lymphocytes and collected from different populations worldwide are available for toxicological research.
- Genetically diverse mouse strains have been created by a multi-institution collaboration (the Complex Trait Consortium; Threadgill and Churchill 2012) and are available for medical and toxicological research. They have been fully genotyped because of the relatively low cost of sequencing today, and the sequence information is publicly available.
- Microarrays and next-generation RNA sequencing can reveal postexposure changes in the simultaneous expression of large numbers of genes (the transcriptome). Technologies are also now available to profile the epigenome (epigenetic changes, such as methylation and histone modifications), the proteome (proteins present in the cell), and metabolome (small molecules).
- Large compilations of a wide variety of biological data are publicly available, as is software for data access, interpretation, and prediction. Text-mining tools applied to scientific-literature databases provide approaches for developing hypotheses on relationships between chemicals, genes, and diseases.
- Automated systems that use multiwell plates provide a high-throughput platform for measuring a wide array of effects in cells and cellular components in response to chemical exposures. Automated, multiwell testing can also be applied for rapid testing of zebrafish, vertebrates that are relatively genetically homologous with humans.
- Computational advances have enabled the development of chemical structure–based methods for predicting toxicity and systems-biology models for evaluating the effects of perturbing various biological pathways.

Some of the advanced tools could be used to address issues in toxicology and ultimately risk assessment (see Chapter 1, Box 1-3). Some of the general risk-assessment questions to which the tools could be applied are the following:

- Planning and scoping: Which chemicals should undergo comprehensive toxicological evaluation first (that is, how should priorities be set among chemicals for testing)?
- Hazard identification: What adverse effects might a chemical have? For example, could it pose a carcinogenic risk or affect kidney or reproductive function? If a data-sparse chemical has a structure or biological activity that is similar to that of a well-studied chemical, can the same types of toxicity be assumed and, if so, at similar exposures? Are cellular-assay responses adaptive (or inconsequential) or harbingers of adverse effects in humans? Does the chemical operate through the same pathways or processes that are associated with cancer, reproductive toxicity, or other adverse human effects?
- Dose–response assessment: How does response change with exposure? At what exposures are risks of

harm inconsequential? Is there a threshold exposure at the population level below which there is no adverse effect?

- Mixtures: What are the hazards and dose–response characteristics of a complex mixture? How does the addition of a chemical to existing exposure contribute to risk?
- Differential susceptibility and vulnerability: Are some populations more at risk than others after exposure to a specific drug or environmental chemical? For example, are some more susceptible because of co-exposures, pre-existing disease, or genetic susceptibility? Are exposures of the young or elderly of greater concern?

Those risk-assessment questions provide the backdrop for considering the recent advances in toxicological tools. Information obtained with the new tools can advance our understanding of the potential health effects of chemical exposures at various points along the exposure-to-outcome continuum, shown in Figure 3-1 below. The starting point along the continuum is the transformation of external exposure to internal exposure, which was discussed in Chapter 2 of this report (see Figure 2-1). The ultimate goal is prediction of the response of the organism or population to exposure, and different tools can be used to probe or inform different places along that continuum. As noted in Chapter 2, although the continuum is depicted as a linear path, the committee recognizes that multiple interconnecting paths are typically involved in the continuum.

This chapter describes a variety of new assays and computational tools that are available for addressing risk-based questions, but it is not meant to be comprehensive. The chapter organization follows the progression along the exposure-to-outcome continuum; the discussion begins with assays and computational tools that are relevant for probing interactions of chemicals with cellular components and ends with ones that are relevant for predicting population-level responses. Understanding of pharmacokinetic relationships is critically important in toxicological evaluations for many reasons—for example, to evaluate whether exposures in in vitro cultures and in vivo assays are similar in magnitude and duration to exposures that result internally in exposed humans; to extrapolate from high to low dose, from one exposure route to another, and between species; and to characterize variability in internal human dose associated with a given exposure. Advances in pharmacokinetic analyses and models were discussed in Chapter 2 and are not elaborated on further

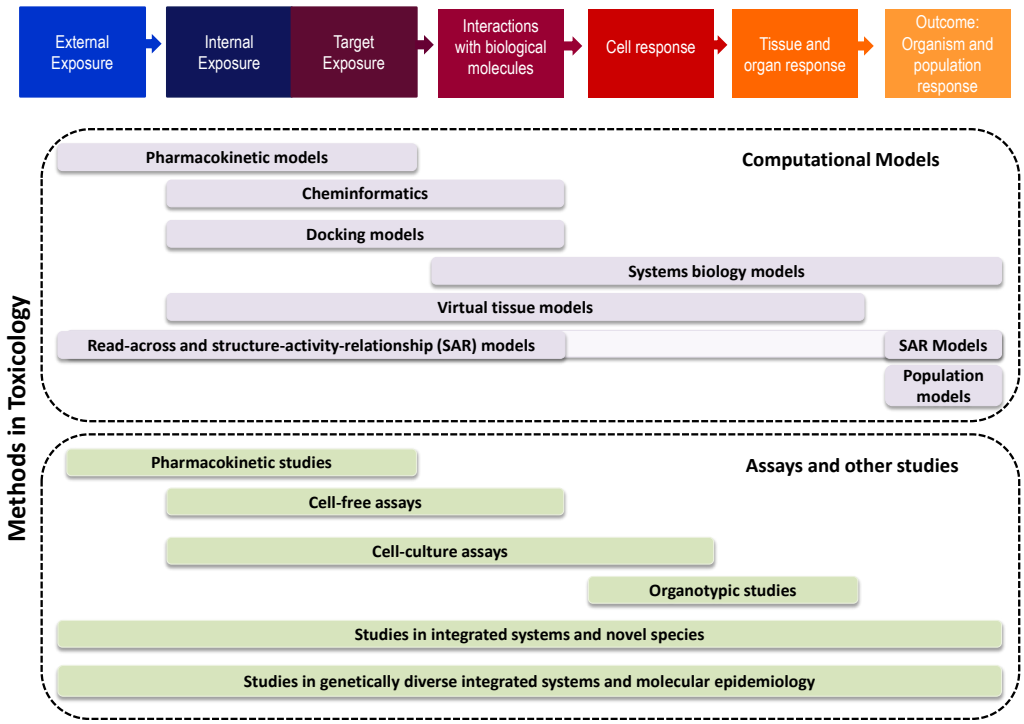



**FIGURE 3-1** Computational models and biological assays are shown with the exposure-to-outcome continuum to illustrate where the models and assays might be used to provide information at various points in the pathway. The clear portion of the bar for read-across and SAR models reflects the fact that connections are typically made between analogous chemicals for either the initial biological effect or the outcome. However, biological tools can also probe the response at the cell or tissue level and provide support for read-across and SAR analyses. If sufficient data are available, read-across and SAR analyses can be performed at various points along the exposure-to-outcome continuum.

here. The chapter concludes with a discussion of challenges and offers recommendations that should help to address the challenges.

The committee emphasizes that most Tox21 assays or systems were not developed with risk-assessment applications as an objective. Therefore, understanding on how best to apply them and interpret data in a toxicology context is evolving. For example, assay systems that were designed to detect agents that have high affinity for or potency against a particular biological target might not be optimized to detect agents that have moderate or low potency or that cause more than one effect. Some risk questions are being addressed as data from high-throughput systems become more available. However, the usefulness or applicability of various assays will need to be determined by continued data generation and critical analysis, and some assays that are highly effective for some purposes, such as pharmaceutical development, might not be as useful for risk assessment of commodity chemicals or environmental pollutants.

## PREDICTING AND PROBING INTERACTIONS OF CHEMICALS WITH CELLULAR COMPONENTS

Chemical interactions with specific receptors, enzymes, or other discrete proteins and nucleic acids and promiscuous interactions, such as those between an electrophile and a protein or DNA, have long been known to have adverse effects on biological systems (NRC 2000, 2007; Bowes et al. 2012). Accordingly, the development of in vitro assays that probe molecular-level interactions of chemicals with cellular components has been rapid, driven partly by the need to reduce high attrition rates in the drug-development process. Although various new assays have been developed, only a single assay—one that evaluates the human potassium channel (hERG channel[1]—has been integrated into new drug applications. Figure 3-2 illustrates some typical interactions with cellular components, and the following sections describe how the interactions are being investigated.

[1]The blockade of hERG channel has been directly implicated in prolongation of the QT interval, which is thought to play a role in the potentially fatal cardiac arrhythmia torsades de pointes.

### Predicting Interaction by Using Chemical Structure

In recent years, predicting chemical interactions with protein targets on the basis of chemical structure has become much more feasible, particularly with the development and availability of open-access data sources (Bento et al. 2014; Papadatos et al. 2015). There are many published examples of computational models that have been developed to predict the interaction of a molecule with a single protein, most notably models for predicting hERG activity (Braga et al. 2014) and interaction with the estrogen receptor (Ng et al. 2015), but prediction of multiple interactions in parallel is now possible given available computational power. For example, Bender et al. (2007) used chemical similarity to predict the protein–chemical interactions associated with a novel chemical structure with a reported average accuracy of over 92% with some proteins and high selectivity; that is, only small numbers of active predictions were later shown to be negative in vitro. Although most of the activities were predicted correctly, it was at the expense of a high false-positive rate (that is, large numbers of inactive chemicals were predicted to be active). Most of the models have been built by using pharmaceutical candidates that have a high affinity for the particular protein, but there are examples in the literature in which the same approaches have been applied to identify chemicals that bind a receptor with low affinity (see, for example, Hornung et al. 2014).

Research to improve the prediction of protein–chemical interactions continues apace. Lounkine et al. (2012) used the similarity-ensemble approach—a method first published by Keiser et al. (2007)—and predicted the activity of 656 marketed drugs with 73 protein targets that were thought to be associated with clinical adverse events. The authors reported that about 50% of the predictions of activity were later confirmed experimentally with binding affinities for the protein targets of 1 nM to 30 μM.

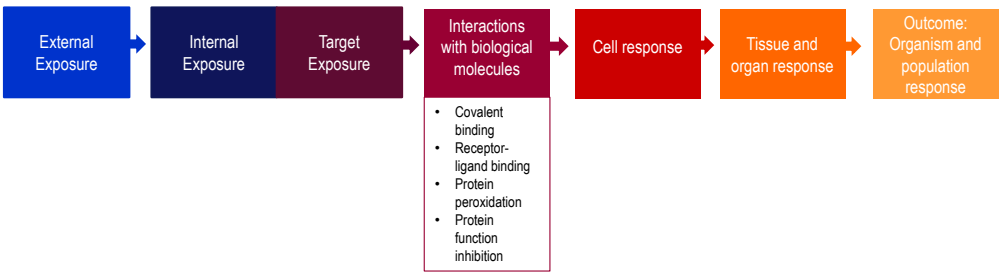



**FIGURE 3-2** Exposure-to-outcome continuum with examples of types of interactions between biological molecules and chemicals.

Cheng et al. (2012) evaluated chemical–protein interaction sets that were extracted from the ChEMBL database[2] by using a computational method, multitarget quantitative structure–activity relationship (QSAR), that evaluates G-protein coupled receptors (GPCRs) and kinase protein targets. Sensitivities were reported to range from 48% to 100% (average, 84.4%), and specificity for the GPCR models (about 99.9%) and the kinases was high.

### Assessing Interactions with Cell-Free Assays

Cell-free or biochemical assays have long been used to probe the interactions of chemicals with biological molecules, such as enzymes and hormone receptors, and their activity with these specific targets (Bhogal et al. 2005). The assays can provide reliable and valid results with high agreement between laboratories and can be applied in low-, medium-, or high-throughput formats (Zhang et al. 2012a).

The US Environmental Protection Agency (EPA) is exploring the use of the commercially available cell-free assays, run in high-throughput format, that were originally developed for preclinical drug evaluation to assess environmental chemicals (Sipes et al. 2013). The panel selected by EPA measures various activities, including binding to GPCRs, steroid-hormone and other nuclear receptors, ion channels, and transporters. The panel also covers activation of kinases, phosphatases, proteases, cytochrome P450, and histone deacetylases (Sipes et al. 2013). Roughly 70% of the assays are derived from human cells, 20% from rat cells, and the remainder from other species.

A wide variety of cell-free assays that evaluate other targets have been developed and are being used in pharmaceutical, biomedical, and academic laboratories (Xia et al. 2011; Mehta et al. 2012; Landry et al. 2015; McKinstry-Wu et al. 2015). They are being used to probe a wide array of protein types and functions, such as nod-like receptors, which are involved in immune and inflammatory responses (Harris et al. 2015), methyltransferases (Dong et al. 2015), and various membrane proteins (Wilcox et al. 2015).

The potency of the chemical's interaction in vitro—measured, for example, as an $IC_{50}$ or KI[3]—provides information on the likelihood of an in vivo concentration high enough to permit observation of the phenotypic response. The degree of inhibition or activation of the protein function that is required for a phenotypic response to be observed can vary widely and will depend partly on the nature and function of the protein or enzyme. For inhibitors of GPCRs, the anticipated pharmacological response has been observed in vivo at plasma concentrations less than or equal to three times the measured $IC_{50}$ of the chemical in question when corrected for plasma-protein binding (McGinnity et al. 2007). As a rule of thumb for pharmaceuticals, a 100-fold margin between the measured $IC_{50}$ or KI in a cell-free assay and the circulating plasma unbound $C_{max}$ has been considered adequate to represent minimal risk of toxicity (N. Greene, AstraZeneca, personal commun., December 14, 2015). However, for environmental chemicals, which are not tested in clinical trials or followed up through medical surveillance, a different rule of thumb might be appropriate. And it is important to remember that toxicity is influenced by many factors, including the required degree of receptor occupancy, the ability of the chemical to reach the site of action (for example, to penetrate the blood–brain barrier), the nature of the modulatory effects (for example, inhibitor, agonist, or allosteric modulator), the kinetics of the binding of the interaction with the receptor, and exposure duration.

## CELL RESPONSE

Cell-based in vitro assays have existed for nearly a century; the first publication of a dissociated cell culture was in 1916 (Rous and Jones 1916). Cell-culture technology has evolved to the point where many cell lines are available and more can be produced with current techniques. Cell cultures provide easy measurement of gene and protein expression and a variety of potentially adverse responses (see Figure 3-3) and can be scaled to a high-throughput format (Astashkina et al. 2012). Additionally, cell-based assays derived from genetically different populations can allow rapid assessment of some aspects of variability in response to chemical exposures that depend on genetic differences (Abdo et al. 2015).

Cell-based assays are being used to inform hazard identification and dose–response assessments, mostly as a complement to data from whole-animal or epidemiological studies to address questions of biological plausibility and mechanisms of toxicity. For example, in evaluations of chemical carcinogenicity, the International Agency for Research on Cancer (IARC) gives weight to functional changes at the cellular level (IARC 2006) and considers the relevance of the mechanistic evidence with regard to key characteristics of cancinogens (Smith et al. 2016). Cell-based assays have been critical in the IARC assessments (IARC 2015a,b). Human-derived and animal-derived cell cultures have also been used to discern dose–response relationships and toxicogenomic profiles, for example, for ethylene oxide responses (Godderis et al. 2012). The assays have potential use in addressing many

[2]ChEMBL is a chemical database of biologically active molecules that is maintained by the European Bioinformatics Institute of the European Molecular Biology Laboratory.

[3]$IC_{50}$ is the concentration required to cause 50% of the maximal inhibitory effect in the assay, and KI is the inhibition constant for a chemical and represents the equilibrium constant of the dissociation of the inhibitor-bound enzyme complex.

of the risk-based questions raised at the beginning of this chapter and as illustrated in Chapter 5.

Cell cultures can be grown in a variety of architectures, including monolayer and 3-D cultures of cell lines, and can be used as indicators of possible tissue, organ, and sometimes organism-level signs of possible toxicity, particularly in integrated systems that consider effects and signaling among cell types (Zhang et al. 2012a). They can be used to evaluate a number of cellular processes and responses, including receptor binding, gene activation, cell proliferation, mitochondrial dysfunction, morphological or phenotypic changes, cellular stress, genotoxicity, and cytotoxicity. Various techniques and measurements—such as impedance, gene transcription, direct staining, reporter-gene output, and fluorescence or bioluminescence resonance energy transfer—can be used to measure cellular responses and processes (An and Tolliday 2010; Song et al. 2011; Asphahani et al. 2012; Smith et al. 2012). Furthermore, simultaneous measurements of multiple toxic phenotypes are possible with high-content imaging and other novel techniques. This section describes some of the recent developments in using cell-based assays to evaluate cellular response and emphasizes advances that can improve toxicology and risk assessment.

The committee notes that cell-based assays have some limitations; one key concern involves metabolic capabilities. Specifically, do the assays capture how exogenous substances are metabolized in the body? That particular limitation might not be a concern for assays that are performed with low-throughput methods in which it might be possible to determine a priori whether metabolism is important for toxicity and, if so, to find ways to test the metabolites in addition to the parent chemicals. However, little or no metabolic capacity is a particular concern for high-throughput systems that are used for priority-setting. Parent chemicals and metabolites can differ substantially in toxicity and potency. If the in vitro assays do not sufficiently capture critical metabolites that form in humans, they might not give valid results for assessment because they are not testing the chemicals that potentially give rise to toxicity. Furthermore, although some assay systems might capture metabolism in the liver, extrahepatic metabolism might be the driver of some chemical toxicity, so the spectrum of relevant in vivo metabolic activation is an important consideration in understanding the validity of in vitro studies and interpreting the results from both in vitro and in vivo studies. EPA, the National Institute of Environmental Health Sciences, and the National Center for Advancing Translational Sciences are awarding research grants to make progress on the issue. For example, a multiagency collaborative announced in 2016 a $1 million competition in the Transform Tox Testing Challenge: Innovating for Metabolism; the challenge called on innovators to identify ways to incorporate metabolism into high-throughput screening assays (EPA/NIH/NCATS/NTP 2016). EPA is also attempting to develop a system that encapsulates microsomal fractions of human liver homogenate in a matrix, such as an alginate, that will allow diffusion of low-molecular-weight chemicals but retain the toxic lipid peroxides. As an alternative approach, EPA is attempting a method that would transfect cells with mRNAs of enzyme-encoding genes to increase metabolic transformation intracellularly. The committee views those initiatives as steps in the right direction and emphasizes the importance of addressing the issue of metabolic capacity.

### Primary Cells

Primary cells are isolated directly from fresh animal or human tissue. They can be obtained from a wide variety of tissues, such as liver, brain, skin, and kidney; and they are amenable to high-content screening and analysis (Xu et al. 2008; Zhang et al. 2011; Thon et al. 2012; Raoux et al. 2013; Tse et al. 2013; Valdivia et al. 2014; Feliu et al. 2015). Although primary cells are more reflective of in vivo cellular and tissue-specific characteristics than are immortalized cells (Bhogal et al. 2005), they can

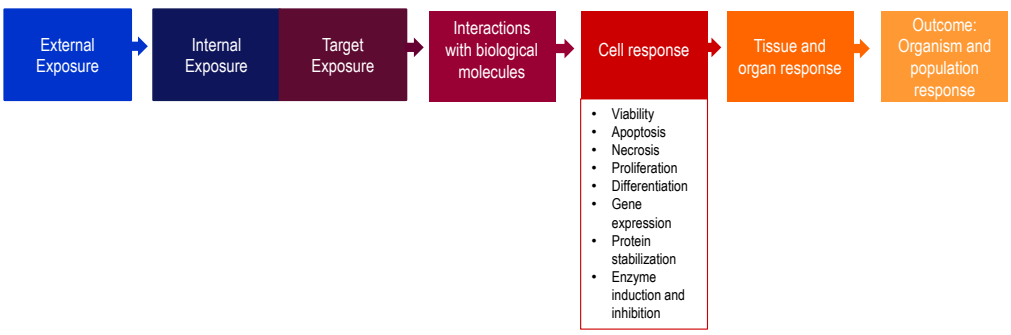



**FIGURE 3-3** The exposure-to-outcome continuum with examples of cell responses.

be short-lived in culture and suffer from rapid dedifferentiation within hours to days.

Several approaches to adapt primary cell culture to a high-throughput format for chemical-toxicity testing have been made (Sharma et al. 2012; Berg et al. 2015). For example, EPA profiled over 1,000 chemicals (Houck et al. 2009; Kleinstreuer et al. 2014) to identify activity in eight primary cell systems, including ones that used fibroblasts, keratinocytes, and endothelial, peripheral blood mononuclear, bronchial epithelial and coronary artery smooth muscle cells. With proprietary software, chemicals were clustered by bioactivity profiles, and some possible mechanisms of chemical toxicity were identified. The lack of publicly available datasets with which to compare the results and the complexity of the resulting data precluded sensitivity and specificity calculations (Kleinstreuer et al. 2014). The standard by which to judge construct validity—that is, whether an assay system as a whole adequately represents the target biological effect—still poses a challenge for these and other assays described in this chapter (see "Challenges and Recommendations for Advancing Toxicology" later in this chapter).

A major advance in primary cell culture over the last decade is the development of 3-D cultures of cell lines.[4] 3-D cell cultures have better behavior and function than the monolayer cultures (van Vliet 2011) and are of increasing interest in the development of cancer drugs because they recapitulate the tumor microenvironment to a much greater extent than do conventional monolayer assays that use a flat layer of cells (Edmondson et al. 2014; Lovitt et al. 2014). A number of assays that use 3-D cultures of primary cells from various tumors have been developed. Several studies (Arai et al. 2013; Chen et al. 2014) have shown some degree of drug resistance to well-characterized cancer drugs, depending on assay type; 3-D assays show greater drug resistance.

Similarly, primary isolated hepatocytes are the most widely used for in vitro testing, and 3-D culture systems with added cofactors are being developed to overcome limitations of conventional monolayer systems (Soldatow et al. 2013), which notably include lack of sensitivity for detection of hepatotoxic drugs. The 3-D cultures that are used, for example, for enzyme induction or inhibition studies, maintain function for a relatively long period (1–3 days) and can be used to re-establish cellular polarity that is lost in monolayer cultures. Advances in liver-culture techniques and technology have led to improvements and greater complexity in 3-D liver-cell culture for use in toxicological evaluations, and the next step is development of a bioartificial liver, commonly referred to as an organ-on-a-chip, discussed in greater detail in "Tissue-Level and Organ-Level Response."

[4] *3-D culture* is a generic term that is used to describe culture systems that are grown on some sort of support or scaffold, such as a hydrogel matrix. 3-D cultures often have two or more cell types.

The examples of tumor-cell and liver-cell cultures discussed in this section highlight the movement from monolayer cultures to improved 3-D cultures of greater complexity and ultimately toward organotypic models for various tissues and organs (Huh et al. 2011; Bulysheva et al. 2013; Guiro et al. 2015).

### Immortalized Cell Lines

Immortalized cell lines can be derived from isolated human cancer cells or from primary cells that have been genetically altered for enhanced longevity and resilience in tissue culture. Immortalized cell lines do not need to be isolated and harvested for each use, are relatively easy to maintain and propagate, are stable when replated multiple times, and can be easily frozen and shared between laboratories and grown in large quantities. Cloning immortalized cells enables testing in genetically identical cells, and immortalized cell lines that are derived from diverse populations allow inquiry into the variability of chemical toxicity among populations (Abdo et al. 2015). However, more than the conventional monolayer cultures of primary cells, immortalized cell lines can lose native in vivo properties and functionality. They can have altered cellular polarity (Prozialeck et al. 2003; Soldatow et al. 2013), non-native genetic content (Yamasaki et al. 2007), and decreased amounts of key cellular features (such as ligands, transporters, and mucin production); and they can be contaminated with other cell lines, such as HeLa and HepG2. Alterations in cellular phenotype can result in insensitivity to and mischaracterization of test chemicals. For example, when testing the difference between mitochondrial toxicity observed in renal proximal tubule cells (primary cells) and that observed in immortalized human renal cells, researchers found that primary cells were capable of identifying more possible toxicants than were immortalized cell lines (Wills et al. 2015).

Many of the assays in the federal government's ToxCast and Tox21 programs use immortalized carcinoma-derived cell lines (T47D breast, HepG2 liver, and HEK293T kidney). The assays have shown potential for identifying chemical carcinogens found in rodents (Kleinstreuer et al. 2014) and for exhibiting some predictive ability in the preliminary classification of hepatotoxic chemicals in guideline and guideline-like animal studies (Liu et al. 2015). However, the assays have also been shown to be unable to predict some well-recognized hazards observed in humans or animals (Silva et al. 2015; Pham et al. 2016).

ToxCast data have been proposed for use in predicting in vivo outcomes of regulatory importance (see Rotroff et al. 2013; Sipes et al. 2013; Browne et al. 2015), such as estrogenic properties of chemicals predicted by the uterotrophic assay, but their use as replacement assays has been the subject of research and discussion. For

example, EPA's Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Science Advisory Panel recommended that the agency not replace the uterotrophic assay with a computational model of estrogen receptor agonist and antagonist activity derived from ToxCast data (EPA 2014a). Although the panel noted a number of strengths of the model, it had concerns about diminished performance of the model for nonreference chemicals and the inability of the model to assess chemicals that had modified toxicity because of pharmacokinetic factors or that had toxicity pathways different from those evaluated in the assays. Thus, the panel found that further research was needed. More recently, EPA reconsidered the results of a high-throughput battery of estrogenicity assays, concluded that the test battery is a satisfactory replacement of the uterotrophic assay for tier 1 endocrine-disrupter screening, and intends to use the results of the test battery to evaluate and screen chemicals in the future (Browne et al. 2015; EPA 2015).

Because immortalized cell lines are limited in the degree to which they can represent cells in intact tissues, alternative approaches of cell immortalization have been developed and are now being made commercially available. "Conditionally immortalized" cell lines that can undergo differentiation are increasingly available for use in biomedical research with potential applications in toxicology (Liu et al. 2015).

## Stem Cells

Advances in stem-cell research have allowed the generation of a wide array of cell types, some of which have metabolic competence, which makes them useful for studying the effects of chemicals on various tissues (Scott et al. 2013; Gieseck et al. 2015). Fit-for-purpose stem-cell–based tests are becoming commercially available (Anson et al. 2011; Kolaja 2014), and research is under way to develop stem cells for application in toxicology (Sjogren et al. 2014; Romero et al. 2015). For example, an in vitro murine neural embryonic stem-cell test has been advanced as an alternative for a neurodevelopmental toxicity test (Theunissen et al. 2012; Tonk et al. 2013). The ability to grow rapidly, manipulate, and characterize an array of cell types makes stem cells potentially useful for chemical-toxicity evaluations. Furthermore, assays that use stem cells harvested from genetically diverse populations show considerable promise for providing information that can help in addressing hazard and risk-assessment questions.

Stem cells of potential use in toxicology research are of three primary types: embryonic, adult, and induced pluripotent stem cells. Embryonic stem cells are harvested from embryos that are less than 5 days old and have unlimited differentiation ability. Adult stem cells are isolated from adult bone marrow, skin, cord blood, heart tissue, and brain tissue. Induced pluripotent stem cells (iPSCs) are produced from adult somatic cells that are genetically transformed into a pluripotent state (Takahashi et al. 2007). iPSCs are similar to embryonic stem cells (pseudoembryonic) and can be grown in monolayer and 3-D structures for multiple generations. They can take on a variety of cell types, including neuronal cells (Efthymiou et al. 2014; Malik et al. 2014; Sirenko et al. 2014a; Wheeler et al. 2015), hepatocytes (Gieseck et al. 2014; Sirenko et al. 2014b; Mann 2015), and cardiomyocytes (Sinnecker et al. 2014; Karakikes et al. 2015). The ability to be derived from adult cells and the capacity to differentiate into multiple cell types also make iPSCs particularly promising for exploring human diversity. Cells can be created from specific individuals to produce personalized biomarkers, and iPSCs derived from large patient populations (Hossini et al. 2015; Mattis et al. 2015) could help to identify pathways involved in disease and susceptibility (Astashkina et al. 2012). Because iPSCs are relatively cost-effective to produce on a large scale (Beers et al. 2015), they have the potential to improve cell-based toxicity testing substantially.

There are some challenges to overcome in using stem cells. They can have different expression profiles, which indicate that they might have altered cellular processes, pathways, and functions. Stem cells generally can be difficult to culture and transfect, and the difficulties could limit their application in high-throughput formats. The lack of systematic approaches for characterizing and standardizing culture practices (such as characterizing cell types, sex origin, and cell function) also presents an obstacle for using stem cells in toxicology applications. Although stem cells (and other cells) have inherent limitations, they are still useful windows into biological processes at the cellular and molecular levels and remain useful for assessing chemical toxicity. A careful evaluation of cell phenotype and properties would help to determine the extent to which human biology is recapitulated in the cellular model.

## Modeling Cellular Response

Over the last decade, numerous mathematical models and systems-biology tools have been advanced to describe various aspects of cell function and response. Considerable progress has been made in describing feedback processes that control cell function. The development of cell-based modeling has benefited greatly from coordinated contributions from the fields of cell biology, molecular biology, biomedical engineering, and synthetic biology.

A few simple structural units that have specific functions and appear repeatedly in different species are referred to as network motifs (Milo et al. 2002; Alon 2007). Molecular circuits are built up from network motifs and carry out specific cellular functions, such as controlling

cell-cycle progression, xenobiotic metabolism, hormone function, and the activation of stress pathways—the major pathways by which cells attempt to maintain homeostasis in response to chemical and other stressors, such as oxidative stress, DNA damage, hypoxia, and inflammation. Computational models are used to examine those circuits, the consequences of their activation, and their dose–response characteristics.

Toxicity pathways defined in NRC (2007) as cellular-response pathways can be thought of as molecular circuits that, when sufficiently perturbed, lead to adverse effects or toxicity. The circuits can be modeled with computational systems-biology approaches. The tools for describing the circuits and function are developing rapidly (Tyson and Novak 2010; Zhang et al. 2010) and should enable study of the dose–response characteristics of the perturbation of toxicity pathways (Simmons et al. 2009; Zhang et al. 2014, 2015). Quantitative descriptions of the pathways hold the promise of characterizing differences in individual susceptibility to chemicals at the cellular level but will require identification of components of signaling pathways that differ among individuals; sensitivity and other analyses can be applied to determine components that most affect human variability in adverse response. Confidence in the models will increase as they are applied to a more diverse suite of signaling pathways. Model refinement coupled with careful collection of data on detailed biological responses to chemical exposure will test model structures, refine experimental strategies, and help to chart new approaches to understanding of the biological basis of cellular dose–response behaviors at low doses.

## TISSUE-LEVEL AND ORGAN-LEVEL RESPONSE

The last decade has seen advances in engineered 3-D models of tissue and computational models for simulating response at the tissue level (see Figure 3-4). This section describes organotypic models, organ-on-a-chip models, and virtual-tissue models that might be particularly applicable for toxicology research.

### Organotypic Models

An organotypic model is a specific type of 3-D culture in which two or more cell types are put together in an arrangement intended to mimic, at least in part, an in vivo tissue and that therefore recapitulates at least some of the physiological responses that the tissue or organ exhibits in vivo. Organotypic models of skin, which contain keratinocytes and fibroblasts, have been developed and validated for use as alternative models for testing skin irritation (Varani et al. 2007), and data from these models are now accepted in Europe for classification and labeling of topically applied products (Zuang et al. 2010). The skin model is being evaluated to improve the specificity of in vitro genotoxicity testing. Organotypic skin cultures appear to have reasonably good concordance with in vivo genotoxicity results (Pfuhler et al. 2014) probably because they retain the ability to metabolize and detoxify chemicals and because the rate of delivery of chemicals to the basal layer is more comparable with the kinetics of dermal absorption in vivo. Other organotypic models include eye, lung epithelium, liver and nervous system tissue (see NASEM 2015). The effects of environmental chemicals have been explored in mouse organoids by using proteomic tools (Williams et al. 2016).

### Organ-on-a-Chip Models

An emerging scientific development is the organ-on-a-chip model (see Figure 3-5), which is a 3-D culture grown in a multichannel microfluidic device (Esch et al. 2015). The models are meant to have the same functionality as organotypic cultures but with the ability to manipulate physiological and pharmacokinetic processes (that is, the rate at which a chemical is introduced via the flow-through channels). Several organ-on-a-chip models have been engineered, including ones for liver, heart, lung, intestine, and kidney. The models allow the study of how chemicals can disrupt an integrated biological system and how the disruption might be influenced by the mechanical

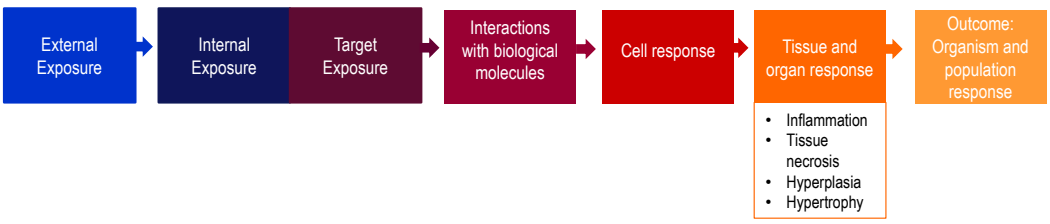



**FIGURE 3-4** Exposure-to-outcome continuum with examples of tissue and organ effects.

forces at play in the intact organ, such as the stretching of the alveolar-capillary barrier in lungs due to the act of breathing.

Attempts have been made to design platforms that have different organ mimics arranged in series or parallel that as a system can recapitulate aspects of tissue interactions and in vivo pharmacokinetics (Sung and Shuler 2010). A long-term goal is to introduce a parent chemical into the system and have it move through a liver compartment where it would be metabolized, flow to compartments that contain responsive cell types or to other compartments that contain hydrophobic materials that represent fat, and finally flow through a kidney compartment where it could be eliminated. To date, microfluidic platforms that have that much complexity have not yet been introduced in practice and have not achieved a realistic metabolite distribution through the various tissues in the system (Andersen et al. 2014).

Researchers face challenges in developing such experimental platforms, for example, with the synthetic materials used in the manufacture of the cell-culture substrates. They often are not good mimics of the extracellular matrix and can even absorb small hydrophobic molecules (Wang et al. 2012); that absorption might exert an undue influence on the physiological system or alter chemical concentrations. Large-scale manufacture and high-throughput operation of organ chips also present challenges to the adoption of the technology. Similarly, access to sustainable sources of human cells presents a substantial hurdle for reproducibility and interpretation of the data produced.

Microsystems that are composed of multiple synthetic organ compartments are in the early stages of development, and a number of initiatives are going on to validate model correlations with in vivo observations. For example, the National Center for Advancing Translational Sciences has a number of efforts in this field (NCATS 2016), and the European Union–funded initiative Mechanism Based Integrated Systems for the Prediction of Drug Induced Liver Injury (EU 2015) has also been exploring the use of liver-chip models to predict adverse effects of drugs. Organ-on-a-chip models are promising, but they are not yet ready for inclusion in risk assessments.

## Virtual Tissues

As discussed earlier, computational systems biology might be used to describe pathway perturbations that are caused by chemical exposures and the resulting cell responses. Such modeling can be applied to multiple processes that operate in sequence or parallel and used to link cellular responses to tissue-level responses. Modeling feedforward and feedback controls through sequential dose-dependent steps also enable the examination of responses to toxicant exposure that require multiple cell types, such as Kupffer cell–hepatocyte interactions

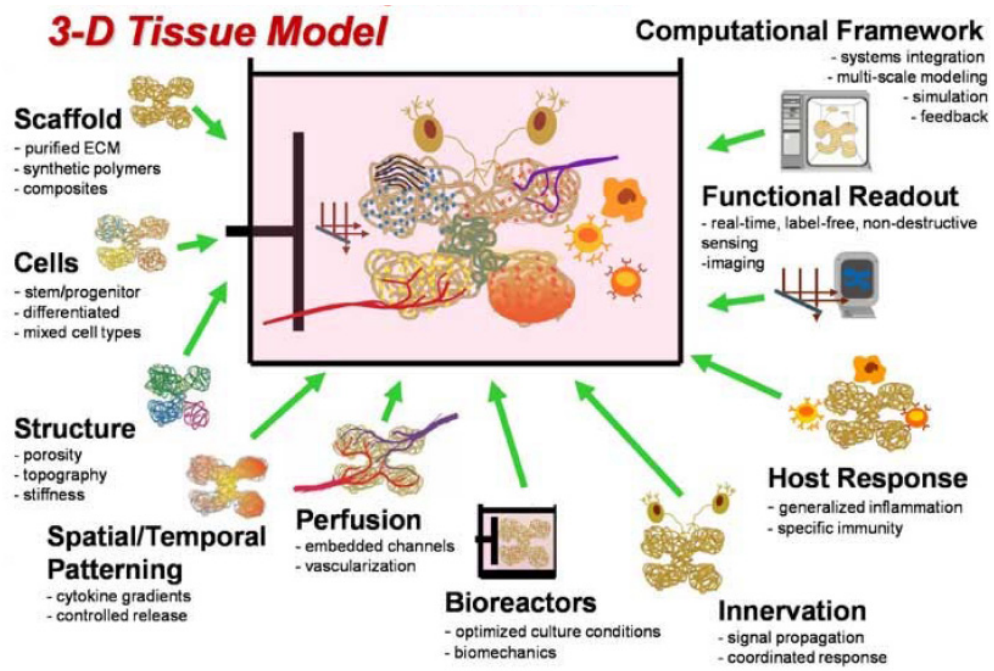



**FIGURE 3-5** Generalized components of an organ-on-a-chip model. Source: Birnbaum 2011.

involved in hepatocyte proliferation. Feedback and feed-forward control might also contribute to intercellular patterns of response that require input from earlier pathway or cellular functions to activate or inhibit integrated multicellular responses. The cellular responses alter tissue function; the quantitative modeling then focuses on the interface between the cellular-level computation models and virtual-tissue models.

EPA's Computational Toxicology Program has developed mathematical models called virtual tissues for the embryo and the liver (Shah and Wambaugh 2010; Wambaugh and Shah 2010). EPA also has developed a model of blood-vessel development. Virtual-tissue models can use "agent-based" modeling of different cells in the tissue, which relies on and mathematically describes key aspects of cellular behavior or other tissue components to derive the properties of the tissue or organ of interest (Swat and Glazier 2013). The EPA models evaluate chemical exposures that alter growth and phenotypic characteristics of the agents in the models, which in this case are the cells. The models can describe cell growth or pattern formation of different structures in the virtual embryo or regional distribution of cell response in the virtual liver.

As with any model, a critical consideration in developing response models is fidelity of biology between the modeled outcome (virtual-tissue responses) and the apical and other responses observed experimentally. Assumptions and predictions of the models can be tested by using information from human cells and co-cultures with different human cell types. Short-term targeted animal studies that use toxicogenomic tools and other approaches can be used to evaluate the model more broadly. Virtual-tissue models have the potential to help in conceptualizing and integrating current knowledge about the factors that affect key pathways and the degree to which pathways must be perturbed to activate early and intermediates responses in human tissues and, when more fully developed, to support risk assessments based on studies of key events and how the key events combine to cause adverse responses at the organism level.

## ORGANISM-LEVEL AND POPULATION-LEVEL RESPONSE

The Tox21 report (NRC 2007) emphasized a future in which routine toxicity testing would rely on in vitro assays with human cells or assays that probe molecular responses of human toxicity pathways and pathway components. But, the report also noted that in some cases testing in whole animals might be necessary, depending on the nature of the risk-assessment questions, although whole-animal studies were not intended to provide routine information for assessing risks. The need for different types of information related to the nature of the question posed was also emphasized in EPA's report on next-generation risk assessment (EPA 2014b; Krewski et al. 2014; Cote et al. 2016). That report considered three types of assessments: screening and priority-setting assessments, limited-scope assessments, and in-depth assessments. The last one would likely involve a wide array of toxicity-testing approaches, including whole-animal studies. Approaches for assessing variability could also benefit from rodent panels that capture population variability and panels of human cells derived from a group of diverse people. As is true of toxicity-testing tools at the molecular and cellular levels, there has been continuing development of new methods for examining responses in whole animals that are likely to provide important information for the limited-scope and especially for the in-depth assessments. The approaches for assessments on different levels emphasize a fit-for-purpose orientation of designing the testing assays or batteries that depend on the risk-assessment question. This section discusses novel animal models that provide opportunities for enhancing the utility and power of whole-animal testing. It also describes recent advances in structure-based computational models and read-across approaches that provide opportunities for predicting response of data-poor chemicals at the organism level. Figure 3-6 highlights some organism-level and population-level responses.

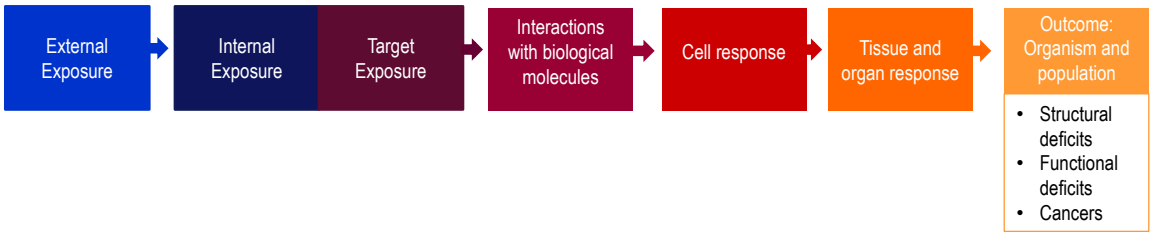



**FIGURE 3-6** Exposure-to-outcome continuum with examples of organism and population-level responses.

## Novel Whole-Animal Models

Advances in genetics, genomics, and model-organism development have led to genetically well-characterized whole-animal models, including transgenic rodent lines, isogenic mouse strains, and alternative species, such as zebrafish and *Caenorhabditis elegans*, which can be studied in a high-throughput format. Those models coupled with toxicogenomics and novel imaging offer improvements over the traditional in-life rodent studies in that they offer new ways to explore chemical interactions at tissue and cellular levels. Isogenic strains also offer new opportunities to identify determinants of human susceptibility, especially when coupled with new interrogation tools, and to define new mechanisms of toxicity. Targeted testing, which is typically hypothesis-driven and more focused than historical testing strategies, can help to develop and enhance the value of the new animal models, as well as traditional ones. It can be used to explore the mechanisms by which a chemical causes toxicity, how outcomes might differ by age and sex, and how susceptibility might vary in the population. It can help to address specific knowledge gaps in risk assessment and can link in vitro observations to molecular, cellular, or physiological effects in the whole animal. Targeted testing will be critically important in evaluating and validating the robustness and reliability of new computational models, in vitro assays, and testing batteries (Andersen and Krewski 2009; Krewski et al. 2009). As this section shows, the new animal models and outcome-interrogation tools might provide broader assessment of hazards in whole organisms.

### Transgenic Rodents

The development of transgenic mouse lines (such as knockin, knockout, conditional knockout, reporter, and humanized lines) advanced biomedical research; a few transgenic rat lines are also available now. Novel gene-editing technologies, such as CRISPR/Cas9, have the potential to generate inducible gene editing in adult animals and the creation of transgenic lines in nontraditional mammalian models (Dow et al. 2015). Gene editing permits the creation of experimental approaches that are more specifically suited for various tasks, including targeted testing of susceptible strains and exploration of gene–environment interactions.

Although transgenic animals have been available for decades (Lovik 1997; Boverhof et al. 2011), testing in transgenic animals and incorporation of data from transgenic models into risk assessment has been limited, partly because of questions about applicability for risk assessment and concerns about the cost to develop the models and evaluate a chemical in multiple strains. The National Toxicology Program (NTP) continues to evaluate and develop such models. For example, NTP is using transgenic mice in the testing of the artificial sweetener aspartame, which generally tested negative in standard assays but showed a slight increase in brain tumors in a more sensitive transgenic-mouse strain. The transgenic p16 model was used because it was thought to be susceptible to brain glial-cell tumors. NTP is also testing aspartame in transgenic strains with knocked-out tumor-suppressor genes and activated oncogenes to improve characterization of susceptibility and risk related to gene–environment interactions. Transgenic-rodent mutation data have been used by EPA to understand carcinogenic mechanisms of several agents, such as acrylamide (EPA 2010), but beyond those applications their incorporation into risk assessment has been limited. They have been somewhat more widely used to test specific hypotheses about mechanism, such as the mechanism of liver-cancer induction by phthalates (Guyton et al. 2009), and to evaluate the depth of biological understanding to apply fully organotypic, computational systems-biology, physiologically based pharmacokinetic (PBPK), or other tools.

### Genetically Diverse Rodents

Historically, toxicity testing has used only a few rodent species and strains. Although there are advantages in using a well-characterized strain of mice or rats to test chemical toxicity, there are many shortcomings, including concerns about inadequately accounting for profound strain differences in chemical sensitivity and metabolism (Kacew and Festing 1996; Pohjanvirta et al. 1999; De Vooght et al. 2010) and inadequate genetic and phenotypic diversity. High rates of spontaneous disease in some strains (outbred and inbred) can sometimes complicate the interpretation of results. For example, the incidence of background cardiomyopathy in the Sprague Dawley rat can be as high as 100% (Chanut et al. 2013), some strains are completely resistant to some toxicants (Shirai et al. 1990; Pohjanvirta et al. 1999), and it is unclear a priori whether the standard strain has sensitivity that is adequate or too high for identifying a potential human hazard.

Assessment in multiple strains that have known genetic backgrounds is one approach to address variable sensitivity among relatively homogeneous test strains and to address questions related to interindividual sensitivity to toxicants. Initiated in 2005, the Collaborative Cross (CC) is a large panel of novel recombinant mouse strains created from an eight-way cross of founder strains that include three wild-derived strains. The CC has a level of genetic variation akin to that of humans and captures nearly 90% of the known variation in laboratory mice (Churchill et al. 2004). Outbred progeny that have completely reproducible genomes can be produced through the generation of recombinant inbred intercrosses (RIX) (Zou et al. 2005). Because the CC strains and, by extension, the RIX lines have a population structure that randomizes exist-

ing genetic variation, these models provide the increased power that is required to explore the genetic underpinnings of interindividual susceptibility. For example, the CC mouse replicated human susceptibility, immunity, and outcome of West Nile virus infection more comprehensively than the standard inbred model (C57BL/6J) (Graham et al. 2015).

There are several examples of the value of the CC in toxicological evaluation. Trichloroethylene (TCE) metabolism, for example, varies considerably among people and among mouse strains, and the metabolites differ in their mechanisms, toxicity, and organ-specific effects (NRC 2006). That variability has been a critical barrier to understanding of the risk that TCE poses to humans. To address the challenge in TCE-toxicity testing, a battery of mouse lines was used to assess interindividual variability in TCE metabolism and toxicity in the liver and kidney (Bradford et al. 2011; Yoo et al. 2015a,b). Significant differences in toxicity and metabolism were observed in the different strains. Population PBPK modeling was applied to the study results to illustrate how data on diverse mouse strains can provide insight into pharmacokinetic variability in the human population (Chiu et al. 2013).

Multistrain approaches have also revealed fundamental mechanisms of hepatotoxicity of acetaminophen and biomarkers of this potentially fatal effect. Harrill et al. (2009) used a panel of 36 inbred mouse strains and found that liver injury induced by acetaminophen was associated with polymorphisms in four genes, but susceptibility to hepatotoxicity was associated with yet another, CD44. Follow-up study of two healthy human cohorts showed that variation in the human CD44 gene conferred susceptibility to acetaminophen liver toxicity. This powerful example shows how a diverse animal population (in this case, mice) can be used to characterize and identify potential susceptibility in humans.

The Diversity Outbred (DO) population is a heterogeneous stock seeded in 2009 from 144 independent lineages from the CC breeding colony. Each DO mouse is unique and has a high level of allelic heterozygosity (Churchill et al. 2012). Because they were derived from the same eight strains as the CC mice, their genome can be reconstructed with a high degree of precision—a feature that facilitates genome-wide association studies and other similar approaches. A 2015 NTP proof-of-concept study that used DO mice to capture variation in benzene susceptibility successfully identified two sulfotransferases that modify and eliminate benzene metabolites that confer resistance to benzene toxicity (French et al. 2015).

One caveat in using genetically diverse rodent models is that their use potentially can increase animal use. The most effective use of such models in toxicology requires acceptance of novel computational approaches, experimental designs, and statistical approaches that are specifically suited for the models and capable of handling the unprecedented amount of data that these studies generate (Festing 2010). For example, factorial designs can maximize genetic diversity and reduce the risk of false negatives without necessarily requiring more animals than traditional rodent studies to address the central question. Additionally, using DO mice requires accepting that each individual is unique and that there is no way to incorporate "biological replicates" in the traditional sense. Researchers and risk assessors need to be aware of and comfortable with the suite of data that results from these studies and to understand how to integrate the data with information from other sources, including more traditional animal models (see Chapter 7). Computational tools uniquely suited for these emerging animal models are available and readily adaptable to toxicological testing (Zhang et al. 2012b; Morgan and Welsh 2015). Tools for data analysis, visualization, and dissemination are also available (Morgan and Welsh 2015). As with any model system, these rodent models should be used only for questions that they are best suited to address. NTP and other groups are developing frameworks and use cases to highlight when it is advantageous to use such models, and the committee supports further discussion on this issue.

### Other Whole-Animal Systems

Advances in genomics, imaging, and instrumentation have made some alternative species—such as *Caenorhabditis elegans* (a nematode), *Drosophila melanogaster* (a fruit fly), and *Danio rerio* (the zebrafish)—useful animal models for hazard identification and pathway discovery. Many technical advantages are shared among the three dominant nonmammalian species, but zebrafish have several useful characteristics not shared by the others. The genomes of zebrafish and humans display remarkable homology with an overall conservation of over 70%. Furthermore, 80% of the genes known to be involved in human disease are expressed in zebrafish (Howe et al. 2013b). The signal-transduction mechanisms, anatomy, and physiology of zebrafish are homologous to those of humans (Dooley and Zon 2000), and zebrafish have all the classical sensory pathways, which are generally homologous to those of humans (Moorman 2001; Colley et al. 2007).

Another important attribute that might make zebrafish particularly well suited for translational research is the capacity to generate transgenic reporter lines that express fluorescent genes in specific cells, tissues and organs. The large collection of transgenic fish lines are curated by the Zebrafish Model Organism Database and maintained by the Zebrafish International Information Network (Howe et al. 2013a). There is also a rich diversity of zebrafish-disease models and drug screens to help to understand, prevent, and develop therapies for human diseases, including various cancers (Feitsma and Cuppen

2008; Nguyen et al. 2012; Gallardo et al. 2015; Gordon et al. 2015), diabetes and obesity (Gut et al. 2013; Dalgin and Prince 2015; Schlegel and Gut 2015), psychiatric conditions (Panula et al. 2010; Norton 2013; Jones and Norton 2015), heart disease (Arnaout et al. 2007; Chico et al. 2008; Arnaout et al. 2014; Asnani and Peterson 2014; Walcott and Peterson 2014), neurodegenerative syndromes (Bretaud et al. 2004; Chapman et al. 2013; Mahmood et al. 2013; Da Costa et al. 2014; Martin-Jimenez et al. 2015; Preston and Macklin 2015), autism (Tropepe and Sive 2003), immunodeficiencies (Meeker and Trede 2008; Cui et al. 2011), and blood disorders (Ablain and Zon 2013). Zebrafish have been used to investigate neurotoxicants (Levin et al. 2007; Egan et al. 2009; Irons et al. 2010), and Box 3-1 provides an example of using zebrafish for behavioral assessments.

The Zebrafish Mutation Project hosted by the Sanger Institute is yet another major effort that will facilitate cross-species studies. The project aims to develop a knockout allele in every protein-coding gene in the zebrafish genome and characterize its morphological phenotype (Kettleborough et al. 2013). Mining of zebrafish gene or phenotype databases should provide powerful opportunities to identify genes involved in chemical-induced phenotypes.

An additional advantage of zebrafish is that the zebrafish genome is fully annotated, so transcriptomic and all other -omics approaches are possible. Repression of gene expression by antisense morpholinos, siRNA, and such gene-editing techniques as CRISPR/Cas9 is routinely used to assess gene functions in the intact fish, and zebrafish embryos and larvae are nearly transparent, so noninvasive observation is possible. Because larvae measure less than a few millimeters, they can be accommodated in multiwell plates, such as 384-well formats (Rennekamp and Peterson 2015). Only small quantities of test chemicals are needed, so exposure–response relationships can be evaluated over a broad concentration range and testing can be replicated to increase data confidence.

Although substantial research is going on with adult zebrafish for translational research (Phillips and Westerfield 2014; Pickart and Klee 2014), early zebrafish life stages are particularly well suited for rapid screening. During the first 5 days of life, nearly all gene products and signal-transduction pathways are expressed (Pauli et al. 2012); thus, as in other vertebrates, development is a period of heightened sensitivity to chemical exposure. Early–life-stage zebrafish also express a full battery of phase I and phase II metabolism systems, whose activities are highly similar to those of humans (Goldstone et al. 2010).

**BOX 3-1** Using Zebrafish to Assess Behavior

A limitation of current in vitro screening is the general paucity of assay coverage to identify neurotoxic chemicals reliably. Observations of zebrafish embryonic and larval photomotor responses provide robust measures of nervous-system deficits based on well-established methods. For example, 18–24 hours after fertilization (embryo stage), the photomotor response is measured as tail flexions before and after a bright-light impulse. That assay has proved to be a highly sensitive chemical-toxicity screening tool (Kokel et al. 2010; Reif et al. 2016). At 5 days after fertilization (larval stage), the photomotor response can be assessed as a change in swimming activity in response to a sudden light–dark transition. Both tasks can be digitally measured in individual wells, so these complex behavioral assays are highly amenable to high-throughput analysis (Padilla et al. 2012; Truong et al. 2014). The adult zebrafish is increasingly used to measure neurobiological end points affected by chemical exposures. An array of behavioral tests have been designed to probe different domains involved in sensorimotor systems, cognition, and responses related to learning, memory, and anxiety. Indeed, zebrafish adults and juveniles display a variety of complex behaviors, such as kin recognition (Mann et al. 2003; Gerlach et al. 2008), shoaling and schooling (Engeszer et al. 2007; Miller and Gerlai 2012), territoriality (Spence and Smith 2005), associative learning (Al-Imari and Gerlai 2008; Fernandes et al. 2014), and nonassociative responses, such as habituation (Best et al. 2008). A number of neurobehavioral tests of anxiety and exploration have been modeled, and there is some evidence of conserved responses that resemble those of rodent models (Panula et al. 2006; Egan et al. 2009; Champagne et al. 2010; Steenbergen et al. 2011). Startle tests have been developed to understand sensorimotor responses in zebrafish exposed to environmental chemicals. Those assays have been used to test chemical effects on zebrafish motor responses, including responses related to fluorinated organics (Chen et al. 2013), vitamin E deficiencies (Lebold et al. 2013), nanoparticles (Truong et al. 2012), and pesticides (Sledge et al. 2011; Crosby et al. 2015). Collectively, the sophisticated assays could be scaled to increase the throughput with which chemicals are assessed for their effects on the nervous system.

Despite the advantages of incorporating the use of early–life-stage zebrafish as part of a strategy for making risk-based decisions, there are some noteworthy limitations. First, test chemicals typically are added directly to the aqueous media, not unlike cells in culture. However, the routes of exposure over the course of development, which can affect chemical uptake and metabolism, can be quite different. During the first 2 days of embryonic development, the primary route of exposure is passive dermal adsorption. Later in development, the gills and oral routes become available, and circulation plays a major role in chemical distribution. For the varied routes of exposure, there is little understanding of tissue concentrations, and this contributes to the challenges in comparing concentration–response results in zebrafish with dose–response studies in other systems directly.

A related potential limitation is that despite metabolic similarities to other vertebrates, subtle differences in metabolic activity could lead to inaccurate toxicity predictions, particularly if metabolic activation or inactivation is mechanistically important for specific test chemicals. Because the developing embryo constitutes a comprehensive integrated system, all potential molecular initiating events are operational during testing. Thus, zebrafish are uniquely sensitive to chemical contaminants present in test solutions in that a contaminant could act on biological targets and disrupt critical molecular events. Finally, as with any animal model, the primary sequences of individual pathway components are not necessarily highly conserved. For example, the zebrafish cyclin-dependent kinase 20 (cdc20) protein is 75% identical with the human protein at the amino acid level, and the zebrafish and human aryl hydrocarbon receptors are only 40% identical. In both cases, the homologous proteins are functionally conserved. Although variable conservation of the genomes is a source for potential discordance between zebrafish and humans, the challenge is not unique to zebrafish inasmuch as individual allelic variations between humans can also result in marked differences in chemical susceptibility.

## Computational Structure-Based Models for Predicting Organism-Level Response

It has long been recognized that chemicals that have similar chemical structures can elicit the same or similar toxicological effects and that, paradoxically, almost identical chemicals can cause dissimilar biological responses. The extent to which similar chemicals or their metabolites interact with critical biological molecules, such as target proteins, and operate by similar mechanisms is a critical element in determining structure–activity relationships. The last decade has seen advances in the development of structure-based computational methods to predict human health effects. Some are computational expert systems that consider structural alerts and underlying mechanisms, others are QSAR models that rely on statistical correlations with molecular fragments, and still others are hybrids of these. Many advances have been supported by large curated databases and increased computational power. Health effects addressed include carcinogenicity (Contrera et al. 2005; Valerio et al. 2007), hepatotoxicity (Greene et al. 2010; Hewitt et al. 2013), reproductive and developmental effects (Matthews et al. 2007; Wu et al. 2013), and skin sensitization (Roberts et al. 2007a,b; Alves et al. 2015).

The structure-based computational models that are probably the most advanced in model performance and regulatory acceptance are QSAR models for genotoxicity or more specifically for mutagenicity as measured in the Ames assay, a reverse-mutation bacterial assay that is commonly used to evaluate the potential of chemicals to induce point mutations. The development of those models has benefited from the quantity and structural diversity of data available in the public domain on chemicals that have been tested in the Ames assay. As a result of performance, computational models are being accepted as surrogates for actual testing and have recently been incorporated into international guidelines for assessing mutagenic impurities in pharmaceuticals to limit potential carcinogenic risk (ICH 2014). Computational approaches for other human health effects are being considered for use in a regulatory setting (Kruhlak et al. 2012), and the Organisation for Economic Co-operation and Development has published guidance that outlines the needed components of a QSAR model in regulatory settings (OECD 2004). They include "a defined end point; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodness of fit, robustness, and predictivity; and, if possible, a mechanistic interpretation" (Gavaghan 2007).

The lack of wide use of QSAR models for end points other than mutagenicity might reflect predictive performance that falls short of that required for practical applications. Most approaches predict only whether a chemical will cause the adverse effect. The inability to predict a plasma concentration that would be expected to elicit toxicity ultimately limits utility for differentiating between closely related structures on which little or no safety information is available for comparison.

## Read-Across Predictions

Read-across is a process that uses two-dimensional chemical-structure information to identify chemicals (analogues) that have been well studied toxicologically that are then used to predict the toxicity of a similar chemical that has inadequate toxicological data or to group chemicals for the purpose of evaluating their toxicity collectively. Structural similarity can be determined by atom-by-atom matching that results, for example, in a chemical-similarity score or by identifying core molecular structures or

functional groups that are thought to be important in conferring toxicity potential. There should also be a consideration of physicochemical similarity among analogues because significant differences in, for example, partition coefficients (such as $\log K_{OW}$, a measure of lipophilicity) will have important effects on pharmacokinetic and often pharmacodynamic behavior of a chemical. Read-across approaches are receiving much attention because they can help to satisfy the information requirements under European Union Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulations; the general concept has been accepted by the European Chemicals Agency (ECHA) and member-state authorities (Patlewicz et al. 2013). When robust toxicological data are available on one or more structurally related chemicals, they can be used to infer the activity of a chemical that has not been adequately tested. ECHA (2015) has recently published a framework by which it evaluates read-across submissions under REACH. ECHA's framework groups the read-across into six categories according to such factors as whether the read-across is for a single analogue or an entire category, whether it is based on metabolism to a common product, and the relative potencies of members of a chemical series.

Phthalate esters provide a well-studied example of the utility of read-across for male reproductive toxicity. Phthalate esters that have chain lengths of four to six carbons (more if branched) cause testicular toxicity (Foster et al. 1980) and adverse effects on male reproductive-system development (Gray et al. 2000; NRC 2008) in rats. Studies of global gene expression in the fetal rat testis show comparable effects of all the developmentally toxic phthalates (Liu et al. 2005) and support a conclusion that these chemicals act via the same mechanism. Phthalate esters with shorter chains, such as dimethyl and diethyl phthalate, do not produce similar effects on gene expression or on testicular function or male reproductive-system development. Thus, well-studied phthalate esters in this group would serve as anchor chemicals for other phthalates that have chains of four to six carbons in a read-across approach.

Read-across can be problematic, and caution is needed before its conclusions are relied on heavily. For example, thalidomide has two stereoisomers, (*S*)-thalidomide and (*R*)-thalidomide, that are virtually identical from a structural perspective in all aspects except for the 3-D orientation of the two ring systems in relation to one another (see Figure 3-7). Their physical characteristics are also identical, so read-across analysis might conclude that the chemicals will have similar or identical safety profiles. However, (*S*)-thalidomide causes birth defects, embryo death or altered development, growth retardation, and functional defects, whereas (*R*)-thalidomide does not. Still, the enantiomers are capable of interconverting in vivo, so it is impossible to eliminate the teratogenic effects by administering only the (*R*)-enantiomer.

Despite the limitations, read-across remains a screening approach for assessing the safety of a molecule in the absence of data on which to base an assessment. The 2015 ECHA framework provides guidance on how protein binding, metabolism, and other data can be used in read-across analyses and potentially overcome the limitations. Furthermore, a recent European study team proposed evaluation of read-across for four basic chemical-group scenarios (Berggren et al. 2015): chemicals that do not undergo metabolism to exert toxicity, that exert their toxicity through the same or structurally similar metabolites, that have low toxicity, or that are structurally similar but have variable toxicity on the basis of their hypothesized mechanism. They have selected chemical groups for case studies in each of the four categories.

Low et al. (2013) extended the concept of similarity in read-across from chemical structure to bioactivity, specifically responses in a variety of in vitro and genomic assays. They proposed a hazard classification and visualization method that draws on both chemical structure and biological features to establish similarity among chemicals in read-across. The approach incorporates mechanistic data to increase the confidence of read-across.

In addition to serving as a screening approach, read-across can be regarded as a hypothesis-generating exercise. The hypotheses can be lumped into two broad cat-

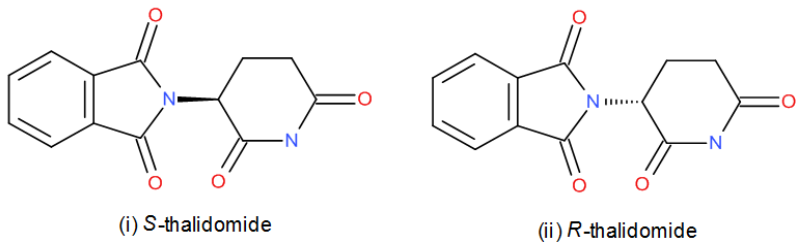



**FIGURE 3-7** Molecular structures of (*S*)-thalidomide and (*R*)-thalidomide.

egories: the new chemical is metabolized to a chemical that has already been tested (or it and its analogue are metabolized to the same chemical), or the new chemical and its analogues are sufficiently similar in chemical structure and properties that their biological activity is the same (that is, they have the same mechanism). In the former case, there are long-standing methods for assessing chemical metabolism that can be applied to support or refute the hypothesis that the new chemical is metabolized to something that has already been tested. In the latter case, if the mechanism of the analogous chemicals is known, it is reasonably straightforward to test for effects on the initial events of the mechanism (for example, receptor occupancy or enzyme inhibition). In most cases, however, mechanisms are not known; in such cases, it is still possible to compare the responses of the chemical and its analogues in screening systems that globally assess toxicological responses. Global gene-expression analysis is likely to provide universal coverage of possible mechanisms. Gene expression in an animal model in which the target tissues (for the tested analogues) are known or in an in vitro system that represents the target tissue is a reasonable way to test the hypothesis of a comparable mechanism among analogues. It still might be possible to use gene expression in in vitro models to identify a mechanism when target tissue is not known, but it will probably require testing in more than one cell type. Lamb et al. (2006) evaluated the gene-expression changes elicited in four cell types by a large number of drugs; they clearly showed the connections between agents that have the same pharmacological action and demonstrated that this approach has high potential for toxicology. High-throughput screening batteries, such as ToxCast, might also have utility for that purpose, but it will need to be determined whether the current battery covers the universe of known toxicity mechanisms. Higher-order models, such as organ-on-a-chip or zebrafish, might also be used for testing hypotheses of biological similarity if it can be shown that these models have the biological machinery that is critical for the mechanism in question. As data streams are added more systematically to the read-across process, integrated approaches, such as Bayesian models, that provide for a more agnostic evaluation and promote consistency in output could be developed. Figure 3-8 illustrates several scenarios for read-across and how it can be used to infer hazard and dose–response relationships.

## INCORPORATING DATA STREAMS

Various chemicals will have multiple data streams along the exposure-to-outcome continuum that can be used to characterize hazard or risk. For example, pharmacokinetic studies might point to tissues that have particularly high concentrations of a chemical that are potentially increased by active transport as indicated in in vitro studies. Cell-free assays might suggest a set of key receptors, with cell-response assays indicating response; the results, when considered in the context of high concentrations of a chemical in tissues, might indicate particular hazards, such as particular cancers or reproductive toxicity. Targeted studies might show early markers of effect histopathologically, and gene expression in the studies might show consistency with the findings of cell-based assays. The results might be supported by findings on similar chemicals that predict the activity through structure–activity analyses. Robust assessments will identify the more influential data streams with which to develop an integrated assessment. Some streams will be more information-rich than others. The integration of multiple data streams is discussed further in Chapter 7.

## CHALLENGES AND RECOMMENDATIONS FOR ADVANCING TOXICOLOGY

This chapter shows how emerging scientific tools generate toxicological evidence on hazard and dose–response relationships of chemicals and other risk issues. It emphasizes how the tools apply to different components in the exposure-to-outcome continuum. Some tools, such as PBPK and systems-biology models, provide a basis for linking components along the continuum. Others, such as high-throughput assays or targeted testing, provide a direct readout of chemical effect within a single component or in multiple components. The tools vary in their maturity for application, their scope of applicability among chemical classes, and the questions that they can address. The committee emphasizes that the level of performance required for the various tools will depend on the question that is being addressed (context) and on agency policies.

There are specific technical and research challenges. Some have been mentioned in preceding sections of this chapter; the challenges related to molecular and cell-based assays are particularly notable. Some important challenges in advancing the tools for risk-assessment application are described below, and some recommendations are offered.

### Advancing the New Testing Paradigm

*Challenge*: Obtaining the vision described in the Tox21 report in which traditional whole-animal testing is replaced with a broad toxicity-testing strategy that uses primarily in vitro assays, computational methods, and targeted animal testing for assessing the biological activity of chemicals is a complex and labor-intensive task that requires focus, commitment, and resources (NRC 2007). The strategy for achieving the vision involves research to understand the spectrum of perturbations that could result in human toxicity and the nature and extent of the toxicity caused by the perturbations and research to understand

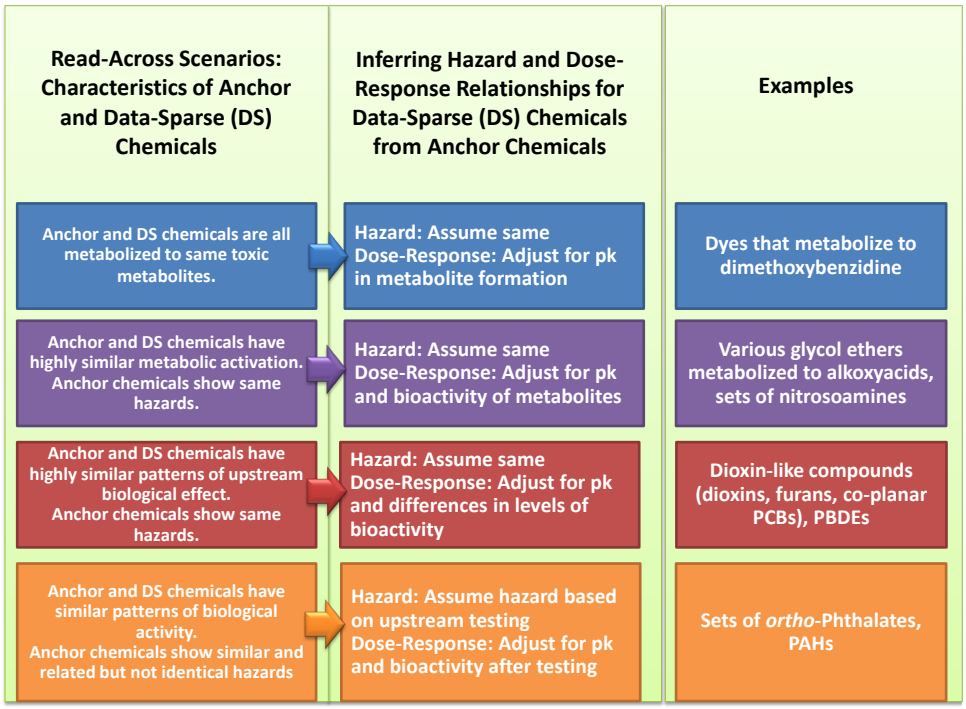



**FIGURE 3-8** Scenarios for conducting read-across.

how determinants of human variability (for example, underlying nutritional, genetic, or disease state or life stage) and exposure duration might affect biological responses or toxicity. The scientific community needs to recognize that the current approach to toxicity testing and data analysis is often compartmentalized, and this prevents a holistic approach in trying to determine toxicity of chemical exposure.

*Recommendation*: Broad consideration of research that is needed to advance the development of a suite of tests that essentially achieves the vision in the Tox21 report is beyond the present committee's charge, but the committee notes that the research described above in the challenge statement should have high priority so that the vision can be achieved. The committee expresses its concurrence with the Tox21 committee and emphasizes that testing should not be limited to the goal of one-to-one replacement but rather should extend toward development of the most salient and predictive assays for the end point or disease being considered.

### Optimizing Tools to Probe Biological Response

*Challenge*: Developing a comprehensive in vitro system that covers the important biological responses to chemical exposure that contribute to human adverse health effects is a considerable challenge. Most assays used in the ToxCast program were developed to meet the needs of the pharmaceutical industry and were not designed to cover the full array of biological response, given the extensive testing in whole animals and humans that is conducted for drug development. Thus, not all major forms of toxicity are captured in the current assays, and correlating tested activities with toxicity-hazard traits has been limited. For example, few or no ToxCast or Tox21 assays test for several of the key characteristics of carcinogenesis (Smith et al. 2016). There is also the question of how short-term assay exposures are related to chronic exposure or developmental exposures in vivo. Responses that depend on higher levels of biological complexity could be missed by cell-based assays. A number of issues for assay development acknowledged in NRC (2007) remain, including coverage of the necessary biological space to ensure that human sensitivity and susceptibility to toxicants are adequately captured.

*Recommendation*: Whole-animal testing should move beyond standard approaches, including those associated with experimental design and statistical methods, to maximize their utility. An array of whole-animal tools are now available, and their adoption could address knowledge gaps in risk assessment more comprehensively and begin to address the breadth of genetic sensitivity in response to chemical exposure and other contributors to human variability in response. Guidance for incorporating these whole-animal tools into risk assessment would likely speed their adoption and use.

*Recommendation*: Use of targeted rodent tests that incorporate the use of -omics technologies, such as sentinel-tissue transcriptomics, should be encouraged. The experimental design should include strategies for data interpretation and analysis, such as Bayesian approaches, that are specifically developed for these studies. Strategic whole-animal testing could help to identify the broader suite of pathways that are beyond the scope of current molecular and cell-based tests, guide the development of in vitro assays that could enhance confidence in extrapolating from in vitro tests to whole-animal responses, and provide a stronger basis of hazard identification and dose–response assessment.

*Recommendation*: Tools for probing genomic, epigenetic, transcriptomic, proteomic, and metabolomic changes in cells should be advanced because they provide an opportunity to assess cellular changes in a nontargeted and non–pathway-specific manner. Because virtually all toxicity is accompanied by specific changes in gene expression (and presumably changes in protein expression and metabolic profile), continued exploration of these in vivo and in vitro approaches as standalone screens or as complements to in vitro screens might be a way to cover more biological space.[5]

### Understanding and Addressing Limitations of Cell Systems

*Challenge*: Substantial progress has been made in developing and adapting a wide array of assays for screening environmental chemicals, but cell cultures have several important limitations. There are challenges in incorporating metabolic capacity into the assays to ensure that assay conditions generate chemical exposures that are representative of the exposures in humans that could lead to toxicity. Cell cultures also tend to be extremely sensitive to environmental conditions; changes in microenvironments can alter cellular phenotypes and responses and result in skewed results of toxicity screens. Furthermore, conventional monolayer cultures are less sensitive than 3-D cultures, and the response obtained from an in vitro assay can depend on the cell type that is used—a liver cell versus a neuron or a primary cell versus an immortalized cell. Current in vitro assays evaluate only chemicals that have particular properties; chemicals typically must be soluble in dimethyl sulfoxide, have low volatility, meet molecular-weight cutoffs, and be available in high enough quantity and purity.

*Recommendation*: Formalized approaches should be developed to characterize the metabolic competence of assays, to determine for which assays metabolic competence is not an essential consideration, and to account for the toxicity of metabolites appropriately. Approaches could include the development and application of better in silico methods for predicting metabolism and elimination and the development of methods for including metabolic capability without compromising other aspects of assay performance. Federal agencies have initiated some research to address the metabolic-capacity issue, and the committee recommends that the research have high priority.

*Recommendation*: Research should be conducted to understand the breadth of cell types needed to capture toxicity that might occur only with specific cell lines. It is possible to identify common pathways of toxicity that exist in all cell types, but biology specific to cell types could be of great use in identifying organ-specific toxicities.

*Recommendation*: Cell batches—even those from established cell lines—should be characterized sufficiently before, during, and after experimentation. Genetic variability, phenotypic characteristics, and purity should be reported in published literature or on publicly accessible Web sites or interfaces.

*Recommendation*: Assay development should be coordinated with development of computational models of cellular responses involved in pathway perturbations to promote deeper understanding of shapes of dose–response curves at the cellular level.

### Addressing the Whole Human and the Human Population

*Challenge*: The exposure-to-outcome continuum in reality can be complex. Chemicals can perturb multiple pathways and lead to various forms of toxicity. Furthermore, toxicity can be influenced by genetics, diet, lifestyle choices, social factors, sex, life stage, health status, and past and present exposures. All those factors can influence responses at different points in the exposure-to-outcome continuum and occur in the exposure milieu and context of human experience.

*Recommendation*: Efforts to capture human variability better in in vitro and in vivo toxicity tests should be explored. Broader testing of multiple cell lines from diverse human populations could find idiosyncratic sensitivity of some populations, as has been seen in in vivo testing of panels of isogenic mouse strains, although this approach addresses only variability due to genetic factors for a single upstream end point. Approaches for better characterization of the variety of possible responses to chemicals in food, drugs, or the environment are needed. Experi-

[5]If in vitro methods are used for this purpose, it will be important to identify the minimum number of cell types necessary for full coverage. Identifying the cell types will require a combination of statistical approaches that retrospectively analyze the available transcriptomic data and prospective experimentation to determine the number of cell types that are responsive to a broad array of mechanisms. High-content imaging techniques that capture effects on multiple cellular-toxicity indicators simultaneously—including mitochondrial integrity, cell viability, lipid accumulation, cytoskeletal integrity, and formation of reactive oxygen species (Grimm et al. 2015)—can also be used for nontargeted screening and offer the potential to integrate multiple aspects of cell function.

mental approaches could be coupled with computational approaches for better characterization.

*Recommendation*: Relatively low-cost, rapid molecular and cellular assays should be used to investigate the toxicity of chemical mixtures. Furthermore, humans are not exposed to single chemicals in isolation but instead are constantly exposed to myriad chemicals in their environment, endogenous chemicals produced in the body or modulated as a consequence of social and behavioral factors, and complex chemical mixtures. Cell-based assays can be used to explore at the molecular and pathway level how the addition of a chemical exposure to existing exogenous and endogenous exposures might contribute to risk.

## REFERENCES


Abdo, N., M. Xia, C.C. Brown, O. Kosyk, R. Huang, S. Sakamuru, Y.H. Zhou, J.R. Jack, P. Gallins, K. Xia, Y. Li, W.A. Chiu, A.A. Motsinger-Reif, C.P. Austin, R.R. Tice, I. Rusyn, and F.A. Wright. 2015. Population-based in vitro hazard and concentration-response assessment of chemicals: The 1000 genomes high-throughput screening study. Environ. Health Perspect. 123(5):458-466.

Ablain, J., and L.I. Zon. 2013. Of fish and men: Using zebrafish to fight human diseases. Trends Cell. Biol. 23(12):584-586.

Al-Imari, L., and R. Gerlai. 2008. Sight of conspecifics as reward in associative learning in zebrafish (Danio rerio). Behav. Brain. Res. 189(1):216-219.

Alon, U. 2007. Network motifs: Theory and experimental approaches. Nat. Rev. Genet. 8(6):450-461.

Alves, V.M., E. Murastov, D. Fourches, J. Strickland, N. Kleinstreuer, C.H. Andrade, and A. Tropsha. 2015. Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds. Toxicol. Appl. Pharmacol. 284(2):262-272.

An, W.F., and N. Tolliday. 2010. Cell-based assays for high-throughput screening. Mol. Biotechnol. 45(2):180-186.

Andersen, M.E., and D. Krewski. 2009. Toxicity testing in the 21st century: Bringing the vision to life. Toxicol. Sci. 107(2):324-330.

Andersen, M.E., K. Betts, Y. Dragan, S. Fitzpatrick, J.L. Goodman, T. Hartung, J. Himmelfarb, D.E. Ingber, A. Jacobs, R. Kavlock, K. Kolaja, J.L. Stevens, D. Tagle, D. Lansing Taylor, and D. Throckmorton. 2014. Developing microphysiological systems for use as regulatory tools—challenges and opportunities. ALTEX 31(3):364-367.

Anson, B.D., K.L. Kolaja, and T.J. Kamp. 2011. Opportunities for use of human iPS cells in predictive toxicology. Clin. Pharmacol. Ther. 89(5):754-758.

Arai, K., R. Sakamoto, D. Kubota, and T. Kondo. 2013. Proteomic approach toward molecular backgrounds of drug resistance of osteosarcoma cells in spheroid culture system. Proteomics 13(15):2351-2360.

Arnaout, R., T. Ferrer, J. Huisken, K. Spitzer, D.Y.R. Stainier, M. Tristani-Firouzi, and N.C. Chi. 2007. Zebrafish model for human long QT syndrome. Proc. Natl. Acad. Sci. US 104(27):11316-11321.

Arnaout, R., S., Reischauer, and D.Y. Stainier. 2014. Recovery of adult zebrafish hearts for high-throughput applications. J. Vis. Exp. 94:e52248.

Asnani, A., and R.T. Peterson. 2014. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. Dis. Model Mech. 7(7):763-767.

Asphahani, F., M. Thein, K. Wang, D. Wood, S.S. Wong, J. Xu, and M. Zhang. 2012. Real-time characterization of cytotoxicity using single-cell impedance monitoring. Analyst. 137(13):3011-3019.

Astashkina, A., B. Mann, and D.W. Grainger. 2012. A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. Pharmacol. Ther. 134(1):82-106.

Beers, J., K.L. Linask, J.A. Chen, L.I. Siniscalchi, Y. Lin, W. Zheng, M. Rao, and G. Chen. 2015. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. Sci. Rep. 5:11319.

Bender, A., J. Scheiber, M. Glick, J.W. Davies, K. Azzaoui, J. Hamon, L. Urban, S. Whitebread, and J.L. Jenkins. 2007. Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. Chem. Med. Chem. 2(6):861-873.

Bento, A.P., A. Gaulton. A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Krüger, Y. Light, L. Mark, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, and J.P. Overington. 2014. The ChEMBL bioactivity database: An update. Nucleic Acids Res. 42:D1083-D1090.

Berg, E.L., M.A. Polokoff, A. O'Mahony, D. Nguyen, and X. Li. 2015. Elucidating mechanisms of toxicity using phenotypic data from primary human cell systems--a chemical biology approach for thrombosis-related side effects. Int. J. Mol. Sci. 16(1):1008-1029.

Berggren, E., P. Amcoff, R. Benigni, K. Blackburn, E. Carney, M. Cronin, H. Deluyker, F. Gautier, R.S. Judson, G.E. Kass, D. Keller, D. Knight, W. Lilienblum, C. Mahony, I. Rusyn, T. Schultz, M. Schwarz, G. Schüürmann, A. White, J. Burton, A.M. Lostia, S. Munn, and A. Worth. 2015. Chemical safety assessment using read-across: Assessing the use of novel testing methods to strengthen the evidence base for decision making. Environ. Health Perspect. 123(12):1232-1240.

Best, J.D., S. Berghmans, J.J. Hunt, S.C. Clarke, A. Fleming, P. Goldsmith, and A.G. Roach. 2008. Non-associative learning in larval zebrafish. Neuropsychopharmacology 33(5):1206-1215.

Bhogal, N., C. Grindon, R. Combes, and M. Balls. 2005. Toxicity testing: Creating a revolution based on new technologies. Trends Biotechnol. 26(6):299-307.

Birnbaum, L. 2011. Presentation at NIEHS Workshop: Engineered Tissue Models for Environmental Health Science Research, June 27-28, 2011, Washington, DC.

Boverhof, D.R., M.P. Chamberlain, C.R. Elcombe, F.J. Gonzalez, R.H. Heflich, L.G. Hernandez, A.C. Jacobs, D. Jacobson-Kram, M. Luijten, A. Maggi, M.G. Manjanatha, J. Benthem, and B.B. Gollapudi. 2011. Transgenic animal models in toxicology: Historical perspectives and future outlook. Toxicol. Sci. 121(2):207-233.

Bowes, J., A.J. Brown, J. Hamon, W. Jarolimek, A. Sridhar, G. Waldron, and S. Whitebread. 2012. Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. Nat. Rev. Drug Discov. 11(12):909-922.

Bradford, B.U., E.F. Lock, O. Kosyk, S. Kim, T. Uehara, D. Harbourt, M. DeSimone, D.W. Threadgill, V. Tryndyak, I.P. Pogribny, L. Bleyle, D.R. Koop, and I. Rusyn. 2011. Interstrain differences in the liver effects of trichloroethylene in a multistrain panel of inbred mice. Toxicol. Sci. 120(1):206-217.

Braga, R.C., V.M. Alves, M.F. Silva, E. Muratov, D. Fourches, A. Tropsha, and C.H. Andrade. 2014. Tuning hERG out: Antitarget QSAR models for drug development. Curr. Top Med. Chem. 14(11):1399-1415.

Bretaud, S., S. Lee, and S. Guo. 2004. Sensitivity of zebrafish to environmental toxins implicated in Parkinson's disease. Neurotoxicol. Teratol. 26(6):857-864.

Browne, P., R.S. Judson, W.M. Casey, N.C. Kleinstreuer, and R.S. Thomas. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environ. Sci. Technol. 49(14):8804-8814.

Bulysheva, A.A., G.L. Bowlin, S.P. Petrova, and W.A. Yeudall. 2013. Enhanced chemoresistance of squamous carcinoma cells grown in 3D cryogenic electrospun scaffolds. Biomed. Mater. 8(5):055009.

Champagne, D.L., C.C. Hoefnagels, R.E. de Kloet, and M.K. Richardson. 2010. Translating rodent behavioral repertoire to zebrafish (Danio rerio): Relevance for stress research. Behav. Brain Res. 214(2):332-342.

Chanut, F., C. Kimbrough, R. Hailey, B. Berridge, A. Hughes-Earle, R. Davies, K. Roland, A. Stokes, A. Casartelli, M. York, H. Jordan, F. Crivellente, P. Cristofori, H. Thomas, J. Klapwijk, and R. Adler. 2013. Spontaneous cardiomyopathy in young Sprague-Dawley rats: Evaluation of biological and environmental variability. Toxicol. Pathol. 41(8):1126-1136.

Chapman, A.L., E.J. Bennett, T.M. Ramesh, K.J. De Vos, and A.J. Grierson. 2013. Axonal transport defects in a mitofusin 2 loss of function model of Charcot-Marie-Tooth disease in zebrafish. PLoS One 8(6):e67276.

Chen, J., S.R. Das, J. La Du, M.M. Corvi, C. Bai, Y. Chen, X. Liu, G. Zhu, R.L. Tanguay, Q. Dong, and C. Huang. 2013. Chronic PFOS exposures induce life stage-specific behavioral deficits in adult zebrafish and produce malformation and behavioral deficits in F1 offspring. Environ. Toxicol. Chem. 32(1):201-206.

Chen, J., J. Wang, Y. Zhang, D. Chen, C. Yang, C. Kai, X. Wang, F. Shi, and J. Dou. 2014. Observation of ovarian cancer stem cell behavior and investigation of potential mechanisms of drug resistance in three-dimensional cell culture. J. Biosci. Bioeng. 118(2):214-222.

Cheng, F., Y. Zhou, J. Li, W. Li, G. Liu, and Y. Tang. 2012. Prediction of chemical-protein interactions: Multitarget-QSAR versus computational chemogenomic methods. Mol. Biosyst. 8(9):2373-2384.

Chico, T.J., P.W. Ingham, and D.C. Crossman. 2008. Modeling cardiovascular disease in the zebrafish. Trends Cardiovasc. Med. 18(4):150-155.

Chiu, W.A., J. Jinot, C.S. Scott, S.L. Makris, G.S. Cooper, R.C. Dzubow, A.S. Bale, M.V. Evans, K.Z. Guyton, N. Keshava, J.C. Lipscomb, S. Barone Jr., J.F. Fox, M.R. Gwinn, J. Schaum, and J.C. Caldwell. 2013. Human health effects of trichloroethylene: Key findings and scientific issues. Environ. Health Perspect. 121(3):303-311.

Churchill, G.A., D.C. Airey, H. Allayee, J.M. Angel, A.D. Attie, J. Beatty, W.D. Beavis, J.K. Belknap, B. Bennett, W. Berrettini, A. Bleich, M. Bogue, K.W. Broman, K.J. Buck, E. Buckler, M. Burmeister, E.J. Chesler, J.M. Cheverud, S. Clapcote, M.N. Cook, R.D. Cox, J.C. Crabbe, W.E. Crusio, A. Darvasi, C.F. Deschepper, R.W. Doerge, C.R. Farber, J. Forejt, D. Gaile, S.J. Garlow, H. Geiger, H. Gershenfeld, T. Gordon, J. Gu, W. Gu, G. de Haan, N.L. Hayes, C. Heller, H. Himmelbauer, R. Hitzemann, K. Hunter, H.C. Hsu, F.A. Iraqi, B. Ivandic, H.J. Jacob, R.C. Jansen, K.J. Jepsen, D.K. Johnson, T.E. Johnson, G. Kempermann, C. Kendziorski, M. Kotb, R.F. Kooy, B. Llamas, F. Lammert, J.M. Lassalle, P.R. Lowenstein, L. Lu, A. Lusis, K.F. Manly, R. Marcucio, D. Matthews, J.F. Medrano, D.R. Miller, G. Mittleman, B.A. Mock, J.S. Mogil, X. Montagutelli, G. Morahan, D.G. Morris, R. Mott, J.H. Nadeau, H. Nagase, R.S. Nowakowski, B.F. O'Hara, A.V. Osadchuk, G.P. Page, B. Paigen, K. Paigen, A.A. Palmer, H.J. Pan, L. Peltonen-Palotie, J. Peirce, D. Pomp, M. Pravenec, D.R. Prows, Z. Qi, R.H. Reeves, J. Roder, G.D. Rosen, E.E. Schadt, L.C. Schalkwyk, Z. Seltzer, K. Shimomura, S. Shou, M.J. Sillanpaa, L.D. Siracusa, H.W. Snoeck, J.L. Spearow, K. Svenson, L.M. Tarantino, D. Threadgill, L.A. Toth, W. Valdar, F.P. de Villena, C. Warden, S. Whatley, R.W. Williams, T. Wiltshire, N. Yi, D. Zhang, M. Zhang, F. Zou, and Complex Trait Consortium. 2004. The collaborative cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36(11):1133-1137.

Churchill, G.A., D.M. Gatti, S.C. Munger, and K.L. Svenson. 2012. The diversity outbred mouse population. Mamm. Genome 23(9-10):713-718.

Colley, H., D. James, K. Diment, and M. Tedder. 2007. Learning as becoming in vocational education and training: Class, gender and the role of vocational habitus. J. Voc. Educ. Train. 55(4):471-498.

Contrera, J.F., P. MacLaughlin, L.H. Hall, and L.B. Kier. 2005. QSAR modeling of carcinogenic risk using discrim-

inant analysis and topological molecular descriptors. Curr. Drug Discov. Technol. 2(2):55-67.

Cote, I., M.E. Andersen, G.T. Ankley, S. Barone, L.S. Birnbaum, K. Boekelheide, F.Y. Bois, L.D. Burgoon, W.A. Chiu, D. Crawford-Brown, K.M. Crofton, M. DeVito, R.B. Devlin, S.W. Edwards, K. Guyton, D. Hattis, R.S. Judson, D. Knight, D. Krewski, J. Lambert, E.A. Maull, D. Mendrick, G.M. Paoli, C.J. Patel, E. Perkins, G. Poje, C.J. Portier, I. Rusyn, P.A. Schulte, A. Simeonov, M.T. Smith, K. Thayer, R.S. Thomas, R. Thomas, R.R. Tice, J.J. Vandenberg, D. Villeneuve, S. Wesselkamper, M. Whelan, C. Whittaker, R. White, M. Xia, C. Yauk, L. Zeise, J. Zhao, and R. DeWoskin. 2016. The next generation of risk assessment multiyear study- highlights of findings, applications to risk assessment and future directions. Environ. Health Perspect. 121(11):1671-1682.

Crosby, E.B., J.M. Bailey, A.N. Oliveri, and E.D. Levin. 2015. Neurobehavioral impairments caused by developmental imidacloprid exposure in zebrafish. Neurotoxicol. Teratol. 49:81-90.

Cui, C., E.L. Benard, Z. Kanwal, O.W. Stockhammer, M. van der Vaart, A. Zakrzewska, H.P. Spaink, and A.H. Meijer. 2011. Infectious disease modeling and innate immune function in zebrafish embryos. Methods Cell Biol.105:273-308.

Da Costa, M.M., C.E. Allen, A. Higginbottom, T. Ramesh, P.J. Shaw, and C.J. McDermott. 2014. A new zebrafish model produced by TILLING of SOD1-related amyotrophic lateral sclerosis replicates key features of the disease and represents a tool for in vivo therapeutic screening. Dis. Model Mech. 7(1):73-81.

Dalgin, G., and V.E. Prince. 2015. Differential levels of Neurod establish zebrafish endocrine pancreas cell fates. Dev. Biol. 402(1):81-97.

De Vooght, V., J. A. Vanoirbeek, K. Luyts, S. Haenen, B. Nemery, and P.H. Hoet. 2010. Choice of mouse strain influences the outcome in a mouse model of chemical-induced asthma. PLoS One 5(9):e12581.

Dong, H., W. Xu, J.K. Pillai, C. Packianathan, and B.P. Rosen. 2015. High-throughput screening-compatible assays of As(III) S-adenosylmethionine methyltransferase activity. Anal. Biochem. 480:67-73.

Dooley, K., and L.I. Zon. 2000. Zebrafish: A model system for the study of human disease. Curr. Opin. Genet. Dev. 10(3):252-256.

Dow, L.E., J. Fisher, K.P. O'Rourke, A. Muley, E.R. Kastenhuber, G. Livshits, D.F. Tschaharganeh, N.D. Socci and S.W. Lowe. 2015. Inducible in vivo genome editing with CRISPR-Cas9. Nat. Biotechnol. 33(4):390-394.

ECHA (European Chemicals Agency). 2015. Read-Across Assessment Framework (RAAF) [online]. Available: http://echa.europa.eu/documents/10162/13628/raaf_en.pdf [accessed July 19, 2016].

Edmondson, R., J.J. Broglie, A.F. Adcock, and L. Yang. 2014. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. Assay Drug Dev. Technol. 12(4):207-218.

Efthymiou, A., A. Shaltouki, J.P. Steiner, B. Jha, S.M. Heman-Ackah, A. Swistowski, X. Zeng, M.S. Rao, and N. Malik. 2014. Functional screening assays with neurons generated from pluripotent stem cell-derived neural stem cells. J. Biomol. Screen. 19(1):32-43.

Egan, R.J., C.L. Bergner, P.C. Hart, J.M. Cachat, P.R. Canavello, M.F. Elegante, S.I. Elkhayat, B.K. Bartels, A.K. Tien, D.H. Tien, S. Mohnot, E. Beeson, E. Glasgow, H. Amri, Z. Zukowska, and A.V. Kalueff. 2009. Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. Behav. Brain Res. 205(1):38-44.

Engeszer, R.E., L.A. da Barbiano, M.J. Ryan, and D.M. Parichy. 2007. Timing and plasticity of shoaling behaviour in the zebrafish, Danio rerio. Anim. Behav. 74(5):1269-1275.

EPA (US Environmental Protection Agency). 2010. Toxicological Review of Acrylamide (CAS No. 79-06-1). EPA/635/R-07/009F [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0286tr.pdf [accessed July 19, 2016].

EPA (US Environmental Protection Agency). 2014a. FIFRA Scientific Advisory Panel Minutes No. 2015-01: A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding Integrated Endocrine Bioactivity and Exposure-Based Prioritization and Screening, FIFRA Scientific Advisory Panel Meeting, December 2-4, 2014, Arlington, VA [online]. Available: https://www.epa.gov/sites/production/files/2015-06/documents/120214minutes.pdf [accessed July 19, 2016].

EPA (US Environmental Protection Agency). 2014b. Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology. EPA/600/R-14/004. National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=286690 [accessed July 19, 2016].

EPA (US Environmental Protection Agency). 2015. Use of High Throughput Assays and Computational Tools in the Endocrine Disruptor Screening Program-Overview [online]. Available: https://www.epa.gov/endocrine-disruption/use-high-throughput-assays-and-computational-tools-endocrine-disruptor [accessed December 1, 2016].

EPA/NIH/NCATS/NTP (US Environmental Protection Agency, National Institutes of Health, National Center for Advancing Translational Sciences, and National Toxicology Program). 2016. Transform Tox Testing Challenge: Innovating for Metabolism-Challenge Overview [online]. Available: https://www.challenge.gov/wp-content/uploads/2016/09/Transform-Tox-Testing-Challenge-Brief.pdf [accessed October 14, 2016].

Esch, E.W., A. Bahinski, and D. Huh. 2015. Organs-on-chips at the frontiers of drug discovery. Nat. Rev. Drug Discov. 14:248-260.

EU (European Union). 2015. Mechanism Based Integrated Systems for the Prediction of Drug Induced Liver Injury [online]. Available: http://www.mip-dili.eu/ [accessed July 19, 2016].

Feitsma, H., and E. Cuppen. 2008. Zebrafish as a cancer model. Mol. Cancer Res. 6(5):685-694.

Feliu, N., P. Kohonen, J. Ji, Y. Zhang, H.L. Karlsson, L. Palmberg, A. Nyström, and B. Fadeel. 2015. Next-generation sequencing reveals low-dose effects of cationic dendrimers in primary human bronchial epithelial cells. ACS Nano. 9(1):146-163.

Fernandes, Y., S. Tran, E. Abraham, and R. Gerlai. 2014. Embryonic alcohol exposure impairs associative learning performance in adult zebrafish. Behav. Brain Res. 265:181-187.

Festing, M.F. 2010. Improving toxicity screening and drug development by using genetically defined strains. Methods Mol. Biol. 602:1-21.

Foster, P.M., L.V. Thomas, M.W. Cook, and S.D. Gangolli. 1980. Study of the testicular effects and changes in zinc excretion produced by some n-alkyl phthalates in the rat. Toxicol. Appl. Pharmacol. 54(3):392-398.

French, J.E., D.M. Gatti, D.L. Morgan, G.E. Kissling, K.R. Shockley, G.A. Knudsen, K.G. Shepard, H.C. Price, D. King, K.L. Witt, L.C. Pedersen, S.C. Munger, K.L. Svenson, and G.A. Churchill. 2015. Diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. Environ. Health Perspect. 123(3):237-245.

Gallardo, V.E., G.K. Varshney, M. Lee, S. Bupp, L. Xu, P. Shinn, N.P. Crawford, J. Inglese, and S.M. Burgess. 2015. Phenotype-driven chemical screening in zebrafish for compounds that inhibit collective cell migration identifies multiple pathways potentially involved in metastatic invasion. Dis. Model Mech. 8(6):565-576.

Gavaghan, C. 2007. Practical Considerations in Using QSARs in Pharmaceutical Safety Assessment [online]. Available: http://www.ukqsar.org/slides/ClaireGavaghan_2007.pdf [accessed January 3, 2017].

Gerlach, G., A. Hodgins-Davis, C. Avolio, and C. Schunter. 2008. Kin recognition in zebrafish: A 24-hour window for olfactory imprinting. Proc. Biol. Sci. 275(1647):2165-2170.

Gieseck, R.L., III, N.R. Hannan, R. Bort, N.A. Hanley, R.A. Drake, G.W. Cameron, T.A. Wynn, and L. Vallier. 2014. Maturation of induced pluripotent stem cell derived hepatocytes by 3D-culture. PLoS One 9(1):e86372.

Gieseck, R.L., III, L. Vallier, and N.R. Hannan. 2015. Generation of hepatocytes from pluripotent stem cells for drug screening and developmental modeling. Methods Mol. Biol. 1250:123-142.

Godderis, L., R. Thomas, A.E. Hubbard, A.M. Tabish, P. Hoet, L. Zhang, M.T. Smith, H. Veulemans, and C.M. McHale. 2012. Effect of chemical mutagens and carcinogens on gene expression profiles in human TK6 cells. PLOS One 7(6):e39205.

Goldstone, J.V., A.G. McArthur, A. Kubota, J. Zanette, T. Parente, M.E. Jonsson, D.R. Nelson, and J.J. Stegeman. 2010. Identification and developmental expression of the full complement of cytochrome P450 genes in zebrafish. BMC Genomics 11:643.

Gordon, M.W., F. Yan, X. Zhong, P.B. Mazumder, Z.Y. Xu-Monette, D. Zou, K.H. Young, K.S. Ramos, and Y. Li. 2015. Regulation of p53-targeting microRNAs by polycyclic aromatic hydrocarbons: Implications in the etiology of multiple myeloma. Mol. Carcinog. 54(10):1060-1069.

Graham, J.B., S. Thomas, J. Swarts, A.A. McMillan, M.T. Ferris, M.S. Suthar, P.M. Treuting, R. Ireton, M. Gale, Jr., and J.M. Lund. 2015. Genetic diversity in the collaborative cross model recapitulates human West Nile virus disease outcomes. MBio 6(3):e00493-15.

Gray, L.E., J. Ostby, J. Furr, M. Price, D.N. Veeramacheni, and L. Parks. 2000. Perinatal exposure to the phthalates DEHP, BBP, and DINP, but not DEP, DMP, or DOTP, alters sexual differentiation of the male rat. Toxicol. Sci. 58(2):350-365.

Greene, N., L. Fisk, R.T. Naven, R.R. Note, M.L. Patel, and D.J. Pelletier. 2010. Developing structure-activity relaitonships for the preduction of hepatotoxicity. Chem. Res. Toxicol. 23(7):1215-1222.

Grimm, F.A., Y. Iwara, O. Sirenko, M. Bittner, and I. Rusyn. 2015. High-content assay multiplexing for toxicity screening in induced pluripotent stem cell-derived cardiomyocytes and hepatocytes. Assay Drug Dev. Technol. 13(9):529-546.

Guiro, K., S.A. Patel, S.J. Greco, P. Rameshwar, and T.L. Arinzeh. 2015. Investigating breast cancer cell behavior using tissue engineering scaffolds. PLoS One 10(3):e0118724.

Gut, P., B. Baeza-Raja, O. Andersson, L. Hasenkamp, J. Hsiao, D. Hesselson, K. Akassoglou, E. Verdin, M.D. Hirschey, and D.Y. Stainier. 2013. Whole-organism screening for gluconeogenesis identifies activators of fasting metabolism. Nat. Chem. Biol. 9(2):97-104.

Guyton, K.Z., W.A. Chiu, T.F. Bateson, J. Jinot, C.S. Scott, R.C. Brown, and J.C. Caldwell. 2009. A reexamination of the PPAR-alpha activation mode of action as a basis forassessing human cancer risks of environmental contaminants. Environ. Health Perspect. 117(11):1664-1672.

Harrill, A.H., P.B. Watkins, S. Su, P.K. Ross, D.E. Harbourt, I.M. Stylianou, G.A. Boorman, M.W. Russo, R.S. Sackler, S.C. Harris, P.C. Smith, R. Tennant, M. Bogue, K. Paigen, C. Harris, T. Contractor, T. Wiltshire, I. Rusyn, and D.W. Threadgill. 2009. Mouse population-guided resequencing reveals that variants in CD44 contribute to

acetaminophen-induced liver injury in humans. Genome Res. (9):1507-1515.

Harris, P.A., C. Duraiswami, D.T. Fisher, J. Fornwald, S.J. Hoffman, G. Hofmann, M. Jiang, R. Lehr, P.M. McCormick, L. Nickels, B. Schwartz, Z. Wu, G. Zhang, R.W. Marquis, J. Bertin, and P.J. Gough. 2015. High throughput screening identifies ATP-competitive inhibitors of the NLRP1 inflammasome. Bioorg. Med. Chem. Lett. 25(14):2739-2743.

Hewitt, M., S.J. Enoch, J.C. Madden, K.R. Przybylak, and M.T. Cronin. 2013. Hepatotoxicity: A scheme for generation chemical categories for read-across, structural alerts and insights into mechanism(s) of action. Crit. Rev. Toxicol. 43(7):537-558.

Hornung, M.W., M.A. Tapper, J.S. Denny, R.C. Kolanczyk, B.R. Sheedy, P.C. Hartig, H. Aladjov, T.R. Henry, and P.K. Schmieder. 2014. Effects-based chemical category approach for prioritization of low affinity estrogenic chemicals. SAR QSAR Environ. Res. 25(4):289-323.

Hossini, A.M., M. Megges, A. Prigione, B. Lichtner, M.R. Toliat, W. Wruck, F. Schröter, P. Nuernberg, H. Kroll, E. Makrantonaki, C.C. Zouboulis, and J. Adjaye. 2015. Induced pluripotent stem cell-derived neuronal cells from a sporadic Alzheimer's disease donor as a model for investigating AD-associated gene regulatory networks. BMC Genomics 16:84.

Houck, K.A., D.J. Dix, R.S. Judson, R.J. Kavlock, J. Yang, and E.L. Berg. 2009. Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. J. Biomol. Screen 14(9):1054-1066.

Howe, D.G., Y.M. Bradford, T. Conlin, A.E. Eagle, D. Fashena, K. Frazer, J. Knight, P. Mani, R. Martin, S.A. Moxon, H. Paddock, C. Pich, S. Ramachandran, B.J. Ruef, L. Ruzicka, K. Schaper, X. Shao, A. Singer, B. Sprunger, C.E. Van Slyke, and M. Westerfield. 2013a. ZFIN, the Zebrafish Model Organism Database: Increased support for mutants and transgenics. Nucleic Acids Res. 41:D854-D860.

Howe, K., M.D. Clark, C.F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J.E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J.C. Barrett, R. Koch, G.J. Rauch, S. White, W. Chow, B. Kilian, L.T. Quintais, J.A. Guerra-Assuncao, Y. Zhou, Y. Gu, J. Yen, J.H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S.F. Maguire, G.K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliot, G. Threadgold, G. Harden, D. Ware, S. Begum, B. Mortimore, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C.M. Dooley, Z. Ersan-Urun, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberlander, S. Rudolph-Geiger, M. Teucke, C. Lanz, G. Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S. C. Schuster, N.P. Carter, J. Harrow, Z. Ning, J. Herrero, S.M. Searle, A. Enright, R. Geisler, R.H. Plasterk, C. Lee, M. Westerfield, P.J. de Jong, L.I. Zon, J.H. Postlethwait, C. Nusslein-Volhard, T.J. Hubbard, H. Roest Crollius, J. Rogers, and D.L. Stemple. 2013b. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 496(7446):498-503.

Huh, D., G.A. Hamilton, and D.E. Ingber. 2011. From 3D cell culture to organs-on-chips. Trends Cell Biol. 21(12):745-754.

IARC (International Agency for Research on Cancer). 2006. Preamble. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon, France: IARC [online]. Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed July 19, 2016].

IARC (International Agency for Research on Cancer). 2015a. Diazinon in Some Organophosphate Insecticides and Herbicides. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 112 [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol112/mono112-08.pdf [accessed May 15, 2016].

IARC (International Agency for Research on Cancer). 2015b. Malathion in Some Organophosphate Insecticides and Herbicides. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 112 [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol112/mono112-07.pdf [accessed May 15, 2016].

ICH (International Conference on Harmonization). 2014. ICH Harmonised Tripartite Guideline: Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk. M7: Current Step 4 Version, June 23, 2014. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use [online]. Available: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_Step_4.pdf [accessed July 22, 2016].

Irons, T.D., R.C. MacPhail, D.L. Hunter, and S. Padilla. 2010. Acute neuroactive drug exposures alter locomotor activity in larval zebrafish. Neurotoxicol. Teratol. 32(1):84-90.

Jones, L.J., and W.H. Norton. 2015. Using zebrafish to uncover the genetic and neural basis of aggression, a frequent comorbid symptom of psychiatric disorders. Behav. Brain Res. 276:171-180.

Kacew, S., and M.F. Festing. 1996. Role of rat strain in the differential sensitivity to pharmaceutical agents and naturally occurring substances. J. Toxicol. Environ. Health 47(1):1-30.

Karakikes, I., M. Ameen, V. Termglinchen, and J.C. Wu. 2015. Human induced pluripotent stem cell-derived cardiomyocytes: Insights into molecular, cellular, and functional phenotypes. Circ. Res. 117(1):80-88.

Keiser, M.J., B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, and B.K. Shoichet. 2007. Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. 25(2):197-206.

Kettleborough, R.N., E.M. Busch-Nentwich, S.A. Harvey, C.M. Dooley, E. de Bruijn, F. van Eeden, I. Sealy, R.J. White, C. Herd, I.J. Nijman, F. Fenyes, S. Mehroke, C. Scahill, R. Gibbons, N. Wali, S. Caruthers, A. Hall, J. Yen, E. Cuppen, and D.L. Stemple. 2013. A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature 496(7446):494-497.

Kleinstreuer, N.C., J. Yang, E.L. Berg, T.B. Knudsen, A.M. Richard, M.T. Martin, D.M. Reif, R.S. Judson, M. Polokoff, D.J. Dix, R.J. Kavlock, and K.A. Houck. 2014. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. Nat. Biotechnol. 32(6):583-591.

Kokel, D., J. Bryan, C. Laggner, R. White, C.Y. Cheung, R. Mateus, D. Healey, S. Kim, A.A. Werdich, S.J. Haggarty, C.A. Macrae, B. Shoichet, and R.T. Peterson. 2010. Rapid behavior-based identification of neuroactive small molecules in the zebrafish. Nat. Chem. Biol. 6(3):231-237.

Kolaja, K. 2014. Stem cells and stem cell-derived tissues and their use in safety assessment. J. Biol. Chem. 289(8):4555-4561.

Krewski, D., M.E. Andersen, E. Mantus, and L. Zeise. 2009. Toxicity testing in the 21st century: Implications for human health risk assessment. Risk Anal. 29(4):474-479.

Krewski, D., M. Westphal, M.E. Andersen, G. Paoli, W. Chiu, M. Al-Zoughool, M.C. Croteau, L. Burgoon, and I. Cote. 2014. A framework for the next generation of risk science. Environ Health Perspect.122(8):796-805.

Kruhlak, N.L., R.D. Benz, H. Zhou, and T.J. Colatsky. 2012. (Q)SAR modeling and safety assessment in regulatory review. Clin. Pharmacol. Ther. 91(3):529-534.

Lamb, J., E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S.A. Armstrong, S.J. Haggerty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, and T.R. Golub. 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929-1935.

Landry, J.P., G. Malovichko, and X.D. Zhu. 2015. High-throughput dose-response measurement using a label-free microarray-in-microplate assay platform. Anal. Chem. 87(11):5640-5648.

Lebold, K.M., C.V. Lohr, C.L. Barton, G.W. Miller, E.M. Labut, R.L.Tanguay, and M.G. Traberet. 2013. Chronic vitamin E deficiency promotes vitamin C deficiency in zebrafish leading to degenerative myopathy and impaired swimming behavior. Comp. Biochem. Phys. C Toxicol. Pharmacol. 157(4):382-389.

Levin, E.D. Z. Bencan, and D.T. Cerutti. 2007. Anxiolytic effects of nicotine in zebrafish. Physiol. Behav. 90(1):54-58.

Liu, J., K. Mansouri, R.S. Judson, M.T. Martin, H. Hong, M. Chen, X. Xu, R.S. Thomas, and I. Shah. 2015. Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem. Res. Toxicol. 28(4):738-751.

Liu, K., K.P. Lehmann, M. Sar, S.S. Young, and K.W.Gaido. 2005. Gene expression profiling following in utero exposure to phthalate esters reveals new gene targets in the etiology of testicular dysgenesis. Biol. Reprod. 73(1):180-192.

Lounkine, E., M.J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J.L. Jenkins, P. Lavan, E.Weber, A.K. Doak, S. Côté, B.K. Shoichet, and L. Urban. 2012. Large-scale prediction and testing of drug activity on side-effect targets. Nature 486(7403):361-367.

Lovik, M. 1997. Mutant and transgenic mice in immunotoxicology: An introduction. Toxicology119(1):65-76.

Lovitt, C.J., T.B. Shelper, and V.M. Avery. 2014. Advanced cell culture techniques for cancer discovery. Biology 3(2):345-367.

Low, Y., A. Sedykh, D. Fourches, A. Golbraikh, M. Whelan, I. Rusyn, and A.Tropsha .2013. Integrative chemical-biological read-across approach for chemical hazard classification. Chem. Res. Toxicol. 26(8):1199-208.

Mahmood, F., S. Fu, J. Cooke, S.W. Wilson, J.D. Cooper, and C. Russell. 2013. A zebrafish model of CLN2 disease is deficient in tripeptidyl peptidase 1 and displays progressive neurodegeneration accompanied by a reduction in proliferation. Brain 136(Pt. 5):1488-1507.

Malik, N., A.G. Efthymiou, K. Mather, N. Chester, X. Wang, A. Nath, M.S. Rao, and J.P. Steiner. 2014. Compounds with species and cell type specific toxicity identified in a 2000 compound drug screen of neural stem cells and rat mixed cortical neurons. Neurotoxicology 45:192-200.

Mann, D.A. 2015. Human induced pluripotent stem cell-derived hepatocytes for toxicology testing. Expert Opin. Drug Metab. Toxicol. 11(1):1-5.

Mann, K.D., E.R. Turnell, J. Atema, and G. Gerlach. 2003. Kin recognition in juvenile zebrafish (Danio rerio) based on olfactory cues. Biol. Bull. 205(2):224-225.

Martin-Jimenez, R., M. Campanella, and C. Russell. 2015. New zebrafish models of neurodegeneration. Curr. Neurol. Neurosci. 15(6):33.

Matthews, E.J., N.L. Kruhlak, R. Daniel Benz, J. Ivanov, G. Klopman, and J.F. Contrera. 2007. A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. Regul. Toxicol. Pharmacol. 47(2):136-155.

Mattis, V.B., C. Tom, S. Akimov, J. Saeedian, M.E. Østergaard, A.L. Southwell, C.N. Doty, L. Ornelas, A. Sahabian, L. Lenaeus, B. Mandefro, D. Sareen, J. Arjomand, M.R. Hayden, C.A. Ross, and C.N. Svendesn. 2015. HD iPSC-derived neural progenitors accumulate in culture and are susceptible to BDNF withdrawal due to glutamate toxicity. Hum. Mol. Genet. 24(11):3257-3271.

McGinnity, D.F., J. Collington, R.P. Austin, and R.J. Riley. 2007. Evaluation of human pharmacokinetics, therapeutic dose and exposure predictions using marketed oral drugs. Curr. Drug Metab. 8(5):463-479.

McKinstry-Wu, A.R., W. Bu, G. Rai, W.A. Lea, B.P. Weiser, D.F. Liang, A. Simeonov, A. Jadhav, D.J. Maloney, and R.G. Eckenhoff. 2015. Discovery of a novel general anesthetic chemotype using high-throughput screening. Anesthesiology 122(2):325-333.

Meeker, N.D., and N.S. Trede. 2008. Immunology and zebrafish: Spawning new models of human disease. Dev. Comp. Immunol. 32(7):745-757.

Mehta, J., E. Rouah-Martin, B. Van Dorst, B. Maes, W. Herrebout, M.L. Scippo, F. Dardenne, R. Blust, and J. Robbens. 2012. Selection and characterization of PCB-binding DNA aptamers. Anal. Chem. 84(3):1669-1676.

Miller, N., and R. Gerlai. 2012. From schooling to shoaling: Patterns of collective motion in zebrafish (Danio rerio). PLoS One 7(11):e48865.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network motifs: Simple building blocks of complex networks. Science 298(5594):824-827.

Moorman, S.J. 2001. Development of sensory systems in zebrafish (Danio rerio). ILAR J. 42(4):292-298.

Morgan, A.P., and C.E. Welsh. 2015. Informatics resources for the Collaborative Cross and related mouse populations. Mamm. Genome 26(9-10):521-539.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2015. Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense. Washington, DC: The National Academies Press.

NCATS (National Center for Advancing Translational Sciences). 2016. About Tissue Chip. Available: http://ncats.nih.gov/tissuechip/about [accessed July 20, 2016].

Ng, H.W., S.W. Doughty, H. Luo, H. Ye, W. Ge, W. Tong, and H. Hong. 2015. Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets. Chem. Res. Toxicol. 28(12):2343-2351.

Nguyen, A.T., A. Emelyanov, C.H. Koh, J.M. Spitsbergen, S. Parinov, and Z. Gong. 2012. An inducible kras(V12) transgenic zebrafish model for liver tumorigenesis and chemical drug screening. Dis. Model Mech. 5(1):63-72.

Norton, W.H. 2013. Towards developmental models of psychiatric disorders in zebrafish. Front. Neural Circuits 7:79.

NRC (National Research Council). 2000. Scientific Frontiers in Developmental Toxicology and Risk Assessment. Washington, DC: National Academy Press.

NRC (National Research Council). 2006. Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues. Washington, DC: The National Academies Press.

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

NRC (National Research Council). 2008. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. Washington, DC: The National Academies Press.

OECD (Organisation for Economic Co-operation and Development). 2004. The Report from the Expert Group on (Quantitative) Structure- Activity Relationships [QSARs] on the Principles for the Validation of (Q)SARs. ENV/JM/MONO(2004)24. OECD Series on Testing and Assessment No. 49. Paris: OECD [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2004)24&doclanguage=en [accessed July 20, 2016].

Padilla, S., D. Corum, B. Padnos, D.L. Hunter, A. Beam, K.A. Houck, N. Sipes, N. Kleinstreuer, T. Knudsen, D.J. Dix, and D.M. Reif. 2012. Zebrafish developmental screening of the ToxCast Phase I chemical library. Reprod. Toxicol. 33(2):174-187.

Panula, P., V. Sallinen, M. Sundvik, J. Kolehmainen, V. Torkko, A. Tiittula, M. Moshnyakov, and P. Podiasz. 2006. Modulatory neurotransmitter systems and behavior: Towards zebrafish models of neurodegenerative diseases. Zebrafish 3(2):235-247.

Panula, P., C.Y. Chen, M. Priyadarshini, H. Kudo, S. Semenova, M. Sundvik, and V. Sallinen. 2010. The comparative neuroanatomy and neurochemistry of zebrafish CNS systems of relevance to human neuropsychiatric diseases. Neurobiol. Dis. 40(1):46-57.

Papadatos, G., A. Gaulton, A. Hersey, and J.P. Overington. 2015. Activity, assay and target data curation and quality in the ChEMBL database. J. Comput. Aided Mol. Des. 29(9):885-896.

Patlewicz, G., N. Ball, E.D. Booth, E. Hulzebos, E. Zvinavashe, and C. Hennes. 2013. Use of category approaches, read-across and (Q)SAR: General considerations. Regular. Toxicol. Pharmacol. 67(1):1-12.

Pauli, A., E. Valen, M.F. Lin, M. Garber, N.L. Vastenhouw, J.Z. Levin, L. Fan, A. Sandelin, J.L. Rinn, A. Regev, and A.F. Schier. 2012. Systematic identification of long non-

coding RNAs expressed during zebrafish embryogenesis. Genome Res. 22(3):577-591.

Pfuhler, S., R. Fautz, G. Ouedraogo, A. Latil, J. Kenny, C. Moore, W. Diembeck, N.J. Hewitt, K. Reisinger, and J. Barroso. 2014. The Cosmetics Europe strategy for animal-free genotoxicity testing: project status up-date. Toxicol. In Vitro 28(1):18-23.

Pham, N., S. Iyer, E. Hackett, B.H. Lock, M. Sandy, L. Zeise, G. Solomon, and M. Marty. 2016. Using ToxCast to explore chemical activities and hazard traits: A case study with ortho-phthalates. Toxicol. Sci. 151(2):286-301.

Phillips, J.B., and M. Westerfield. 2014. Zebrafish models in translational research: Tipping the scales toward advancements in human health. Dis. Model Mech. 7(7):739-743.

Pickart, M.A., and E.W. Klee. 2014. Zebrafish approaches enhance the translational research tackle box. Transl. Res. 163(2):65-78.

Pohjanvirta, R., M. Viluksela, J.T. Tuomisto, M. Unkila, J. Karasinska, M.A. Franc, M. Holowenko, J.V. Giannone, P.A. Harper, J. Tuomisto, and A.B. Okey. 1999. Physicochemical differences in the AH receptors of the most TCDD-susceptible and the most TCDD-resistant rat strains. Toxicol. Appl. Pharmacol. 155(1):82-95.

Preston, M.A., and W.B. Macklin. 2015. Zebrafish as a model to investigate CNS myelination. Glia 63(2):177-193.

Prozialeck, W.C., P.C. Lamar, and S.M. Lynch. 2003. Cadmium alters the localization of N-cadherin, E-cadherin, and beta-catenin in the proximal tubule epithelium. Toxicol. Appl. Pharmacol. 189(3):180-195.

Raoux, M., N. Azorin, C. Colomban, S. Rivoire, T. Merrot, P. Delmas, and M. Crest. 2013. Chemicals inducing acute irritant contact dermatitis mobilize intracellular calcium in human keratinocytes. Toxicol. In Vitro. 27(1):402-408.

Reif, D.M., L. Truong, D. Mandrell, S. Marvel, G. Zhang, and R.L. Tanguay. 2016. High-throughput characterization of chemical-associated embryonic behavioral changes predicts teratogenic outcomes. Arch. Toxicol. 90(6):1459-1470.

Rennekamp, A.J., and R.T. Peterson. 2015. 15 years of zebrafish chemical screening. Curr. Opin. Chem. Biol. 24:58-70.

Roberts, D.W., A.O. Aptula, and G. Patlewicz. 2007a. Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published data set of 106 chemicals tested in the mouse local lymph node assay. Chem. Res. Toxicol. 20(1):44-60.

Roberts, D.W., G. Patlewicz, S.D. Dimitrov, L.K. Low, A.O. Aptula, P.S. Kern, G.D. Dimitrova, M.I.H. Comber, R.D. Phillips, J. Niemelä, C. Madsen, E.B. Wedebye, P.T. Bailey, and O.G. Mekenyan. 2007b. TIMES-SS-A mechanistic evaluation of an external validation study using reaction chemistry principles. Chem. Res. Toxicol. 20(9):1321-1330.

Romero, A.C., E. Del Río, E. Vilanova, and M.A. Sogorb. 2015. RNA transcripts for the quantification of differentiation allow marked improvements in the performance of embryonic stem cell test (EST). Toxicol. Lett. 238(3):60-69.

Rotroff, D.M., D.J. Dix, K.A. Houck, T.B. Knudsen, M.T. Martin, K.W. McLaurin, D.M. Reif, K.M. Crofton, A.V. Singh, M. Xia, R. Huang, and R.S. Judson. 2013. Using in vitro high throughput screening assays to identify potential endocrine-disrupting chemicals. Environ. Health Perspect. 121(1):7-14.

Rous, P., and F.S. Jones. 1916. A method for obtaining suspensions of living cells from the fixed tissues, and for the plating out of individual cells. J. Exp. Med. 23(4):549-555.

Schlegel, A., and P. Gut. 2015. Metabolic insights from zebrafish genetics, physiology, and chemical biology. Cell. Mol. Life Sci. 72(12):2249-2260.

Scott, C.W., M.F. Peters, and Y.P. Dragan. 2013. Human induced pluripotent stem cells and their use in drug discovery for toxicity testing. Toxicol. Lett. 219(1):49-58.

Shah, I., and J. Wambaugh. 2010. Virtual tissues in toxicology. J. Toxicol. Environ. Health B. Crit. Rev. 13(2-4):314-328.

Sharma, P., D.M. Ando, A. Daub, J.A. Kaye, and S. Finkbeiner. 2012. High-throughput screening in primary neurons. Methods Enzymol. 506:331-360.

Shirai, T., A. Nakamura, S. Fukushima, A. Yamamoto, M. Tada, and N. Ito. 1990. Different carcinogenic responses in a variety of organs, including the prostate, of five different rat strains given 3,2'-dimethyl-4-aminobiphenyl. Carcinogenesis 11(5):793-797.

Silva, M., N. Pham, C. Lewis, S. Iyer, E. Kwok, G. Solomon, and L. Zeise. 2015. A comparison of ToxCast test results with In vivo and other In vitro endpoints for neuro, endocrine, and developmental toxicities: A case study using endosulfan and methidathion. Birth Defects Res. B Dev. Reprod. Toxicol. 104(2):71-89.

Simmons, S.O., C.Y. Fan, and R. Ramabhadran. 2009. Cellular stress response pathway system as a sentinel ensemble in toxicological screening. Toxicol. Sci. 111(2):202-225.

Sinnecker, D., K.L. Laugwitz, and A. Moretti. 2014. Induced pluripotent stem cell-derived cardiomyocytes for drug development and toxicity testing. Pharmacol. Ther. 143(2):246-252.

Sipes, N.S., M.T. Martin, P. Kothiya, D.M. Reif, R.S. Judson, A.M. Richard, K.A. Houck, D.J. Dix, R.J. Kavlock, and T.B. Knudsen. 2013. Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. Chem. Res. Toxicol. 26(6):878-895.

Sirenko, O., J. Hesley, I. Rusyn, and E.F. Cromwell. 2014a. High-content high-throughput assays for characterizing the viability and morphology of human iPSC-derived neuronal cultures. Assay Drug Dev. Technol. 12(9-10):536-547.

Sirenko, O., J. Hesley, I. Rusyn, and E. Cromwell. 2014b. High-content assays for hepatotoxicity using induced pluripotent stem cell (iPSC)-derived cells. Assay Drug Dev. Technol. 12(1):43-54.

Sjogren, A.K., M. Liljevald, B. Glinghammar, J. Sagemark, X.Q. Li, A. Jonebring, I. Cotgreave, G. Brolén, and T.B. Andersson. 2014. Critical differences in toxicity mechanisms in induced pluripotent stem cell-derived hepatocytes, hepatic cell lines and primary hepatocytes. Arch. Toxicol. 88(7):1427-1437.

Sledge, D., J. Yen, T. Morton, L. Dishaw, A. Petro, S. Donerly, E. Linney, and E.D. Levin. 2011. Critical duration of exposure for developmental chlorpyrifos-induced neurobehavioral toxicity. Neurotoxicol. Teratol. 33(6):742-751.

Smith, A.J., M.K. Hancock, K. Bi, J. Andrews, P. Harrison, and T.J. Vaughan. 2012. Feasibility of implementing cell-based pathway reporter assays in early high-throughput screening assay cascades for antibody drug discovery. J. Biomol. Screen. 17(6):713-726.

Smith, M.T., K.Z. Guyton, C.F. Gibbons, J.M. Fritz, C.J. Portier, I. Rusyn, D.M. DeMarini, J.C. Caldwell, R.J. Kavlock, P. Lambert, S.S. Hecht, J.R. Bucher, B.W. Stewart, R. Baan, V.J. Cogliano, and K. Straif. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ Health Perspect. 124(6):713-721.

Soldatow, V.Y., E.L. Lecluyse, L.G. Griffith, and I. Rusyn. 2013. In vitro models for liver toxicity testing. Toxicol. Res. (Camb). 2(1):23-39.

Song, Y., V. Madahar, and J. Liao. 2011. Development of FRET assay into quantitative and high-throughput screening technology platforms for protein-protein interactions. Ann. Biomed. Eng. 39(4):1224-1234.

Spence, R., and C. Smith. 2005. Male territoriality mediates density and sex ratio effects on oviposition in the zebrafish. Anim. Behav. 69(6):1317-1323.

Steenbergen, P. J., M.K. Richardson, and D.L. Champagne. 2011. The use of the zebrafish model in stress research. Prog. Neuropsychopharmacol. Biol. Psychiatry 35(6):1432-1451.

Sung, J.H., and M.L. Shuler. 2010. In vitro microscale systems for systematic drug toxicity study. Bioprocess Biosyst. Eng. 33(1):5-19.

Swat, M., and J.A. Glazier. 2013. Agent-based virtual-tissue simulations. Biomed. Comput. Rev. (Fall):28-29.

Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131(5):861-872.

Theunissen, P.T., J.F. Robinson, J.L. Pennings, M.H. van Herwijnen, J.C. Kleinjans, and A.H. Piersma. 2012. Compound-specific effects of diverse neurodevelopmental toxicants on global gene expression in the neural embryonic stem cell test (ESTn). Toxicol. Appl. Pharmacol. 262(3):330-340.

Thon, J.N., M.T. Devine, A. Jurak Bgeonja, J. Tibbitts, and J.E. Italiano Jr. 2012. High-content live-cell imaging assay used to establish mechanism of trastuzumab emtansine (T-DM1)–mediated inhibition of platelet production. Blood 120(10):1975-1984.

Threadgill, D.W., and G.A. Churchill. 2012. Ten years of the Collaborative Cross. Genetics 190(2):291-294.

Tonk, E.C., J.F. Robinson, A. Verhoef, P.T. Theunissen, J.L. Pennings, and A.H. Piersma. 2013. Valproic acid-induced gene expression responses in rat whole embryo culture and comparison across in vitro developmental and non-developmental models. Reprod. Toxicol. 41:57-66.

Tropepe, V., and H.L. Sive. 2003. Can zebrafish be used as a model to study the neurodevelopmental causes of autism? Genes Brain. Behav. 2(5):268-281.

Truong, L., K.S. Saili, J.M. Miller, J.E. Hutchison, and R.L. Tanguay. 2012. Persistent adult zebrafish behavioral deficits results from acute embryonic exposure to gold nanoparticles. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 155(2):269-274.

Truong, L., D.M. Reif, L. St Mary, M.C. Geier, H.D. Truong, and R.L. Tanguay. 2014. Multidimensional in vivo hazard assessment using zebrafish. Toxicol. Sci. 137(1):212-233.

Tse, A.C., K.Y. Lau, W. Ge, and R.S. Wu. 2013. A rapid screening test for endocrine disrupting chemicals using primary cell culture of the marine medaka. Aquat. Toxicol. 144-145:50-58.

Tyson, J.J., and B. Novák. 2010. Functional motifs in biochemical reaction networks. Annu. Rev. Phys. Chem. 61:219-240.

Valdivia, P., M. Martin, W.R. LeFew, J. Ross, K.A. Houck, and T.J. Shafer. 2014. Multi-well microelectrode array recordings detect neuroactivity of ToxCast compounds. Neurotoxicology 44:204-217.

Valerio, L.G., K.B. Arvidson, R.F. Chanderbhan, and J.F. Contrera. 2007. Prediction of rodent carcinogenic potential of naturally occurring chemicals in the human diet using high-throughput QSAR predictive modeling. Toxicol. Appl. Pharmacol. 222(1):1-16.

Van Vliet, E. 2011. Current standing and future prospects for the technologies proposed to transform toxicity testing in the 21st century. ALTEX 28(1):17-44.

Varani, J., P. Perone, D.M. Spahlinger, L.M. Singer, K.L. Diegel, W.F. Bobrowski, and R. Dunstan. 2007. Human skin in organ culture and human skin cells (keratinocytes and fibroblasts) in monolayer culture for assessment of chemically induced skin damage. Toxicol. Pathol. 35(5):693-701.

Walcott, B.P., and R.T. Peterson. 2014. Zebrafish models of cerebrovascular disease. J. Cereb. Blood Flow Metab. 34(4):571-577.

Wambaugh, J., and I. Shah. 2010. Simulating microdosimetry in a virtual hepatic lobule. PLoS Comput. Biol. 6(4):e1000756.

Wang, J.D., N.J. Douville, S. Takayama, and M. El Sayed. 2012. Quantitative analysis of molecular absorption into PDMS microfluidic channels. Ann. Biomed. Eng. 40(9):1862-1873.

Wheeler, H.E., C. Wing, S.M. Delaney, M. Komatsu, and M.E. Dolan. 2015. Modeling chemotherapeutic neurotoxicity with human induced pluripotent stem cell-derived neuronal cells. PLoS One 10(2):e0118020.

Wilcox, K.C., M.R. Marunde, A. Das, P.T. Velasco, B.D. Kuhns, M.T. Marty, H. Jiang, C.H. Luan, S.G. Sligar, and W.L. Klein. 2015. Nanoscale synaptic membrane mimetic allows unbiased high throughput screen that targets binding sites for Alzheimer's-associated Aβ oligomers. PLoS One 10(4):e0125263.

Williams, K.E., G.A. Lemieux, M.E. Hassis, A.B. Olshen, S.J. Fisher, and Z. Werb. 2016. Quantitative proteomic analyses of mammary organoids reveals distinct signatures after exposure to environmental chemicals. Proc. Natl. Acad. Sci. US 113(10):E1343-E1351.

Wills, L.P., G.C. Beeson, D.B. Hoover, R.G. Schnellmann, and C.C. Beeson. 2015. Assessment of ToxCast Phase II for mitochondrial liabilities using a high-throughput-respirometric assay. Toxicol Sci. 146(2):226-234.

Wu, S., J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, and K. Blackburn. 2013. Framework for identifying chemicals with structural features associated with the potential act as developmental or reproductive toxicants. Chem. Res. Toxicol. 26(12):1840-1861.

Xia, W., Y.J. Wan, X. Wang, Y.Y. Li, W.J. Yang, C.X. Wang, and S.Q. Xu. 2011. Sensitive bioassay for detection of PPARα potentially hazardous ligands with gold nanoparticle probe. J. Hazard Mater. 192(3):1148-1154.

Xu, J.J., P.V. Henstock, M.C. Dunn, A.R. Smith, J.R. Chabot, and D. de Graaf. 2008. Cellular imaging predictions of clinical drug-induced liver injury. Toxicol. Sci. 105(1):97-105.

Yamasaki, K., S. Kawasaki, R.D. Young, H. Fukuoka, H. Tanioka, M. Nakatsukasa, A.J. Quantock, and S. Kinoshita. 2007. Genomic aberrations and cellular heterogeneity in SV40-immortalized human corneal epithelial cells. Invest. Ophthalmol. Vis. Sci. 50(2):604-613.

Yoo, H.S., B.U. Bradford, O. Kosyk, S. Shymonyak, T. Uehara, L.B. Collins, W.M. Bodnar, L.M. Ball, A. Gold, and I. Rusyn. 2015a. Comparative analysis of the relationship between trichloroethylene metabolism and tissue-specific toxicity among inbred mouse strains: liver effects. J. Toxicol. Environ. Health A. 78(1):15-31.

Yoo, H.S., B.U. Bradford, O. Kosyk, T. Uehara, S. Shymonyak, L.B. Collins, W.M. Bodnar, L.M. Ball, A. Gold, and I. Rusyn. 2015b. Comparative analysis of the relationship between trichloroethylene metabolism and tissue-specific toxicity among inbred mouse strains: Kidney effects. J. Toxicol. Environ. Health A. 78(1):32-49.

Zhang, Q., S. Bhattacharya, M.E. Andersen, and R.B. Conolly. 2010. Computational systems biology and dose-response modeling in relation to new directions in toxicity testing. J. Toxicol. Environ. Health B Crit. Rev. 13(2-4):253-276.

Zhang, X., S. Wiseman, H. Yu, H. Liu, J.P. Giesy, and M. Hecker. 2011. Assessing the toxicity of napthenic acids using a microbial genome wide live cell reporter array system. Environ. Sci. Technol. 45(5):1984-1991.

Zhang, Z., N. Guan, T. Li, D.E. Mais, and M. Wang. 2012a. Quality control of cell-based high-throughput drug screening. Acta. Pharma. Sin. B 2(5):429-438.

Zhang, W., R. Korstanje, J. Thaisz, F. Staedtler, N. Harttman, L. Xu, M. Feng, L. Yanas, H. Yang, W. Valdar, G. A. Churchill, and K. Dipetrillo. 2012b. Genome-wide association mapping of quantitative traits in outbred mice. G3 (Bethesda) 2(2):167-174.

Zhang, Q., S. Bhattacharya, R.B. Conolly, H.J. Clewell, N.E. Kaminski, and M.E. Andersen. 2014. Molecular signaling network motifs provide a mechanistic basis for cellular threshold responses. Environ. Health Perspect. 122(12):1261-1270.

Zhang, Q., S. Bhattacharya, J. Pi., R.A. Clewell, P.L. Carmichael, and M.E. Andersen. 2015. Adaptive posttranslational control in cellular stress response pathways and its relationship to toxicity testing and safety assessment. Toxicol. Sci. 147(2):302-316.

Zou, F., J.A. Gelfond, D.C. Airey, L. Lu, K.F. Manly, R.W. Williams, and D.W. Threadgill. 2005. Quantitative trait locus analysis using recombinant inbred intercrosses: Theoretical and empirical considerations. Genetics 170(3):1299-1311.

Zuang, V., J. Barroso, S. Bremer, S. Casati, M. Ceridono, S. Coecke, R. Corvi, C.Eskes, A. Kinsner, C. Pellizzer, P. Prieto, A. Worth, and J. Kreysa. 2010. ECVAM Technical Report on the Status of Alternative Methods for Cosmetics Testing (2008-2009). Luxembourg: Publications Office of the European Union [online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-status-reports/files/ecvam-report-2008-2009 [accessed July 19, 2016].

# 4

# Advances in Epidemiology

Epidemiology is the study of health and disease in populations. Standard definitions of epidemiology emphasize a descriptive component that captures patterns of disease by person, place, and time and an etiological component that identifies causes of disease (Gordis 2013). The descriptive element of epidemiology comprises tracking of health and disease indicators and population risk factors (surveillance). The etiological activities—searching for the causes and determinants of disease—involve primarily case-control and cohort studies. The span of epidemiological research also includes intervention studies, both randomized and nonrandomized in the assignment of preventive measures, such as vaccinations, or other interventions.

This chapter addresses the evolving approaches used by epidemiologists to investigate the associations between environmental factors and human disease and the role of epidemiology in the context of the committee's charge regarding 21st century science related to risk-based decision-making. It does not give an overall introduction to the science of epidemiology; such material is readily available in textbooks and elsewhere. It briefly discusses, however, the role of epidemiology in risk assessment, the evolution of epidemiology, data opportunities now available, and types of biases to consider given the use of Tox21 and ES21 tools and methods. The chapter then focuses on the use of -omics technologies in epidemiology and concludes with some challenges and recommendations.

## RISK ASSESSMENT AND EPIDEMIOLOGY

The role of epidemiological evidence has long been established within the risk-assessment paradigm originally described in the report *Risk Assessment in the Federal Government: Managing the Process* (NRC 1983) and in various later reports (Samet et al. 1998). Identification of risk factors for disease and inference of causal associations from epidemiological studies provide important information for the hazard-identification component. Evidence on hazard obtained from epidemiological studies is given precedence in evidence-evaluation guidelines, including those of the US Environmental Protection Agency and the International Agency for Research on Cancer (IARC). Convincing epidemiological evidence that indicates a hazard is considered sufficient to establish causation, for example, in the IARC carcinogen classification scheme. However, human data are available on only a relatively small number of agents, particularly in comparison with the large number of environmental agents to which people are potentially exposed. In the absence of natural experiments, observational epidemiological studies are the only scientific approach available and ethically acceptable for studying possible effects of potentially harmful agents directly in human populations.

In addition to providing evidence for hazard identification, epidemiological studies can provide understanding of the exposure–response relationship. For some agents, the effects of exposure have been investigated primarily in particular groups of workers, such as asbestos workers, at exposure magnitudes typically much higher than those of the general population, and exposure–response relationships are extrapolated downward, introducing uncertainty. If the needed exposure data on a general population are available, epidemiological studies can provide key information on risk at exposure concentrations relevant to the population at large. For example, air-pollution exposures of participants in large cohort studies, including the American Cancer Society's Cancer Prevention Study 2 and the multiple studies involved in the European Study of Cohorts for Air Pollution Effects (ESCAPE 2014), have been estimated. Although some exposure misclassification is inherent in the case of most environmental and occupational exposures, there are numerous examples of successful incorporation of epidemiologically based exposure–response relationships into risk assessments: ionizing radiation and cancer, particulate-matter air pollution and mortality, arsenic exposure and cancer, and childhood lead exposure and neuropsychological development. Methods of addressing or correcting for measurement error have been developed; such corrections generally lead to exposure-response curves with steeper slopes (Hart et al. 2015).

Epidemiological studies can also contribute to understanding the exposure–response relationship by identifying determinants of susceptibility if information on characteristics of study participants (such as their age, sex, and now genomes) is available. Data collected for epidemiological research or for population surveillance can be useful for describing exposure distributions on the basis of questionnaires, monitoring, models, and analyses of biological specimens.

Epidemiological research might also provide information on overall population risk that fits into the risk-characterization component of risk assessment. The population attributable risk statistic, originally developed to estimate the burden of lung cancer caused by smoking, provides an estimate of the burden of disease resulting from a causal factor (Levin 1953). Thus, data on human populations can contribute to all four components of the risk-assessment paradigm described in Chapter 1.

## EPIDEMIOLOGY IN THE 21st CENTURY

### The Evolution of Epidemiology

The methods of epidemiological research have not been static. Initially, epidemiological research on the etiology of noncommunicable diseases—primarily cancer, cardiovascular diseases, pulmonary diseases, and metabolic diseases—focused on particular risk factors; exposure assessment was accomplished largely by using self-report questionnaires, measurement and estimation methods in the case of occupational studies, and relatively crude indicators in the case of environmental exposures. Some studies incorporated measurements from biological samples, such as lead or cadmium concentrations, and some estimated exposures with models that used extensive data. For example, in the study of survivors of the Hiroshima and Nagasaki atomic bombings, radiation dose was estimated with an elaborate algorithm that incorporated such information as location and body position at the time of the blast. Epidemiological studies of noncommunicable disease, carried out beginning in the 1950s, focused on risk factors at the individual level; some later studies began to incorporate risk determinants at higher levels of social or organizational structure, including the family, the places of residence and work, and the state and country. Efforts were made to build the studies around conceptual frameworks that reflected understanding of structural, sociological, and cultural factors driving health status and disease risk, and recent decades have seen increasing emphasis on life-course approaches that acknowledge the importance of early life exposures, even in utero and transgenerational, for disease risk. Furthermore, many later studies of the environment and health have been designed to reflect the variation in environmental exposures among and within communities.

Most recently, epidemiological research has been greatly affected by advances in other fields. The start of the 21st century was characterized by rapid advances in technology, medical sciences, biology, and genetics pertinent to epidemiology (Hiatt et al. 2013). Enhanced computing and data-storage capacity have been critical. The advent of genomics and genome-wide association studies (GWASs), for example, has played an important role in promoting the transformation of the practice of epidemiology.

The need to achieve samples large enough to provide studies that have adequate statistical power and the need to replicate novel findings in independent study populations facilitated the evolution of large epidemiological research teams, multicenter studies and consortia, meta-analytical tool development, and data-sharing etiquette. Recent decades have seen an evolution from single investigative teams that have proprietary control of individual datasets and specimens to the establishment of research consortia that have adopted a team-based science and a reproducibility culture through greater sharing of data, protocols, and analytical approaches (Guttmacher et al. 2009; Tenopir et al. 2011). Indeed, some funding agencies have sought to catalyze the transformation further by supporting the development and dissemination of validated state-of-the-science protocols designed to ascertain a broad array of phenotypic measures so that individual research teams (when designing new studies) might be positioned better to share and harmonize data among multiple studies (PhenX Toolkit NHGRI).

Case-control and cohort studies—the traditional workhorses of epidemiology—will continue to make strong contributions. Case-control studies, in particular, will continue to contribute to timely in-depth examination of people that have specific rare outcomes, such as rare cancers or reproductive outcomes, including specific birth defects. Cohort studies will continue to play an important role in aiding in the delineation of early antecedents of disease and the identification of preclinical biomarkers and risk factors and contribute to the foundation for translational research and precision medicine. Cohort studies, if started early enough, can be informative on the importance of early life exposures and their influence throughout the life course. The committee anticipates an increasing number of cohort studies that integrate treatment and health-outcome information from multiple sources, including information from health-care delivery systems. Studies that incorporate analysis of samples from companion biobanks will become key resources for connecting mechanisms identified in -omics and other assessments to pathogenesis in humans. Availability of more extensive geographical location information would allow incorporation of new and emerging data streams that document physical and social environments of populations on small scales into existing and new studies.

In summary, the factors reshaping the field of epidemiology in the 21st century include expansion of the interdisciplinary nature of the discipline; the increasing complexity of scientific inquiry that involves multilevel analyses and consideration of disease etiology and progression throughout the life course; emergence of new sources and technologies for data generation, such as new medical and environmental data sources and -omics technologies; advances in exposure characterization; and increasing demands to integrate new knowledge from basic, clinical, and population sciences (Lam et al. 2013). There is also a movement to register past and present datasets so that on particular issues data can be identified and combined. There are already models for data aggregation across studies (for example, National Cancer Institute Cohort Consortium and Agricultural Health cohorts), and researchers recognize the need for harmonizing data collection to facilitate future dataset aggregation (PhenX Toolkit NHGRI; Fortier et al. 2010). They are also considering how to create global biobanks (Harris et al. 2012).

## New Data Opportunities

Epidemiology has always been a discipline that uses large quantities of information with the goal of identifying risk factors that can be targeted in individuals or populations ultimately to reduce disease morbidity and mortality. Today, modern technologies—including genomic, proteomic, metabolomic, epigenomic, and transcriptomic platforms and sophisticated sensor and modeling techniques—facilitate the generation and collection of new types of data. The data can be used to generate hypotheses, but they can also be used to supplement data from legacy studies to strengthen their findings (see Box 4-1). New data opportunities have arisen from changes in how medicine is practiced, how health care is delivered, and how systems store and monitor health-care data (AACR 2015). Biobanks are being constructed by a variety of institutions that provide clinical care and potentially constitute new data sources.[1] They typically include collections of biological specimens (blood, urine, and surgical and biopsy specimens), clinical patient information that provides demographic and lifestyle information, perhaps a questionnaire on lifestyle and environmental and occupational exposures, and ascertainment of health outcomes from clinical records. Thus, human data and biosamples potentially available for application of various -omics and other technologies might come from opportunistic studies that rely on data sources that might have been collected and stored for nonresearch purposes. However, evidence from studies that use human tissue and medical data gained through convenience sampling from special populations might not be readily generalized. Furthermore, such studies carry the same potential for bias as other nonexperimental research data, but there is no opportunity with these studies to address some biases via a well-thought out study design, data collection, and protocols for obtaining biospecimens. Thus, new data streams and technologies, although promising, raise important methodological concerns and challenges and are driving the need to develop new study designs and analytical methods to account for technology-specific peculiarities (Khoury et al. 2013). Investigators have cautioned about the increasing possibility of false leads and dead ends with each new assay and have called for careful evaluation of analytical performance, reproducibility, concept

[1]The committee notes that biobanks are not a new creation. For example, the National Health and Nutrition Examination Survey, which is conducted for surveillance purposes, collects and analyzes specimens, and the data generated have proved invaluable for exposure assessment. Many other population-based biobanks have been created, usually by enrolling healthy subjects; the largest ones include the European Prospective Investigation into Cancer and Nutrition (IARC 2016) and the UK Biobank (2016).

**BOX 4-1** Using Legacy Studies

"Legacy" studies have accumulated substantial information on various environmental exposures, such as tobacco use, occupational exposures, and air pollution; personal factors, including genetic data; and disease events that have occurred over decades of follow-up. Some include biological-specimen banks and measures of disease phenotype and intermediate outcomes that were obtained by imaging, physiological testing, and other assessment methods. Some studies have already been used for application of -omics technologies (EXPoSOMICS 2016). Various cohorts have been used to address the association of ambient air pollution with disease incidence and mortality by adding estimates of air pollution at residence locations that were generated by new exposure models that have sufficient spatial resolution. Combining data from multiple studies provides an opportunity to gain statistical power and make results more precise while increasing the variety of exposures and the heterogeneity of study participants.

validity, and ethical and legal implications (Alsheikh-Ali et al. 2011; Khoury et al. 2013).

The tsunami of data spanning the spectrum of genomic, molecular, clinical, epidemiological, environmental, and digital information is already a reality of 21st century epidemiology (Khoury et al. 2013). There are challenges in using current methods to process, analyze, and interpret the data systematically and efficiently or to find relevant signals in potential oceans of noise. To address those issues, the US government in 2012 announced the "Big Data" Initiative and committed funds to support research in data science in multiple agencies (Mervis 2012). Epidemiologists are poised to play a central role in shaping the directions and investment in building infrastructures for the storage and robust analysis of massive and complex datasets. Given experience with multidisciplinary teams, epidemiologists are also equipped to direct the interpretation of the data in collaboration with experts in clinical and basic health sciences, biomedical informatics, computational biology, mathematics and biostatistics, and exposure sciences. Adaptation of technological advances, such as cloud computing, and strategic formation of new academic–industry partnerships to facilitate the integration of state-of-the-art computing into biomedical research and health care (Pechette 2012) are only some of the initial challenges that must be confronted before new data opportunities can be properly and effectively integrated into future epidemiological studies.

### Types of Biases and Challenges Related to External Validity

As noted, contemporary epidemiology is faced with an unprecedented proliferation of clinical and health-care administrative data, -omics data, and social and environmental data. The biases that generally affect epidemiological evidence can be grouped into three broad categories: *information bias* that arises from error in measurements of exposure or outcome variables and co-variates, *selection bias* that arises from the ways in which participants are chosen to take part in epidemiological studies, and *confounding* that arises from the mingled effects of exposures of interest and other exposures. *External validity* refers to the generalizability of findings and is a key consideration in risk assessment. Understanding the selection processes, measurement accuracy, and interpretation of analyses is critical for using epidemiological data in risk assessment, including the new and perhaps large cohorts that will be created from health-care databases and combined with exposure estimates.

The multiplicity, diversity, and size of data sources have generated widespread enthusiasm in researchers about the new possibilities (Roger et al. 2015a,b). There will, however, be some challenges in using the data. For example, reliance on electronic medical records as a sole basis for assembling cohorts might accentuate sample-selection biases because of health-care–seeking behaviors of patients; promote misclassification or incomplete documentation of phenotypes, clinical diagnoses, and procedures because of vagaries in clinical coding incentives and practices; and lead to confounding because key factors needed to evaluate confounding are not routinely collected in medical records, particularly those associated with environmental exposures. Although electronic record systems might support the generation of large cohorts for investigations, having a large sample size does not mitigate the potential for biases, and it increases the likelihood of statistically significant false-positive findings. Furthermore, electronic medical records typically contain little information on occupational and environmental exposures, linkage to exposure databases might be problematic, and information on important potential confounders, such as tobacco use, might be sparse and not collected in the standardized fashion needed for research.

In evaluating risks posed by environmental agents, epidemiologists and exposure scientists typically work together to enhance exposure estimates used in epidemiological studies by broadening the variety of exposures considered, increasing precision of exposure measures, and providing insights into errors that inevitably affect exposure estimates. The full array of advances in exposure science that are described in the ES21 report (NRC 2012) and in Chapter 2 of the present report have application in epidemiological studies. When exposure methods are appropriately incorporated into the study design, they facilitate exploration of measurement error in exposure variables and covariates. Such error has long been considered a serious limitation of epidemiological evidence in risk-assessment contexts; nonrandom errors can bias apparent effects upward or downward, and random error generally obscures associations and dose–response relationships. Measurement-error corrections can be made by using data from validation studies and statistical models that have been developed over the last 2 decades and applied, for example, to studies on diet and disease risk, radiation and cancer, and air pollution and health (Li et al. 2006; Freedman et al. 2015; Hart et al. 2015).

## EPIDEMIOLOGY AND -OMICS DATA

Historically, epidemiological research has incorporated emerging technologies into new and current studies. The need to incorporate new science, however, accelerated several decades ago with the introduction of the paradigm of molecular epidemiology. The new paradigm emerged as a replacement of "black box" epidemiology, an approach that examined associations of risk factors with disease while not addressing the intervening mechanisms. The molecular-epidemiology paradigm opens the black boxes through the incorporation of biomarkers of

exposure, susceptibility, and disease. It stresses the importance of pathways and their perturbation, which is highly relevant to the opportunities provided by 21st century science and specifically -omics technologies. The approach also strengthens the evidence base for one of Bradford Hill's guidelines for causality: understanding of biological plausibility (see Chapter 7). For example, carcinogenesis is thought to be a multifactorial process in which mutations and selective microenvironments play critical roles, and key steps of the process can be explored with biomarkers. The molecular-epidemiology paradigm is a general one and conceptually accommodates emerging methods for generating biomarker data.

As indicated, molecular-epidemiology research is focused on underlying biology (exposure and disease pathogenesis) rather than on empirical observation. Thus, as -omics technologies have emerged, they have been integrated into current studies and have affected study design, particularly specimen collection and management. The incorporation of -omics approaches dates back about 2 decades, beginning with the genomic revolution. In some of the current cohort studies, blood samples that had been appropriately stored were analyzed for single-nucleotide polymorphisms (SNPs) and other markers to search for genes associated with disease risk, including those modifying risk associated with environmental agents.

The utility of bringing -omics technologies into epidemiological research is already clear as exemplified by many studies that have incorporated genomics. One well-known starting point for exploring the genetic basis of disease has been GWAS, which involves the comparison of genomic markers in people who have and people who do not have a disease or condition of interest. The list of -omics approaches applied in epidemiological research has now expanded beyond genomics to include epigenomics, proteomics, transcriptomics, and metabolomics (see Box 1-1). Table 4-1 lists advantages and disadvantages of their use. Examples of their use in a specific context are provided in Appendix B, which describes the meaning and limitations of -omic approaches in the context of epidemiological research on air pollution. Although the new methods have the potential to bring new insights from epidemiological research, there are many challenges in applying them. Some new studies are being designed with the intent of prospectively storing samples that can be used for existing and future -omics technologies, for example, in the case of the EU-funded projects Helix and EXPOsOMICS described in Chapter 1. Obtaining data from human population studies that are parallel to data that can be obtained from in vitro and in vivo toxicity assessments is already possible and offers the possibility of harmonizing comparisons of exposure and dose.

In principle, the -omics approaches now support nontargeted explorations of genes with genomics, mRNA with transcriptomics, proteins with proteomics, and metabolites with metabolomics. With the exception of genomics, the measurements usually reflect changes within cells at one or a few points in time only, and the tissues that are used in humans are primarily surrogates, such as blood, urine, and saliva. Combining different -omics tools, however, increases the possibility for a better understanding of how different external exposures interact with internal molecules, for example, by inducing mutations (genomics), causing epigenetic changes (epigenom-

**TABLE 4-1** Advantages and Limitations of -Omics Technologies

| | |
|---|---|
| Advantages | Use in large, hypothesis-free investigations of the whole complement of relevant biological molecules. |
| | Better understanding of phenotype–genotype relations. |
| | Might provide insights into the effects of interactions between environmental conditions and genotypes and mechanistic insights into disease aetiology. |
| Limitations | There are limitations arising from cost of assays, quality of biological material available (such as instability of RNAs), and the amount of labor needed. |
| | Techniques that are still in their discovery state and new leads need to be carefully investigated and compared with existing biological information from in vivo and in vitro tests. |
| | New leads in the discovery of novel intermediate markers need to be confirmed in other independent studies preferably with different platforms. |
| | Moving from promising techniques to successful application of biomarkers in occupational and environmental medicine requires not only standardizing and validating techniques, but also appropriate study designs and sophisticated statistical analyses for interpreting study results especially for untargeted approaches (the issue of multiple comparisons and false positives). |

Source: Adapted from Vineis et al. 2009.

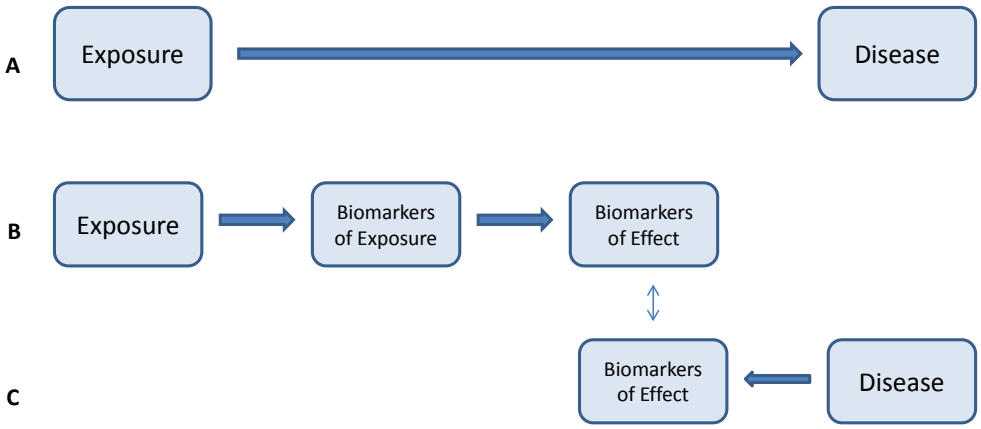



**FIGURE 4-1** The meet-in-the-middle approach centers on investigating (A) the association between exposure and disease, (B) the relationship between exposure and biomarkers of exposure or effect, and (C) the relationship between disease and biomarkers of exposure or effect.

ics), or modifying the internal cell environment in more complex ways. The latter changes might be monitored with proteomics, transcriptomics, or metabolomics.

### Meet-in-the-Middle Approach

One informative strategy for the integration of -omics technologies into epidemiological research is the meet-in-the-middle approach (Vineis et al. 2013). The approach provides insights into biological plausibility that can bolster causal inference. In the context of a population study, the approach generally involves a prospective search for intermediate biomarkers that are linked to the underlying disease and are increased in those who eventually develop disease, and a retrospective search that links the intermediate biomarkers to past exposures of the environmental agent of concern. As illustrated in Figure 4-1, the approach can be considered as three steps: an investigation into the association between exposure and disease, an assessment of the relationship between exposure and biomarkers of exposure and early effects, and an assessment of the relationship between the disease outcome and intermediate biomarkers. Inference of a causal relationship between exposure and disease is strengthened if associations are documented for each of the three key relationships in Figure 4-1, corresponding to A, B, and C.

A recent study of epigenetics and lung cancer (Fasanelli et al. 2015) is illustrative. The biomarkers are methylation status of the AHRR gene and the F2RL gene, which are hypomethylated in smokers (exposure in Figure 4-1B) (Vineis et al. 2013; Guida et al. 2015). Hypomethylation of the genes is also associated with lung cancer (disease in Figure 4-1C). The question is, Are those biomarkers on the causal pathway for lung cancer caused by smoking? Fasanelli et al. (2015) showed by using the statistical technique of mediation analysis that 37% of lung cancers could be explained by the methylation status of the two genes. Thus, the two genes are biomarkers that are likely to be on the causal pathway and illustrate the "meeting in the middle" of the exposure and the disease, the middle being the biomarker. The committee notes, however, that fully assessing causality requires additional steps beyond statistical analysis.

### Exposome-Wide Association Studies

As defined in Chapter 1, *exposome* refers to the totality of exposures from conception to death. Some have questioned whether the exposome as defined defies practical measurement and is therefore not amenable to scientific methods (Miller and Jones 2014). In an attempt to define the exposome as a measurable entity, Rappaport and Smith (2010) proposed to consider first the body's internal chemical environment and how the body responds to these chemical exposures.[2] They referred to the exposures as the internal exposome and distinguished it from the external exposome—exposures external to the body—and suggested that the internal and external exposomes are complementary. For example, internal assessment might identify environmental health associations (that is, generate new hypotheses on disease etiology), but external exposure assessments are needed to identify sources, consider exposure routes, and address spatial and

[2]The inclusion of biological response in the concept helps to expand beyond external chemical exposures to many types of exposures—including psychological or physical stress, infections, and gut flora—that produce endogenous chemicals, such as oxidative molecules, and disease-producing responses, such as inflammation, oxidative stress, and lipid peroxidation.

temporal variability of exposures (Turner et al. in press). Consequently, an external-exposome assessment can take place after hypotheses have been generated, and the environmental sources of internal changes can be sought. The two study designs—one that looks for internal changes starting from external measurements (external-exposome assessment) and one that looks for external sources on the basis of internal signals (internal-exposome assessment)—are complementary and have been defined as "bottom-up" and "top-down" approaches, respectively.

The -omics tools that can be used to capture the internal exposome make nontargeted analyses that parallel GWASs in concept and approach possible. Studies of that design have been referred to as exposome-wide association studies (EWASs).[3] Specifically, the EWAS approach involves the investigation of associations of a large number of small molecules, proteins, or lipids with disease or intermediate phenotypes to identify biomarkers of exposure or disease. One general EWAS approach to generate new hypotheses on disease causation has been described by Rappaport and Smith (2010). Figure 4-2 shows a study design that can lead to the generation of new hypotheses about chemical hazards in the context of a case–control study. Targeted and nontargeted metabolomics approaches are used to compare exposures of cases that have a specific disease with exposures of ones that do not (controls). After the initial discovery phase, the experimental design can be improved by a testing (replication) phase with a prospective context (a case-control study that is nested in a prospective cohort). That approach takes temporality into account by using biological samples collected before disease manifestation to avoid or to reduce the potential for reverse causation. Unidentified features that are significantly associated with the outcomes of interest would next be chemically identified by using methods described in Chapter 2, for example, by using NMR, IMS-MS/MS, or cheminformatics or by synthesizing and evaluating chemical standards for candidate chemicals. In the next step, validation of the association and a final causal assessment would be attempted through replication in more than one cohort, and biological plausibility would be evaluated.

Biological plausibility could be evaluated with a targeted analysis of available human tissues by using proteomics, metabolomics, or other methods to search for biological responses related to the disease. Alternatively, novel animal models or high-throughput in vitro assays described in Chapter 3 could be used to test candidate chemicals and generate biological-response data that could be compared with responses related to the EWAS-identified association with disease. Evaluation of biologi-

[3]The committee notes that the acronym EWAS was originally proposed by Patel et al. (2010) to refer to environment-wide association studies, but others, such as Rappaport (2012), have used EWAS to refer more specifically to exposome-wide association studies, as used here by the committee.

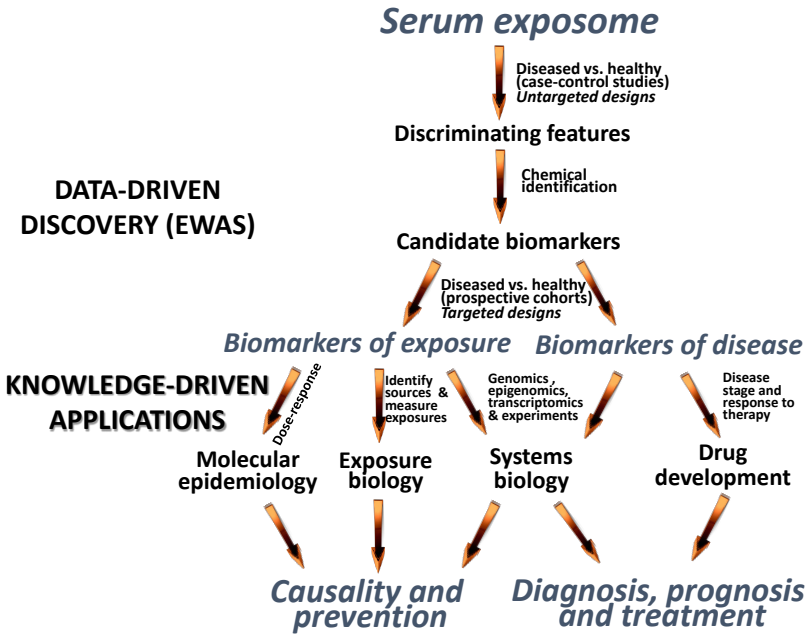



**FIGURE 4-2** A study design for developing new hypotheses on causation of disease by exposure. The committee notes that the approach and tools used to investigate exposures and biological pathways for "causality and prevention" are not necessarily different from those used to investigate biological pathways relevant for drug development. Source: Rappaport 2012.

cal plausibility would ideally also include refinement of exposure, if necessary, and a systematic comparison of human exposures to exposures in test systems that are used to produce the supporting biological-response data. If similar toxicity data and models are used, responses to exposures in cohort members could be directly compared with those in test systems; the comparison would provide additional evidence on the likelihood of biological plausibility, which would be greater if responses to exposure were similar, and smaller if they were not. An example of the approach described was used to investigate colon cancer. The research began with three cross-sectional case–control studies and found an association between an unidentified metabolomic feature (analyte) and colon cancer (Ritchie et al. 2013). The association was later confirmed prospectively in the European Prospective Investigation into Cancer and Nutrition cohort, and the metabolic feature was identified as belonging to a group of ultra–long-chain fatty acids (Perttula et al. 2016).

The EWAS approach offers exciting opportunities, but there are challenges that need to be addressed. The challenges in using tools that produce "big data" are similar to those encountered in all multiexposure studies. The study design and analysis have to be chosen carefully and assessed in terms of all classic biases to establish causality, that is, using principles that apply to targeted designs that focus on a single exposure and outcome. The EWAS approach adds the challenge of determining which exposures among many correlated ones have a causal role and which reflect a biological perturbation caused by other agents. The temporal dynamics of the exposures need to be addressed with the stability of media concentrations. An additional premise of the EWAS approach is that useful, biologically informative biomarkers can be identified, that is, that the chemicals in question are not too short-lived and exposure not too sporadic to be captured by only one or a few biospecimens obtained in a cross-sectional survey or cohort study.

The committee notes that use of retrospective case–control design for EWAS makes it impossible to be certain if associations observed reflect a causal relationship between exposures and the outcome investigated or if the associations are a consequence of the disease or its treatment. As summarized by Thomas et al. (2012), the technique of Mendelian randomization (Davey Smith et al. 2004) is one way to address reverse causation and uncontrolled confounding; a gene is used as an instrumental variable (Greenland 2000) to evaluate the causal effect of a biomarker on disease risk. In an approach that parallels the meet-in-the-middle approach, a novel two-step extension of this idea has been proposed for methylation studies that uses two genes as instrumental variables: one estimates the exposure–methylation association, and the other the methylation–disease association (Cortessis et al. 2012; Relton and Davey Smith 2012). There is an inherent assumption in that approach that the instrumental variable is indeed an appropriate instrument for exposure.

## New Analytical Challenges

There are formidable challenges in integrating the -omics technologies and data into epidemiological research, and robust high-dimensional analytical techniques will be required to integrate and analyze all the data. For example, statistical analyses that consider many exposure variables simultaneously without strong priors, such as in EWASs, greatly increase the risk of observing random associations (false positives) because of multiple testing. Therefore, statistical tools for the analysis of multiple exposures have motivated investigators to draw on important lessons learned from the analysis of GWAS data (Shi and Weinberg 2011; Thomas et al. 2012); some are described below. In general, statistical techniques for high-dimensional data—such as those noted and others, including machine learning, dimension reduction, and variable-selection techniques—must be adapted to the longitudinal-data-accrual context to account for such issues as time-varying exposure and delayed effects (Buck Louis and Sundaram 2012).

Multistep analytical approaches have been used to estimate health risks associated with different types or combinations of exposures. For example, estimates from EWAS analytical approaches with no a priori information might be quantified by using classical regression models while controlling for false discovery rate, as is done in GWASs (Patel et al. 2010, 2013; Vrijheid et al. 2014). Furthermore, flexible and smoothing modeling techniques (Slama and Werwatz 2005) might be used to identify and characterize possible thresholds or exposure–response relationships.

Pathway analytical approaches are increasingly used for integrating and interpreting high-dimensional data generated by multiple -omics techniques; these approaches have enabled analyses of relationships between multiple exposures and multiple health outcomes. It is noteworthy that pathway analytical approaches have been used to identify molecular signatures associated with environmental agents through exploratory analyses of metabolites, proteins, transcripts, and DNA methylation in biological samples (Jennen et al. 2011; Vrijheid et al. 2014). As summarized by Vrijheid et al. (2014), once biomarkers have been identified, available libraries of biological pathways—such as Gene Ontology (Ashburner et al. 2000), Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto 2000), Reactome (Fabregat et al. 2016), and Comparative Toxicogenomics Database (Davis et al. 2015)—can be searched and used to identify relevant biological pathways affected by exposures whether alone or in combination. Furthermore, biological pathways can be grouped and described using available soft-

ware, such as Ingenuity Pathway Analysis (Krämer et al. 2014), Cytoscape (Saito et al. 2012), and Impala (Kamburov et al. 2011). For example, those analytical approaches have been applied to several types of -omics data from systems that respond to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and to a broader set of environmental and pharmacological agents (Jennen et al. 2011; Kamburov et al. 2011).

Other methods are also available to address the new analytical challenges. First, analysis of covariance techniques has been used to integrate individual exposures (obtained, for example, from personal wearable devices) and outdoor exposures (obtained, for example, from environmental monitoring) by exploring the variance components of key exposures arising from multiple sources before creating exposure groups or clusters. Second, factor analysis and latent class analysis have proved useful for creating reduced sets of exposure indexes on the basis of commonly occurring exposures while allowing people who share similar exposure profiles to be grouped. Third, to address the high-dimensional nature of epigenetic data, cluster-analysis techniques developed by Siegmund et al. (2006) can be applied to exposome-wide association-genomic studies; these techniques treat the cluster rather than individual epigenetic marks as a latent risk factor for disease (Cortessis et al. 2012). Fourth, structural equation modeling approaches might be used to define combined exposure variables on the basis of knowledge summarized by directed acyclic graphs (Budtz-Jørgensen et al. 2010).

Bayesian profile regression models might be used to identify groups of people who have a similar exposome but show marked differences in the health-outcome variable of interest (Molitor et al. 2010; Papathomas et al. 2011; Vrijheid et al. 2014). Model-based clustering would be applied to the exposure data while allowing the outcome of interest to influence cluster membership. The Bayesian model–based clustering technique has been used, for example, to identify a cluster in a high-risk set for lung cancer—a group who has the characteristics of living near a main road, having high exposure to $PM_{10}$ (particulate matter with aerodynamic diameter $\leq 10$ μm) and to nitrogen dioxide, and carrying out manual work (Papathomas et al. 2011; Vrijheid et al. 2014).

The general need for caution in contending with the potential for false-positive associations that arise from analysis of large datasets is generally recognized among those handling such data. In addition to analytical approaches, such as correcting p values for multiplicity and using such parameters as the false-discovery rate, the committee notes that epidemiological findings are interpreted holistically in the context of other relevant evidence. In the context of risk assessment, hazard identification would rarely, if ever, be based on an association found in a single epidemiological study, absent additional evidence.

## CHALLENGES AND RECOMMENDATIONS FOR ADVANCING EPIDEMIOLOGY

With the emergence of Tox21 and ES21 approaches, the committee anticipates new connections between biomarkers and human health outcomes. Epidemiological studies have an implicit role in providing the population counterpart that is needed to interpret biomarkers measured in laboratory studies through the general paradigm of molecular epidemiology and the meet-in-the-middle approach. For that purpose, epidemiologists need to generate human data (1) to harmonize doses used in in vitro high-throughput assays with those associated with the exposures experienced in the population setting, (2) to explore the relevance of pathways identified in assay systems to human responses to the same agents and validate the predictive value of pathways detected in vitro assays for the occurrence of human disease, (3) to develop and validate models of human susceptibility, and (4) to compare and corroborate exposure–response relationships obtained from in vitro assays and in human populations.

The overall goal of gaining new insights by connecting -omics data generated in laboratory with data gathered in population contexts will not be achieved without consideration of the needed research infrastructure and the logistical barriers to bringing together datasets from disparate sources. The committee concludes by highlighting some challenges that face epidemiological research and recommendations for addressing them. The committee notes that several recommendations below call for developing or expanding databases. In all cases, data curation and quality evaluation should be routine in database development and maintenance.

### Developing the Infrastructure and Methods Needed to Advance the Science

*Challenge*: When used in epidemiological studies, particularly ones with large biobank cohorts that might reach a million or more participants, -omics assays can generate large databases that need to be managed and curated in ways that will facilitate access and analysis. There is an additional challenge of analyzing extremely large datasets by using a hypothesis-driven or exploratory approach.

*Recommendation*: Resources should be devoted to accelerating development of database management systems that will accommodate extremely large datasets, support analyses for multiple purposes, and foster data-sharing and development of powerful and robust statistical techniques for analyzing associations of health outcomes with -omics data and exploring such complex problems as gene–environment interactions. Such efforts are already under way in a number of fields, such as clinical research

that involves health-care data, and should be extended to epidemiological research.

*Challenge*: Standard methods are needed to describe the data that have been generated and that are shared among disciplines. The problem has been recognized in genomics and has led to the development of annotated gene ontologies, and similar approaches could be extended to other types of -omics data.

*Recommendation*: Ontologies should be developed and expanded so that data can be harmonized among investigative groups, internationally, and among -omic platforms. Such ontologies generally do not incorporate data collected by epidemiologists. Such tools as STROBE should be expanded and adapted to the new generation of epidemiological studies; STROBE has already been expanded to encompass molecular epidemiology (Gallo et al. 2011). The Framework Programme 7 EU Initiative—coordination of standards in metabolomics (COSMOS)—is developing "a robust data infrastructure and exchange standards for metabolomics data and other metadata" (Salek et al. 2015); this type of approach should be extended to other -omics data.

### Data-Sharing

*Challenge*: Data-sharing involves many complexities, particularly when the data are from human studies. However, data-sharing could be particularly beneficial if data could be accessed in a way that would support uniform analyses and integration through hierarchical analyses or meta-analysis. Data-sharing could also lead to more powerful assessments of hazard and of exposure–response relationships. One useful example is the pooling of data from studies of radon-exposed underground miners that supported the development of risk models for indoor radon (Lubin et al. 1995).

The same issues surrounding data-sharing arise in other domains in which big-data approaches are emerging, and a general culture of data-sharing will be needed. Regarding genomics, posting of sequencing data has become the norm but with attention to anonymity. Similar sharing will ideally extend to other -omics data and lead to the development of a culture of data-sharing, pragmatic solutions to the inherent ethical problems, and standardized ontologies and databases. The committee notes that discussion around data-sharing is moving rapidly with regard to clinical trials; similar efforts around observational data are needed (Mascalzoni 2015).

*Recommendation*: Steps should be taken to ensure sharing of observational data relevant to risk assessment so that, for example, biomarkers can be validated among populations. As noted above, to achieve that goal, standard ontologies should be developed and used for capturing and coding key variables. There is also a need for systematic exploration of possible logistical and ethical barriers to sharing potentially massive datasets drawn from human populations.

### Collaborating and Training the Next Generation of Scientists

*Challenge*: New research models based on biobanks and large cohorts derived from clinical populations will become a valuable resource for applying -omics and other biomarker assays, but there are intrinsic limitations related to biases and the scope of data available in electronic records. There are also complicated issues related to access to private and confidential medical records and to sharing of such data.

*Recommendation*: As biobanks and patient-based cohorts are developed, those developing them should engage with epidemiologists and exposure scientists on the collection of exposure data to ensure that the best and most comprehensive data possible are collected in this context. Finding ways to capture exposure information will be particularly challenging and will likely require ancillary data collection in nested studies.

*Challenge*: A wide array of biospecimens is being collected and stored on the assumption that they will be useful in the future for a variety of purposes, including assays that cannot be anticipated. Storage methods and consent procedures need to support future use.

*Recommendation*: Epidemiologists should anticipate future uses of biospecimens that are collected in the course of epidemiological research or other venues, such as screening or surveillance, and ensure that the array of specimens and their handling and storage will support multiple assays in the future. Such future-looking collections should be a design consideration, and input should be obtained from scientists who are developing new assays.

*Challenge*: A new generation of researchers who can conduct large-scale population studies and integrate -omics and other emerging technologies into population studies is needed. The next generation also needs sufficient multidisciplinary training to be able to interact with exposure and data scientists.

*Recommendation*: The training of epidemiologists should be enriched with the addition of more in-depth understanding of the biological mechanisms underlying human diseases and of the biomarker assays used to probe them.

*Challenge*: The landscape of epidemiological research is changing quickly with a move away from the fixed legacy cohorts of the past, such as the Nurses' Health Study, to pragmatically developed cohorts that are grounded in new and feasible ways of cohort identification

and follow-up. There are also likely to be large national cohorts, such as the cohort already under development for the Precision Medicine Initiative. Those cohorts are intended as platforms for a wide array of research questions; they are designed as large banks of biospecimens but will have inherent limitations regarding the exposure information available.

*Recommendation*: Epidemiologists, exposure scientists, and laboratory scientists should collaborate closely to ensure that the full potential of 21st century science is extended to and incorporated into epidemiological research. Multidisciplinarity should be emphasized and sought with increasing intensity. As the new cohorts are developed, the opportunity to ensure that they will be informative on the risks posed by environmental exposures should not be lost.

## REFERENCES


AARC (American Association for Cancer Research). 2015. AACR Cancer Progress Report 2015 [online]. Available: http://cancerprogressreport.org/2015/Documents/AACR_CPR2015.pdf [accessed July 21, 2016].

Alsheikh-Ali, A.A., W. Qureshi, M.H. Al-Mallah, and J.P. Ioannidis. 2011. Public availability of published research data in high-impact journals. PLoS One 6(9):e24357.

Ashburner, M., C.A. Ball, J.A Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25(1):25-29.

Buck Louis, G.M., and R. Sundaram. 2012. Exposome: Time for transformative research. Stat. Med. 31(22):2569-2575.

Budtz-Jørgensen, E., F. Debes, P. Weihe, and P. Grandjean. 2010. Structural equation models for meta-analysis in environmental risk assessment. Environmetrics 21(5):510-527.

Cortessis, V.K., D.C. Thomas, A.J. Levine, C.V. Breton, T.M. Mack, K.D. Siegmund, R.W. Haile, and P.W. Laird. 2012. Environmental epigenetics: Prospects for studying epigenetic mediation of exposure–response relationships. Hum. Genet. 131(10):1565-1589.

Davey Smith, G., R. Harbord, and S. Ebrahim. 2004. Fibrinogen, C-reactive protein and coronary heart disease: Does Mendelian randomization suggest the associations are non-causal? QJM 97(3):163-166.

Davis, A.P., C.J. Grondin, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B.L. King, T.C. Wiegers, and C.J. Mattingly. 2015. The Comparative Toxicogenomics Database's 10th year anniversary: Update 2015. Nucleic Acids Res. 43(Database issue):D914-D920.

ESCAPE (European Study of Cohorts for Air Pollution Effects). 2014. ESCAPE Project [online]. Available: http://www.escapeproject.eu/index.php [accessed July 21, 2016].

EXPOsOMICS. 2016. About EXPOsOMICS [online]. Available: http://www.exposomicsproject.eu/ [accessed July 21, 2016].

Fabregat, A., K. Sidiropoulos, P. Garapati, M. Gillespie, K. Hausmann, R. Haw, B. Jassal, S. Jupe, F. Korninger, S. McKay, L. Matthews, B. May, M. Milacic, K. Rothfels, V. Shamovsky, M. Webber, J. Weiser, M. Williams, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. 2016. The Reactome pathway knowledgebase. Nucleic Acids Res. 44(D1):D481-D487.

Fasanelli, F., L. Baglietto, E. Ponzi, F. Guida, G. Campanella, M. Johansson, K. Grankvist, M. Johansson, M.B. Assumma, A. Naccarati, M. Chadeau-Hyam, U. Ala, C. Faltus, R. Kaaks, A. Risch, B. De Stavola, A. Hodge, G.G. Giles, M.C. Southey, C.L. Relton, P.C. Haycock, E. Lund, S. Polidoro, T.M. Sandanger, G. Severi, and P. Vineis.2015. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat. Commun. 6:10192.

Fortier, I., P.R. Burton, P.J. Robson, V. Ferretti, J. Little, F. L'Heureux, M. Deschênes, B.M. Knoppers, D. Doiron, J.C. Keers, P. Linksted, J.R. Harris, G. Lachance, C. Boileau, N.L. Pedersen, C.M. Hamilton, K. Hveem, M.J. Borugian, R.P. Gallagher, J. McLaughlin, L. Parker, J.D. Potter, J. Gallacher, R. Kaaks, B. Liu, T. Sprosen, A. Vilain, S.A. Atkinson, A. Rengifo, R. Morton, A. Metspalu, H.E. Wichmann, M. Tremblay, R.L. Chisholm, A. Garcia-Montero, H. Hillege, J.E. Litton, L.J. Palmer, M. Perola, B.H. Wolffenbuttel, L. Peltonen, and T.J. Hudson. 2010. Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies. Int. J. Epidemiol. 39(5):1383-1393.

Freedman, L.S., J.M. Commins, J.E. Moler, W. Willett, L.F. Tinker, A.F. Subar, D. Spiegelman, D. Rhodes, N. Potischman, M.L. Neuhouser, A.J. Moshfegh, V. Kipnis, L. Arab, and R.L. Prentice. 2015. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. Am. J. Epidemiol. 181(7):473-487.

Gallo, V., M. Egger, V. McCormack, P.B. Farmer, J.P. Ioannidis, M. Kirsch-Volders, G. Matullo, D.H. Phillips, B. Schoket, U. Stromberg, R. Vermeulen, C. Wild, M. Porta, and P. Vineis. 2011. STrengthening the Reporting of OBservational studies in Epidemiology-Molecular Epidemiology (STROBE-ME): An extension of the STROBE statement. PLoS Med. 8(10):e1001117.

Gordis, L. 2013. Epidemiology, 5th Ed. Philadelphia: Elesevier and Saunders. 416 pp.

Greenland, S. 2000. An introduction to instrumental variables for epidemiologists. Int. J. Epidemiol. 29(4):722-729.

Guida, F., T.M. Sandanger, R. Castagné. G. Campanella, S. Polidoro, D. Palli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, G. Severi, S.A. Kyrtopoulos, P. Georgiadis, R.C. Vermeulen, E. Lund, P. Vineis, and M. Chadeau-Hyam. 2015. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. Hum. Mol. Genet. 24(8):2349-2359.

Guttmacher, A.E., E.G. Nabel, and F.S. Collins. 2009. Why data-sharing policies matter. Proc. Natl. Acad. Sci. US 106(40):16894.

Harris, J.R., P. Burton, B.M. Knoppers, K. Lindpaintner, M. Bledsoe, A.J. Brookes, I. Budin-Ljøsne, R. Chisholm, D. Cox, M. Deschênes, I. Fortier, P. Hainaut, R. Hewitt, J. Kaye, J.E. Litton, A. Metspalu, B. Ollier, L.J. Palmer, A. Palotie, M. Pasterk, M. Perola, P.H. Riegman, G.J. van Ommen, M. Yuille, and K. Zatloukal. 2012. Toward a roadmap in global biobanking for health. Eur. J. Hum. Genet. 20(11):1105-1111.

Hart, J.E., X. Liao, B. Hong, R.C. Pruett, J.D. Yanosky, H. Suh, M.A. Kiomourtzoglou, D. Spiegelman, and F. Laden. 2015. The association of long-term exposure to PM 2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environ. Health. 14:38.

Hiatt, R.A., S. Sulsky, M.C. Aldrich, N. Kreiger, and R. Rothenberg. 2013. Promoting innovation and creativity in epidemiology for the 21st century. Ann. Epidemiol. 23(7):452-454.

IARC (International Agency for Research on Cancer). 2016. The European Prospective Investigation into Cancer and Nutrition (EPIC) Study [online]. Available: http://epic.iarc.fr/ [accessed July 21, 2016].

Jennen, D., A. Ruiz-Aracama, C. Magkoufopoulou, A. Peijnenburg, A. Lommen, J. van Delft, and J. Kleinjans. 2011. Integrating transcriptomics and metabonomics to unravel modes-of-action of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in HepG2 cells. BMC Syst. Biol. 5:139.

Kamburov, A., R. Cavill, T.M. Ebbels, R. Herwig, and H.C. Keun. 2011. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics 27(20):2917-2918.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28(1):27-30.

Khoury, M.J., T.K. Lam, J.P. Ioannidis, P. Hartge, M.R. Spitz, J.E. Buring, S.J. Chanock, R.T. Croyle, K.A. Goddard, G.S. Ginsburg, Z. Herceg, R.A. Hiatt, R.N. Hoover, D.J. Hunter, B.S. Kramer, M.S. Lauer, J.A. Meyerhardt, O.I. Olopade, J.R. Palmer, T.A. Sellers, D. Seminara, D.F. Ransohoff, T.R. Rebbeck, G. Tourassi, D.M. Winn, A. Zauber, and S.D. Schully. 2013. Transforming epidemiology for 21st century medicine and public health. Cancer Epidemiol. Biomarkers Prev. 22(4):508-516.

Krämer, A., J. Green, J. Pollard, and S. Tugendreich. 2014. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30(4):523-530.

Lam, T.K., M. Spitz, S.D. Schully, and M.J. Khoury. 2013. "Drivers" of translational cancer epidemiology in the 21st century: Needs and opportunities. Cancer Epidemiol. Biomarkers Prev. 22(2):181-188.

Levin, M.L. 1953. The occurrence of lung cancer in man. Acta. Unio. Int. Contra. Cancrum. 9(3):531-541.

Li, R., E. Weller, D.W. Dockery, L.M. Neas, and D. Spiegelman. 2006. Association of indoor nitrogen dioxide with respiratory symptoms in children: Application of measurement error correction techniques to utilize data from multiple surrogates. J. Expo. Sci. Environ. Epidemiol. 216(4):342-350.

Lubin, J.H., J.D. Boice, Jr., C. Edling, R.W. Hornung, G.R. Howe, E. Kunz, R.A. Kusiak, H.I. Morrison, E.P. Radford, J.M. Samet, M. Tirmarche, A. Woodward, S.X. Yao, and D.A. Pierce. 1995. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. J. Natl. Cancer Inst. 87(11):817-827.

Mascalzoni, D., E.S. Dove, Y. Rubinstein, H.J.S. Dawkins, A. Kole, P. McCormack, S. Woods, O. Riess, F. Schaefer, H. Lochmüller, B.M. Knoppers, and M. Hansson. 2015. International Charter of principles for sharing bio-specimens and data. Eur. J. Hum. Genet. 23:721-728.

Mervis, J. 2012. US science policy. Agencies rally to tackle big data. Science 336(6077):22.

Miller, G.W., and D.P. Jones. 2014. The nature of nurture: Refining the definition of the exposome. Toxicol. Sci. 137(1):1-2.

Molitor, J., M. Papathomas, M. Jerrett, and S. Richardson. 2010. Bayesian profile regression with an application to the National Survey of Children's Health. Biostatistics 11(3):484-498.

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.

NRC (National Research Council). 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

Papathomas, M., J. Molitor, S. Richardson, E. Riboli, and P. Vineis. 2011. Examining the joint effect of multiple risk factors using exposure risk profiles: Lung cancer in nonsmokers. Environ. Health Perspect. 119(1):84-91.

Patel, C.J., J. Bhattacharya, and A.J. Butte. 2010. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS One 5(5):e10746.

Patel, C.J., D.H. Rehkopf, J.T. Leppert, W.M. Bortz, M.R. Cullen, G.M. Chertow, and J.P. Ioannidis. 2013. Systematic evaluation of environmental and behavioral factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. Int. J. Epidemiol. 42(6):1795-1810.

Pechette, J.M. 2012. Transforming health care through cloud computing. Health Care Law Mon. 5:2-12.

Perttula, K., W.M. Edmands, H. Grigoryan, X. Cai, A.T. Iavarone, M.J. Gunter, A. Naccarati, S. Polidoro, A. Hubbard, P. Vineis, and S. Rappaport. 2016. Evaluating ultra-long chain fatty acids as biomarkers of colorectal cancer risk. Cancer Epidemiol. Biomarkers Prev. 25(8):1216-1223.

Rappaport, S.M. 2012. Biomarkers intersect with the exposome. Biomarkers 17(6):483-489.

Rappaport, S.M., and M.T. Smith. 2010. Environment and disease risks. Science 330(6003):460-461.

Relton, C.L., and G. Davey Smith. 2012. Two-step epigenetic Mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. Int. J. Epidemiol. 41(1):161-176.

Ritchie, S.A., J. Tonita, R. Alvi, D. Lehotay, H. Elshoni, S. Myat, J. McHattie, and D.B. Goodenowe. 2013. Low-serum GTA-446 anti-9 inflammatory fatty acid levels as a new risk factor for colon cancer. Int. J. Cancer. 132(2):355-362.

Roger, V.L., E. Boerwinkle, J.D. Crapo, P.S. Douglas, J.A. Epstein, C.B. Granger, P. Greenland, I. Kohane, and B.M. Psaty. 2015a. Roger et al. respond to "future of population studies." Am. J. Epidemiol.181(6):372-373.

Roger, V.L., E. Boerwinkle, J.D. Crapo, P.S. Douglas, J.A. Epstein, C.B. Granger, P. Greenland, I. Kohane and B.M. Psaty. 2015b. Strategic transformation of population studies: Recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. Am. J. Epidemiol. 181(6):363-368.

Saito, R., M.E. Smoot, K. Ono, J. Ruscheinski, P.L. Wang, S. Lotia, A.R. Pico, G.D. Bader, and T. Ideker. 2012. A travel guide to Cytoscape plugins. Nat. Methods 9(11):1069-1076.

Salek, R.M., S. Neumann, D. Schober, J. Hummel, K. Billiau, J. Kopka, E. Correa, T. Reijmers, A. Rosato. L. Tenori, P. Turano, S. Marin, C. Deborde, D. Jacob, D. Rolin, B. Dartigues, P. Conesa, K. Haug, P. Rocca-Serra, S. O'Hagan, J. Hao, M. van Vliet, M. Sysi-Aho, C. Ludwig, J. Bouwman, M. Cascante, T. Ebbels, J.L. Griffin, A. Moing, M. Nikolski, M. Oresic, S.A. Sansone, M.R. Viant, R. Goodacre, U.L. Günther, T. Hankemeier, C. Luchinat, D. Walther, and C. Steinbeck. 2015. COordination of Standards in MetabOlomicS (COSMOS): Facilitating integrated metabolomics data access. Metabolomics 11(6):1587-1597.

Samet, J.M., R. Schnatter, and H. Gibb. 1998. Invited commentary: Epidemiology and risk assessment. Am. J. Epidemiol. 148(10):929-936.

Shi, M., and C.R. Weinberg. 2011. How much are we missing in SNP-by-SNP analyses of genome-wide association studies? Epidemiology 22(6):845-847.

Siegmund, K.D., A.J. Levine, J. Chang, and P.W. Laird. 2006. Modeling exposures for DNA methylation profiles. Cancer Epidemiol. Biomarkers Prev. 15(3):567-572.

Slama, R., and A. Werwatz. 2005. Controlling for continuous confounding factors: Non- and semiparametric approaches. Rev. Epidemiol. Sante Publique 53(Spec. No. 2):2S65-2S80.

Tenopir, C., S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. 2011. Data sharing by scientists: Practices and perceptions. PLoS One 6(6):e21101.

Thomas, D.C., J.P. Lewinger, C.E. Murcray, and W.J. Gauderman. 2012. Invited commentary: GE-Whiz! Ratcheting gene-environment studies up to the whole genome and the whole exposome. Am. J. Epidemiol. 175(3):203-207.

Turner, M.C., M. Nieuwenhuijsen, K. Anderson, D. Balshaw, Y. Cui, G. Dunton, J.A. Hoppin, P. Koutrakis, and M. Jerrett. In press. Assessing the exposome with external measures: Commentary on the State of the Science and Research Recommendations. Annual Review of Public Health.

UK Biobank. 2016. Biobank [online]. Available: http://www.ukbiobank.ac.uk/ [accessed July 21, 2016].

Vineis, P., A.E. Khan, J. Vlaanderen, and R. Vermeulen. 2009. The impact of new research technologies on our understanding of environmental causes of disease: The concept of clinical vulnerability. Environ. Health 8:54.

Vineis, P., K. van Veldhoven, M. Chadeau-Hyam, and T.J. Athersuch. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology. Environ. Mol. Mutagen. 54(7):461-467.

Vrijheid, M., R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, C. Thomsen, J. Wright, T.J. Athersuch, N. Avellana, X. Basagaña, C. Brochot, L. Bucchini, M. Bustamante, A. Carracedo, M. Casas, X. Estivill, L. Fairley, D. van Gent, J.R. Gonzalez, B. Granum, R. Gražulevičienė, K.B. Gutzkow, J. Julvez, H.C. Keun, M. Kogevinas, R.R. McEachan, H.M. Meltzer, E. Sabidó, P.E. Schwarze, V. Siroux, J. Sunyer, E.J. Want, F. Zeman, and M.J. Nieuwenhuijsen. 2014. The human early-life exposome (HELIX): Project rationale and design. Environ. Health Perspect. 122(6):535-544.

# 5

# A New Direction for Risk Assessment and Applications of 21st Century Science

The scientific and technological advances described in Chapters 2–4 offer opportunities to improve the assessment or characterization of risk for the purpose of environmental and public-health decision-making. To facilitate appreciation of the new opportunities, this chapter first discusses the new direction envisioned for risk assessment and then highlights applications (see Box 1-3) of 21st century science that can be used to improve decision-making. It provides concrete examples of pragmatic approaches for using 21st century science along with long-standing toxicological and epidemiological approaches to improve the evidence used in decision-making. The chapter next addresses communication of the new approaches to stakeholders. It concludes with a brief discussion of the challenges that they pose and recommendations for addressing the challenges.

## A NEW DIRECTION FOR RISK ASSESSMENT

The seminal 1983 National Research Council (NRC) report *Risk Assessment in the Federal Government: Managing the Process* (NRC 1983) defined risk assessment as "the use of the factual base to define the health effects of exposure of individuals or populations to hazardous materials and situations." The report noted that risk assessment had four components—hazard identification, exposure assessment, dose–response assessment, and risk characterization—and that risk assessments contain some or all of them. It stated that various data streams from, for example, toxicological, clinical, epidemiological, and environmental research need to be integrated to provide a qualitative or quantitative description of risk to inform risk-based decisions. It recognized explicitly the uncertainty that arises when information on a particular substance is missing or ambiguous or when there are gaps in current scientific theory, and it called for inferential bridges or inferential guidelines to bridge such gaps to allow the assessment process to continue. Risk assessment then (as now) relied heavily on the measurement of apical responses, such as tumor incidence and developmental delays, in homogeneous animal models, often with little exposure or epidemiological information.

Although today's risk assessments generally support the same types of decisions as those in 1983, the tools available for asking and answering relevant risk-based questions have evolved substantially. As outlined in Chapters 2–4 of the present report, modern tools in exposure assessment, toxicology, and epidemiology have increased the speed at which information can be collected and the scope of the data available for risk assessment. The focus has also shifted from observing apical responses to understanding biological processes or pathways that lead to the apical responses or to disease. The tools are designed to investigate or measure molecular changes that give insight into the biological pathways. Thus, a "factual base" is being created that is increasingly upstream of the adverse health effects that federal agencies seek to prevent or minimize.

The Tox21 report (NRC 2007) fixed the new direction for risk assessment with its focus on discerning toxicity pathways, which were defined as "cellular response pathways that, when sufficiently perturbed in an intact animal, are expected to result in adverse health effects." Since publication of that report, the understanding of biological processes underlying disease has increased dramatically and has provided an opportunity to understand the biological basis of how different environmental stressors can affect the same pathway, each potentially contributing to the risk of a particular disease. To operationalize a risk-assessment approach that relies on mechanistic understanding, it will be necessary to understand the critical steps in the pathways, but beginning to apply the approach does not require knowing all pathways. For example, the results of a subchronic rat study might indicate a failure of animals to thrive, which is manifested as decreased weight gain and some deaths over the course of the study, but no obvious target-organ effects. Studies on the molecular effect of the chemical indicate that it is an uncoupler of oxidative phosphorylation. Epidemiological studies could then focus on biological processes that are

energy-intensive, such as heart muscle under stress. Exposure science could be used to measure or estimate population exposure to the stressor over space and time and could align toxicity data with environmental exposures for use in epidemiological studies. Assays to screen for the perturbation along with chemical-structure considerations might help to characterize risks posed by similarly acting chemicals, and exposure estimates could be generated for other chemicals hypothesized to exert a similar response.

Today, there is an appreciation of the multifactorial nature of disease, that is, a recognition that a single adverse outcome might result from multiple mechanisms that can have multiple components. (See further discussion in Chapter 7.) Thus, the question shifts from whether A causes B to whether A increases the risk of B. Figure 5-1 provides an illustration of that concept, and Box 5-1 provides a concrete example. In the figure, four mechanisms ($M_1$–$M_4$) and various combinations of six components ($C_1$–$C_6$) are involved in producing two outcomes ($O_1$ and $O_2$). For example, three components ($C_1$, $C_2$, and $C_3$) are involved in activating mechanism $M_1$, which leads to outcome $O_1$, and $C_1$ is a component in several mechanisms. Here, a component is defined as a biological factor, event, or condition that when present with other components produces a disease or other adverse outcome; mechanism is considered to be comprised of one or more components that cause disease or other adverse outcomes when they co-occur; and pathways are considered to be components of mechanisms. The model can incorporate societal factors that drive exposure or susceptibility, such as poverty, and that might ultimately lead to cellular responses that activate various mechanisms. For example, in mechanism $M_1$, societal factors could perturb component $C_1$, the same one that the chemical under consideration perturbs. Alternatively, societal factors could perturb components $C_2$ and $C_3$ of mechanism $M_1$, which in combination with the chemical's direct perturbation of component $C_1$ could fully activate the mechanism. The ability to identify the contribution of various components of a given mechanism and to understand the significance of changes in single components of a mechanism is critical for risk-based decision-making based on 21st century science.

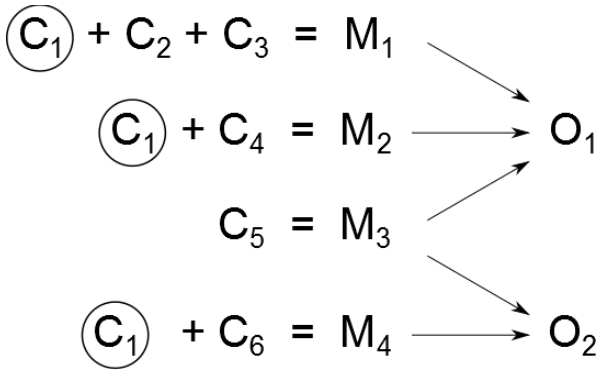



**FIGURE 5-1** Multifactorial nature of disease illustrated with four mechanisms (M) that have various components (C) and lead to two outcomes (O).

In the challenging context of multifactorial diseases, the 21st century tools allow implementation of the new direction for risk assessment that acknowledges the complexity of the determinants of risk. They can enable the identification of multiple disease contributors and advance understanding of how identified mechanisms, pathways, and components contribute to disease. They can be used to probe specific chemicals for their potential to perturb pathways or activate mechanisms and thereby increase risk. And the new tools provide critical biological information on how a chemical might add to a disease process and how individuals might differ in response; thus, they can provide insight on the shape of the population dose-response curve and on individual susceptibility to move toward the risk characterizations envisioned in the report *Science and Decisions: Advancing Risk Assessment* (NRC 2009). As noted by the NRC (2007, 2009), people differ in predisposing factors and co-exposures, so the extent to which any particular chemical perturbs a pathway and contributes to disease varies in the population. A challenge for the dose–response assessment is to characterize the extent to which the whole population and sensitive groups might be affected or, at a minimum, whether the perturbation exceeds some de minimis level.

Although the discussion above focuses primarily on the toxicological and epidemiological aspects of the new direction, exposure science will play a critical role. The exposure data arising from the technological advances in exposure science will provide much needed and increasingly rich information. For example, comprehensive exposure assessments that use targeted and nontargeted analyses of environmental and biomonitoring samples or that use computational exposure methods will help to identify chemical mixtures to which people are exposed. Such comprehensive assessments will support evaluating risks of groups of similarly acting chemicals for single end points or investigating chemical exposures that might activate multiple mechanisms that contribute to a specific disease. Advancing our understanding of the pharmacokinetics will further the ability to translate exposure-response relationships observed in in vitro systems to humans, characterize susceptible populations, and ultimately reduce uncertainty in risk assessment. Personalized exposure assessment will provide critical information on individual variability in exposure to complement pharmacodynamic variability assessed in pathway-based biological test systems. Ultimately, these and other advances in exposure

**BOX 5-1** Example of Multifactorial Nature of Disease

Sonic hedgehog (shh) is a signaling protein that is synthesized in mammalian embryos by the notochord and floor plate of the neural tube. Its function is to establish the ventral midline for the developing central nervous system. Interference with shh signaling during early embryonic development leads to the birth defect holoprosencephaly, in which the cerebrum fails to develop into two hemispheres. A number of events ("components" in Figure 5-1) can interfere with shh functioning. They include point mutations in the shh gene that lead to a partial loss of function (Roessler et al. 1997); mutations in the 7-dehydrocholesterol reductase gene that prevent the post-translational modification of shh in which cholesterol is added to the protein (a step that is essential for signaling—the mutation can lead to a condition described as the Smith-Lemli-Opitz syndrome) (Battaile and Steiner 2000); cholesterol synthesis-inhibiting drugs, such as BM15,766, that act on the same enzyme (Kolf-Clauw et al. 1997); and some plant alkaloids, such as cyclopamine, that inhibit the post-translational modification of shh (Incardona et al. 1998). Any component at a high enough dose or rate is sufficient to cause holoprosencephaly, but there are probably cases in which the dose or rate of one or more of the components is insufficient to disrupt shh signaling, but added together can perturb function.

In this example, all the components are acting on the same target, shh, but in different ways: some affect the integrity of the protein (point mutations in the gene), some affect its post-translational modification, and some affect its ability to interact with its receptor. Regardless, the result is the same: signaling by shh secreted by the notochord or ventral neural tube that is insufficient to establish a ventral field. The disruption of shh signaling is the "mechanism" in Figure 5-1.

Shh is expressed elsewhere in the embryo where it has a role in limb development and tooth development. Limb abnormalities, such as extra digits or fused digits, are often observed in the Smith-Lemli-Opitz syndrome. In Figure 5-1, that syndrome would represent a second outcome of the same mechanism. As indicated in Figure 5-1, different mechanisms can produce the same outcome. For example, retinoic acid is also an important morphogenetic factor in limb development, and retinoic acid excess or deficiency can produce limb defects. That would represent a separate mechanism that would involve other components (for example, nutritional vitamin A deficiency and inhibition of the enzyme that converts retinol to retinoic acid) but lead to the same adverse outcome (digit defects). The figure below illustrates this example in terms of Figure 5-1. Abbreviations: CSI, cholesterol synthesis-inhibiting; DHCR7, 7-dehydrocholesterol reductase; R, retinol; RA, retinoic acid; shh, sonic hedgehog.

science in combination with advances in toxicology and epidemiology will provide an even stronger foundation for the new direction for risk assessment.

## APPLICATIONS

Full implementation of the new direction for risk assessment or the visions described in the NRC report *Science and Decisions* and the Tox21 and ES21 reports (NRC 2007, 2009, 2012) is not yet possible, but the data being generated today can be used to improve decision-making in several areas. As noted in Chapter 1 (see Box 1-3), priority-setting, chemical assessment, site-specific assessments, and assessments of new chemistries are risk-related tasks that can all benefit from incorporating 21st century science. The methods and data required to support the various tasks will probably differ, and confidence in them will depend to some extent on the context. For example, scientists have a great deal of experience in using laboratory data to support biological plausibility in epidemiology studies, and the new data can be relatively easily applied in that context. In contrast, methods used to support definitive chemical assessments will likely need extensive evaluation, and risk assessors will need to be trained in how to use them. In the following sections, the committee describes approaches that can use the new scientific approaches in specific applications.

### Priority-Setting

Tens of thousands of chemicals are used in commerce in the United States (Muir and Howard 2006; Egeghy et al. 2012) in various items—including building materials, consumer products, and craft supplies—and can cause exposure through product use and environmental releases associated with manufacture and disposal. Although the number of chemicals in the environment is large, the number of chemicals for which toxicity, exposure, and epidemiology datasets are complete remains small. Given the finite resources of government agencies and other stakeholders for investigating the risks associated with the wide array of chemicals present in people, places, and goods, mechanisms for setting priorities for chemical evaluation and determining appropriate risk-management strategies—reduction of use, replacement, or removal—are essential.

Some categories of chemicals that are intended to have biological activity, such as drugs and pesticides, are routinely subjected to a suite of toxicity tests as required by law. However, extensive toxicity testing of most chemicals is not required, and the need for testing is determined by priority-setting schemes. For example, the National Toxicology Program (NTP 2016) sets testing priorities on the basis of the extent of human exposure, suspicion of toxicity, or the need for information to fill data gaps in an assessment, and the European Union's Registration, Evaluation, and Authorization of Chemicals (REACH) testing requirements are based predominantly on production volume (chemical quantity produced per annum) and the potential for widespread exposure or human use, such as would occur with a consumer product (NRC 2006; Rudén and Hansson 2010). Considerations of potential toxicity have generally been limited to alerts based on the presence of specific chemical features, such as a reactive epoxide moiety, or similarity to known potent toxicants. Using only those considerations to set priorities is clearly limited; additional hazard information that covers more biological space and exposure information that provides more detailed estimates of exposure from multiple sources and routes would improve the priority-setting process.

As Tox21 tools—such as high-throughput screening, toxicogenomics,[1] and cheminformatics—have become available, priority-setting has been seen as a principal initial application. High-throughput platforms, such as the US Environmental Protection Agency (EPA) ToxCast program described in Chapter 1, have produced data on thousands of chemicals. Toxicogenomic analyses have the potential to increase the biological coverage of in vitro cell-based assays and might be a useful source of data for priority-setting. For example, efforts are under way to assess transcriptomic responses in a suite of human cells by using positive control chemicals ultimately to determine whether biological pathways can be identified on the basis of select patterns of gene expression (Lamb et al. 2006) or whole-genome transcriptomics (de Abrew et al. 2016). Mismatches between in vitro and in vivo results might occur for several reasons, such as a lack of metabolism in the in vitro assays. As discussed in Chapter 3, lack of or low-level metabolic activation of an agent is widely recognized as a potential problem in in vitro studies, and development of methods to introduce metabolic systems into assays that can be run in high-throughput format is under active research.

Cheminformatic approaches can also be used to set priorities for chemical testing by evaluating series of chemicals for the presence of chemical features that are associated with toxicity—for example, through the use of such proprietary tools as DEREK[2]—or by using decision trees that evaluate whether there are precedents in the literature for specific chemical features to be associated with a particular toxicity outcome, such as developmental toxicity (Wu et al. 2013). Those methods have been automated and allow for rapid identification of chemicals that have specific chemical features that have been identified as potentially problematic, such as reactive functional groups, or that have a reasonably high similarity to chemicals that are potent toxicants, such as steroid-like substances (Wu et al. 2013).

[1]Toxicogenomics is transcriptomic analysis of responses to chemical exposure.

[2]See http://www.lhasalimited.org/.

Several new high-throughput methods—for example, ExpoCast (Wambaugh et al. 2013) or ExpoDat (Shin et al. 2015)—have been developed to provide quantitative exposure estimates for exposure-based and risk-based priority-setting. The new technologies can estimate exposures more explicitly than older simpler models by taking into account chemical properties, chemical production amounts, chemical use and human behavior (likelihood of exposure), potential exposure routes, and possible chemical intake rates. Information produced via high-throughput exposure calculations could be used to refine priority-setting schemes.

Depending on the context, hazard and exposure information could be used in various ways for priority-setting. For example, screening based only on hazard could be particularly useful in situations, such as those involving changes in product composition, in which exposure information is unknown or evolving and there is an assumption that the product would be used in the same way with roughly the same exposure. Methods have been proposed for risk-based priority-settting that use a combination of high-throughput exposure and hazard information in which the highest estimated exposure and the lowest-measured-effect concentration are identified, and margins of exposure (differences between toxicity and exposure metrics) are calculated (see Figure 5-2). Refinement of the margins of exposure by using reverse pharmacokinetic techniques to estimate exposure has also been proposed (Wetmore et al. 2013). Confidence in the approach should increase with broader biological coverage of the in vitro assays, innovations that add metabolic activation to the assays, methods that take into account toxicity that is associated with a particular route of exposure (such as inhalation), and improved accuracy of computational exposure models to predict human and ecosystem exposures.

## Chemical Assessment

Chemical assessments encompass a broad array of analyses, from Integrated Risk Information System assessments that include hazard and dose–response assessments to ones that also incorporate exposure assessments to produce risk characterizations. Moreover, chemical assessments performed by the federal agencies cover chemicals on which there are few data to use in decision-making (data-poor chemicals) and chemicals on which there is a substantial database for decision-making (data-rich chemicals). The following sections address how 21st century data could be used in the contrasting situations.

### Assessments of Data-Poor Chemicals

Assessments of some data-poor chemicals might begin by evaluating outcomes whose mechanisms are known. That is, mechanisms of a few toxicity outcomes, such as genotoxicity and skin sensitization, are sufficiently well known for it to be possible to rely on mechanistically based in vitro assays—for example, the Ames assay and direct peptide reactivity assay—for which the Organisation for Economic Co-operation and Development guidelines already exist as the starting point for hazard assessment. For such well-defined outcomes for which in vitro assays are sufficient for characterization, the process of hazard assessment is relatively straightforward. Rather than using animal data as the starting point for establishing hazard, one replaces the animal data with data from the alternative method. In most cases, conclusions are qualitative and binary—for example, the chemical is or is not a genotoxicant. However, efforts are under way to provide quantitative ways of using in vitro test information to describe the dose–response characteristics of chemicals and ultimately to calculate a health reference value, such as a reference dose or a reference concentration (see Figure 5-3). In the approach that uses animal data and in the approach that relies on in vitro results, uncertainty factors (UFs) are typically included to address interindividual differences in human response and the uncertainty associated with extrapolating from a test system to people. Alternatively, a model can be used to extrapolate to low doses. Box 5-2 provides further discussion on uncertainty, variability, UFs, and extrapolation.

Most toxicity outcomes involve multiple pathways by which chemicals can exert an adverse influence, and not all pathways have been determined for many outcomes, such as organ toxicity and developmental toxicity. For those outcomes, simple replacement of animal-derived information with in vitro information might not be possible. Another possible approach to evaluating chemicals is to use toxicity data on previously well-tested chemicals that are structurally similar to the chemical of interest (see Figure 5-4). Analogues are selected on the basis of similarities in chemical structure, physical chemistry, and biological activity in in vitro assays. Comparisons of analogues with the chemical of interest are based on the premise that the chemical of interest and its analogues are metabolized to common or biologically similar metabolites or that they are sufficiently similar in structure to have the same or similar biological activity (for example, they activate receptors similarly). The similarity supports the inference that the chemical will induce the same type of hazard as the analogues although not necessarily at similar doses.

The method described in Figure 5-4 depends on having a comprehensive database of toxicity data that is searchable by curated and annotated chemical structure (such as ACToR or DSSTox) and a consistent decision process for selecting suitable analogues. Wu et al. (2010) published a set of rules for identifying analogues and categorizing them as suitable, suitable with interpretation, suitable with precondition (such as metabolism), or un-

suitable for analogue-based assessment. The rules consider physical chemistry, potential chemical reactivity, and potential metabolism of the chemical.

In many cases, a close similarity based on atom-by-atom matching is sufficient to classify two or more chemicals as suitable analogues for each other. However, atom-by-atom matching is not sufficient in every case. Small differences can sometimes alter the chemical activity in such a way that one metabolic pathway is favored over another or the chemical reactivity with various biological molecules changes. In practice, analogue-based assessment can be greatly facilitated by expert-rule–based considerations with molecular similarity. The approach was tested in a case study that used a blinded approach and found to be robust (Blackburn et al. 2011). Given that the total dataset for traditional animal toxicity data is large (millions of entries in ACToR and probably tens of thousands of entries for each toxicity outcome), the analogue-based approach could have great utility. Similar approaches are being developed and used for read-across assessment of chemicals submitted under the European REACH regulation.

A structure–activity assessment can be thought of as a testable hypothesis that can be addressed with a variety of methods, such as those described in Chapter 3. Comparable metabolism can be assessed by using established methods for testing xenobiotic metabolism in vivo and in vitro with the recognition that metabolism can be complex for even simple molecules, such as benzene (McHale et al. 2012). Testing for similar biological activity can be based on what is understood about the primary pathways by which the chemicals in the class exert toxicity. If the mechanisms are not known, it is possible to survey some (for example, using ToxCast assays) or all (for example, by using global gene-expression analysis) of the universe of possible pathways that are affected by the chemical to determine the extent to which the biological activities of the chemical of interest and its analogues are comparable. Toxicogenomic analyses have been found to be useful for identifying a mechanism in both in vivo and in vitro models (see, for example, Daston and Naciff 2010). With lower-cost methods now available, large datasets of gene-expression responses for small molecules have become available (for example, the National Institutes of Health's Library of Network-based Cellular Signatures, LINCS), and these data can support determination of the extent to which chemicals of similar structure are sufficiently comparable for read-across (Liu et al. 2015).

Combining cheminformatic and rapid laboratory-based approaches makes it possible to arrive at a surrogate point of departure for risk assessment that uses analogue data. The surrogate can then be adjusted for pharmacokinetic differences and bioactivity (see Figure 5-5). The committee explored that approach in a case study on alkylphenols (see Box 5-3 and Appendix B).

Eventually, it might be possible to conduct similar assessments of chemicals without adequate analogue data. Cheminformatic and laboratory methods could be used to generate hypotheses about the possible activities of a new chemical, and the hypotheses could be tested virtually in

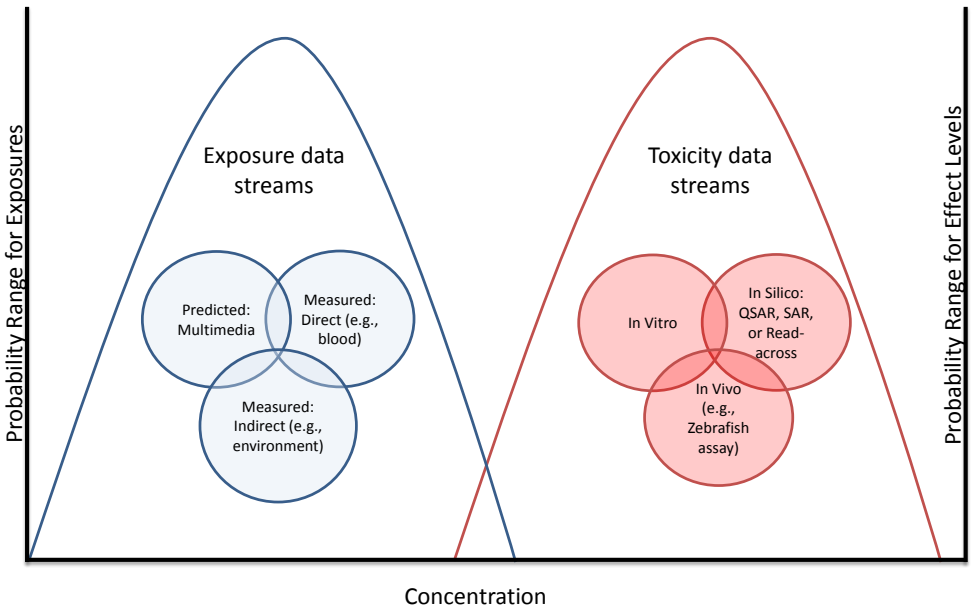



**FIGURE 5-2** Screening assessments could be used to estimate toxicity or predict exposure to rank chemicals for further testing or assessment. Chemicals that have the smallest margins of exposure (that is, upper bounds of exposure that are closest to or overlap with effect concentrations of toxicity) would be given the highest priority for further evaluation.

systems-biology models and verified in higher-order in vitro models. As discussed in Chapter 3, computational models, such as the cell-agent-based model used in the EPA virtual-embryo project, have done a reasonable job of predicting the effects of potent antiangiogenic agents on blood vessel development by using high-throughput screening data and information on key genes in the angiogenic pathway as starting points for model development (Kleinstreuer et al. 2013). The model can be run thousands of times—the virtual equivalent of thousands of experiments—and adjusted on the basis of the simulation results. The outcome of the model was evaluated in in vitro vascular-outgrowth assays and in zebrafish (Tal et al. 2014) and was found to be a good predictor of outcome in the assays. Such an approach clearly depends on a deep understanding of the biology underlying a particular process and how it can be perturbed and on sophisticated laboratory models that will support evaluation of the virtual model. This approach will require some knowledge of the key events that connect the initial interaction of an exogenous chemical with its molecular target and the ultimate adverse outcome.

Regardless of whether the risk assessment is conducted with the read-across approaches depicted in Figures 5-4 and 5-5 or the pathway approach just described, there will be circumstances in which the uncertainty in the assessment needs to be reduced to support decision-making. That situation can arise because the margin of exposure is too small, the possible mechanisms have still not been adequately defined, or the quantitative relationship between effects measured at the molecular or cellular level and adverse outcome have not been adequately defined. In such cases, one might use increasingly complex models—for example, zebrafish or targeted rodent testing—to assess biological activity and the outcomes of a chemical exposure.

### Assessment of Data-Rich Chemicals

Some chemicals are the subjects of substantive databases that leave no question regarding the causal relationship between exposure and effect; that is, hazard identification is not an issue for decision-making. However, there might still be unanswered questions that are relevant to regulatory decision-making, such as questions concerning the effects of exposure at low doses, susceptible populations, possible mechanisms for the observed effects, and new outcomes associated with exposure. The advances described in Chapters 2–4 have the potential to reduce uncertainty around such key issues. The committee explores how 21st century science can be used to address various questions in a case study that uses air pollution as an example (see Box 5-4 and Appendix B).

### Cumulative Risk Assessment

Cumulative risk assessment could benefit from the mechanistic data that are being generated. It is well understood that everyone is exposed to multiple chemicals simultaneously in the environment, for example, through the air we breathe, the foods we eat, and the products we use. However, risk assessment is still conducted largely on individual chemicals even though chemicals that have a similar mechanism for an outcome or that are associated with similar outcomes are considered as posing a cumulative risk when they are encountered together (EPA 2000; NRC 2008). Cumulative risk assessment of carcinogens is somewhat common in agencies, but cumulative risk assessment of noncarcinogens is not so common. One example of cumulative assessment is that of organophosphate pesticides whose mechanism is known to be acetylcholinesterase inhibition.

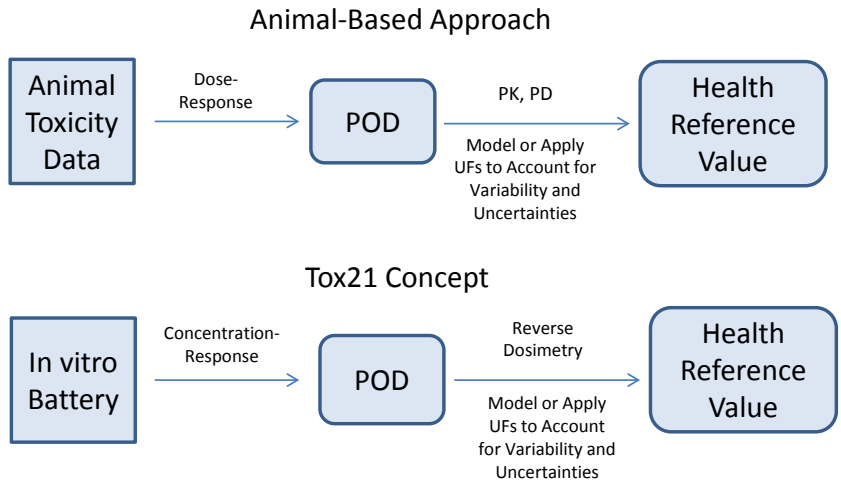



**FIGURE 5-3** A comparison of the animal-based approach to derive reference values compared with an approach under development that uses in vitro batteries where a biological pathway for a specific outcome has been elucidated. The UFs (or models) for the approaches would differ but are used to make adjustments on the basis of uncertainty or variability or to extrapolate across doses. Abbreviations: PD, pharmacodynamics; PK, pharmacokinetics; POD, point of departure; UF, uncertainty factor.

**BOX 5-2** Uncertainty and Variability in Assessment

Risk assessment involves the estimation of risk associated with a particular exposure and characterization of the inherent uncertainties associated with the estimate. For human risk estimates based on animal data, the uncertainties include ones associated with possible species differences (between laboratory animals and humans) in pharmacokinetics and sensitivity, human population variability, and prediction of lifetime exposures from less-than-lifetime testing protocols, and others. Although the magnitude of each uncertainty can be approached experimentally (given enough resources and time), they have typically been addressed in noncancer assessment by assigning uncertainty factors (UFs) that have a specific value (usually 1, 3, or 10) to derive a toxicity or risk estimate. Using 21st century science will require new thinking about the uncertainties associated with risk assessment and their magnitude. Some aspects of uncertainty will be eliminated; for example, using human-derived cells and receptors will eliminate the need to account for interspecies differences in pharmacodynamic sensitivity. However, using an in vitro approach introduces new uncertainties, such as how an in vitro concentration is related to an exposure scenario in an intact human or how an upstream molecular-level response is related quantitatively to a downstream disease outcome. Quantitative methods of combining information from multiple assays or data streams into integrated testing strategies (see, for example, Jaworska et al. 2013; Rovida et al. 2015) have been used to represent the key steps of diseases to overcome the uncertainty associated with using molecular-level responses.

It might also be possible to use biologically based dose–response modeling or other empirical modeling to replace a UF-based approach for extrapolation; this would agree with the NRC (2009) recommendation that dose–response modeling be based on a "formal, systematic assessment of background disease processes, possible vulnerable populations, and modes of action." A modeling approach has been used to determine a dose–response relationship for a toxicity pathway that involves DNA damage and repair (Bhattacharya et al. 2011) that could be developed further to address human heterogeneity in response. Another approach to estimating interindividual variability is large-scale in vitro profiling of multiple human cell lines (Abdo et al. 2015a,b; Eduati et al. 2015), but this addresses only variability due to genetic differences, which are expected to be a minor contributor in many cases. The range of human population variability in exposure and response is poorly understood, but new technologies should improve our ability to quantify some uncertainties, including human heterogeneity in vulnerability to exposures. Characterizing the new uncertainties and estimating their magnitude will be important as the new approaches are integrated into risk assessment.

Testing systems that evaluate more fundamental levels of biological organization (effects at the cellular or molecular level) might be useful in identifying agents that act via a common mechanism and in facilitating the risk assessment of mixtures. Identifying complete pathways for chemicals (from molecular initiating events to individual or population-level disease) could also be useful in identifying chemicals that result in the same adverse health outcome through different molecular pathways. High-throughput screening systems and global gene-expression analysis are examples of technologies that could provide the required information. The techniques applied in support of cumulative risk assessment will also support multifactorial risk evaluations discussed further in Chapter 7.

### Site-Specific Assessments

Understanding the risks associated with a chemical spill or the extent to which a hazardous-waste site needs to be remediated depends on understanding exposures to various chemicals and their toxicity. The assessment problem contains three elements: identifying and quantifying chemicals present at the site, characterizing their toxicity, and characterizing the toxicity of chemical mixture. Thus, one might consider this situation to be an exposure-initiated assessment in which exposure information is a starting point as illustrated in Figure 5-6. In this context, *exposure information* means information on newly identified chemicals and more complete characterization of exposure to chemicals previously identified at a site. Box 5-5 provides two specific examples of exposure-initiated assessments.

The advances described in Chapters 2–4 can address each element involved in site-specific assessments. Targeted analytical-chemistry approaches, particularly ones that use gas or high-performance liquid chromatography coupled with mass spectrometry, can identify and quantify chemicals for which standards are available. Nontargeted analyses can help to assign provisional identities to previously unidentified chemicals. The committee ex-

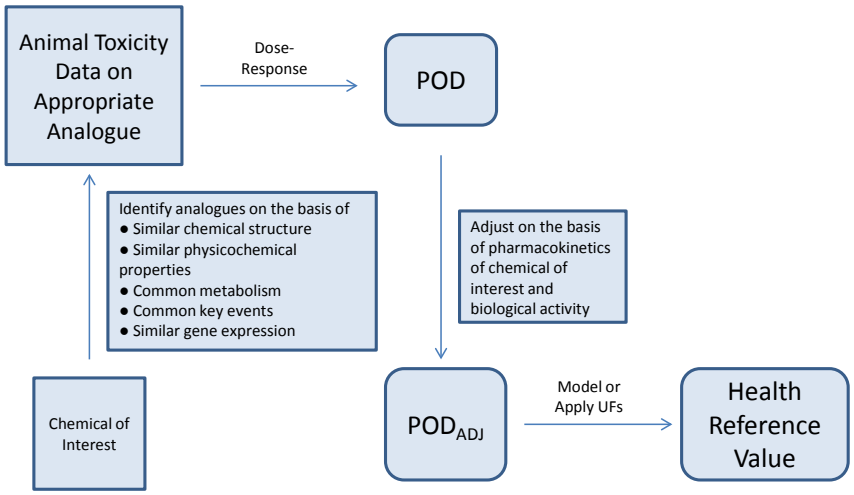



**FIGURE 5-4** Approach to deriving health reference values when data on structurally similar chemicals are available. Similarity can be based on such characteristics as chemical structure, physicochemical properties, metabolism, key events in biological pathways, or gene expression; similarity of several characteristics increases confidence in the analogy. The point of departure (POD) of the appropriate analogue would be adjusted on the basis of pharmacokinetic differences between the chemical of interest and the analogue and other important biological factors, such as receptor activation; relevant uncertainty factors would then be applied or models would be used. Accounting for uncertainty could include a determination of the degree of confidence in the read-across, including the number of analogues identified, the degree of similarity of the analogues to the chemical of interest, and the extent of the dataset on the analogues.

**BOX 5-3** Case Study: Alkylphenols

This case study illustrates the use of read-across for derivation of a health reference value. As detailed in Appendix B, a data-poor alkylphenol (*p*-dodecylphenol) is compared with two data-rich alkylphenols (*p*-octylphenol and *p*-nonylphenol). Comparisons are made on the basis of two-dimensional chemical structure and physicochemical properties. High-throughput in vitro data from ToxCast are used to add confidence to the selection of the analogues. Data from in vivo rat multigeneration studies of the data-rich alkylphenols are used as a starting point for derivation of a health reference value and adjustments are suggested on the basis of the ToxCast data. Limitations of the analysis are discussed, and information that would add confidence to the results of the analysis is identified.

plored the application of advances in exposure science to a case study of a large historically contaminated site (see Box 5-6 and Appendix C).

As for toxicity characterization, assessments of waste sites and chemical releases often involve chemicals on which few toxicity data are available. In the case of waste sites, EPA assigns provisional reference values for a number of chemicals by using the Provisional Peer Reviewed Toxicity Value (PPRTV) process. However, because of the amount or quality of the data available, the PPRTV values tend to entail large uncertainties. Analogue-based methods coupled with high-throughput or high-content screening methods have the potential to improve the PPRTV process. Identification of well-tested appropriate analogues to an untested chemical at clean-up sites can provide more certain estimates of the hazard and potency of the chemical, and the appropriateness of the analogues can be confirmed with high-throughput screening or high-content data that show comparability of biological targets or other end points and relative potency. Although the high-throughput or high-content models still require validation, the read-across approach could be applied immediately.

In the case of chemical releases, few data might be available on various chemicals—a situation similar to waste sites—but decisions might need to be made quickly. The committee uses the scenario of a chemical release as a case study to examine how Tox21 approaches can be used to provide data on a data-poor chemical quickly (see Box 5-6 and Appendix C).

As for understanding the toxicity of chemical mixtures, high-throughput screening methods provide infor-

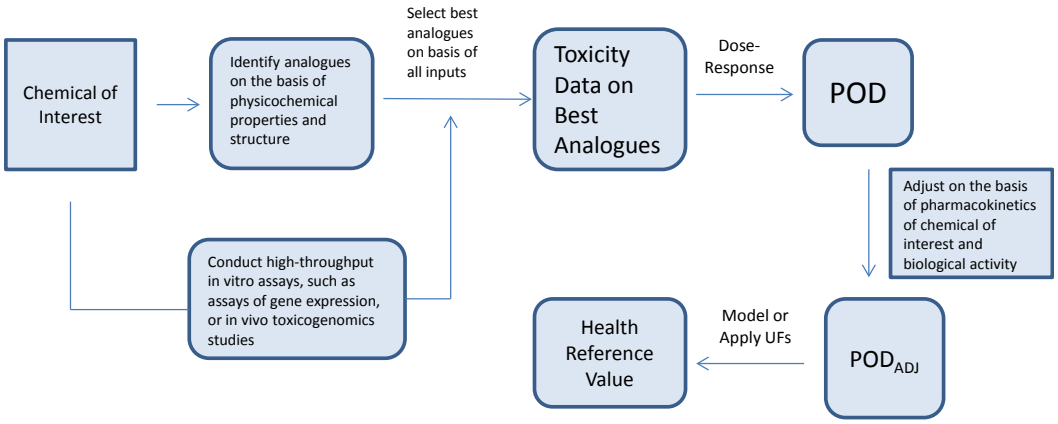



**FIGURE 5-5** Approach for deriving acceptable values when an appropriate analogue cannot be identified solely through comparisons of structure and physicochemical data. In such a case, data from high-throughput in vitro assays of the chemical of interest can be used as an additional source of information to identify the best analogue that can then be used to derive acceptable values.

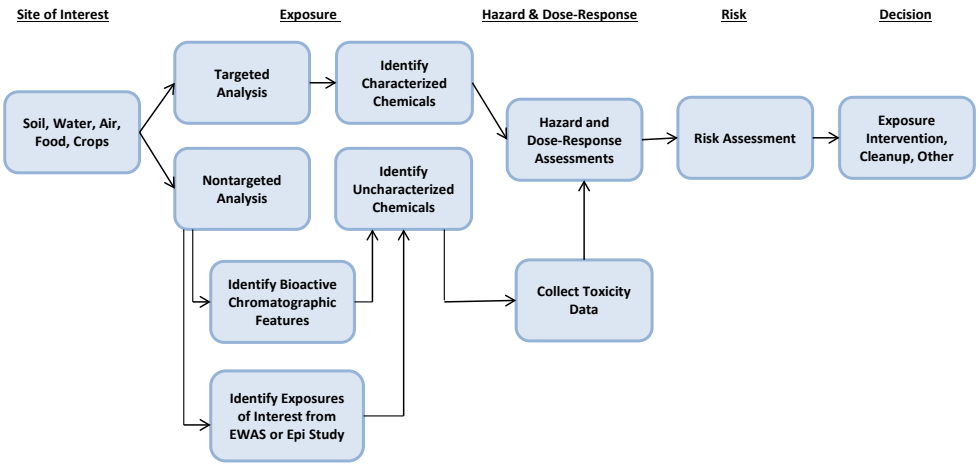



**FIGURE 5-6** Overview of approach and decisions for an exposure-initiated assessment. Abbreviations: Epi, epidemiological; EWAS, exposome-wide association study.

mation on mechanisms that can be useful in determining whether any mixture components might act via a common mechanism, affect the same organ, or cause the same outcome and thus should be considered as posing a cumulative risk (EPA 2000; NRC 2008). High-throughput methods can also be used to assess the toxicity of mixtures that are present at specific sites empirically rather than assessing individual chemicals. Such real-time generation of hazard data was conducted on the dispersants that were used to treat the crude oil released during the *Deepwater Horizon* disaster (Judson et al. 2010) to determine whether some had greater endocrine activity or cytotoxicity than others. Endocrine assays were the focus because of the known estrogenic activity of nonylphenol ethoxylates; nonylphenol (the degradation product of nonylphenol ethoxylates) is known to be estrogenic.

It is possible to use high-throughput assay data as the basis of a biological read-across for complex mixtures. For example, an uncharacterized mixture could be evaluated in high-throughput or high-content testing, and the results could be compared with existing results for individual chemicals or well-characterized mixtures. That process is similar to the connectivity mapping approach (Lamb et al. 2006) in which the biological activity of a single chemical entity is compared with the fingerprint of other chemicals in a large dataset, and it is assumed that chemicals with like biological activity have the same mechanism. That approach for single chemicals can be used for uncharacterized mixtures. One would still not know whether the biological activity was attributable to a single chemical entity or to multiple chemicals, but it would not matter if one were concerned only about char-

**BOX 5-4** Case Study: Air Pollution

The consequences of exposure to air pollution have been extensively investigated, the evidence concerning a causal relationship between air pollution and lung cancer is strong, and various agencies, including the International Agency for Research on Cancer, have concluded that outdoor air pollution is carcinogenic. However, there are still unanswered questions, such as which components are primarily responsible for carcinogenicity, whether there are interactions or synergies among the various components, what effects might occur at low exposures, and which groups might be at greater risk because of particular characteristics, such as smoking tobacco. As detailed in Appendix B, the first part of this case study describes advances in exposure science and toxicology, specifically -omics technologies that can help to characterize adverse effects, refine exposure further, and identify mechanisms and groups at risk.

The second part of the case study (see Appendix B) examines the situation in which a new outcome is associated with a well-studied substance. In this case, recent evidence has emerged concerning an association between neurodevelopmental outcomes in children and air-pollution exposure. Here the question concerns mainly hazard identification because causal associations between air pollution and any specific neurodevelopmental outcome are not yet established. Advances in exposure science that could augment or improve new or continuing epidemiological investigations are described. Advances in toxicology that could be used to assess the developmental neurotoxicity risk associated with air pollution are also described.

acterizing the risk associated with that particular mixture. The committee notes that it is possible that a mixture will exhibit more than one biological activity, particularly at high concentrations, but it should be possible to gain a better understanding of the biological activity by testing multiple concentrations of the mixture. The committee explores a biological read-across approach for complex mixtures further in a case study that considers the hypothetical site imagined in the first case study (see Box 5-6 and Appendix C).

Finally, new methods in exposure science, -omics technologies, and epidemiology provide another approach to generate hypotheses about the role of chemicals and chemical mixtures in specific disease states and to gather information about potential risks associated with specific sites. Information generated on chemical mechanisms, particularly of site-specific chemical mixtures, might be useful for identifying highly specific biomarkers of effect that can be measured in people who work or reside near a site of concern. Measurement of biomarkers has advantages over collection of data on disease outcome because many diseases of concern, such as cancer, are manifested only after chronic exposure or after a long latency period. Such measurement could also be of value in determining the effectiveness of remediation efforts at the site if biomarkers can be measured before and after mitigation. Real-time individualized measurements of exposure of people near a site are also possible and could provide richer data about peak exposures or exposure durations.

### Assessment of New Chemistries

Green chemistry involves the design of molecules and products that are optimized to have minimal toxicity and limited environmental persistence, are (ideally) derived from renewable sources, and perform comparably with or better than the chemicals that they are replacing. The green-chemistry approach often involves synthesis of new molecules on which there are no toxicity data and that might not have close analogues. Green-chemistry design is another case in which the use of modern in vitro toxicology methods could have great utility by providing guidance on which molecular features are associated with greater or less toxicity and by identifying chemicals that do not affect biological pathways that are known to be relevant for toxicity (Voutchkova et al. 2010). There are a few examples of the use of in vitro toxicity methods to determine whether potential replacement chemicals are less toxic. For example, Nardelli et al. (2015) evaluated the effects of a series of potential replacements for phthalate plasticizers on Sertoli cell function, and high-throughput testing has been used to evaluate alternatives to bisphenol A in the manufacture of can linings (Seltenrich 2015). Using high-throughput methods in this context is not conceptually different from screening prospective therapeutic agents for maximal efficacy and minimal off-target effects. Box 5-7 and Appendix D describe a case study of assessment of new chemistries.

**BOX 5-5** Two Examples of Exposure-Initiated Assessment

In the first example of exposure-initiated assessment, scientists who were investigating Superfund sites around the Portland Harbor in Oregon recently found novel environmental degradation products of common polycyclic aromatic hydrocarbons (PAHs) (O'Connell et al. 2013). Thirty-eight oxygenated PAHs were identified as toxicologically uncharacterized members of a PAH mixture at the site. Given the urgent need for testing, novel high-throughput toxicity testing in zebrafish has been conducted on representative mixtures of PAHs that were found in soil and water media of Portland Harbor (Knecht et al. 2013), and passive sampling devices have been deployed to characterize concentrations in species of the aquatic ecosystem that are used as human food (Paulik et al. 2016).

In the second example, nontargeted chemical analysis of dust samples that were collected as part of the US American Healthy Homes Survey was conducted (Rager et al. 2016). Nontargeted analysis revealed a spectrum of chromatographic features (elution time, exact mass, and isotopic signature) that could not initially be assigned to distinct chemicals. Some features were later identified by using analytical standards that were selected on the basis of probable matches to chemical structures in EPA's Distributed Structure-Searchable Toxicity database. Initial screening of the group of identified chemicals—including pesticides, nicotine, and perfluorooctanoic acid—was completed by using exposure and bioactivity estimates from ExpoCast and ToxCast, respectively, and information on detection frequency and abundance; the information was presented in ToxPi format. The authors also reported the presence of large numbers of features that remain unidentified and untested. The approach could be applied to other environmental media, such as soil and water at Superfund sites or water streams that are used as public drinking-water supplies but have been tested only for small numbers of chemicals.

One could use the same methods as described above to evaluate the toxicity of newly discovered chemicals in the environment, for example, from unexpected breakdown products of a widely used pesticide. If breakdown products are chemically related to their parent molecules, cheminformatics (read-across) methods could also be appropriate for estimating their toxicity.

## COMMUNICATING THE NEW APPROACHES

Many of the approaches introduced in this chapter will be unfamiliar to some stakeholder groups. Communicating the strengths and limitations of the approaches in a transparent and understandable way will be necessary if the results are to be applied appropriately and will be critical for the ultimate acceptance of the approaches. The information needs and communication strategies will depend on the stakeholder group. The discussion here focuses on four stakeholder groups: risk assessors, risk managers and public-health officials, clinicians, and the lay public.

Risk-assessment practitioners who are responsible for generating health reference values need to have information on the details of the new approaches and on how to apply their results to predict human risk. They probably need formal training in the interpretation and application of new data streams emerging from exposure science, toxicology, and epidemiology. Read-across, for example, is perhaps the most familiar of the alternative approaches described in this chapter, but most risk assessors still need a great deal of training in identifying appropriate chemical analogues on which to base a read-across and in accounting for decreased confidence in the assessment if there are few analogues or less than perfect structural matches. They also need to develop new partnerships that can help them with their tasks, for example, with computational and medicinal chemists who develop strategies for analogue searching, gauge the suitability of each analogue, or determine the likely metabolic pathway of a chemical of interest and its analogues to see whether they become more or less alike as they are biotransformed.

Most risk assessors are already familiar with the integration of traditional data for risk assessment, but they will need help in understanding how to integrate novel data streams and how much confidence they can have in the new data. One approach will be to compare the results from new methods with more familiar data sources, particularly in vivo toxicology studies. For example, EPA recently concluded that a high-throughput battery of estrogenicity assays is an acceptable alternative to the uterotrophic assay for tier 1 endocrine-disrupter screening (Browne et al. 2015; EPA 2015). The communication strategy in this case involved a description of the purpose of the assay battery, an explanation of the biological space covered by the battery (that is, the extent of the estrogen-signaling pathway being evaluated and the redundancy of the assays), a description of a computational model that integrates the data from all the assays and discriminates

between a true response and noise, and a comparison with an existing method that showed the new way working in most cases. Papers like the one cited provide useful models for further technical communication to risk assessors.

Risk managers and public-health officials do not need information that provides details on the assays or how they are applied to risk assessment; they do need to know the uncertainty associated with risk estimates and the confidence that they should place in them. Communication to this group will need to address those issues. There will be cases in which the new approaches will provide information that was heretofore unavailable to them, and the new information will assist them in making decisions about site remediation or acceptable exposure levels. This chapter discussed the possibility of using read-across to increase the number of chemicals evaluated in the PPRTV process, and Appendix C highlights a case study that uses cheminformatic approaches to address the developmental-toxicity potential of 4 methylcyclohexanemethanol, a chemical for which there was no experimental data on that outcome. Both examples illustrate how new approaches can provide information that would not have been available in any other way. However, the uncertainties associated with the new approaches need to be communicated.

As scientists advance the vision of identifying the many components that are responsible for multifactorial diseases, it will be necessary to communicate with clinicians and the public about how the factors have been identified, how each is related to others, and whether it is possible to reduce exposure to one or more factors to decrease disease risk. Physicians are beginning to embrace new methods as genomic information on individual patients becomes more available and personalized medicine becomes more of a reality, but there will still need to be communication to physicians in venues that they are likely to read and with diagnostic and treatment approaches that they are likely to be able to implement.

As for the general public, although many people get their health information from their doctors, some are far more likely to get medical information from the Internet and the popular press. The information that those media outlets require about new approaches is not qualitatively different from what clinicians need, but it needs to be presented in a format that is digestible by educated laypeople.

Finally, enhanced communication among the scientific community both nationally and internationally is vitally important for fully achieving the goals outlined in the Tox21 and ES21 reports and for gaining consensus regarding the utility of the new approaches and their incorporation into decision-making. The communication should include enhanced and more transparent integration of data and technology generated from multiple sources, including academic laboratories. Universities could serve as a communication conduit for multiple stakeholders, particularly clinicians and the lay public; thus, their engagement should be strategically leveraged. Ultimately, a more multidisciplinary and inclusive strategy for scientific discourse will help attain broad understanding and confidence in the new tools.

## CHALLENGES AND RECOMMENDATIONS

As noted earlier in this report, there are challenges to achieving the new direction for risk assessment fully. Some, such as model and assay validation, are addressed in later chapters. Here, the committee highlights a few challenges that are specific to the applications and ap-

**BOX 5-6** Case Studies: Site-Specific Assessments

The committee created three case studies related to site-specific assessment that explore each element of the problem and how to incorporate 21st century science into the evaluations. Appendix C provides the details of the case studies described below.

- *Identifying chemicals present*. The committee considers a large historically contaminated site with land and surface water near a major population center and describes how targeted and nontargeted analyses of chemicals can be used at the site.
- *Characterizing toxicity*. The committee considers the release of 4-methylcyclohexanemethanol into the Elk River about 1 mile upstream of a water intake facility for the city of Charleston, West Virginia, in 2014 and describes exposure and toxicity screening tools that help to understand the human risk.
- *Characterizing mixture toxicity*. The committee considers a toxicity assessment of complex mixtures observed in environmental samples, tissues, and biofluids and illustrates how a biological read-across approach could be used to conduct an assessment.

proaches described in the present chapter and offers some recommendations to address them.

*Challenge*: For risk assessment of individual chemicals, various approaches, such as cheminformatics and read-across, are already being applied because existing approaches are insufficient to meet the backlog of chemicals that need to be assessed. However, methods for grouping chemicals, assessing the suitability of analogues, and accounting for data quality and confidence in assessment are still being developed or are being applied inconsistently.

*Recommendation*: Read-across and cheminformatic approaches should be developed further and integrated into environmental-chemical risk assessments. High-throughput, cell-based assays and high–information-content approaches, such as gene-expression analysis, provide a large volume of data that can be used to test the assumptions made in read-across that analogues have the same biological targets and effects. Read-across and cheminformatics approaches depend on high-quality databases that are well curated; data curation and quality assurance should be a routine part of database development and maintenance. New case studies that use cheminformatic and read-across approaches could demonstrate new applications and encourage their use.

*Challenge*: Approaches that use large data streams to evaluate the potential for toxicity present a challenge in synthesizing information in a way that supports decision-making.

*Recommendation*: Statistical methods that can integrate multiple data streams and that are easy for risk assessors and decision-makers to use should be developed further and made transparent and user-friendly.

*Challenge*: Measuring biological events that are far upstream of disease states will introduce new sources of uncertainty into the risk-assessment process. Using data on those events as the starting point for risk assessment will require new approaches for risk assessment that are different from the current methods, which identify a point of departure and apply default uncertainty factors or extrapolate by using mathematical models.

*Recommendation*: New types of uncertainty will arise as the 21st century tools and approaches are used, and research should be conducted to identify these new sources and their magnitude. Some traditional sources of uncertainty will disappear as scientists rely less on animal models to predict toxicity, and these should also be identified.

## REFERENCES

Abdo, N., B.A. Wetmore, G.A. Chappell, D. Shea, F.A. Wright, and I. Rusyn. 2015a. In vitro screening for population variability in toxicity of pesticide-containing mixtures. Environ. Int. 85:147-155.

Abdo, N., M. Xia, C.C. Brown, O. Kosyk, R. Huang, S. Sakamuru, Y.H. Zhou, J.R. Jack, P. Gallins, K. Xiam Y. Li, W.A. Chiu, A.A. Motsinger-Reif, C.P. Austin, R.R. Tice, I. Rusyn, and F.A. Wright. 2015b. Population-based in vitro hazard and concentration-response assessment of chemicals: The 1000 genomes high-throughput screening study. Environ. Health Perspect. 123(5):458-466.

Battaile, K.P, and R.D. Steiner. 2000. Smith-Lemli-Opitz syndrome: The first malformation syndrome associated with defective cholesterol synthesis. Mol. Genet. Metab. 71(1-2):154-162.

Bhattacharya, S., Q. Zhang, P.L. Carmichael, K. Boekelheide, and M.E. Andersen. 2011. Toxicity testing in the 21 century: Defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. PLoS One 6(6):e20887.

Blackburn,K., D. Bjerke, G. Daston, S. Felter, C. Mahony, J. Naciff, S. Robison, and S. Wu. 2011. Case studies to test: A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. Regul. Toxicol. Pharmacol. 60(1):120-135.

Browne, P., R.S. Judson, W.M. Casey, N.C. Kleinstreuer, and R.S. Thomas. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environ. Sci. Technol. 49(14):8804-8814.

**BOX 5-7** Case Study: Assessment of New Chemistries

This case study describes a hypothetical example in which there are three choices of "new" chemicals for use in a manufacturing process that will result in human exposure. In Appendix D, the committee describes in vitro high-throughput data available on the chemicals and what those data might mean. It then considers several scenarios in which human exposure could occur and calculates indoor air releases that correspond to the in vitro bioassay data. The committee concludes with a discussion of how the data could be used in the decision-making process.

Daston, G., and J.M. Naciff. 2010. Predicting developmental toxicity through toxicogenomics. Birth Defects Res. C. Embryo Today 90(2):110-117.

De Abrew, K.N., R.M. Kainkaryam, Y.K. Shan, G.J. Overmann, R.S. Settivari, X. Wang, J. Xu, R.L. Adams, J.P. Tiesman, E.W. Carney, J.M. Naciff, and G.P. Daston. 2016. Grouping 34 chemicals based on mode of action using connectivity mapping. Toxicol. Sci. 151(2):447-461.

Eduati, F. L.M. Mangravite, T. Wang, H. Tang, J.C. Bare, R. Huang, T. Norman, M. Kellen, M.P. Menden, J. Yang, X. Zhan, R. Zhong, G. Xiao, M. Xia, N. Abdo, O. Kosyk, S. Friend, A. Dearry, A. Simeonov, R.R. Tice, I. Rusyn, F.A. Wright, G. Stolovitzky, Y. Xie, and J. Saez-Rodriguez. NIEHS-NCATS-UNC DREAM Toxicogenetics Collaboration. 2015. Prediction of human population responses to toxic compounds by a collaborative competition. Nat. Biotechnol. 33(9):933-940.

Egeghy, P.P., R. Judson, S. Gangwal, S. Mosher, D. Smith, J. Vail, and E.A Cohen Hubal. 2012. The exposure data landscape for manufactured chemicals. Sci. Total Environ. 414(1):159-166.

EPA (US Environmental Protection Agency). 2000. Supplementary Guidance for Conducting Risk Assessments of Chemical Mixtures. EPA/630/R-00/002. Risk Assessment Forum Technical Panel, US Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/raf/pdfs/chem_mix/chem_mix_08_2001.pdf [accessed September 30, 2016].

EPA (US Environmental Protection Agency). 2015. Use of High Throughput Assays and Computational Tools in the Endocrine Disruptor Screening Program-Overview [online]. Available: https://www.epa.gov/endocrine-disruption/use-high-throughput-assays-and-computational-tools-endocrine-disruptor [accessed December 1, 2016].

Incardona, J.P., W. Gaffield, R.P. Kapur, and H. Roelink. 1998. The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. Development 125(18):3553-3562.

Jaworska, J., Y. Dancik, P. Kern, F. Gerberick, and A. Natsch. 2013. Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. J. Appl. Toxicol. 33(11):1353-1364.

Judson, R.S., M.T. Martin, D.M. Reif, K.A. Houck, T.B. Knudsen, D.M. Rotroff, M. Xia, S. Sakamuru, R. Huang, P. Shinn, C.P. Austin, R.J. Kavlock, and D.J. Dix. 2010. Analysis of eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity. Environ. Sci. Technol. 44(15):5979-5985.

Kleinstreuer, N., D. Dix, M. Rountree, N. Baker, N. Sipes, D. Reif, R. Spencer, and T. Knudsen. 2013. A computational model predicting disruption of blood vessel development. PLoS Comput. Biol. 9(4):e1002996.

Knecht, A.L., B.C. Goodale, L. Troung, M.T. Simonich, A.J. Swanson, M.M. Matzke, K.A. Anderson, 1 and R.L. Tanguay. 2013. Comparative developmental toxicity of environmentally relevant 2 oxygenated PAHs. Toxicol. Appl. Pharmacol. 271(2):266-275.

Kolf-Clauw, M, F. Chevy, B. Siliart, C. Wolf, N. Mulliez, and C. Roux. 1997. Cholesterol biosynthesis inhibited by BM15.766 induces holoprosencephaly in the rat. Teratology 56(3):188-200.

Lamb, J., E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S.A. Armstrong, S.J. Haggarty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, and T.R. Golub. 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795):1929-1935.

Liu, C., J. Su, F. Yang, K. Wei, J. Ma, and X. Zhou. 2015. Compound signature detection on LINCS L1000 big data. Mol. Biosyst. 11(3):714-722.

McHale, C.M., L. Zhang, and M.T. Smith. 2012. Current understanding of the mechanism of benzene-induced leukemia in humans: Implications for risk assessment. Carcinogenesis 33(2):240-252.

Muir, D.C., and P.H. Howard. 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. Environ. Sci. Technol. 40(23):7157-7166.

Nardelli, T.C., H.C. Erythropel, and B. Robaire. 2015. Toxicogenomic screening of replacements for di(2-ethylhexyl) phthalate (DEHP) using the immortalized TM4 Sertoli cell line. PLoS One 10(10):e0138421.

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press.

NRC (National Research Council). 2006. Human Biomonitoring for Environmental Chemicals. Washington, DC: The National Academies Press.

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

NRC (National Research Council). 2008. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. Washington, DC: The National Academies Press.

NRC (National Research Council). 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

NRC (National Research Council). 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: National Academies Press.

NTP (National Toxicology program). 2016. Nominations to the testing program [online]. Available: http://ntp.niehs.nih.gov/testing/noms/index.html [accessed July 22, 2016].

O'Connell, S.G., T. Haigh, G. Wilson, and K.A. Anderson. 2013. An analytical investigation of 24 oxygenated-PAHs (OPAHs) using liquid and gas chromatography-mass spectrometry. Anal Bioanal Chem. 405(27):8885-8896.

Paulik, L.B., B.W. Smith, A.J. Bergmann, G.J. Sower, N.D. Forsberg, J.G. Teeguarden, and K.A. 1 Anderson. 2016.

Passive samplers accurately predict PAH levels in resident crayfish. Sci. Total Environ. 544:782-791.

Rager, J.E., M.J. Strynar, S. Liang, R.L. McMahen, A.M. Richard, C.M. Grulke, J.F. Wambaugh, K.K. Isaacs, R. Judson, A.J. Williams, and J.R. Sobus. 2016. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environment Int. 88:269-280.

Roessler, E., E. Belloni, K. Gaudenz, F. Vargas, S.W. Scherer, L.C. Tsui, and M. Muenke. 1997. Mutations in the C-terminal domain of Sonic Hedgehog cause holoprosencephaly. Hum. Mol. Genet. 6(11):1847-1853.

Rovida, C., N. Alépée, A.M. Api, D.A. Basketter, F.Y. Bois, F. Caloni, E. Corsini, M. Daneshian, C. Eskes, J. Ezendam, H. Fuchs, P. Hayden, C. Hegele-Hartung, S. Hoffmann, B. Hubesch, M.N. Jacobs, J. Jaworska, A. Kleensang, N. Kleinstreuer, J. Lalko, R. Landsiedel, F. Lebreux, T. Luechtefeld, M. Locatelli, A. Mehling, A. Natsch, J.W. Pitchford, D. Prater, P. Prieto, A. Schepky, G. Schüürmann, L. Smirnova, C. Toole, E. van Vliet, D. Weisensee, and T. Hartung. 2015. Integrated testing strategies (ITS) for safety assessment. ALTEX 32(1):25-40.

Rudén, C., and S.O. Hansson. 2010. Registration, Evaluation, and Authorization of Chemicals (REACH) is but the first step. How far will it take us? Six further steps to improve the European chemicals legislation. Environ. Health Perspect. 118(1):6-10.

Seltenrich, N. 2015. A hard nut to crack: Reducing chemical migration in food-contact materials. Environ. Health Perspect. 123(7):A174-A179.

Shin, H.M., A. Ernstoff, J.A. Arnot, B.A. Wetmore, S.A. Csiszar, P. Fantke, X. Zhang, T.E. McKone, O. Jolliet, and D.H. Bennett. 2015. Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays. Environ. Sci. Technol. 49(11):6760-6771.

Tal, T.L., C.W. McCollum, P.S. Harris, J. Olin, N. Kleinstreuer, C.E. Wood, C. Hans, S. Shah, F. A. Merchant, M. Bondesson, T.B. Knudsen, S. Padilla, and M.J. Hemmer. 2014. Immediate and long-term consequences of vascular toxicity during zebrafish development. Reprod. Toxicol. 48:51-61.

Voutchkova, A.M., T.G. Osimitz, and P.T. Anastas. 2010. Toward a comprehensive molecular design framework for reduced hazard. Chem. Rev. 110(10):5845-5882.

Wambaugh, J.F., R.W. Setzer, D.M. Reif, S. Gangwal, J. Mitchell-Blackwood, J.A. Arnot, O. Joliet, A. Frame, J. Rabinowitz, T.B. Knudsen, R.S. Judson, P. Egeghy, D. Vallero, and E.A. Cohen Hubal. 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ. Sci. Technol. 47(15):8479-8488.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, L. Li, H.J. Clewell, III, R.S. Judson, K. Freeman, W. Bao, M.A Sochaski, T.M. Chu, M.B. Black, E. Healy, B. Allen, M.E. Andersen, R.D. Wolfinger, and R.S. Thomas. 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol. Sci. 132(2):327-346.

Wu, S., K. Blackburn, J. Amburgey, J. Jaworska, and T. Federle. 2010. A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. Regul. Toxicol. Pharmacol. 56(1):67-81.

Wu, S., J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, and K. Blackburn. 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem. Res. Toxicol. 26(12):1840-1861.

# 6

# Model and Assay Validation and Acceptance

Models and test systems for toxicity testing have evolved over past decades. Their strengths and weaknesses have been debated, and most agree that no inherently perfect model could exist (Cunningham 2002). Gradually, however, regulatory agencies in the United States and elsewhere have come to accept data from mathematical models and from assay systems that use mammalian and other experimental organisms, cultured cells, and bacteria for evaluating potential hazards and quantifying risks posed by chemical exposures. Some model systems have become nearly indispensable for risk assessment even though inherent shortcomings and imperfections have been widely acknowledged. Such systems include rodent cancer bioassays, multigeneration tests of reproductive and developmental outcomes in rodents, and bacterial mutagenicity tests. Such tests and resulting data have become commonly accepted for use in human-health assessments and often serve as a benchmark or comparator for new assays and data types that are emerging (Thomas et al. 2012).

Before new assays are used in particular regulatory-decision contexts, such as pesticide registration, their relevance, reliability, and fitness for purpose are established and documented. Such characterization of assays has evolved into elaborate processes that are commonly referred to as *validation* of alternative methods. Formal mechanisms for validation have been established in the United States, Europe, and many Asian countries. In addition, an international standardization of validation methods is emerging to ensure reciprocity and uniformity of outcomes (Burden et al. 2015). According to the Organisation for Economic Co-operation and Development (OECD), validation is "the process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose" (OECD 2005). In that context, the term *reliability* refers to the reproducibility of the method "within and between laboratories over time, when performed using the same protocol." The term *relevance* is meant to ensure the scientific underpinning of the test and of the outcome that it is meant to evaluate so that it tests "the effect of interest and whether it is meaningful and useful for a particular purpose." The Institute of Medicine (IOM 2010) defined the process of validation as "assessing [an] assay and its measurement performance characteristics [and] determining the range of conditions under which the assay will give reproducible and accurate data."

In plain language, a *validation process* is used to establish for developers and users of an assay that it is ready and acceptable for its intended use. Although the purpose and principles of validation remain generally constant, the underlying process must evolve to reflect scientific advances. Indeed, the availability of new tests has increased dramatically; many are attractive in cost, time, or use of animals and animal-welfare considerations. The number of chemicals that have been evaluated with new test methods has also increased dramatically (Kavlock et al. 2009; Tice et al. 2013). The reliability of the new tests is of general concern given that existing validation processes cannot match the pace of development of new tests.

The new tests are being developed by scientists in academe, private companies, and government laboratories; sometimes, the utility of a particular marker, assay, or model for decision-making is not immediately recognized by the original developer. Likewise, the resources, time, and effort that are invested in the development can be vastly different and not reflect the ultimate utility of a particular test. Thus, the original developers might not be involved in determining whether a test is fit for purpose for a particular application or provides the degree of certainty that is required to provide information necessary in a particular decision-making context.

In this chapter, the committee describes existing frameworks and efforts for validation of new alternative or nontraditional methods, assays, and models and provides recommendations on the key elements of validation for toxicity testing. The committee emphasizes that validation, although important, is not the only factor involved in achieving regulatory acceptance of new alternative test methods. Furthermore, the committee notes that although assay and model validation for toxicity testing is already an established process, other important disciplines, such

as exposure science, have yet to develop formal criteria and processes for validation, although some have developed approaches to establish best practices.

## GUIDANCE ON THE VALIDATION OF IN VITRO AND OTHER NEW TEST METHODS

### United States

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) was established by the National Institute of Environmental Health Sciences (NIEHS) in 1997 as an ad hoc federal interagency committee to address the growing need for obtaining regulatory acceptance of new toxicity-testing methods (NIEHS 1997). The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) was also established in NIEHS to support ICCVAM in "the development and evaluation of new, revised, and alternative methods to identify potential hazards to human health and the environment with a focus on replacing, reducing, or refining animal use" (Casey 2016). Since 2000, ICCVAM activities have been governed by the ICCVAM Authorization Act (2000), which specifies that 15 agencies of the federal government—including the US Food and Drug Administration, the US Environmental Protection Agency, the Consumer Product Safety Commission, the US Department of Transportation, the Occupational Safety and Health Administration, and the US Department of Agriculture—be represented on ICCVAM.

ICCVAM established the *Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods* (NIEHS 2003) and has successfully evaluated and recommended numerous alternative test methods for regulatory use. Test methods that have been evaluated and recommended for use by NICEATM and ICCVAM are aimed at acute systemic toxicity, dermal corrosivity and irritation, developmental toxicity, endocrine disruption, genetic toxicity, immunotoxicity (allergic contact dermatitis), biologics and nanomaterials, pyrogenicity, and ocular toxicity. The evaluation process includes not only individual test methods but computational and integrated testing strategies (Pirone at al. 2014).

ICCVAM-recommended methods, however, have not always been implemented, and this has caused increasing concern. A potential solution for the near term has been to integrate some activities of NICEATM with those of the federal government's Tox21 consortium (Birnbaum 2013). Specifically, the revised charge to NICEATM now consists of supporting ICCVAM; providing bioinformatics and computational toxicology support to NTP and NIEHS projects, especially those related to Tox21; conducting and publishing analyses of data from new, revised, and alternative testing approaches; and providing information to test-method developers, regulators, and regulated industries (Casey 2016).

Another highly relevant activity that was conducted under the auspices of IOM was the report of the Committee on the Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease (IOM 2010). Specifically, that committee recommended a three-part framework for biomarker evaluation consisting of analytical validation (Is the biomarker able to be accurately measured?), qualification (Is the biomarker associated with the clinical end point of concern?), and use (What is the specific context of the proposed use?). Although the primary users of the IOM framework are stakeholders that are concerned with evidence-based decision-making in medicine and public health, the framework has great relevance to the process for validating any new test method (see Box 6-1).

### European Union

In the European Union, formal activities for validating alternative approaches to animal testing started in 1991 with creation of the European Centre for the Validation of Alternative Methods (ECVAM). Since 2011, ECVAM's tasks have been subsumed by the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), part of the European Commission's Joint Research Centre. The general aims and approaches of EURL ECVAM are similar to those of ICCVAM and include activities to advance the scientific and regulatory acceptance of nonanimal tests that are important to biomedical sciences through research, test development, and validation and maintaining databases (Gocht and Schwarz 2013) and to co-ordinate at the European level the independent evaluation of the relevance and reliability of tests for specific purposes. The guiding principles of the EURL ECVAM work are based on ECVAM recommendations concerning the practical and logistical aspects of validating alternative test methods in prospective studies (Balls 1995; Hartung et al. 2004; EC 2016a); the recommendations are in internal guidelines and strategy papers, for example, ECVAM *Guidance on Good Cell Culture Practice* (Coecke et al. 2005), the OECD guidelines (see Box 6-2), and relevant parts of the EU Test Methods Regulation (EC 2008). ECVAM and the European Partnership for Alternative Approaches to Animal Testing (Kinsner-Ovaskainen et al. 2012) have also made conclusions and offered recommendations on the validation of integrated approaches.

### International

At the international level, OECD has been active, especially in the last 5 years, in coordinating the development of formal guidelines for validation of individual tests, alternative methods, and computational models (see Box 6-2). The 1981 Mutual Acceptance of Data Deci-

**BOX 6-1** Summary of the Institute of Medicine Recommendations for Effective Biomarker Evaluation

1. The biomarker evaluation process should consist of the following three steps:
   a. Analytical validation: analyses of available evidence on the analytical performance of an assay;
   b. Qualification: assessment of available evidence on associations between the biomarker and disease states, including data showing effects of interventions on both the biomarker and clinical outcomes; and
   c. Utilization: contextual analysis based on the specific use proposed and the applicability of available evidence to this use. This includes a determination of whether the validation and qualification conducted provide sufficient support for the use proposed.

2a. For biomarkers with regulatory impact, the Food and Drug Administration (FDA) should convene expert panels to evaluate biomarkers and biomarker tests.

2b. Initial evaluation of analytical validation and qualification should be conducted separately from a particular context of use.

2c. The expert panels should reevaluate analytical validation, qualification, and utilization on a continual and a case-by-case basis.

Source: IOM 2010.

sion for the Assessment of Chemicals including Pesticides C(81)30(Final) stipulated that "data generated in the testing of chemicals in an OECD Member country in accordance with OECD Test Guidelines and OECD Principles of Good Laboratory Practice shall be accepted in other Member countries for purposes of assessment and other uses relating to the protection of man and the environment." It created an impetus for establishing a formal international process for validating test methods. A formal process now exists for development and adoption of OECD test guidelines, part being a formal validation, where the nomination usually begins at the national level, proceeds through the expert committees (from the Working Group of National Coordinators of the Test Guidelines Programme to OECD Chemicals and Environmental Policy Committees), and ultimately is approved by the OECD Council.

### Opinions of the Broader Scientific Community on Validation

Because of the importance of validating novel toxicity-testing methods and the reality of the rapid proliferation of new tests, many opinions have been voiced in the last decade on how the validation process needs to evolve. Although there are various degrees of formality in the suggested changes, all authors agree that the existing frameworks are not optimal and could be improved. Hartung (2007) argued for a move away from validating by comparison with existing "gold standards,"[1] a common testing approach that might not reflect molecular and physiological realities of the human body and argued that tests should be developed to provide more mechanistic information and thus help to establish causality.

Judson and colleagues (Judson et al. 2013) suggested the following general principles: follow current validation practice to the extent possible and practical, increase the use of reference compounds to demonstrate assay reliability and relevance better, de-emphasize the need for cross-laboratory testing, and implement a Web-based, transparent, and expedited peer-review process.

Patlewicz and colleagues (Patlewicz et al. 2013) argued that standard steps of validation practice should still apply and that the validation process for any new test must articulate the scientific and regulatory rationale for the test, the relationship between what the test measures and the resulting biological effect of interest, a detailed protocol for the test, the domain of applicability, criteria for describing the results of the test, known limitations, and standards for determining good performance (positive and negative standards).

Finally, the International Life Sciences Institute Health and Environmental Sciences Institute, an industry-funded nonprofit organization, has recently begun a new project on developing a "Framework for Intelligent Non-Animal

[1]A gold standard is defined as a reference standard that is regarded as the best available to determine a particular condition. The gold standard is the benchmark with which a new procedure is compared. Data from clinical trials and epidemiological studies provide the best examples of benchmarks for the potential effects of drugs or chemicals on the human body. In toxicology, there are cases in which the currently used methods are regarded as inadequate to predict human toxicity. In such cases, other validation methods need to be considered.

**BOX 6-2** Sources of OECD Guidance on Validation of Alternative Test Methods and Models

- Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (OECD 2005)
- Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models (OECD 2007)
- Guidance Document for Describing Non-Guideline In Vitro Test Methods (OECD 2014)

Methods for Safety Assessment."[2] This activity is pursuing a mission to bring together the collective knowledge of scientists from academe, industry, and government with an eye to the development of criteria to establish confidence in using nonanimal methods to support regulatory decisions and to develop a framework organized around IOM (2010) principles noted above.

## CHALLENGES AND RECOMMENDATIONS

The following sections describe what the committee views as the most important aspects of the validation process and challenges associated with them. The committee provides some recommendations for overcoming the challenges and for moving the validation process forward to meet the needs of assessing novel test methods.

### Defining the Scope and Purpose of New Assays as an Essential Element in the Process of Validation and Acceptance

Most of the existing guidance deals with the technical aspects of the process for assay validation, but it is equally important to determine whether a new assay or test battery is meant to replace an existing one or is a novel approach that aims to improve decision-making and provide information that is critical but previously unavailable.

*Recommendation*: A clear definition of the purpose of the new test should be considered before a specific validation process is defined. One must establish the fitness of the test for a particular decision context, select appropriate comparators (for example, a gold standard, mechanistic events, or biomarkers), and delineate the scope of the validation exercise to be commensurate with the proposed use. For example, can a new assay or test battery be used to characterize subchronic or chronic adverse health end points? Test performance characteristics (specificity, sensitivity, and coverage) might need to be adjusted, depending on the decision type and context. Ultimately, it should be clear whether the validation process is aimed at testing reliability, validity, or both.

[2]See http://old.hesiglobal.org/i4a/pages/index.cfm?pageid=3687.

### Enabling Fit-for-Purpose Validation

The challenge of finding an appropriate comparator to enable fit-for-purpose validation of new test methods is considerable because disagreements about the quality of a gold standard or about whether there is one are common. If it is the case of validating a new assay as a replacement for an existing one, one must determine what gold standard is to be used as a comparator. Expert judgment will be needed to determine the validity of an existing method or model to be used as the comparator. If it is the case of validating a novel approach, the decision context for which the information can be used and the availability of other data need to be clearly defined. Statisticians have addressed the question of how to assess the validity of test methods when there is no gold standard (Rutjes et al. 2007). Some of the methods involve correction of imperfect reference standards through the use of additional information or imputed values. Other methods construct a reference standard by using the results of multiple test methods. Each approach has merits for the purpose of replacing animal tests for toxicity.

Two important issues on which there is still no consensus in the scientific community are evaluation of the validity of assays that are not intended as one-to-one replacements for in vivo toxicity assays and assessment of the concordance of data from assays that use cells or proteins of human origin and toxicity data that are virtually all derived from animal models. Judson et al. (2013) have provided ideas on how to validate assays that are intended to be used in a high-throughput context and to be interpreted only in the context of the results of many other assays that evaluate the same biological effect or pathway. Those ideas need to be debated, modified, and tested. As to the concordance issue, it is likely that lack of concordance among species is due not to large differences in the function of highly conserved proteins, such as steroid receptors, but to differences in pharmacokinetics and metabolism. Selected investigation of interspecies concordance at a molecular level will prove or disprove that hypothesis. Data already exist in the literature that will allow comparisons, and the results will support decisions

on what modifications, if any, are needed to accommodate species differences in validation efforts.

*Recommendation*: Workshops or other mechanisms that can be used to develop consensus opinions among scientific experts on defining appropriate reference standards should be considered. Appropriate disclaimers about author affiliations should be included in any reports or opinions that might result from the activities; conflicts of interest need to be carefully managed.

### Establishing the Utility and Domain of New Assays

Another important aspect of validation is establishing the assay utility and clearly defining its domain of applicability,[3] its capacity for chemical biotransformation, its ability to establish a concentration–response relationship, its mechanistic relevance, and the applicability of its results. It is necessary to ensure that negative test results are not negative because of the lack of chemical metabolism, insufficient concentration tested, chemical volatility, chemical binding to plastic, or other factors. Determining the validity of negative results is an important challenging issue because the stakeholders inherently weigh positive data more than negative data or vice versa, depending on the decision context. Likewise, understanding the mechanistic relevance of a result of a new assay is important; it should be clear whether the test is assessing an initiating event, a key event, or an adverse outcome.

*Recommendation*: A description of the utility and domain of the test should be provided to inform the validation process and the ultimate use and interpretation of the data. There should be a clear statement concerning what a positive response or a negative (no) response from the assay means and what controls are appropriate or should be used.

### Establishing Performance Standards

Data quality is a key determinant of acceptance of any test method. Assay performance guidelines that include quality-assurance metrics and quality control of day-to-day operation are well defined (for example, OECD Performance Based Test Guideline TG455), and it is widely recognized that such information needs to be documented. Performance standards[4] are critical in a validation context and are a step toward regulatory acceptance, such as development into an OECD test guideline; however, performance standards are not equally well defined for all types of assays. For example, OECD provides performance standards primarily on estrogen-receptor activity and skin irritation, corrosion, and sensitization.[5]

*Recommendation*: Performance standards should be developed for all types of assays that evaluate relevant adverse health outcomes with relevance being determined by a particular decision context.

Another important part of testing assay performance is establishing reference-chemical lists. A validation reference-chemical list for a number of end points to guide assay developers should help to mitigate disagreements among stakeholders. Engagement of stakeholders—such as regulatory-agency staff, nongovernment organizations, and industry—in establishing the lists will contribute to acceptance of the data produced by assays that are validated using the lists. Some effort has been invested in addressing this challenge, and some valuable lists have been created (Brown 2002; Eskes et al. 2007; Casati et al. 2009; Pazos et al. 2010; EC 2016b). However, there are few molecular targets for which there is a diverse set of specifically defined reference chemicals that can aid in determining both positive and negative performance of a test.

*Recommendation*: Common chemical lists that are fit for different purposes and can evolve should be defined and used for validation of assays and models where possible. That will help the scientific community to establish specificity and potential redundancy among new assays.

Validation or testing in multiple laboratories is one common element of current practice; however, it is recognized that ring trials[6] take too long and are difficult to accomplish if the assays are proprietary, use ultrahigh throughput, or require specialized equipment or expertise. There might not be enough qualified laboratories in the world to perform the test. In the European Union, a network of vetted laboratories that can conduct validation reliably has been established as one way to address the challenge (European Union Network of Laboratories for the Validation of Alternative Methods). Judson et al. (2013) offered another possible solution and proposed performance-based validation: one validates the performance of a new test against the results of previously validated tests for the same end point (for example, a "gold-

[3]The domain of applicability defines what substances can be reliably tested in the assay. For example, can substances that have limited solubility or are volatile be tested using the assay?

[4]Performance standards "provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (1) essential test method components; (2) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of reference chemicals" (OECD 2005).

[5]See http://www.oecd.org/chemicalsafety/testing/performance-standards.htm.

[6] In a ring trial, a given assay is tested in established laboratories to determine its reliability.

standard" test that might have undergone the formal OECD-like validation). Yet another alternative is to use a consensus resulting from multiple tests as a benchmark against which each test is evaluated and to assess variation about the consensus by using resampling techniques or meta-analysis (see Chapter 7). However, there is a real challenge in that many protocols that are used by contract research laboratories to conduct guideline tests are proprietary. Patlewicz et al. (2013) emphasize that any new validation approaches need to allow proprietary tests. In one solution for validating proprietary tests, an outside body provides blinded samples to the testing laboratory and then independently evaluates the accuracy of the test.

*Recommendation*: Government agencies should provide explicit incentives to academic, government, or commercial laboratories to participate in validation.

An alternative (or additional consideration) to technical ring trials is peer review of the methods and of data from new assays. However, more accessible and consistently formatted data are needed for validation through peer review. Data transparency and current agency-specific practices for releasing data to the public pose many challenges. For example, although ToxCast and Tox21 programs have established practices for releasing data in various formats, other agencies in the United States and abroad are not as advanced. Legal challenges involved in data access are many; not only might assays be proprietary but data from nonproprietary assays might be considered confidential business information.

*Recommendation*: Data collected through coordinated validation or screening programs in government laboratories or under contract to government agencies, especially with respect to novel test methods, should be made publicly available as soon as possible, preferably through user-friendly Web-based dashboards. If data are subject to human-subject protections or raise privacy concerns, appropriate measures should be taken to de-identify the information that is being released.

### Establishing Clear Reporting Standards for Assay Results and Testing Conditions

It is widely recognized that the level of detail on methods and experimental conditions reported in scientific publications can be limited by manuscript length restrictions and other factors. It is critical, however, that sufficient information be included in the documentation of assay- or model-validation exercises. It might appear to assay or model developers that some details are obvious and not needed in the documentation, but reproducibility and validity of results might be critically affected by the omission or incompleteness of information. Results might also be misinterpreted in application if incorrect inferences are drawn.

*Recommendation*: Government agencies and organizations involved in assay and model validation should develop clear guidance documents and training materials to support validation, such as training materials that cover various technical aspects of good in vitro method development and practices and cover reporting of methods. All technical aspects of the assay—such as number of cells; media, serum, or additives used; incubation length; readout description; equipment needed; and positive and negative controls—should be described as completely as possible and with the degree of detail needed for replication. The committee acknowledges that for proprietary reasons some information might need to be withheld, but best practice should include disclosure of the nature of and reason for withholding information.

*Recommendation*: Because the chemical or particle concentrations can be different from the administered (nominal or assumed) concentrations, depending on the chemical or particle properties (such as partitioning coefficients and metabolic rates) and the assay system (test materials), efforts should be made to quantify the concentrations in the test system that correspond with the response in the assays either through measurement or through mass-balance model estimation.

### Establishing Clear Guidelines for Evaluating Data Integration and Computational Predictive Modeling in a Common Framework

In the 21st century toxicity-testing paradigm, the results of particular assays are likely to be integrated with data from other sources to obtain the most confident assessment of risk possible. Such integration is the topic of Chapter 7. In anticipation of that chapter, the committee addresses performance issues around models here.

The integrated analysis of data from multiple sources will be increasingly required for making regulatory decisions, and the collective use of these data can be viewed as a new, comprehensive "assay." However, the multiple aspects of an integrated decision process present challenges in reliability and evaluation. The framework underlying integrated approaches to testing and assessment (OECD 2008) provides one example of a structured strategy for combining information for hazard identification and assessment. Here, the focus is on the quality and reliability of the computational aspects of data integration, which are often used in concert with traditional assays. Many of the validation principles of relevance and reliability that were developed for quantitative structure–activity relationship (QSAR) models by OECD (2007) apply to any statistical and integrated model (see Chapter 7 for further discussion). The OECD principles for QSAR model development call for (a) a defined end point, (b) an unambiguous algorithm, (c) a defined domain of (chemical) applicability, (d) measures of goodness of fit, robustness,

and predictivity, and ideally (e) a mechanistic interpretation. Items (b) and (d) often pose the greatest challenge for QSAR or any statistical model, in that complicated modeling schemes are often difficult to reproduce precisely. It has also been recognized and confirmed through systematic reviews of external validation studies of multivariable prediction models that most studies report key details poorly and lack clarity on whether validation was truly external to the information on which the model was based (Collins et al. 2014). Recent efforts by the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) initiative resulted in recommendations for the reporting of studies that develop, validate, or update a prediction model, whether for diagnostic or prognostic purposes (Collins et al. 2015).

Integrated assessment strategies can also benefit from redundancies and weighting of similar assays because a single in vitro assay will probably not provide a "perfect" result. Even assays that are similar mechanistically will likely have some degree of discordance because biological processes are complex, and some test chemicals might be unsuitable for certain assays. In addition, many environmental chemicals are likely to have low potency. As a result, there will be variation from assay to assay in what would be considered a positive response. Multiple assays for critical targets are likely to be needed and can be combined by using computational models (Browne et al. 2015). Any weighting scheme that is data-driven should be carefully cross-validated to avoid optimistic or overfitted final schemes.

As noted, data from assays might be combined with other lines of data to guide decision-making, and issues of documentation and transparency that arise when assay data are combined are similar to those involved when data from a single assay are used.

*Recommendation*: Technical aspects of a statistical predictive model should be described with enough detail for all major steps to be independently reproduced and to ensure the utility and reliability of the predictive models. Statistical predictive models often result in implicit weighting schemes for various features, such as chemical descriptors in QSAR models. Where possible, the final features used and relative model contributions should be published to open the "black box" for future investigators.

*Recommendation*: Weighting schemes for combining assays should be cross-validated if predictive performance or another criterion is driven by the current data and is used in developing a scheme.

*Recommendation*: A culture of independent reproduction of statistical and integrative models should be fostered, ideally with reliability of models assessed by multiple computational groups working independently.

*Recommendation*: Software tools and scripts should be validated by duplicative review by multiple investigators, and where possible software should be made available by open-source mechanisms for continual quality control.

## REFERENCES

Balls, M. 1995. Defining the role of ECVAM in the development, validation and acceptance of alternative tests and testing strategies. Toxicol. In Vitro 9(6):863-869.

Birnbaum, L.S. 2013. 15 years out: Reinventing ICCVAM. Environ. Health Perspect.121(2):A40.

Brown, N.A. 2002. Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. European Centre for the Validation of Alternative Methods. Altern. Lab. Anim. 30 (2):177-198.

Browne, P., R.S. Judson, W.M. Casey, N.C. Kleinstreuer, and R.S. Thomas. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environ. Sci. Technol. 49(14):8804-8814.

Burden, N., C. Mahony, B.P. Müller, C. Terry, C. Westmoreland, and I. Kimber. 2015. Aligning the 3Rs with new paradigms in the safety assessment of chemicals. Toxicology 330:62-66.

Casati, S., P. Aeby, I. Kimber, G. Maxwell, J.M. Ovigne, E. Roggen, C. Rovida, L. Tosti, and D. Basketter. 2009. Selection of chemicals for the development and evaluation of in vitro methods for skin sensitisation testing. Altern. Lab. Anim. 37(3):305-312.

Casey, W.M. 2016. Advances in the development and validation of test methods in the United States. Toxicol. Res. 32(1):9-14.

Coecke, S., M. Balls, G. Bowe, J. Davis, G. Gstraunthaler, T. Hartung, R. Hay, O.W. Merten, A. Price, L. Schechtman, G. Stacey, and W. Stokes. 2005. Guidance on good cell culture practice: A report of The Second ECVAM Task Force on Good Cell Culture Practice. ATLA 33(3):261-287.

Collins, G.S., J.A. de Groot, S. Dutton, O.Omar, M. Shanyinde, A. Tajar, M. Voysey, R. Wharton, L.M. Yu, K.G. Moons, and D.G. Altman. 2014. External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. BMC Med. Res. Methodol. 14:40.

Collins, G.S., J.B. Reitsma, D.G. Altman, and K.G. Moons. 2015. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD statement. J. Clin. Epidemiol. 68(2):134-143.

Cunningham, M.L. 2002. A mouse is not a rat is not a human: Special differences exist. Toxicol. Sci. 70(2):157-158.

EC (European Commission). 2008. Council Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration,

Evaluation, Authorisation and Restriction of Chemicals (REACH). OJEU 51(L142):1-739.

EC (European Commission). 2016a. Validation and Regulatory Acceptance. Joint Research Centre [online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/validation-regulatory-acceptance [accessed January 3, 2017].

EC (European Commission). 2016b. EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database of AMES Positive Chemicals. Joint Research Centre [online]. Available: https://eurl-ecvam.jrc.ec.europa.eu/databases/genotoxicity-carcinogenicity-db [accessed October 24, 2016].

Eskes, C., T. Cole, S. Hoffmann, A. Worth, A. Cockshott, I. Gerner, and V. Zuang. 2007. The ECVAM international validation study on in vitro tests for acute skin irritation: Selection of test chemicals. Altern. Lab. Anim. 35(6):603-619.

Gocht, T., and M. Schwarz, eds. 2013. Implementation of the Research Strategy [online]. Available: http://www.detect-iv-e.eu/wp-content/uploads/2013/09/SEURAT-1v3_LD.pdf [accessed January 3, 2017].

Hartung, T. 2007. Food for thought …on validation. ALTEX 24(2):67-80.

Hartung, T., S. Bremer, S. Casati, S. Coecke, R. Corvi, S. Fortaner, L. Gribaldo, M. Halder, S. Hoffmann, A.J. Roi, P. Prieto, E. Sabbioni, L. Scott, A. Worth, and V. Zuang. 2004. A modular approach to the ECVAM principles on test validity. ATLA 32(5):467-472.

IOM (Institute of Medicine). 2010. Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease. Washington, DC: The National Academies Press.

Judson, R., R. Kavlock, M. Martin, D. Reif, K. Houck, T. Knudsen, A. Richard, R.R. Tice, M. Whelan, M. Xia, R. Huang, C. Austin, G. Daston, T. Hartung, J.R. Fowle, III, W. Wooge, W. Tong, and D. Dix. 2013. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. ALTEX 30(1):51-56.

Kavlock, R.J., C.P. Austin, and R.R. Tice. 2009. Toxicity testing in the 21st century: Implications for human health risk assessment. Risk Anal. 29(4):485-487.

Kinsner-Ovaskainen, A., G. Maxwell, J. Kreysa, J. Barroso, E. Adriaens, N. Alépée, N. Berg, S. Bremer, S. Coecke, J.Z. Comenges, R. Corvi, S. Casati, G. Dal Negro, M. Marrec-Fairley, C. Griesinger, M. Halder, E. Heisler, D. Hirmann, A. Kleensang, A. Kopp-Schneider, S. Lapenna, S. Munn, P. Prieto, L. Schechtman, T. Schultz, J.M. Vidal, A. Worth, and V. Zuang. 2012. Report of the EPAA-ECVAM workshop on the validation of Integrated Testing Strategies (ITS). Altern. Lab. Anim. 40(3):175-181.

NIEHS (National Institute of Environmental Health Sciences). 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No. 97-3981. NIEHS, Research Triangle Park, NC [online]. Available: https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf [accessed July 29, 2016].

NIEHS (National Institute of Environmental Health Sciences). 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Prepared by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) [online]. Available: https://ntp.niehs.nih.gov/iccvam/suppdocs/subguidelines/sd_subg034508.pdf [accessed July 29, 2016].

OECD (Organisation for Economic Co-operation and Development). 2005. Guidance Document on the Validation of and International Acceptance of New or Updated Test Methods for Hazard Assessment. ENV/JM/MONO(2004)14. OECD Series on Testing and Assessment No. 34. Paris: OECD [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2005)14&doclanguage=en [accessed July 29, 2016].

OECD (Organisation for Economic Co-operation and Development). 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. ENV/JM/MONO(2007)2. OECD Series on Testing and Assessment. Paris: OECD [online]. Available: http://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm [accessed July 29, 2016]

OECD (Organisation for Economic Co-operation and Development). 2008. Guidance Document on Magnitude of Pesticide Residues in Processed Commodities. ENV/JM/MONO(2008)23. OECD Series on Testing and Assessment No. 96. Paris: OECD [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2008)23&doclanguage=en [accessed July 29, 2016].

OECD (Organisation for Economic Co-operation and Development). 2014. Guidance Document for Describing Non-guideline in Vitro Test Methods. ENV/JM/MONO(2014)35. OECD Series on Testing and Assessment No. 211. Paris: OECD [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2014)35&doclanguage=en [accessed July 29, 2016].

Patlewicz, G., T. Simon, K. Goyak, R.D. Phillips, J.C. Rowlands, S.D. Seidel, and R.A. Becker. 2013. Use and validation of HT/HC assays to support 21st century toxicity evaluations. Regul. Toxiol. Pharmacol. 65(2):259-268.

Pazos, P., C. Pellizzer, T. Stummann, L. Hareng, and S. Bremer. 2010. The test chemical selection procedure of the European Centre for the Validation of Alternative Methods for the EU Project ReProTect, Reprod. Toxicol. 30(1):161-199.

Pirone, J.R., M. Smith, N.C. Kleinstreuer, T.A. Burns, J. Strickland, Y. Dancik, R. Morris, L.A. Rinckel, W. Casey, and J.S. Jaworska. 2014. Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. ALTEX 31(3):336-340.

Rutjes, A.W., J. B. Reitsma, A. Coomarasamy, K.S. Khan, and P.M. Bossuyt. 2007. Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol. Assess. 11(50):iii, ix-51.

Thomas, R.S., M.B. Black, L. Li, E. Healy, T.M. Chu, W. Bao, M.E. Andersen, and R.D. Wolfinger. 2012. A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. Toxicol. Sci. 128(2):398-417.

Tice, R.R., C.P. Austin, R.J. Kavlock and J.R. Bucher. 2013. Improving the human hazard characterization of chemicals: A Tox21 update. Environ. Health Perspect. 121(7):756-765.

# 7

# Interpretation and Integration of Data and Evidence for Risk-Based Decision-Making

Chapters 2–4 highlighted major advances in exposure science, toxicology, and epidemiology that will enable a better understanding of pathways, components, and mechanisms that contribute to disease. As described in those chapters, the new tools and the resulting data will improve the assessment of exposures that are associated with incremental increases in risk and will enhance the characterization of the spectrum of hazards that can be caused by chemicals. Chapter 5 described the new direction of risk assessment that is based on biological pathways and processes. That approach acknowledges the multifactorial and nonspecific nature of disease causation—that is, stressors from multiple sources can contribute to a single disease, and a single stressor can lead to multiple adverse outcomes. The new direction offers great promise for illuminating how various agents cause disease, but 21st century science—with its diverse, complex, and potentially large datasets—poses challenges related to analysis, interpretation, and integration of the data that are used in risk assessment. For example, transparent, reliable, and vetted approaches are needed to analyze toxicogenomic data to detect the signals that are relevant for risk assessment and to integrate the findings with results of traditional whole-animal assays and epidemiological studies. Approaches will also be needed to analyze and integrate different 21st century data streams and ultimately to use them as the basis of inferences about, for example, chemical hazard, dose–response relationships, and groups that are at higher risk than the general population. Agencies have systems of practice, guidelines, and default assumptions to support consistent and efficient approaches to risk assessment in the face of underlying uncertainties, but their practices will need to be updated to accommodate the new data.

In this chapter, the committee offers some recommendations for improving the use of the new data in reaching conclusions for the purpose of decision-making. Steps in the process include analyzing the data to determine what new evidence has been generated (data-analysis step), combining new data with other datasets in integrated analyses (data-integration step), and synthesizing evidence from multiple sources, for example, for making causal inferences, characterizing exposures and dose–response relationships, and gauging uncertainty (evidence-integration step). The three steps should be distinguished from each other. The purpose of data analysis is to determine what has been learned from the new data, such as exposure data or results from individual toxicity assays. The new data might be combined with similar or complementary data in an integrative analysis, and the resulting evidence might then be integrated with prior evidence from other sources. Because the terminology in the various steps has varied among reports from agencies and organizations, the committee that prepared the present report adopts the concepts and terminology in Box 7-1.

The committee begins by considering data interpretation when using the new science in risk assessment and next discusses some approaches for evaluating and integrating data and evidence for decision-making. The committee briefly discusses uncertainties associated with the new data and methods. The chapter concludes by describing some challenges and offering recommendations to address them.

## DATA INTERPRETATION AND KEY INFERENCES

Interpreting data and drawing evidence-based inferences are essential elements in making risk-based decisions. Whether for establishing public-health protective limits for air-pollution concentrations or for determining the safety of a food additive, the approach used to draw conclusions from data is a fundamental issue for risk assessors and decision-makers. Drawing inferences about human-health risks that are based on a pathway approach can involve answering the following fundamental questions:

- Can an identified pathway, alone or in combination with other pathways, when sufficiently perturbed, increase the risk of an adverse outcome or disease in humans, particularly in sensitive or vulnerable individuals?

- Do the available data—in vitro, in vivo, computational, and epidemiological data—support the judgment that the chemical or agent perturbs one or more pathways linked to an adverse outcome?
- How does the response or pathway activation change with exposure? By how much does a chemical or agent exposure increase the risk of outcomes of interest?
- Which populations are likely to be the most affected? Are some more susceptible because of co-exposures, pre-existing disease, or genetic susceptibility? Are exposures of the young or elderly of greater concern?

To set the context for the discussion of inference and data interpretation to address the above questions, the committee begins by considering a useful causal model of disease. As discussed in Chapter 5, the focus of toxicological research has shifted from observing apical responses to understanding biological processes or pathways that lead to the apical responses or disease. There is also the recognition that a single adverse outcome might result from multiple mechanisms, which can have multiple components (see Figure 5-1). The 21st century tools, which can be used to determine the degree to which exposures perturb pathways or activate mechanisms, facilitate a new direction in risk assessment that acknowledges the multifactorial nature of disease.

One way in which to consider the multifactorial nature of disease is to use the sufficient-component-cause model (Rothman 1976; Rothman and Greenland 2005). The sufficient-component-cause model is an extension of the counterfactual notion[1] and considers sets of actions, events, or states of nature that together lead to the outcome under consideration. The model provides a way to account for multiple factors that can combine to result in disease in an individual or population. It addresses the question, What are the various events that might have caused a particular effect? For example, a house caught fire because of a constellation of events—fire in the fireplace, wooden house, strong wind, and alarm not functioning—that together formed a sufficient causal complex, but no component was sufficient in itself (Mackie 1980). The model leads to the designation of causes or events as necessary, sufficient, or neither.

Figure 7-1 illustrates the sufficient-component-cause concept and shows that the same outcome can result from more than one causal complex or mechanism; each "pie" has multiple components and generally involves the joint action of multiple components. Although most components are neither necessary (contained in every pie) nor sufficient (single-component-cause pie) to produce outcomes or diseases, the removal of any one component will prevent some outcomes. If the component is part of a common complex or part of most complexes, removing it would be expected to result in prevention of a substantial amount of disease or possibly of all disease (IOM 2008). It is important to note that not every component in a complex has to be known or removed to prevent cases of disease. And exposures to each component of a pie do not have to occur at the same time or in the same space, depending on the nature of the disease-producing process. Relevant exposures might accumulate over the life span or occur during a critical age window. Thus, multiple exposures (chemical and nonchemical) throughout the life span might affect multiple components in multiple mechanisms. Moreover, variability in the exposures received by the population and in underlying susceptibility and the multifactorial nature of chronic disease imply that multiple mechanisms can contribute to the disease burden in a population.

[1]The counterfactual is the state that is counter to the facts; for example, what would the risk of lung cancer have been if cigarette-smoking did not exist?

**BOX 7-1** Data Analysis and Integration Terminology Used in This Report

*Data*: the quantitative or qualitative values generated by a measurement process or modeling.

*Evidence*: the accumulated body of knowledge on a particular topic.

*Data analysis*: the application of mathematical and statistical techniques to a dataset to investigate hypotheses, perform estimation, and assess the evidence.

*Data integration*: analytical processes that combine data from multiple sources.

*Evidence integration*: the consideration, whether qualitative or quantitative, of evidence from multiple sources.

*Causal inference*: the evaluation of evidence from all relevant sources to judge whether an association is causal.

The definitions of component, mechanism, and pathway are the same as those provided in Chapter 5 in the discussion of the new direction in risk assessment. Box 7-2 provides the definitions in the context of the sufficient-component-cause model and is a reminder of the general definitions provided in Chapter 1. Given Figure 7-1 and Box 7-2, a mechanism of a disease will typically involve more than one component or pathway; multiple pathways will likely be involved in the production of disease.

The sufficient-component-cause model is a useful construct for considering methods for interpreting data and drawing inferences for risk assessment on the basis of 21st century data. It can be used to interpret mechanistic data for addressing the four critical questions above. And it is useful for considering whether a mechanism is complete (that is, whether all the necessary components are present or activated sufficiently to produce disease) and for considering the degree to which elimination or suppression of one component might be preventive.

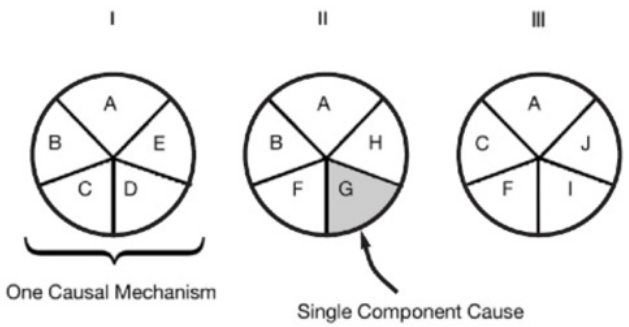



**FIGURE 7-1** Multifactorial nature of disease illustrated by the sufficient-component-cause model in which various overall mechanisms (I, II, and III) of a disease are represented as causal pies of various components (A–J).

## Identifying Components, Mechanisms, and Pathways That Contribute to Disease

Research on the causes of cancer provides a concrete example of the uses of the multifactorial disease concept and consideration of upstream biological characteristics. Ten characteristics of carcinogens have been proposed (IARC 2015; Smith et al. 2016) on the basis of mechanisms associated with chemicals that are known to cause cancer in humans (see Table 7-1). The International Agency for Research on Cancer (IARC) is using the characteristics as a way to organize mechanistic data relevant to agent-specific evaluations of carcinogenicity (IARC 2016a). The committee notes that key characteristics for other hazards, such as cardiovascular and reproductive toxicity, could be developed as a guide for evaluating the relationship between perturbations observed in assays, their potential to pose a hazard, and their contribution to risk.

The IARC characteristics include components and pathways that can contribute to a cancer. For example, "modulates receptor-meditated effects" includes activation of the aryl hydrocarbon receptor, which can initiate downstream events, many of which are linked to cancer, such as thyroid-hormone induction, xenobiotic metabolism, pro-inflammatory response, and altered cell-cycle control. Ones that are linked often fall under other IARC characteristics—for example, "cell-cycle control" falls under "alters cell proliferation"—and therefore are components of other characteristics. At the molecular level, some specific pathways that are ascribed to particular cancers (for example, the "chromosome unstable pathway" for pancreatic cancer) and fall within the IARC characteristics of carcinogens have been curated in the Kyoto Encyclopedia of Genes and Genomes[2] databases.

[2]See http://www.genome.jp/kegg/disease/.

**BOX 7-2** Definitions of Component, Mechanism, and Pathway for This Report

*Component*: In the sufficient-component-cause model, a biological factor, event, or condition that when present with other components produces a disease or other adverse health outcome.

*Mechanism*: Generally, a detailed description of the process by which an agent causes an effect. In the sufficient-component-cause model, the committee considers mechanisms to be comprised of components that cause disease or other adverse health outcome when they co-occur.

*Pathway*: The sequence of events or network of biological processes that make up mechanisms. In applying the sufficient-component-cause model, the committee considers pathways to be components of mechanisms.

**TABLE 7-1** Characteristics of Carcinogens

| Characteristic[a] | Example of Relevant Evidence |
|---|---|
| Is electrophilic or can be metabolically activated | Parent compound or metabolite with an electrophilic structure (e.g., epoxide or quinone), formation of DNA and protein adducts |
| Is genotoxic | DNA damage (DNA-strand breaks, DNA-protein crosslinks, or unscheduled DNA synthesis), intercalation, gene mutations, cytogenetic changes (e.g., chromosome aberrations or micronuclei) |
| Alters DNA repair or causes genomic instability | Alterations of DNA replication or repair (e.g., topoisomerase II, base-excision, or double-strand break repair) |
| Induces epigenetic alterations | DNA methylation, histone modification, microRNA expression |
| Induces oxidative stress | Oxygen radicals, oxidative stress, oxidative damage to macromolecules (e.g., DNA or lipids) |
| Induces chronic inflammation | Increased white blood cells, myeloperoxidase activity, altered cytokine, or chemokine production |
| Is immunosuppressive | Decreased immunosurveillance, immune-system dysfunction |
| Modulates receptor-mediated effects | Receptor activation or inactivation (e.g., ER, PPAR, AhR) or modulation of endogenous ligands (including hormones) |
| Causes immortalization | Inhibition of senescence, cell transformation |
| Alters cell proliferation, cell death, or nutrient supply | Increased proliferation, decreased apoptosis, changes in growth factors, energetics and signaling pathways related to cellular replication or cell cycle |

[a]Any characteristic could interact with any other (such as oxidative stress, DNA damage, and chronic inflammation), and a combination provides stronger evidence of a cancer mechanism than one would alone.
Sources: IARC 2016; Smith et al. 2016.

One challenge is to evaluate whether a component or specific biological pathway contributes to a particular adverse outcome or disease. The challenge is not trivial given that inferences must be drawn from evidence that is far upstream of the apical outcome. The ability to identify the contributions of various components and pathway perturbations to disease and to understand the importance of changes in them can be critical to 21st century risk-based decision-making. However, the need for such an understanding will be specific to the decision context. In some contexts, the lack of any observable effect on biological processes in adequate testing at levels much above those associated with any human exposure might be sufficient; thus, there is not always the need to associate biological processes directly with potential human health effects. In other cases, it will be critical to understand whether a pathway contributes to disease, for example, in conducting a formal hazard identification or in deciding which whole-animal assays should be used when a chemical is highly ranked in a priority-setting exercise for further testing.

The committee proposes a possible starting point for linking components, pathways, and, more generally, mechanisms to a particular disease or other adverse outcome. The question is whether the components or pathways and other contributing factors cause the disease. The committee draws on and adapts a causal-inference approach to guide the evaluation of the new types of data. *Causal inference* refers to the process of judging whether evidence is sufficient to conclude that there is a causal relationship between a putative cause (such as a pathway perturbation) and an effect of interest (such as an adverse outcome). The causal guidelines that were developed by Bradford Hill (1965) and by the committee that wrote the 1964 Surgeon General's report on smoking and health (DHEW 1964) have proved particularly useful for interpreting epidemiological findings in the context of experimental and mechanistic evidence. Those guidelines have

been proposed by others for evaluating adverse-outcome pathways (OECD 2013). Box 7-3 presents the Hill–Surgeon General guidelines and suggests how they can be used to evaluate causal linkages between health effects and components, pathways, and mechanisms.

Only one element of the guidelines, that cause precedes effect (temporality), is necessary, although not sufficient. The remaining elements are intended to guide evaluation of a particular body of observational evidence (consistency and strength of association) and to assess the alignment of that evidence with other types of evidence (coherence). The guidelines were not intended to be applied in an algorithmic or check-list fashion, and operationalizing the guidelines for various applications has not been done (for example, defining how many studies are needed to achieve consistency). Use of the guidelines inherently acknowledges the inevitable gaps and uncertainties in the data considered and the need for expert judgment for synthesis. Guidance and best practices should evolve with increased experience in linking pathways, components, and mechanisms to health effects.

Other approaches have been proposed to link outcomes to pathways or mechanisms. The adverse-outcome-pathway and network approaches represent efforts to map pathways that are associated with various outcomes (see, for example, Knapen et al. 2015), and they are based on general guidance (OECD 2013) similar to that described above. A complementary approach that deserves consideration is the meet-in-the-middle concept described in Chapter 4, in which one tries to link the biomarkers of exposure and early effect with the biomarkers of intermediate effect and outcome (see Figure 4-1). Different scientific approaches—traditional epidemiology at the population level, traditional toxicology at the organism level, and 21st century tools at the mechanistic level—will be used to address the challenge of linking effects with pathways or mechanisms. The multiple data streams combined with expert-judgement–based systems for causal inference (see Box 7-3; DHEW 1964; EPA 2005, 2015; IARC 2006) will probably serve as bridges between effects seen in assay systems and those observed in animal models or in studies of human disease. Expert judgments should ultimately involve assessments by appropriate multidisciplinary groups of experts, whether external to or in an agency.

### Linking Agents to Pathway Perturbations

For drawing conclusions about whether a substance contributes to disease by perturbing various pathways or activating some mechanism, the committee finds the practice of IARC to be a reasonable approach. In evaluating whether an agent has one or more of the 10 characteristics of carcinogens noted above, IARC (2016) conducts a broad, systematic search of the peer-reviewed in vitro and in vivo data on humans and experimental systems for each of the 10 characteristics and organizes the specific mechanistic evidence by these characteristics. That approach avoids a narrow focus on specific pathways and hypotheses and provides for a broad, holistic consideration of the mechanistic evidence (Smith et al. 2016). IARC rates the evidence on a given characteristic as "strong," "moderate," or "weak" or indicates the lack of substantial data to support an evaluation. The evaluations are incorporated into the overall determinations on the carcinogenicity of a chemical. More recently, after providing the evidence on each of the 10 characteristics, IARC summarized the findings from the Tox21 and ToxCast high-throughput screening programs related to the 10 characteristics with the caveat that "the metabolic capacity of the cell-based assays is variable, and generally limited" (IARC 2015, 2016a).

Integrative approaches are being developed to evaluate high-throughput data in the Tox21 and ToxCast databases for the activity of a chemical in pathways associated with toxicity. Qualitative and quantitative approaches for scoring pathway activity have been applied. For example, scoring systems have been developed for "gene sets" of assays that are directed at activity in receptor-activated pathways, such as pathways involving androgen, estrogen, thyroid-hormone, aromatase, aryl-hydrocarbon, and peroxisome proliferator-activated receptors (Judson et al. 2010; Martin et al. 2010, 2011; EPA 2014) and for "bioactivity sets" that are directed at activity in other general pathways, such as acute inflammation, chronic inflammation, immune response, tissue remodeling, and vascular biology (Kleinstreuer et al. 2014). Chemical mechanisms that are inferred from high-throughput findings do not always match the knowledge of how a chemical affects biological processes that is gained from in vivo and mechanistic studies (Silva et al. 2015; Pham et al. 2016). That discordance underscores the importance of a broad review in associating chemicals to pathways or mechanisms that contribute to health effects. Appendix B provides a case study for a relatively data-sparse chemical that appears to activate the estrogenicity pathway as shown in high-throughput assays; a read-across inference could be drawn by comparing the data-sparse chemical to chemicals in the same structural class that have been studied better.

The causal-guidance topics provided in Box 7-3 can be adapted to guide expert judgments in establishing causal links between chemical exposure and pathway perturbations on the basis of broad, systematic consideration of the evidence from the published literature and government databases. Temporality often is not an issue in the context of experimental assays because the effects are measured after exposure. For epidemiological studies, temporality might be a critical consideration inasmuch as biological specimens that are used to assess exposure might have been collected at times of uncertain relevance

**BOX 7-3** Causal Guidelines for Evaluating Associations of Health Effects and Components, Pathways, and Mechanisms

*Temporality*: Interpretation of evidence on temporality is essential for causal inference: cause must come before effect. Assessment of temporality might be complicated by uncertainty because the full sequence of events that leads to health effects is typically not known, and the suite of possible pathways or components involved in the mechanism is rarely completely understood.

*Strength of association*: The size of an effect related to the exposure in question can be important in identifying causality, although a strong signal in single or multiple assays is not a prerequisite for concluding that a true causal association exists. Nonetheless, strong associations of pathway-perturbation measures (such as thyroid-hormone status) with outcome (such as IQ deficit) weigh against other factors that might have led to the association.

*Consistency*: In the original causal-inference guidelines, consistency referred to the reproducibility of a finding, that is, whether findings from multiple observational studies conducted by different investigators in different populations were comparable. Replication is the basis for scientific progress, and replication in multiple studies increases confidence in the new findings. Another consideration of consistency in the context of 21st century data is related to outcomes that have been linked to suites of chemicals tested in assays that evaluate the same perturbations. Do chemicals that affect similar pathways and mechanisms lead to related outcomes and provide consistent results? Can differences in outcome be explained by population or context differences? Variability in assay performance and in domain applicability can result in inconsistent results that do not necessarily exclude the possibility of a causal relationship.

*Plausibility*: The question here is whether activation of a proposed mechanism or perturbation of a pathway can be plausibly linked to a health effect. Is the association consistent with what is known generally about the chemicals or conditions that perturb various pathways and the outcome of concern? The concept of meet-in-the-middle that was described in Chapter 4 is useful in addressing this question. How are the data related to what has been observed in human populations (if studied) regarding some intermediate biomarker that in turn predicts the probability of disease? A cautionary note in incorporating that criterion into guidelines is that plausibility is intrinsically grounded in the state of knowledge, and mechanisms that lead to health effects might act in ways that reside outside of current biological understanding.

*Specificity*: Specificity—generally interpreted as a singular relationship between an exposure and a disease—is often set aside. For example, tobacco smoke, a complex mixture, causes multiple malignancies, cardiovascular diseases, and respiratory diseases, and these conditions have other causes. With the powerful 21st century tools, the specificity could be explored by answering this question: Does the interference or blocking of a pathway (for example, by using knockout mice) block or otherwise change the occurrence of the outcome?

*Coherence*: Coherence, an element of plausibility, generally refers to the complementarity of different lines of evidence of cause and effect. With 21st century tools, coherence acquires a new dimension. Vertical coherence would be related to consistency over several levels of biological organization; for example, one might consider the effect of an inhibitor of histone deacetylation at different levels of organization. Horizontal coherence would be related to the presence of more than one effect at the same level of organization; for example, one might consider the increased rate of apoptosis and the decreased proliferative rate at the cellular level in the case of inhibition of histone deacetylation.

to the underlying disease pathogenesis and biomarkers of effect, and the development of disease might influence exposure patterns. Regarding strength of outcome in the context of Tox21 data, strong responses in multiple assays that are designed to evaluate a specific pathway or mechanism would provide greater confidence that the tested chemical has the potential to perturb the pathway or activate the mechanism. Assessment of the relative potency of test chemicals in activating a mechanism or perturb a pathway will be informed by running assays with carefully selected and vetted positive and negative reference chemicals that have known in vivo effects. As discussed further below, methods or technologies that produce enormous datasets pose special challenges. Procedures to sift through the data to determine signals of importance are needed. As the scientific community develops experience, quantitative criteria and procedures that reflect best practices can be incorporated into guidelines for judging the significance of signals from such data. Regarding consistency, consideration should be given to findings from the same or similar assays in the published literature and government programs and from assays that use appropriately selected reference chemicals. Caution should be exercised in interpreting consistency of results from multiple assays and chemical space because assays might vary in the extent to which they are "fit for purpose" (see Chapter 6). Regarding plausibility and coherence, there are considerations regarding consistency between what is known generally about a chemical or structurally similar chemicals and the outcome of concern and between findings from different types of assays and in different levels of biological organization. In considering the possible applicability of practices adapted from the Bradford Hill guidelines for evaluating the evidence of pathway perturbations by chemicals, the committee emphasizes that the guidelines are not intended to be applied as a checklist.

### Assessing Dose–Response Relationships

Chapter 5 and the case studies described in the appendixes show how some of the various 21st century data might be used in understanding dose–response relationships for developing a quantitative characterization of risks posed by different exposures. As noted in Chapter 5, it is not necessary to know all the pathways or components involved in a particular disease for one to begin to apply the new tools in risk assessment, and a number of types of analyses that involve dose–response considerations can incorporate the new data. Table 7-2 lists some of those analyses and illustrates the type of inferences or assumptions that would typically be required in them.

Given that most diseases that are the focus of risk assessment have a multifactorial etiology, it is recognized that some disease components result from endogenous processes or are acquired by the human experience, such as background health conditions, co-occurring chemical exposures, food and nutrition, and psychosocial stressors (NRC 2009). Those additional components might be independent of an environmental stressor under study but nonetheless influence and contribute to the risk and incidence of disease (NRC 2009; Morello-Frosch et al. 2011). They also can increase the uncertainty and complexity of dose–response relationships—a topic discussed at length in the NRC (2009) report *Science and Decisions: Advancing Risk Assessment*, and the reader is referred to that report for details on deriving dose–response relationships for apical outcomes by using mechanistic and other data. The 21st century tools provide the mechanistic data to support those deviations.

The committee emphasizes the importance of being transparent, clear, and, to the greatest extent appropriate, consistent about the explicit and implicit biological assumptions that are used in data analysis, particularly dose–response analysis. Best practices will develop over time and should be incorporated into formal guidance to ensure the consistent and transparent use of procedures and assumptions in an agency. The development and vetting of such guidance through scientific peer-review and public-comment processes will support the best use of the new data in dose–response practices. The guidelines should address statistical and study-selection issues in addition to the assumptions that are used in the biological and physical sciences for analyzing such data. For example, studies that are used to provide the basis of the dose–response description should generally provide a better quantitative characterization of human dose–response relationships than the studies that were not selected. Some issues related to statistical analyses in the context of large datasets are considered below. Various dose–response issues presented in Table 7-2 involve integration of information in and between data domains, and tools for such integration and the possible implicit biological assumptions needed for their use are discussed later in this chapter.

### Characterizing Human Variability and Sensitive Populations

People differ in their responses to chemical exposures, and variability in exposure and response is a critical consideration in risk assessment. For example, protection of susceptible populations is a critical aim in many risk-mitigation strategies, such as the setting of National Ambient Air Quality Standards for criteria air pollutants under the Clean Air Act. Variability in response drives population-level dose–response relationships (NRC 2009), but characterizing variability is particularly challenging given the number of sources of variability in response related to such inherent factors as genetic makeups, life stage, and sex and such extrinsic factors as psychosocial stressors, nutrition, and exogenous chemical exposures. Genetic

makeup has often been seen as having a major role in determining variability, but research indicates that it plays only a minor role in determining variability in response related to many diseases (Cui et al. 2016). Thus, in considering use and integration of 21st century science data, the weight given to data that reflect genetic variability needs to be considered in the context of the other sources of human variability.

Figure 7-2 illustrates how a wide array of factors—each potentially varying in a population—can combine to affect the overall degree of interindividual variability in a population (Zeise et al. 2013). Variability is shown in the context of the source-to-outcome continuum that has been expanded and elaborated on in Chapters 2 and 3. As described in Chapter 2, environmental chemical exposure at particular concentrations leads to an internal exposure that is modified by pharmacokinetic elements. As described in Chapter 3, internal exposure results in some molecular changes that progress in later steps to outcomes. Figure 7-2 shows how variability in other exposures and in biological factors can affect different points along the source-to-outcome pathway and lead to different outcomes in individuals. Modern exposure, toxicology and epidemiology tools—including biomarkers and measures of physiological status—can all provide indications of susceptibility status. The same indicators can be observed experimentally and used in models to help in drawing inferences about variability that are relevant to humans.

**TABLE 7-2** Examples of Inferences or Assumptions Needed to Use 21st Century Data in Various Analyses

| Analysis That Involves Dose–Response Considerations | Examples of Inferences or Assumptions Needed |
|---|---|
| Read-across: health reference values derived from structurally or biologically similar anchor chemicals | • Sufficiency of chemical similarities for read-across on the basis, for example, of biological, chemical-structure, metabolic, or mechanistic similarities<br>• Comparison of chemical activity on the basis, for example, of pharmacokinetics and biological activity in assays |
| Toxicogenomic screening to determine whether environmental exposures are of negligible concern or otherwise | • Generalizability of results to susceptible and general human populations<br>• Consequence or importance of toxicogenomic effects seen at exposures greater than environmental exposures<br>• Sufficiency of procedure to filter and analyze genomics data; assumptions as to which pathway-related indicators are important |
| Extrapolation of effect or benchmark doses in vitro to human exposures to establish health reference values[a] | • Sufficiency of understanding about human pharmacokinetic and pharmacodynamic variability<br>• Generalizability of results to susceptible and general human populations. |
| Priority-setting of chemicals for testing on the basis of in vitro screens | • Sufficiency of metabolic capacity and biological coverage of cell systems in domains of interest for chemicals that are being ranked<br>• Adequacy of pharmacokinetic adjustments in the context of human exposures and population variability |
| Clarification of low end of dose–response curve (for rich datasets) | • Sufficiency of understanding of mechanisms<br>• Extent to which sensitive elements of involved pathways have been evaluated by mechanistic studies |
| Construction of dose–response curve from population variability characteristics (NRC 2009) | • Sources of pharmacokinetic and pharmacodynamic variability sufficiently captured and integrated into a population-variability characterization |
| Selection of method or model for dose–response characterization | • Choice of a low-dose linear model or a low-dose non-linear or threshold model on the basis of consideration of mechanisms, population vulnerability, and background exposures (NRC 2009) |

[a]For most outcomes, it is not possible simply to replace a value derived from a whole-animal assay with a value derived from an in vitro assay. The lack of understanding of all the pathways involved makes such direct replacement premature. The lack of metabolic capacity in cell systems and the limitations of biological coverage pose further challenges to the free-standing use of in vitro approaches for derivation of guidance values in most contexts (see Chapters 3 and 5).

Chapter 2 describes pharmacokinetic models of various levels of complexity that can be used to evaluate human interindividual variability in an internal dose that results from a fixed external exposure. Chapter 3 describes relatively large panels of lymphoblastoid cell lines derived from genetically diverse human populations that can be used to examine the genetic basis of interindividual variability in a single pathway. The chapter also describes how genetically diverse panels of inbred mice strains can be used to explore variability and how various studies that use such strains have been able to identify genetic factors associated with liver injury from acetaminophen (Harrill et al. 2009) and tetrachloroethylene (Cichocki et al. in press). The combination of such experimental systems with additional stressors can be used to study other aspects of variability. Chapter 4 covers epidemiological approaches used to observe variability in human populations.

Data-driven variability characterizations have been recommended as a possible replacement for standard defaults used by agencies, in specific cases and in general. Data-driven variability factors can be considered in light of the guidance for departure from defaults provided in NRC (2009), the degree to which the full array of sources of variability have been adequately explored, and the reliability of the evidence integration. The modified causal guidance provided in Box 7-3 can be used to assess the emerging qualitative and quantitative evidence on human variability, and the analysis and integration approaches described later in the chapter are also relevant here.

## APPROACHES FOR EVALUATING AND INTEGRATING DATA AND EVIDENCE

The volume and complexity of 21st century data pose many challenges in analyzing them and integrating them with data from other (traditional) sources. As noted earlier, the necessary first step is the analysis of the toxicity-assay results and exposure data. That stage of analysis is followed by the data-integration step in which the new data are combined with other datasets (the combination of similar or complementary data in an integrative analysis). The results of such analyses might then be integrated with prior evidence from other sources (evidence integration). The discussion below first addresses the issues associated

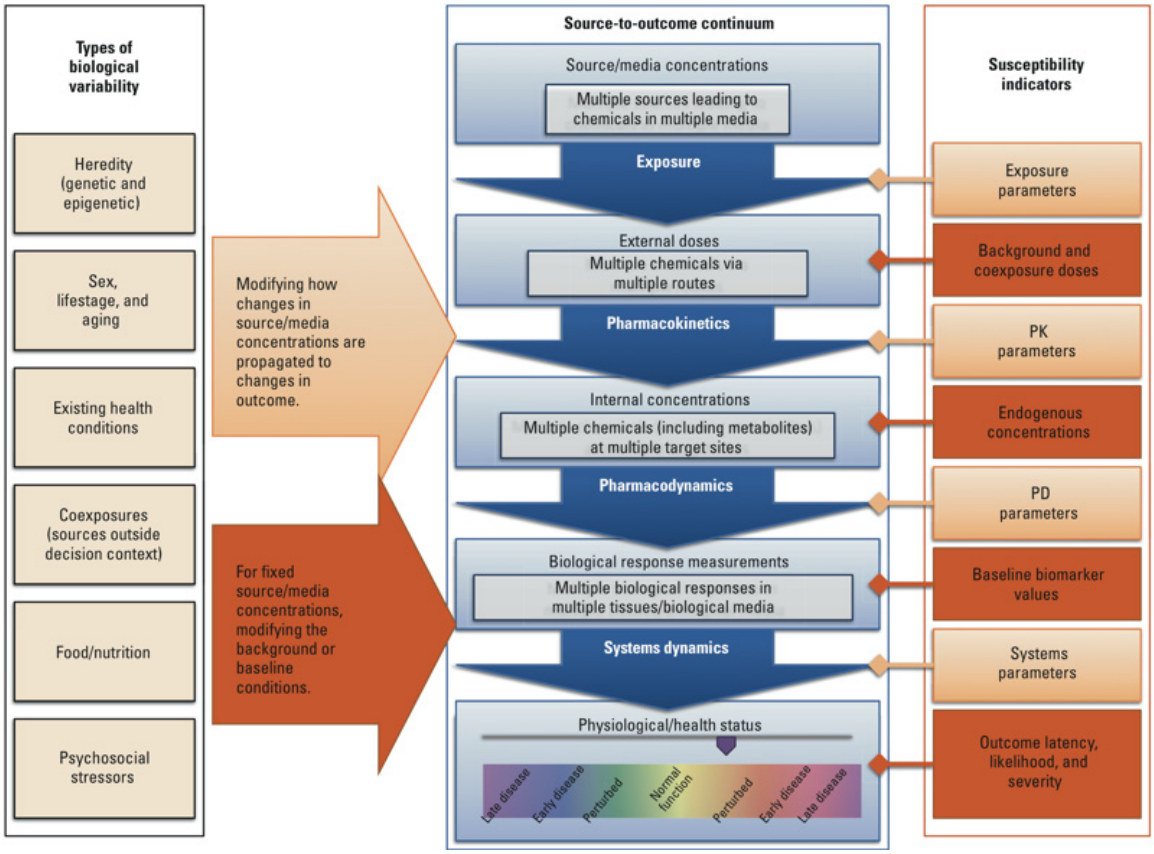



**FIGURE 7-2** Determinants of variability in human response result from inherent and extrinsic factors that influence propagation of dose and responses along the source-to-outcome continuum. Source: Zeise et al. 2013.

with analyzing individual datasets and studies—that is, evaluating individual study quality and tackling the challenge of big data. Next, approaches for interpreting and integrating data from various studies, datasets, and data streams are described, and some suggestions are provided for their use with 21st century data. The committee notes that recent reports of the National Research Council and the National Academies of Sciences, Engineering, and Medicine have dealt extensively with the issues of data and evidence integration (see, for example, NRC 2014 and NASEM 2015). The committee notes that although formal methods receive emphasis below, findings could be sufficiently compelling without the use of complex analytical and integrative methods. In such cases, decisions might be made on direct examination of the findings.

## Analyzing Individual Datasets and Studies

### Evaluating Individual Studies

Several NRC reports have emphasized the need to use standardized or systematic procedures for evaluating individual studies and described some approaches for evaluating risk of bias and study quality (see, for example, NRC 2011, Chapter 7; NRC 2014, Chapter 5). Those reports, however, acknowledged the need to develop methods and tools for evaluating risk of bias in environmental epidemiology, animal, and mechanistic studies. Since release of those reports, approaches for assessing risk of bias in environmental epidemiology and animal studies have been advanced (Rooney et al. 2014; Woodruff and Sutton 2014; NTP 2015a). Approaches for assessing risk of bias in mechanistic studies, however, are still not well developed, and there are no established best practices specifically for high-throughput data. The committee emphasizes the need to develop best practices for systematically evaluating 21st century data and for ensuring transparency when a study or -omics dataset is excluded from analysis. There is also a need for data-visualization tools to aid in interpreting and communicating findings. The committee notes that evaluating the quality of an individual study is a step in systematic review, discussed below.

### Tackling the Challenge of Big Data

The emerging technologies of 21st century science that generate large and diverse datasets provide many opportunities for improving exposure and toxicity assessment, but they pose some substantial analytical challenges, such as how to analyze data in ways that will identify valid and useful patterns and that limit the potential for misleading and expensive false-positive and false-negative findings. Although the statistical analysis and management of such data are topics of active research, development, and discussion, the committee offers in Box 7-4 some practical advice regarding several statistical issues that arise in analyzing large datasets or evaluating studies that report such analyses.

To illustrate one of the statistical issues, the winner's curse correction, consider an in vitro assay that is used to measure chemicals in a class for a particular activity, such as binding to the estrogen receptor alpha. The application might call for identifying the least or most potent chemical or the range of activity for the class. Figure 7-3 shows how a group of chemicals can appear to differ considerably in an assay—by more than two orders of magnitude in this example. However, if the results of the assay are measured with a comparable degree of error, conclusions can be misleading. After correction for error by using a simple Bayesian approach with a hierarchical model for variation of true effects, chemicals in the group differ from one another in potency by less than 1 order of magnitude, and the chemical that originally was observed to have the highest potency in the assay moves to the second position.

Another illustration is offered by the case study for 4-methylcyclohexanemethanol (MCHM, the chemical spilled into the Elk River in West Virginia) that is discussed in Appendix C. In addition to a number of in vitro and in vivo studies, the National Toxicology Program performed 5-day toxicogenomic studies in rats on MCHM and other chemicals spilled into the river. Initial findings of toxicogenomic signals—referred to as "molecular biological processes" that were indicative of liver toxicity—were made at around 100 mg/kg, which was a dose just below the apical observations of liver toxicity at 300 and 500 mg/kg, for example, for increased triglycerides (NTP 2015b). However, a refined analysis that sought to limit false discovery and maximize reproducibility (S. Auerbach, National Toxicology Program, personal communication, November 1, 2016) found changes in measures of dose-related toxicogenomic activity—activity of at least five genes that are associated with, for example, cholesterol homeostasis by the liver—at doses lower by nearly a factor of 10 (median benchmark dose of 13 mg/kg-day; NTP 2016) than doses previously thought to be the lowest doses to show activity. The example illustrates the challenge of developing approaches to evaluate toxicogenomic data that, while not excluding important biological signals, address the issue of false positives. With the increasing generation and analysis of toxicogenomic data in animal experiments, the additional experience should facilitate the development of best practices. Similar considerations apply to the use of toxicogenomic data from in vitro and epidemiological studies.

Aside from the statistical approaches used for data analysis, other considerations are involved in judging the quality and potential bias of studies that use 21st century data, particularly regarding applicability or generalizability of a study for addressing the question at hand. Such

**BOX 7-4** Development of Best Statistical Practices for Analyzing Large Toxicity Datasets

The following topics are applicable to large datasets, such as activity measurements from high-throughput screening (HTS) assays for chemicals. Some are applicable for analyzing associations of single-nucleotide polymorphisms (SNPs) with disease or exposure conditions or for analyzing dose–response relationships of gene expression in HTS studies.

*Multiple comparisons*: The total number of statistical tests performed and the false-negative and false-positive (error) control procedures should be clearly stated. Error-control procedures include ones that control the family-wise error rate, the false-discovery rate, or a Bayesian posterior probability of the null hypothesis (Efron 2011; Gelman et al. 2012). Overly conservative approaches for controlling family-wise error, such as Bonferroni control, can hide important biological signals.

*Filtering*: Assays or chemicals might be excluded before comparisons are made. For example, assays might be dropped if they show no activity for any chemical, lack statistical power to detect an association, or are otherwise uninformative. Care should be taken to avoid bias in the assessment of an association when the associations are themselves used for filtering

*Covariate correction*: Correction for covariates unrelated to the primary hypotheses improves statistical power and reduces the potential for confounding. For high-throughput platforms, the data are often rich enough to provide evidence of latent variation due to technical or batch artifacts (Leek and Storey 2007); failure to account for this variation can result in spurious findings, often dramatically (Leek et al. 2010). Known confounders can be controlled for by regression or stratified analysis, and unobserved confounders can be controlled for by latent or surrogate variable analysis.

*Feature or pathway enrichment*: These methods attempt to identify features that together have stronger or more biologically interpretable results than individual features alone. For example, collections of assays for a receptor target associated with an in vivo end point can be grouped. Ideally, the group tests use methods to address the correlation in the data to control false-positive findings (Hosack et al. 2003; Gatti et al. 2010).

*Network and module analysis*: Networks or modules of predictors or features might be identified by using correlation or co-expression analyses (Langfelder and Horvath 2008). The methods are still being developed to identify how networks change in response to a measured exposure or a toxicity end point. One approach is to derive a summary measure from the network and then to measure the correlation of the summary measure with the end point.

*Integration of hypothesis testing*: When aggregating multiple assays or replicated studies, one might use meta-analysis or empirical Bayes approaches if the assays are on the same scale and are measuring the same quantities. Independent p values might be combined by using Fisher's combined p-value or other method (Zaykin et al. 2007) to test, for example, that a chemical has no effect on any of a large number of -omics outcomes. However, an integrated analysis of multiple separate datasets violates independence assumptions when some portions of the data are shared in conducting analyses. For example, comparison of genome-wide association studies for two or more diseases might use the same set of controls (Wellcome Trust Case Control Consortium 2007), and this could bias the integrated analysis (Zaykin and Kozbur 2010).

*Shrinkage and winner's curse correction*: Measurement error can affect output from multiple assays or conditions in such a way that the measured outcome values are more varied than the underlying true variation. The same principle applies to multiple effect-size estimates; for example, in a genome-wide association study of numerous SNPs, the apparent association of the most significant SNPs with a trait or disease might tend to be greater than the true association. Correction by shrinkage techniques or by winner's curse correction methods will provide more realistic estimates.

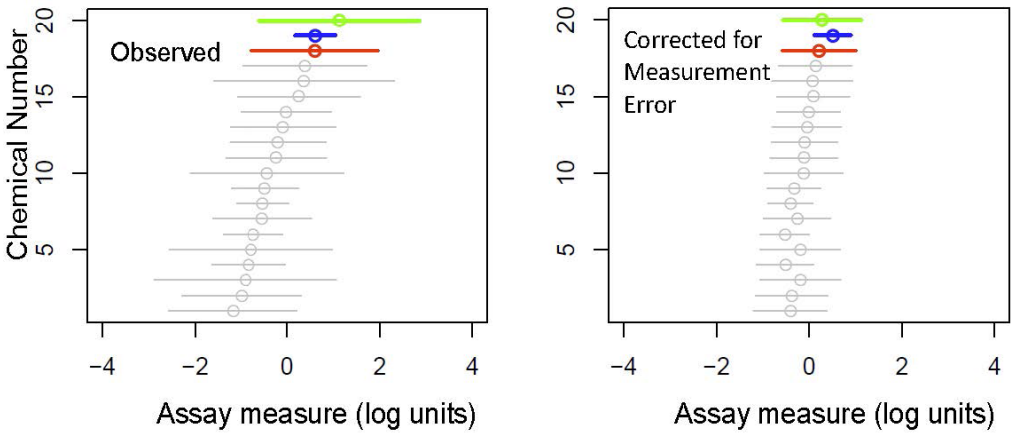



**FIGURE 7-3** Correction for assay measurement error. Left, observed values (circle) for 20 chemicals in an in vitro assay ± 2 standard errors (error bar). Right, values corrected for measurement error.

considerations raised in earlier chapters include the metabolic competence of in vitro assays, the nature of the cells used in in vitro assays, and the representativeness of the nominal dose in in vitro systems.

## Approaches for Integrating Information from Studies, Datasets, and Data Streams

### Systematic Review

As defined by the Institute of Medicine (IOM 2011, p. 1), systematic review "is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies." Specifically, it is an approach that formulates an a priori question that specifies a population or participants (the exposed group under study), an exposure (the substance and exposure circumstance), a comparator (subjects who have lower exposures), and outcomes of interest; conducts a comprehensive literature search to identify all relevant articles; screens the literature according to prespecified exclusion and inclusion criteria; evaluates study quality and study bias according to prespecified methods; and summarizes the results. The summary might or might not provide a quantitative estimate (see meta-analysis discussion below), and transparency is emphasized in the overall approach. Systematic review has been used extensively in the field of comparative-effectiveness research in which one attempts to identify the best treatment option in the clinical setting. In that field, the systematic-review process is relatively mature (Silva et al. 2015); guidance is provided in the Cochrane handbook (Higgins and Green 2011). Although there are some challenges in using systematic review in risk assessment, such as formulating a sufficiently precise research question and obtaining access to primary data, its application in human health risk assessments is a rapidly developing field in which frameworks (Rooney et al. 2014; Woodruff and Sutton 2014) and examples (Kuo et al. 2013; Lam et al. 2014; Chappell et al. 2016) are available. The report *Review of EPA's Integrated Risk Information System (IRIS) Process* (NRC 2014) provides an extensive discussion of systematic review as applied to the development of IRIS assessments (hazard and dose–response assessments). As indicated in that report, systematic review integrates the data within one data stream (human, animal, or mechanistic), and other approaches are then used to integrate the collective body of evidence. As noted above, one challenge for systematic reviews that address environmental risks to human health has been in developing methods to assess bias in mechanistic studies and their heterogeneity. Practical guidance for systematic review focused on human health risk has recently been developed (NTP 2015a).

### Meta-Analysis

*Meta-analysis* is a broad term that encompasses statistical methods used to combine data from similar studies. Its goal is to combine effect estimates from similar studies into a single weighted estimate with a 95% confidence interval that reflects the pooled data. If there is heterogeneity among the results of different studies, another goal is to explore the reasons for the heterogeneity. Two models—the fixed-effect model and the random-effects model—are typically used to pool data from different studies; each model makes different assumptions about the nature of the studies that contributed the data and

therefore uses different mechanisms for estimating the variance of the pooled effect. As noted in NRC (2014), "although meta-analytic methods have generated extensive discussion (see, for example, Berlin and Chalmers 1988; Dickersin and Berlin 1992; Berlin and Antman 1994; Greenland 1994; Stram 1996; Stroup et al. 2000; Higgins et al. 2009; Al Khalaf et al. 2011), they can be useful when there are similar studies on the same question."

As one might expect, meta-analyses are often applied to epidemiological studies to assess hazard (for example, does the pooled relative risk differ significantly from 1.0?) or to characterize dose–response relationships (for example, relative risk per unit concentration). They have not seen much use for evaluating animal datasets because of difficulty in assessing and identifying sources of heterogeneity of the data. Similarly, their application to 21st century data streams is expected to be uncommon given the heterogeneity of the data and the need to integrate data from different types of measures even when evaluating the same mechanisms or pathways.

**Bayesian Approaches**

Bayesian methods provide a natural paradigm for integrating data from various sources while accommodating uncertainty. The method is based on the Bayes theorem and involves representing the state of knowledge about a variable or phenomenon, such as the slope of a dose–response curve or how people differ from one another in their ability to metabolize a chemical, as captured by a probability distribution. As further information is generated about the variable, the "prior" probability distribution is "updated" to a new "posterior" probability distribution that reflects the updated state of knowledge.

Early applications of Bayesian approaches were by DuMouchel and Harris (1983) to evaluate the carcinogenicity of diesel exhaust by combining evidence from human, animal, and mechanistic studies, and by DuMouchel and Groer (1989) to estimate the rate of bone cancer caused by deposited plutonium from data on humans and dogs. Those examples involved strong assumptions about relevance and equivalence of different data streams (for example, human versus animal). Hierarchical, population Bayesian methods have been used to integrate different lines of evidence on metabolism and its variability in risk assessments of tetrachloroethylene (Bois et al. 1996; OEHHA 2001) and trichloroethylene (EPA 2011; Chiu et al. 2014). Bayesian approaches have been used to estimate values of model parameters for physiologically based pharmacokinetic models and to characterize uncertainty and variability in exposure estimates (Bois 1999, 2000; Liao et al. 2007; Wambaugh et al. 2013; Dong et al. 2016). They have also been applied to fate and transport modeling of chemicals at contaminated sites, of natural estrogens from livestock operations, and of bacteria from nonpoint sources (Thomsen et al. 2016) and have been shown to be broadly applicable for evidence integration (NRC 2014; Linkov et al. 2015).

The starting point for a Bayesian analysis is the determination of a prior probability distribution that characterizes the uncertainty in the variable of interest (or hypothesis) before observation of new data. The prior might be elicited on the basis of general knowledge in the literature and the state of scientific knowledge in the field. The process of summarizing information into a prior probability distribution is referred to as prior elicitation. It can be difficult, particularly when little information is available, and it is inherently imperfect in many kinds of applications; there is no best way to obtain and summarize potentially disparate information from the literature and from related studies. Some examples of prior elicitation in environmental risk assessment are provided in Wolfson et al. (1996).

Several strategies have been used to manage the uncertainty in prior elicitation. One involves choosing a prior that is vague. Vague priors can lead to posteriors that are erratic, including posterior densities that have many local bumps and might oscillate as data accumulates between widely divergent values. Gelman et al. (2008) provide some concrete examples of defining probability distributions with weakly informative priors. Another strategy is to estimate parameter values in a prior on the basis of data from related studies. For example, one might be studying a new chemical for which there is not much direct information on mechanism or exact dose–response shapes for different end points; however, there might be much to learn from a collection of the same type of data for similar chemicals. Learning from past data is a version of "empirical Bayes" and can be more easily justified than "subjective Bayes" methods articulated earlier. Potentially, a panel of experts could provide their own priors, which could be combined into a single prior (Albert et al. 2012). However, any one expert tends to be over confident about his or her knowledge and to choose a prior with a variance that is too small. Methods for addressing the over confidence of experts and other deficiencies in expert elicitation are important to consider (NRC 1996, Chapter 4; Morgan 2014). Regardless of the method of elicitation, it is important to assess the plausibility of a selected prior and to conduct sensitivity analyses to understand changes in priors.

Once the prior distribution has been defined, the prior can be updated with information in the likelihood function for each data source. Each time a data source is added, the prior is updated to obtain a posterior distribution that summarizes the new state of knowledge. The posterior distribution can then be used as a prior distribution in future analyses. Bayesian updating can thus be viewed

as a natural method for synthesizing data from different sources.

Sensitivity analysis provides a valuable approach to identify which data uncertainties are the most important in the Bayesian analyses. As noted by NRC (2007), sensitivity analysis can help to set priorities for collecting new data and contribute to a process for systematically managing uncertainties that can improve reliability.

The development of general-purpose, robust, and interpretable Bayesian methods for 21st century data is an active field of research, although hybrid approaches that reduce dimensionality before applying the Bayesian paradigm for synthesis of evidence from different data sources are favored at this point for risk-assessment purposes. The committee provides an example of using Bayesian approaches in a high-dimensional setting in Appendix E.

### Guided Expert Judgment

Guided expert judgment is a process that uses the experience and collective judgment of an expert panel to evaluate what is known on a topic, such as whether the overall evidence supports a hazard finding on a chemical (for example, whether a chemical is a carcinogen). Predetermined protocols for judging evidence generally guide the expert panel. The panel might be asked to judge whether the evidence falls into one of several broad categories, such as strong, moderate, or weak. Such approaches are used by the US Environmental Protection Agency (EPA) in its process for evaluating the evidence gathered for the Integrated Science Assessments for the evaluation of National Ambient Air Quality Standards for selected pollutants. Expert-judgment approaches are often criticized because they can lack transparency and reproducibility in that the processes used to synthesize evidence and the resulting judgments made by the experts might be obscured and because different groups of experts can come to different conclusions after reviewing the same data. Furthermore, because modern risk assessments increasingly involve complex, diverse, and large datasets, the use of a guided-expert-judgment approach can be challenging.

The IARC monograph program (IARC 2006; Pearce et al. 2015) uses guided expert judgment for its causal assessment of carcinogenicity that integrates observational human studies, experimental animal data, and other biological data, such as in vitro assays that contribute mechanistic insights. For several agents on which there are few or no human data to assess carcinogenicity, complementary experimental animal data and mechanistic data have been used to support an overall conclusion that a chemical is carcinogenic in humans. The carcinogenicity assessment of ethylene oxide (EO) for which studies in humans are limited by the use of small cohorts of exposed workers is one example. The high mutagenicity and genotoxicity of EO, clear evidence of such activity in humans, and the similarity of the damage induced in animals and humans led IARC working groups (IARC 1994, 2008, 2012) to classify it as a human carcinogen (Group 1) in spite of the limited epidemiological evidence. The most recent review (IARC 2012) noted that "There is strong evidence that the carcinogenicity of ethylene oxide, a direct-acting alkylating agent, operates by a genotoxic mechanism…. Ethylene oxide consistently acts as a mutagen and clastogen at all phylogenetic levels, it induces heritable translocations in the germ cells of exposed rodents, and a dose-related increase in the frequency of sister chromatid exchange, chromosomal aberrations and micronucleus formation in the lymphocytes of exposed workers." Box 7-5 provides details on the current IARC process.

Some have advocated quantitative approaches to weighting evidence from different sources even if any such weighting approaches can be criticized. A fundamental challenge in evaluating such approaches is that there is often no gold-standard weighting scheme; that is, there are no consensus approaches that are recognized as state-of-the-practice for optimally weighting results obtained from observational epidemiology, laboratory animal studies, in vitro assays, and computational systems for human health risk assessment. Within each of those lines of evidence are studies that vary widely in quality and relevance, and a priori weights established by experts on the basis of general characteristics (for example, animal versus human) fail to account for the scientific nuances. Experts differ as to the best weighting strategy, and formal decision-theory methods do not avoid the need for subjective choices and judgments. Thus, the committee declines to advance quantitative weighting schemes as an approach to integrating evidence from different sources.

Weighting in some cases, however, might be useful in a given data stream or evidence class, such as data from high-throughput assays or in vivo assays with common end points. Weighting typically would follow principles based on statistics and expert judgment. For example, in the absence of additional information, assays that are intended to interrogate the same pathway, mechanism, or end point and are on similar scales can be weighted by using the inverse of sampling variation; this is essentially the approach used in meta-analysis. Assays that evaluate the same end point can be weighted on the basis of prediction accuracy.

Given the current practices, the committee recommends that guided expert judgment be the approach used in the near term to integrate diverse data streams for drawing causal conclusions. Guided expert judgment is not as easily applied to other elements of the risk-assessment process because of the variety of data types and the complexity of decision points in the analyses. Considerable expert review and consultation are recommended for de-

**BOX 7-5** Integrating and Evaluating Mechanistic Data in the International Agency for Research on Cancer

Evaluation of mechanistic information in IARC begins with a systematic search of the mechanistic literature (IARC 2016b). The literature is screened for relevance and organized by mechanistic topic, guided by the 10 key characteristics of carcinogens and data type (human or experimental systems and in vivo or in vitro). There is no expectation that most or all key characteristics are operative for any specific carcinogen.

The working group evaluates data from relevant sources and pays special attention to data gaps and evidence that suggests that multiple mechanisms might be operating (IARC 2006). It evaluates evidence of changes in cell, tissue, or organ physiology after exposure, including alterations in inflammation, hyperplasia, and cell adhesion ability. The working group evaluates functional changes at the cellular level, such as shifts in the abundance of various components of key cellular machinery, increases or decreases in post-translational protein modifications, and effects on xenobiotic metabolism. The working group also evaluates modifications of molecular architecture (changes at the molecular level), including global DNA methylation, the formation of DNA adducts, and gene mutations.

Mechanistic information obtained from in vitro and nonmammalian in vivo systems (such as prokaryotes, cell cultures, and lower eukaryotes) can strengthen the biological plausibility of links to cancer. In addition, high-throughput assays that measure the effects for a single end point, high-content assays that measure multiple end points for a single agent or mixture, and structure–activity relationship information can support consistency among study types, populations, and species. High-throughput assays, especially ones that have metabolic capacity and native cellular environments, can be useful in analyzing plausible mechanisms for chemical classes, as can consistent changes among multiple genes in high-content assays, such as microarrays.

The absence of an effect in narrowly created datasets (such as ones that use specific tissues or cell types) does not necessarily support a finding that there is no effect (IARC 2006). For example, substances can act through multiple mechanisms and pathways, and cell type, developmental stage, genetic background, and co-exposures make null findings difficult to interpret.

For each of the 10 characteristics, the evidence can be labeled strong, moderate, weak, or insufficient to evaluate. The mechanistic evidence is then integrated with the evidence from other data streams to support conclusions about carcinogenicity. As cited in IARC (2016b), the conclusions are as follows:

Group 1: Carcinogenic to humans
- Sufficient evidence in humans OR
- Sufficient evidence in animals AND ***strong evidence in exposed humans*** *that the agent acts through a relevant mechanism* OR
- *Clearly belongs, based on mechanistic considerations, to a class of agents for which one or more members have been classified in Group 1*

Group 2A: Probably carcinogenic to humans
- Limited in humans AND sufficient in animals OR
- Inadequate in humans AND sufficient in animals AND ***strong evidence*** *that the carcinogenesis is mediated by a mechanism that also operates in humans* OR
- *Clearly belongs, based on mechanistic considerations, to a class of agents for which one or more members have been classified in Group 2A*

Group 2B: Possibly carcinogenic to humans
- Limited in humans AND less than sufficient in animals OR
- Inadequate in humans BUT sufficient in animals OR
- Inadequate in humans AND less than sufficient in animals AND *supporting evidence from mechanistic and other relevant data*

Group 3: Not classifiable as to its carcinogenicity to humans
- Inadequate in humans AND inadequate/limited in animals OR
- Inadequate in humans AND sufficient in animals AND ***strong evidence*** *that the mechanism of carcinogenicity in animals does not operate in humans*

velopment of guidance for those activities to be followed by expert scientific peer review of the final product.

## UNCERTAINTIES

Uncertainty accompanies all methods used to generate data inputs for risk assessments. In the case of data from new testing methods, there is the inherent variability of the assays and the qualitative uncertainty associated with their use (see Chapter 3). Such uncertainty arises with other types of assays, such as rodent bioassays, for which standard uncertainty factors are in place and accepted. The Tox21 report acknowledged the need to evaluate "test-strategy uncertainty," that is, the uncertainty associated with the introduction of a novel series of testing methods. For new assay methods, the quantification of uncertainty and its handling in practice remain to be addressed.

With regard to dealing with analytical uncertainties, the committee notes that the 1983 NRC report *Risk Assessment in the Federal Government: Managing the Process* remains enlightening. As discussed in Chapter 5, that report laid out the iconic four steps in risk assessment: hazard identification, dose–response assessment, exposure assessment, and risk characterization. The report noted that in each step a number of decision points occur in which "risk to human health can only be inferred from the available evidence." For each decision point, the 1983 committee recommended the adoption of predetermined choices or inference options ultimately to draw inferences about human risk from data that are not fully adequate. The preferred inference options were also called default options and were to be based on scientific understanding and risk-assessment policy and to be used in the absence of compelling evidence to the contrary. Other NRC committees have reiterated the importance of what have been come to be known simply as defaults and have noted that those used by EPA typically have a relatively strong scientific basis (NRC 1994, 2009). The 1983 committee also called for the establishment of uniform inference guidelines to ensure uniformity and transparency in agency decision-making and called for flexibility in providing for departure from defaults in the presence of convincing scientific evidence. EPA developed a system of guidelines that cover a wide array of risk-assessment topics. The 1983 recommendations have also been reinforced in other NRC reports (NRC 1994, 2009), and the present committee reiterates the importance of establishing uniform guidelines and a system of defaults in the absence of clear scientific understanding and the importance of enhancing the default system as described in *Science and Decisions: Advancing Risk Assessment* (NRC 2009). The enhancements include making explicit or replacing missing and unarticulated assumptions in risk assessment and developing specific criteria and standards for departing from defaults. The current committee notes, however, that the volume and complexity of 21st century data and the underlying science pose particularly difficult challenges. Systems of defaults and approaches to guide assessment should be advanced once best practices develop, as elaborated in the dose–response section above.

The Tox21 report used *test-strategy uncertainty* to refer to the overall uncertainty associated with the testing strategy and commented that "formal methods could be developed that use systematic approaches to evaluate uncertainty in predicting from the test battery results the doses that should be without biologic effect in human populations." Until such methods are developed, judgments as to the strength of evidence on pathway activation will continue to be based on expert judgment that draws on such guidelines as discussed above.

## CHALLENGES AND RECOMMENDATIONS

The new direction for risk assessment advanced in this report is based on data from 21st century science on biological pathways and approaches that acknowledge that stressors from multiple sources can contribute to a single disease and that a single stressor can lead to multiple adverse outcomes. The new techniques of 21st century science have emerged quickly and have made it possible to generate large amounts of data that can support the new directions in exposure science, toxicology, and epidemiology. In fact, the technology has evolved far faster than have approaches for analyzing and interpreting data for the purposes of risk assessment and decision-making. This chapter has addressed the challenges related to data interpretation, analysis, and integration; evidence synthesis; and causal inference. The challenges are not new but are now amplified by the scope of the new data streams. The committee lists some of the most critical challenges below with recommendations to address them.

### A Research Agenda for Data Interpretation and Integration

*Challenge*: Insufficient attention has been given to data interpretation and integration as the development of new methods for data generation has outpaced the development of approaches for interpreting the data that they generate. The complexity was recognized in the Tox21 and ES21 reports, but those reports did not attempt to develop an approach for evidence integration and interpretation to make determinations concerning hazards, exposures, and risks.

*Recommendation*: The committee recommends greater attention to the problem of drawing inferences and proposes the following empirical research agenda:

(1) The development of case studies that reflect various scenarios of decision-making and data availability. The case studies should reflect the types of data typically available for interpretation and integration in each element of risk assessment—hazard identification, dose–response assessment, exposure assessment, and risk characterization—and include assessing interindividual variability and sensitive populations.

(2) Testing of the case studies with interdisciplinary and multidisciplinary panels, using best practices and the guided-expert-judgment approaches, such as described above. There is a need to understand how such panels of experts will evaluate the case studies and how various data elements might drive the evaluation process. Furthermore, communication between people from different disciplines, such as Bayesian statisticians and mechanistic toxicologists, will be essential for successful and reliable use of new data; case studies will provide a means of testing how interactions might best be accomplished in practice.

(3) A comprehensive cataloging of evidence evaluations and decisions that have been taken on various agents so that expert judgments can be tracked and evaluated and the expert processes calibrated. The cataloging should capture the major gaps in evidence and attendant uncertainty that might have figured into evidence evaluation.

(4) More intensive and systematic consideration of how statistically based tools for combining data and integrating evidence, such as Bayesian approaches, can be used for incorporating 21st century science into hazard, dose–response, exposure, and interindividual-variability assessments and ultimately into the overall risk characterization.

### Advancing the Use of Data on Disease Components and Mechanisms in Risk Assessment

*Challenge*: Data generated from tools that probe components of disease are difficult to use in risk assessment partly because of incomplete understanding of the linkages between disease and components and because of uncertainty around the extent to which mitigation of exposure changes expression of a component and consequently changes the associated risk.

*Recommendation*: The sufficient-component-cause model should be advanced as an approach for conceptualizing the pathways that contribute to disease risk.

*Recommendation*: The committee encourages the cataloging of pathways, components, and mechanisms that can be linked to particular hazard traits, similar to the IARC characteristics of carcinogens. This work should draw on existing knowledge and current research in the biomedical fields related to mechanisms of disease that are outside the traditional toxicant-focused literature that has been the basis of human-health risk evaluations and of assessments and toxicology. The work should be accompanied by research efforts to describe the series of assays and responses that provide evidence on pathway activation and to establish a system for interpreting assay results for the purpose of inferring pathway activation from chemical exposure.

*Recommendation*: High priority should be given to the development of a system of practice related to inferences for using read-across for data-sparse chemicals; that practice area provides great opportunities for advancing various tools and incorporating their use into risk assessment. High priority should also be given to using multiple data streams to evaluate low-dose risk, as elaborated on in NRC (2009).

### Developing Best Practices for Data Integration and Interpretation

*Challenge*: The emergence of new data streams clearly has complicated the long-standing problem of integrating data for hazard identification, which the committee views as analogous to inferring a causal relationship between a putative causal factor and an effect. The committee considers that two challenges are related to data integration and interpretation for hazard identification: (1) using the data from the methods of 21st century science to infer a causal association between a chemical or other exposure and an adverse effect, particularly if it is proximal to an apical effect, and (2) integrating new lines of evidence with those from conventional toxicology and epidemiological studies. Although much has been written on this topic, proposed approaches rely largely on guided expert judgment.

*Recommendation*: The committee sees no immediate alternative to the use of guided expert judgment as the basis of judgment and recommends its continued use for the time being. Expert judgment should be guided and calibrated in interpreting data on pathways and mechanisms. Specifically, in these early days, the processes of expert judgment should be documented to support the elaboration of best practices, and there should be periodic reviews of how evidence is being evaluated so that the expert-judgment processes can be refined. Those practices will support the development of guidelines with explicit default approaches to ensure consistency in application within particular decision contexts.

*Recommendation*: In the future, pathway-modeling approaches that incorporate uncertainties and integrate multiple data streams might become an adjunct and perhaps a replacement. Methodological research to advance those approaches is needed.

*Challenge*: The size of some datasets and the number of outcomes covered complicate communication of findings to the scientific community and to those who use

results for decision-making. There might be distrust because of the need to use methods that are complex and possibly difficult to understand for the large datasets.

*Recommendation*: Data integration should be complemented by visualization tools to enable effective communication of analytical findings from complex datasets to decision makers and other stakeholders. Transparency of the methods, statistical rigor, and accessibility to the underlying data are key elements for promoting the use and acceptance of the new data in decision-making.

*Challenge*: Given the complexities of 21st century data and the challenges associated with their interpretation, there is a potential for a decision to be based ultimately on a false-positive or false-negative result. The implications of such an erroneous conclusion are substantial. The challenge is to calibrate analytical approaches to optimize their sensitivity and specificity for identifying true associations. If public-health protection is the underlying goal, an approach that generates more false-positive than false-negative conclusions might be appropriate in some decision contexts. A rigid, algorithmic approach might prove conservative but lead to false-negatives or at least to a delay in decision-making because more evidence is required.

*Recommendation*: This challenge merits the development of guidelines and best practices that use processes that involve direct discussion among researchers, decision-makers, and other stakeholders who might have different views as to where the balance between sensitivity and specificity should be placed.

### Addressing Uncertainties in Using 21st Century Tools in Dose–Response Assessment

*Challenge*: There are multiple potential complications in moving from in vitro testing and in vivo toxicogenomic studies to applying the resulting dose–response estimates to human populations. Uncertainties are introduced that parallel and might exceed those associated with extrapolation from animal studies to humans. Sources of uncertainty include chemical metabolism, the relevance of pathways, and the generalizability of dose–response relationships that are observed in vitro. There is also the challenge of integration among datasets and multiple lines of evidence.

*Recommendation*: The challenges noted should be explored in case studies for which the full array of data is available: high-throughput testing, animal studies, and human studies. Bayesian methods need to be developed and evaluated for combining dose–response data from multiple test systems. And a system or practice and default-data integration approaches need to be developed that promote consistent, transparent, and reliable application that explain and account for uncertainties.

### Developing Best Practices for Analyzing Big Data for Application in Risk Assessment

*Challenge*: Enormous datasets that pose substantial analytical challenges are being generated, particularly in relation to identifying biologically relevant signals given the possibility of false-positives resulting from multiple comparisons.

*Recommendation*: Best practices should be developed through consensus processes to address the statistical issues listed in Box 7-4 that complicate analyses of very large datasets. Those practices might differ by decision context or data type. Adherence to best practices sets a consistent approach for weighing false positives against false negatives and maintaining high integrity in reporting. Analyses should be carried out in transparent and replicable ways to ensure credibility and to enhance review and acceptance of findings for decision-making. Open data access might be critical for ensuring transparency.

## REFERENCES

Al Khalaf, M.M., L. Thalib, and S.A. Doi. 2011. Combining heterogeneous studies using the random-effects model is a mistake and leads to inconclusive meta-analyses. J. Clin. Epidemiol. 64(2):119-123.

Albert, I., S. Donnet, C. Guihenneuc-Jouyaux, S. Low-Choy, K. Mengersen, and J. Rousseau. 2012. Combining expert opinions in prior elicitation. Bayesian Anal. 7(3):503-512.

Berlin, J.A., and E.M. Antman. 1994. Advantages and limitations of metaanalytic regressions of clinical trials data. Online J. Curr. Clin. Trials, Document No. 134.

Berlin, J., and T.C. Chalmers. 1988. Commentary on meta-analysis in clinical trials. Hepatology 8(3):690-691.

Bois, F.Y. 1999. Analysis of PBPK models for risk characterization. Ann. NY Acad. Sci. 895:317-337.

Bois, F.Y. 2000. Statistical analysis of Clewell et al. PBPK model of trichloroethylene kinetics. Environ. Health Perspect. 108(Suppl. 2):307-3016.

Bois, F.Y., A. German, J. Jiang, D.R. Maszle, L. Zeise, and G. Alexeeff. 1996. Population toxicokinetics of tetrachloroethylene. Arch. Toxicol. 70(6):347-355.

Chappell, G., I.P. Pogribny, K.Z. Guyton, and I. Rusyn. 2016. Epigenetic alterations induced by genotoxic occupational and environmental human chemical carcinogens: A systematic literature review. Mutat. Res. Rev. Mutat. Res. 768:27-45.

Chiu, W.A., J.L. Campbell, Jr., H.J. Clewell, III, Y.H. Zhou, F.A. Wright, K.Z. Guyton, and I. Rusyn. 2014. Physiologically based pharmacokinetic (PBPK) modeling of interstrain variability in trichloroethylene metabolism in the mouse. Environ.Health Perspect. 122(5):456-463.

Cichocki, J.A., S. Furuya, A. Venkatratnam, T.J. McDonald, A.H. Knap, T. Wade, S. Sweet, W.A. Chiu, D.W. Thread-

gill, and I. Rusyn. In press. Characterization of variability in toxicokinetics and toxicodynamics of tetrachloroethylene using the Collaborative Cross mouse population. Environ. Health Perspect.

Cui, Y., D.M. Balshaw, R.K. Kwok, C.L. Thompson, G.W. Collman, and L.S. Birnbaum. 2016. The exposome: Embracing the complexity for discovery in environmental health. Environ. Health Perspect. 124(8):A137-A140.

Dickersin, K., and J.A. Berlin. 1992. Meta-analysis: State-of-the-science. Epidemiol. Rev. 14(1):154-176.

DHEW (US Department of Health, Education, and Welfare). 1964. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service. Public Health Service Publication No. 1103. Washington, DC: US Government Printing Office.

Dong, Z., C. Liu, Y. Liu, K. Yan, K.T. Semple, and R. Naidu. 2016. Using publicly available data, a physiologically-based pharmacokinetic model and Bayesian simulation to improve arsenic non- cancer dose-response. Environ. Int. 92-93:239-246.

DuMouchel, W., and P.G. Groër. 1989. Bayesian methodology for scaling radiation studies from animals to man. Health Phys. 57(Suppl. 1):411-418.

DuMouchel, W.H. and J.E. Harris. 1983. Bayes methods for combining the results of cancer studies in humans and other species. J. Am. Stat. Assoc. 78(382):293-308.

Efron, B. 2011. Tweedie's formula and selection bias. J. Am. Stat. Assoc. 106(496):1602-1614.

EPA (US Environmental Protection Agency). 2005. Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Risk Assessment Forum, US Environmental Protection Agency, Washington DC. March 2005 [online]. Available: https://www.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf [accessed August 1, 2016].

EPA (US Environmental Protection Agency). 2011. Toxicological Review of Trichloroethylene. EPA/635/R-09/011F. US Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0199tr/0199tr.pdf [accessed November 1, 2016].

EPA (US Environmental Protection Agency). 2014. Integrated Bioactivity and Exposure Ranking: A Computational Approach for the Prioritization and Screening of Chemicals in the Endocrine Disruptor Screening Program. EPA-HQ-OPP-2014-0614-0003. US Environmental Protection Agency Endocrine Disruptor Screening Program (EDSP). FIFRA SAP December 2-5, 2014 [online]. Available: https://www.regulations.gov/document?D=EPA-HQ-OPP-2014-0614-0003 [accessed November 1, 2016].

EPA (US Environmental Protection Agency). 2015. Preamble to Integrated Science Assessments. EPA/600/R-15/067. National Center for Environmental Assessment, Office of Research and Development, US Environmental Protection Agency, Research Triangle Park, NC. November 2015 [online]. Available: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=310244 [accessed August 1, 2016].

Gatti, D.M., W.T. Barry, A.B. Nobel, I. Rusyn, and F.A. Wright. 2010. Heading down the wrong pathway: On the influence of correlation within gene sets. BMC Genomics 11:574.

Gelman, A., A. Jakulin, M.G. Pittau, and Y.S. Su. 2008. A weakly informative default prior distribution for logistic and other regression models. Ann. Appl. Stat. 2(4):1360-1383.

Gelman, A., J. Hill, and M. Yajima. 2012. Why we (usually) don't have to worry about multiple comparisons. J. Res. Edu. Effect. 5(2):189-211.

Greenland, S. 1994. A critical look in some popular meta-analytical methods. Am. J. of Epidemiol. 140(3):290-296.

Harrill, A.H., P.B. Watkins, S. Su, P.K. Ross, D.E. Harbourt, I.M. Stylianou, G.A. Boorman, M.W. Russo, R.S. Sackler, S.C. Harris, P.C. Smith, R. Tennant, M. Bogue, K. Paigen, C. Harris, T. Contractor, T. Wiltshire, I. Rusyn, and D.W. Threadgill. 2009a. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. Genome Res. (9):1507-1515.

Higgins, J.P., S.G. Thompson, and D.J. Spiegelhalter. 2009. A re-evaluation of random-effects meta-analysis. J. R. Stat. Soc. Ser. A 172(1):137-159.

Higgins, J.P.T., and S. Green, eds. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration [online]. Available: http://handbook.cochrane.org/ [accessed August 1, 2016].

Hill, A.B. 1965. The environment and disease: Association or causation? Proc. R. Soc. Med. 58:295-300.

Hosack, D.A., G. Dennis, Jr., B.T. Sherman, H.C. Lane, and R.A. Lempicki. 2003. Identifying biological themes within lists of genes with EASE. Genome Biol. 4(10):R70.

IARC (International Agency for Research on Cancer). 1994. Ethylene oxide. Pp. 73-159 in Some Industrial Chemicals. IARC Monograph on the Evaluation of Carcinogenic Risk to Human vol. 60 [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol60/mono60-7.pdf [accessed November 2, 2016].

IARC (International Agency for Research on Cancer). 2006. Preamble. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon: IARC [online]. Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed August 2, 2016].

IARC (International Agency for Research on Cancer). 2008. Ethylene oxide. Pp. 185-309 in 1,3-Butadiene, Ethylene Oxide and Vinyl Halides (Vinyl Fluoride, Vinyl Chloride and Vinyl Bromide). IARC Monograph on the Evaluation of Carcinogenic Risk to Humans vol. 97. Lyon, France: IARC [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol97/mono97-7.pdf [accessed November 8, 2016].

IARC (International Agency for Research on Cancer). 2012. Ethylene oxide. Pp. 379-400 in Chemical Agents and Related Occupations. IARC Monograph on the Evaluation of Carcinogenic Risk to Humans vol. 100F. Lyon, France: IARC [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol100F/mono100F-28.pdf [accessed November 8, 2016].

IARC (International Agency for Research on Cancer). 2015. Some Organophosphate Insecticides and Herbicides: Diazinon, Glyphosate, Malathion, Parathion, and Tetrachlorvinphos. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 112 [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol112/ [accessed November 1, 2016].

IARC (International Agency for Research on Cancer). 2016a. 2,4-Dichlorophenoxyacetic acid (2,4-D) and Some Organochlorine Insecticides. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 113 [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol113/index.php [accessed November 1, 2016].

IARC (International Agency for Research on Cancer). 2016b. Instructions to Authors for the Preparation of Drafts for IARC Monographs [online]. Available: https://monographs.iarc.fr/ENG/Preamble/previous/Instructions_to_Authors.pdf [accessed November 1, 2016].

IOM (Institute of Medicine). 2008. Improving the Presumptive Disability Decision-Making Process for Veterans. Washington, DC: The National Academies Press.

IOM (Institute of Medicine). 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press.

Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. Environ. Health Perspect. 118(4):485-492.

Kleinstreuer, N.C., J. Yang, E.L. Berg, T.B. Knudsen, A.M. Richard, M.T. Martin, D.M. Reif, R.S. Judson, M. Polokoff, D.J. Dix, R.J. Kavlock, and K.A. Houck. 2014. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. Nat. Biotechnol. 32:583-591.

Knapen, D., L. Vergauwen, D.L. Villeneuve, and G.T. Ankley. 2015. The potential of AOP networks for reproductive and developmental toxicity assay development. Reprod. Toxicol. 56:52-55.

Kuo, C.C., K. Moon, K.A. Thayer, and A. Navas-Acien. 2013. Environmental chemicals and type 2 diabetes: An updated systematic review of the epidemiologic evidence. Curr. Diab. Rep. 13(6):831-849.

Lam, J., E. Koustas, P. Sutton, P.I. Johnson, D.S. Atchley, S. Sen, K.A. Robinson, D.A. Axelrad, and T.J. Woodruff. 2014. The Navigation Guide—evidence-based medicine meets environmental health: Integration of animal and human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122(10):1040-1051.

Langfelder, P., and S. Horvath. 2008. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9:559.

Leek, J.T., and J.D. Storey. 2007. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 3(9):1724-1735.

Leek, J.T., R.B. Scharpf, H.C. Bravo, D. Simcha, B. Langmead, W.E. Johnson, D. Geman, K. Baggerly, and R.A. Irizarry. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat. Rev. Genet. 11(10):733-739.

Liao, K.H., Y.M. Tan, R.B. Connolly, S.J. Borghoff, M.L. Gargas, M.E. Andersen, and J.H. Clewell, III. 2007. Bayesian estimation of pharmacokinetic and pharmacodynamic parameters in a mode-of- action-based cancer risk assessment for chloroform. Risk Anal. 27(6):1535-1551.

Linkov, I., O. Massey, J. Keisler, I. Rusyn, and T. Hartung. 2015. From "weight of evidence" to quantitative data integration using multicriteria decision analysis and Bayesian methods. ALTEX 32(1):3-8.

Mackie, J.L. 1980. The Cement of the Universe: A Study of Causation. New York: Oxford University Press.

Martin, M.T., D.J. Dix, R.S. Judson, R.J. Kavlock, D.M. Reif, A.M. Richard, D.M. Rotroff, S. Romanov, A. Medvedev, N. Poltoratskaya, M. Gambarian, M. Moeser, S.S. Makarov, and K.A. Houck. 2010. Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem. Res. Toxicol. 23(3):578-590.

Martin, M.T., T.B. Knudsen, D.M. Reif, K.A. Houck, R.S. Judson, R.J. Kavlock, and D.J. Dix. 2011. Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol. Reprod. 85(2):327-339.

Morello-Frosch, R., M. Zuk, M. Jerrett, B. Shamasunder, and A.D. Kyle. 2011. Understanding the cumulative impacts of inequalities in environmental health: Implications for policy. Health Aff. 30(5):879-887.

Morgan, M.G. 2014. Use (and abuse) of expert elicitation in support of decision making for public policy. Proc. Natl. Acad. Sci. US 111(20):7176-7184.

NASEM (National Academies of Sciences, Engineering and Medicine). 2015. Application of Modern Toxicology Approaches for Predicting Acute Toxicity for Chemical Defense. Washington, DC: The National Academies Press.

NRC (National Research Council). 1983. Science and Judgment in Risk Assessment. Washington, DC: National Academy Press.

NRC (National Research Council). 1994. Understanding Risk: Informing Decisions in a Democratic Society. Washington, DC: National Academy Press.

NRC (National Research Council). 1996. Understanding Risk: Informing Decisions in a Democratic Society. Washington, DC: National Academy Press.

NRC (National Research Council). 2007. Models in Environmental Regulatory Decision Making. Washington, DC: The National Academies Press.

NRC (National Research Council). 2009. Science and Decisions: Advancing Risk Assessment. Washington, DC: The National Academies Press.

NRC (National Research Council). 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Washington, DC: The National Academies Press.

NRC (National Research Council). 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2015a. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration, NTP Office of Health Assessment and Translation, Division of NTP, National Institute of Environmental Health Sciences [online]. Available: http://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf [accessed November 1, 2016].

NTP (National Toxicology Program). 2015b. West Virginia Chemical Spill: 5-Day Rat Toxicogenomic Studies, June 2015 NTP Update [online]. Available: http://ntp.niehs.nih.gov/ntp/research/areas/wvspill/micronucleus_update_508.pdf [accessed November 1, 2016].

NTP (National Toxicology Program). 2016. West Virginia Chemical Spill: 5-Day Rat Toxicogenomic Studies, June 2016 NTP Update [online]. Available: http://ntp.niehs.nih.gov/ntp/research/areas/wvspill/tgmx_update_july2016_508.pdf [accessed November 1, 2016].

OECD (Organisation for Economic Co-operation and Development). 2013. Guidance Document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment No. 184 [online]. Available: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en [accessed November 1, 2016].

OEHHA (Office of Environmental Health Hazard Assessment). 2001. Public Health Goal for Tetrachloroethylene in Drinking Water. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency [online]. Available: http://oehha.ca.gov/media/downloads/pesticides/report/pceaug2001.pdf [accessed November 1, 2016].

Pearce, N., A. Blair, P. Vineis, W. Ahrens, A. Andersen, J.M. Anto, B.K. Armstrong, A.A. Baccarelli, F.A. Beland, A. Berrington, P.A. Bertazzi, L.S. Birnbaum, R.C. Brownson, J.R. Bucher, K.P. Cantor, E. Cardis, J.W. Cherrie, D.C. Christiani, P. Cocco, D. Coggon, P. Comba, P.A. Demers, J.M. Dement, J. Douwes, E.A. Eisen, L.S. Engel, R.A. Fenske, L.E. Fleming, T. Fletcher, E. Fontham, F. Forastiere, R. Frentzel-Beyme, L. Fritschi, M. Gerin, M. Goldberg, P. Grandjean, T.K. Grimsrud, P. Gustavsson, A. Haines, P. Hartge, J. Hansen, M. Hauptmann, D. Heederik, K. Hemminki, D. Hemon, I. Hertz-Picciotto, J.A. Hoppin, J. Huff, B. Jarvholm, D. Kang, M.R. Karagas, K. Kjaerheim, H. Kjuus, M. Kogevinas, D. Kriebel, P. Kristensen, H. Kromhout, F. Laden, P. Lebailly, G. LeMasters, J.H. Lubin, C.F. Lynch, E. Lynge, A. 't Mannetje, A.J. McMichael, J.R. McLaughlin, L. Marrett, M. Martuzzi, J. A. Merchant, E. Merler, F. Merletti, A. Miller, F.E. Mirer, R. Monson, K. Nordby, A.F. Olshan, M. Parent, F.P. Perera, M.J. Perry, A.C. Pesatori, R. Pirastu, M. Porta, E. Pukkala, C. Rice, D.B. Richardson, L. Ritter, B. Ritz, C.M. Ronckers, L. Rushton, J.A. Rusiecki, I. Rusyn, J.M. Samet, D.P. Sandler, S. de Sanjose, E. Schernhammer, A.S. Costantini, N. Seixas, C. Shy, J. Siemiatycki, D.T. Silverman, L. Simonato, A.H. Smith, M.T. Smith, J.J. Spinelli, M.R. Spitz, L. Stallones, L.T. Stayner, K. Steenland, M. Stenzel, B.W. Stewart, P.A. Stewart, E. Symanski, B. Terracini, P.E. Tolbert, H. Vainio, J. Vena, R. Vermeulen, C.G. Victora, E.M. Ward, C.R. Weinberg, D. Weisenburger, C. Wesseling, E. Weiderpass, and S.H. Zahm. 2015. IARC Monographs: 40 years of evaluating carcinogenic hazards to humans. Environ. Health Perspect. 123(6):507-514.

Pham, N., S. Iyer, E. Hackett, B.H. Lock, M. Sandy, L. Zeise, G. Solomon, M. Marty. 2016. Using ToxCast to explore chemical activities and hazard traits: A case study with ortho-phthalates. Toxicol. Sci. 151(2):286-301.

Rooney A.A., A.L. Boyles, M.S. Wolfe, J.R. Bucher, and K.A. Thayer. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ. Health Perspect. 122(7):711-718.

Rothman, K.J. 1976. Causes. Am. J. Epidemiol. 104(6):587-592.

Rothman, K.J., and S. Greenland. 2005.Causation and causal inference in epidemiology. Am. J. Public Health 95(Suppl. 1):S144-S150.

Silva, M., N. Pham, C. Lewis, S. Iyer, E. Kwok, G. Solomon, and L. Zeise. 2015. A comparison of ToxCast test results with in vivo and other in vitro endpoints for neuro, endocrine, and developmental toxicities: A case study using endosulfan and methidathion. Birth Defects Res. B Dev. Reprod. Toxicol. 104(2):71-89.

Smith, M.T., K.Z. Guyton, C.F. Gibbons, J.M. Fritz, C.J. Portier, I. Rusyn, D.M. DeMarini, J.C. Caldwell, R.J. Kavlock, P. Lambert, S.S. Hecht, J.R. Bucher, B.W. Stewart, R. Baan, V.J. Cogliano, and K. Straif. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ. Health Perspect.124(6):713-721.

Stram, D.O. 1996. Meta-analysis of published data using a linear mixed-effects model. Biometrics 52(2):536-544.

Stroup, D.F., J.A. Berlin, S.C. Morton, I. Olkin, G.D. Williamson, D. Rennie, D. Moher, B.J. Becker, T.A. Sipe, and S.B. Thacker. 2000. Meta-analysis of observational studies in epidemiology: A proposal for reporting. JAMA 283(15):2008-2012.

Thomsen, N.I., P.J. Binning, US McKnight, N. Tuxen, P.L. Bjerg, and M. Troldborg. 2016. A Bayesian belief network approach for assessing uncertainty in conceptual site models at contaminated sites. J. Contam. Hydrol. 188:12-28.

Wambaugh, J.F., R.W. Setzer, D.M. Reif, S. Gangwal, J. Mitchell-Blackwood, J.A. Arnot, O. Joliet, A. Frame, J. Rabinowitz, T.B. Knudsen, R.S. Judson, P. Egeghy, D. Vallero, and E.A. Cohen-Hubal. 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ. Sci. Technol. 47(15):8479-8488.

Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661-678.

Wolfson, L.J., J.B. Kadane, and M.J. Small. 1996. Bayesian environmental policy decisions: Two case studies. Ecol. Appl. 6(4):1056-1066.

Woodruff, T.J., and P. Sutton. 2014. The Navigation Guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. Environ. Health Perspect. 122(10):1007-1014.

Zaykin, D.V., and D.O. Kozbur. 2010. P-value based analysis for shared controls design in genome-wide association studies. Genet. Epidemiol. 34(7):725-738.

Zaykin, D.V., L.A. Zhivotovsky, W. Czika, S. Shao, and R.D. Wolfinger. 2007. Combining p-values in large-scale genomics experiments. Pharm. Stat. 6(3):217-226.

Zeise, L., F.Y. Bois, W.A. Chiu, D. Hattis, I. Rusyn, and K.Z. Guyton. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environ. Health Perspect. 121(1):23-31.

# A

# Biographical Information on the Committee on Incorporating 21st Century Science into Risk-Based Evaluations

**Jonathan M. Samet** (*Chair*) is a pulmonary physician and epidemiologist. He is Distinguished Professor and Flora L. Thornton Chair for the Department of Preventive Medicine at the Keck School of Medicine at the University of Southern California (USC) and director of the USC Institute for Global Health. Dr. Samet's research has focused on the health risks posed by inhaled pollutants. He has served on numerous committees concerned with public health: the US Environmental Protection Agency Science Advisory Board; committees of the National Research Council, including chairing the Committee on Health Risks of Exposure to Radon (BEIR VI), the Committee on Research Priorities for Airborne Particulate Matter, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, the Committee to Review the IRIS Process, and the Board on Environmental Studies and Toxicology; and committees of the Institute of Medicine. He is a member of the National Academy of Medicine. Dr. Samet received his MD from the University of Rochester School of Medicine and Dentistry.

**Melvin E. Andersen** is Distinguished Research Fellow at ScitoVation. His research career has focused on developing biologically based dose–response models and applying them to human-health risk assessments for many environmental chemicals. Before joining ScitoVation in 2016, he held positions at the Hamner Institutes for Health Sciences, the Chemical Industry Institute of Toxicology, Colorado State University, the US Environmental Protection Agency, and the US Department of Defense with both the Navy and the Air Force. He has served on several National Research Council Committees, including the Committee on Toxicity Testing and Assessment of Environmental Agents. Dr. Andersen is a fellow of the Academy of Toxicological Sciences and a diplomate of both the American Board of Toxicology and the American Board of Industrial Hygiene. He earned a PhD in biochemistry and molecular biology from Cornell University.

**Jon A. Arnot** is the president of ARC Arnot Research & Consulting and an adjunct professor in the Department of Physical and Environmental Science and in the Department of Pharmacology and Toxicology of the University of Toronto. He has 15 years of research experience in the development, application, and evaluation of methods and models to assess the exposure, hazard, and risk posed by organic chemicals. His research has focused on the application of high-throughput screening methods for prioritizing chemicals for risk assessment. He is the principal investigator or co-investigator on various international projects, including collaborations in the United States, Europe, and Canada. He was the recipient of the James M. McKim III Innovative Student Research Award (2008) from the International QSAR Foundation to Reduce Animal Testing and the Society of Environmental Toxicology and Chemistry (SETAC) Best Student Paper Award (2009). Dr. Arnot earned his PhD in environmental and life sciences from Trent University.

**Esteban Burchard** is professor of medicine and biopharmaceutical sciences at the University of California, San Francisco (UCSF). His research interests center on identifying genetic, social, and environmental risk factors for asthma in ethnically diverse populations. Dr. Burchard has created the largest gene–environment study of asthma in minority-group children in the United States. He directs the UCSF Center for Genes, Environment & Health. Dr. Burchard received his MD from the Stanford University School of Medicine and his clinical training at Harvard's Brigham and Women's Hospital and UCSF. He also completed epidemiology training at the Harvard and University of California, Berkley, School of Public Health.

**George P. Daston** is the Victor Mills Society Research Fellow at the Procter & Gamble Company. He has published more than 100 articles and book chapters and edited 5 books in toxicology and risk assessment. His current re-

search efforts are in toxicogenomics and mechanistic toxicology, particularly in addressing how findings in these fields can improve risk assessment of chemicals and the development of nonanimal alternatives. Dr. Daston has served as president of the Teratology Society, as councilor and treasurer-elect of the Society of Toxicology, and on the US Environmental Protection Agency Science Advisory Board, the Board on Scientific Counselors of the National Toxicology Program, the National Research Council's Board of Environmental Studies and Toxicology, and the National Children's Study Advisory Committee. He is editor-in-chief of *Birth Defects Research: Developmental and Reproductive Toxicology*. With scientists at the US Humane Society, Dr. Daston manages the AltTox website, which is devoted to the exchange of scientific information leading to the development of in vitro replacements for toxicity assessments. Dr. Daston has been awarded the Teratology Society's Josef Warkany Lectureship and Distinguished Service Award, the Toxicology Forum's George H. Scott Award, and the Society of Toxicology's Best Paper of the Year Award, and he is an elected fellow of American Association for the Advancement of Science. He is an adjunct professor of pediatrics of the University of Cincinnati. Dr. Daston earned his PhD in developmental biology from the University of Miami.

**David B. Dunson** is Arts and Sciences Distinguished Professor in the Department of Statistical Science of Duke University. His research interests include Bayesian statistics, complex hierarchical and latent variable modeling, and nonparametric statistical modeling. His methodological research focuses on nonparametric Bayes, latent variable methods, big data, scalable Bayesian inferences, functional and object data, and dimensionality reduction. He is a member of the International Society for Bayesian Analysis, the Institute of Mathematical Statistics, the American Statistical Association, and the International Biometrics Society. Dr. Dunson received his PhD in biostatistics from Emory University.

**Nigel Greene** is the director of predictive compound ADME and safety at AstraZeneca and specializes in the application of computational and in vitro approaches to assess compound liabilities. His specific duties include establishing and managing a group of PhD scientists that profile chemicals for off-target pharmacology. His group uses computational modeling and analysis of chemical properties and in vitro assay profiles to help to predict the safety profile of chemicals in early discovery programs and to aid in chemical series and compound selection before in vivo studies are conducted. Dr. Greene's other activities include mining internal and public databases of gene-expression data to explore biological mechanisms of toxicity and developing new in vitro assays for safety profiling on the basis of findings from the mining exercises. He recently served as a member of the National Research Council Committee on the Design and Evaluation of Safer Chemical Substitutions. Dr. Greene received his PhD in organometallic chemistry from the University of Leeds.

**Heather B. Patisaul** is an associate professor in the Department of Biology of North Carolina State University. Her research examines the steroid-dependent mechanisms through which sexually dimorphic behaviors and brain circuits arise. She also explores the mechanisms by which sexually dimorphic systems and behaviors can be disrupted by environmental estrogens. Her laboratory is interested in the mechanisms by which exposure to environmental estrogens can advance puberty and impair fertility in females. Dr. Patisaul served on the World Health Organization expert panel that assessed the risks associated with bisphenol A in 2010 and recently served on the National Research Council Committee to Review EPA's Draft, State of the Science Paper on Nonmonotonic Dose Response. She received her PhD in population biology, ecology, and evolution from Emory University.

**Kristi Pullen Fedinick** is a staff scientist with the Natural Resources Defense Council (NRDC) Health and Environment Program. Her multidisciplinary training spans nearly 20 years and includes work in molecular biology, biochemistry, structural biology, computational biology, and population health. Dr. Pullen Fedinick's work at NRDC has focused on the application of high-throughput technologies in predictive toxicology and chemical risk assessment. Before joining NRDC, she worked at a small environmental nonprofit in Chicago where she focused on air and drinking-water quality, science communication, and environmental-justice projects. Dr. Pullen Fedinick received a PhD in molecular and cell biology from the University of California, Berkeley, and was a Robert Wood Johnson Foundation Health and Society Scholar at the Harvard T.H. Chan School of Public Health.

**Beate R. Ritz** is a professor in the Department of Epidemiology of the University of California, Los Angeles (UCLA), Fielding School of Public Health. Her research focuses on the health effects of occupational and environmental toxicants, such as pesticides, ionizing radiation, and air pollution; on chronic diseases, including neurodegenerative and neurodevelopmental disorders, and cancers; on adverse birth outcomes; and on asthma. In her research, she uses geographic information system (GIS) modeling of environmental exposures, including pesticide use and traffic-related air pollution in California, and investigates links between genetic susceptibility factors and environmental exposures in populations. Dr. Ritz is a member of the Center for Occupational and Environmental Health and the Southern California Environmental Health Science Center and co-directed the UCLA Center

for Gene-Environment Studies of Parkinson's disease, funded by the National Institute of Environmental Health Sciences. She received her MD and a PhD in medical sociology from the University of Hamburg, Germany, and an MPH and a PhD in epidemiology from UCLA.

**Ivan Rusyn** is a professor in the Department of Veterinary Integrative Biosciences of the College of Veterinary Medicine and Biomedical Sciences of Texas A&M University. Before joining the university, he was a professor of environmental sciences and engineering at the University of North Carolina at Chapel Hill. Dr. Rusyn's laboratory has an active research portfolio with a focus on the mechanisms of action of environmental toxicants, the genetic determinants of susceptibility to toxicant-induced injury, and computational toxicology. His studies on health effects of environmental agents have resulted in more than 150 peer-reviewed publications. He has served on several National Research Council committees and was a member of the Standing Committee on Use of Emerging Science for Environmental Health Decisions and the Committee on Toxicology. Dr. Rusyn received his MD from Ukrainian State Medical University in Kiev and his PhD in toxicology from the University of North Carolina at Chapel Hill.

**Robert L. Tanguay** is Distinguished Professor of Molecular Toxicology in the Department of Environmental and Molecular Toxicology of Oregon State University. His research interests include exploiting the advantages of the zebra fish (*Danio rerio*) model to improve human and environmental health; evaluating biological interactions and responses to environmental chemicals, pharmaceuticals, and nanoparticles by using rapid-throughput approaches; and understanding the mechanisms underlying the toxicity of chemicals, such as 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, polycyclic aromatic hydrocarbons (PAHs), ethanol, pharmaceuticals, and pesticides. Dr. Tanguay directs the Oregon State's Superfund Research Program as the project leader for an investigation into PAH-induced developmental toxicity, as a co-investigator in work involving biological response indicator devices, and as a research coordinator in the program. He received his PhD in biochemistry from the University of California, Riverside.

**Justin G. Teeguarden** is a staff scientist and chief exposure scientist in the Environmental and Biological Sciences Directorate of the Pacific Northwest National Laboratory (PNNL). He holds a joint faculty position with the Oregon State University (OSU) Department of Environmental and Molecular Toxicology, where he serves as the director of the OSU–PNNL–Superfund Center Research Translation Core. Dr. Teeguarden's research focuses on computational and experimental exposure assessment in humans, animals, and cell-culture systems. Over the last decade, his research teams have focused on using emerging technologies, novel experimental data, and computational methods for solving public-health challenges related to human exposure to chemicals. He is the director of the PNNL Exposure Surveillance and Health Optimization Consortium in which he leads efforts to develop nontargeted analytical methods for characterizing the exposome. Dr. Teeguarden has received several awards from the Society of Toxicology for his work in computational and experimental exposure science as they are related to translating exposures across cell-culture, human, and animal test systems. He has served as the president of the Dose–Response Specialty Section of the Society for Risk Analysis and as president of the Nanotoxicology Specialty Section of the Society of Toxicology. Dr. Teeguarden served on the National Research Council Committee on Human and Environmental Exposure Science in the 21st Century. He received his PhD in toxicology from the University of Wisconsin–Madison.

**James M. Tiedje** is the University Distinguished Professor of Microbiology and Molecular Genetics and of Plant, Soil and Microbial Sciences and is director of the Center for Microbial Ecology of Michigan State University. His research focuses on ecology, physiology, and genetics underlying important microbial processes in nature, including biodegradation of pollutants. He has made notable contributions to the use of genomics and metagenomics to understand ecological functions, speciation, and niche adaptation. He has served as editor-in-chief of *Applied and Environmental Microbiology and as editor of Microbial and Molecular Biology Reviews*. He has more than 500 refereed papers, including seven in Science and Nature. He shared the 1992 Finley Prize of UNESCO for research contributions of international significance in microbiology; is a fellow of the American Association for the Advancement of Science, of the American Academy of Microbiology, and of the Soil Science Society of America; and is a member of the US National Academy of Sciences. He was president of the American Society for Microbiology in 2004–2005. He received his PhD from Cornell University.

**Paolo Vineis** is professor and chair of environmental epidemiology at Imperial College London, School of Public Health. He is a leading researcher in the field of molecular epidemiology and his latest research focuses on examining biomarkers of disease risk, complex exposures, and intermediate biomarkers by using omic platforms in large epidemiological studies. He also studies the effects of climate change on noncommunicable diseases. Dr. Vineis is coordinating the European Commission–funded Exposomics Project and is a principal investigator or co-investigator on numerous international projects. He has

more than 700 publications, including papers in *Nature, Nature Genetics, Lancet, and Lancet Oncology*. He is a member of various international scientific and ethics committees and vice-chair of the Ethics Committee of the International Agency for Research on Cancer. Dr. Vineis received his MD from the University of Torino, Italy.

**Michelle Williams** is dean of the faculty and professor of epidemiology of the Harvard T.H. Chan School of Public Health. Her research interests lie principally in reproductive and perinatal epidemiology, in which she focuses on integrating epidemiological, biological, and molecular approaches into rigorously designed clinical epidemiology research projects. Her overarching goal is to use biological and molecular biomarkers as objective measures of exposure and as validated preclinical proximal determinants (such as oxidative stress, systemic inflammation, and endothelial dysfunction) of discrete outcomes of clinical, public, and global health importance. She is the principal investigator on three large projects funded by the National Institutes of Health and previously served on the National Research Council Committee on Evaluation of Children's Health: Measures of Risks, Protective and Promotional Factors for Assessing Child Health in the Community. Dr. Williams received her ScD in epidemiology from Harvard University.

**Fred Wright** is professor of statistics and biological sciences and director of the Bioinformatics Research Center of North Carolina State University. He is an internationally known statistical geneticist who has wide-ranging research interests, including genomics, bioinformatics, toxicogenomics, and the statistical principles underlying high-dimensional data analysis. Dr. Wright has been principal investigator on numerous grants with activities ranging from development of new methods of gene mapping to expression-quantitative trait mapping for multiple tissues. He was principal investigator of a US Environmental Protection Agency–funded STAR Center to apply genomics principles to long-standing problems in toxicology. He is an elected fellow of the American Statistical Association and of the Delta Omega Honor Society for Public Health. Dr. Wright received his PhD in statistics from the University of Chicago.

**Lauren Zeise** is director of the California Environmental Protection Agency's Office of Environmental Health Hazard Assessment. She oversees the department's activities, which include the development of risk assessments, hazard evaluations, toxicity reviews, cumulative impact analyses, frameworks and methods for assessing toxicity and cumulative effects of vulnerability and environmental exposures on communities, and the department's activities in the California Environmental Contaminant Biomonitoring Program. Dr. Zeise was the 2008 recipient of the Society for Risk Analysis' Outstanding Practitioners Award. She has served on advisory boards and committees of the US Environmental Protection Agency (EPA), the Office of Technology Assessment, the World Health Organization, and the National Institute of Environmental Health Sciences. Dr. Zeise has served on numerous National Research Council and Institute of Medicine committees, including the Committee on Toxicity Testing and Assessment of Environmental Agents and the Committee on Improving Risk Analysis Approaches Used by EPA. Dr. Zeise received a PhD from Harvard University.

# B

# Case Studies on Chemical Assessments

This appendix provides case studies that show how 21st century science can be used for chemical assessment, including any component of the risk-assessment process (hazard identification, dose–response assessment, exposure assessment, or risk characterization). The first case study illustrates the use of read-across methods to address gaps in information on a data-poor chemical. The second uses air pollution as a topic to illustrate how 21st century science can be used to address unanswered questions about well-defined hazards or data-rich chemicals and to evaluate emerging concerns about those hazards or chemicals.

## APPLICATIONS OF READ-ACROSS FOR A DATA-POOR CHEMICAL

As discussed in Chapters 3 and 5, read-across involves the assessment of a chemical on the basis of its structural similarities to chemicals that have already been tested and takes into account any differences that might influence pharmacokinetics, metabolism, or toxicodynamics. The approach can be coupled with computational and high-throughput data to support or refute the read-across results (see Figure 5-5). Alkylphenols are used as example chemicals for this case study.

Alkylphenols are metabolites or persistent environmental breakdown products of alkylphenol ethoxylates, chemicals that were formerly used in detergents. A few of the more widely used alkylphenols, particularly *p*-octylphenol and *p*-nonylphenol, have a rich toxicology dataset. In this case study, *p*-octylphenol and *p*-nonylphenol are used as analogues to support the assessment of *p*-dodecylphenol, a data-poor chemical that has been tested in ToxCast. Both *p*-octylphenol and *p*-nonylphenol have weak affinity for estrogen receptors in vitro (Laws et al. 2000). In vivo reproductive-toxicity data on the two chemicals have conflicting results. Multigeneration studies run under good-laboratory-practice (GLP) conditions by National Toxicology Program (NTP) indicate a few effects on reproduction with lowest observed-adverse-effect levels in the oral-intake range of about 30–100 mg/kg-day (*p*-nonylphenol, Chapin et al. 1999; *p*-octylphenol, Tyl et al. 1999). Other studies show effects on the reproductive system, although by different routes, such as parenteral injection, or at higher oral doses (see, for example, Hossaini et al. 2003; Mikkilä et al. 2006). Thus, the critical end point for the read-across is reproductive toxicity with estrogenicity as the presumed mechanism.

*p*-Dodecylphenol is a related chemical on which there are few in vivo toxicity data. The $K_{OW}$ for *p*-dodecylphenol is higher than those of the other alkylphenols, but all are very hydrophobic (see Table B-1). The chemical structure of *p*-dodecylphenol is similar to those of *p*-octylphenol and *p*-nonylphenol; the difference is that it has four or three more carbons, respectively, on the alkyl chain. Chemical-similarity scores for straight-chain *p*-octylphenol or *p*-nonylphenol are in the range of 55–65%. The chemical similarity score is a measure of molecular similarity that is based on atom-by-atom matching and is a good starting point for molecular comparisons. However, there is no bright-line chemical-similarity score for analogue suitability; it should be considered with other factors, such as physical chemistry and specific molecular features that can dramatically change potential reactivity or biological activity. Wu et al. (2010) provide a series of heuristics for determining the suitability of analogues for read-across. The committee notes that the chemical-similarity scores in Table B-1 suggest that the branched *p*-nonylphenol might be inappropriate for read-across for *p*-dodecylphenol. However, it is included here because most models of estrogenicity would consider *para*-substituted phenol moieties to have a potential to interact with the estrogen-receptor binding site—see, for example, the decision-tree scheme of Wu et al. (2013).

ToxCast has data on the chemicals in Table B-1. In each case, the most sensitive assay (the assay that has the lowest $AC_{50}$[1]) was one that measured estrogenic activity, and all chemicals were active in several estrogen–response assays at concentrations below 10 μM. Estrogen response (such as binding to the receptor or activation of an estrogen response element) was by far the most preva-

[1] $AC_{50}$ is the concentration at which a 50% response is elicited in an in vitro assay.

lent response to all four chemicals in ToxCast. Those results are consistent with the predictions from a qualitative structure–activity relationship (SAR) program developed by the US Environmental Protection Agency (EPA) that classifies all the chemicals as having weak estrogenic activity on the basis of the presence of a *para*-substituted phenol and the known estrogenic activity of *p*-alkylphenols as a class. A few other assays had a strong positive concentration response and an $AC_{50}$ at or below 10 μM (see Table B-2). Activity also included interactions with a retinoid X receptor (RXR) isoform, pregnane X receptor (PXR), a vitamin D receptor, and peroxisome proliferator-activated receptor gamma (PPAR-γ), and mitochondrial toxicity (see Table B-2).

In summary, the SAR and ToxCast data support grouping *p*-dodecylphenol with the other phenols as chemicals that appear to have a common mechanism, weak estrogenicity. The minor bioactivity observed with the other receptors (RXR, PXR, vitamin D receptor, and PPAR-γ) is not unexpected and emphasizes that the toxicant activity is typically multimodal. Even endogenous hormones that are considered to have high specificity for a particular receptor have comparable nonspecificity (Kelce and Gray 1997), and high-throughput assays provide the basis for evaluating other potential or unsuspected toxicities. The interactions at higher concentrations are probably not involved in toxicity. The overall in vitro potency of *p*-dodecylphenol as an estrogen appears to be higher by a factor of roughly 15 than that of the *p*-octyl and *p*-nonyl analogues, and it was active in 3 times as many estrogen-receptor assays. Because *p*-dodecylphenol is the most hydrophobic of the alkylphenols, its lower $AC_{50}$ could be inaccurate (see references and discussion in Chapter 2 on challenges in interpreting in vitro test data), but the data indicate that its estrogenicity in vitro is in the range of the other alkylphenols tested.

Estrogenic responses of *p*-octylphenol and *p*-nonylphenol have been reported in numerous studies, including in vivo rat multigeneration studies conducted by NTP (Chapin et al. 1999; Tyl et al. 1999). For the present case study, the no-observed-adverse-effect levels (NOAELs)[2] identified by the two NTP studies could be used as a starting point to derive a reference dose of *p*-dodecylphenol, although it should be noted that other published studies reported effects at lower doses. The studies were both feeding studies in which a dietary concentration of 200 mg/kg had no reproductive effects. Because the animals' growth and food consumption changed over time, a range of doses

[2]The committee notes that a point of departure identified through benchmark-dose modeling could also be used.

**TABLE B-1** Octanol–Water Partition Coefficients ($K_{OW}$s) and Chemical Similarity Scores (CSSs) of Selected Alkylphenols

| Chemical | CAS Number | Log $K_{OW}$[a] | CSS[b] |
|---|---|---|---|
| *p*-Octylphenol | 1806-26-4 | 5.5 | 0.55 |
| *p*-Nonylphenol | 104-40-5 | 5.76 | 0.64 |
| Branched *p*-nonylphenol | 84852-15-3 | 5.77 | 0.15 |
| *p*-Dodecylphenol | 104-43-8 | 7.91 | – |

[a]Log $K_{OW}$s are from EPA's EPI Suite database and prediction program (EPA 2011).
[b]The CSSs of analogues to test chemical (*p*-dodecylphenol) were calculated by using the Tanimoto coefficient from an online source (http://chemmine.ucr.edu). CSSs provide another line of evidence (quantitative) for using (or not using) visual read-across (qualitative) data.

**TABLE B-2** Activity in ToxCast Assays for Selected Alkylphenols

| | Protein Interactions; $AC_{50}$ values in μM[a] | | | | | Mitochondrial |
|---|---|---|---|---|---|---|
| Chemical | ER | RXR | PXR | Vitamin D Receptor | PPAR-γ | Toxicity |
| *p*-Octylphenol | 1.44 (4) | – | 1.71 | – | – | 9.23 |
| *p*-Nonylphenol | 1.35 (3) | 8.19 | – | – | 7.36 | – |
| Branched *p*-nonylphenol | 0.517 (14) | 1.4 | 2.29 | 1.98 | – | 6.3 |
| *p*-Dodecylphenol | 0.084 (13) | 2.74 | 1.45 | – | – | 3.28 |

[a]Number in parentheses is the number of estrogen-responsiveness assays with an $AC_{50}$ less than 10 μM.
Abbreviations: ER, estrogen receptor; PPAR-γ, peroxisome proliferator-activated receptor gamma; PXR, pregnane X receptor; RXR, retinoid X receptor.

(9-36 mg/kg-day) was associated with that concentration. Using the NOAELs as surrogates for *p*-dodecylphenol could require an adjustment for potency: the lowest $AC_{50}$ for *p*-dodecylphenol was about one-twentieth of the lowest $AC_{50}$ for *p*-octylphenol and *p*-nonylphenol, and this could require a comparable revision of the NOAEL.

Several limitations were identified in this read-across exercise. Improved estimations of the $AC_{50}$ data by using in vitro mass-balance models could be prudent before adjusting the NOAEL. Adjustments of the NOAEL on the basis of possible differences in the pharmacokinetics of the chemicals should also be considered. Differences in $logK_{OW}$ of 2 orders of magnitude are likely to be important in the rate and extent of absorption and clearance, although in this case the hydrophobicity of all the chemicals is high enough that one would expect high oral absorption of all chemicals. Predicted estimates of absorption and clearance and NOAELs for estrogenic effects could be obtained from targeted testing or similarly focused studies to corroborate the inferences based on read-across. Finally, the uncertainty in read-across should be assessed to ensure consistency and appropriate conservatism (Blackburn and Stuard 2014).

An outcome of this read-across exercise could be classification of *p*-dodecylphenol as an estrogenic compound potentially more potent than the other alkylphenols. Establishment of a reference dose would be plausible, but additional information on metabolism, absorption, and developmental effects on estrogen-sensitive organs would improve confidence.

## AIR-POLLUTION CASE STUDY

There is long-standing concern that exposure to air pollution might lead to chronic health effects, but only in the last several decades have epidemiological studies convincingly linked air-pollution exposure to premature mortality and increased risk of cardiovascular disease and cancer (EPA 2009). Beyond demonstrating hazard, recent studies have refined the characterization of the exposure–response relationship (Beelen et al. 2014). The new evidence reflects the increasing computing power that has enabled refinements in epidemiological methods, especially data-intensive exposure assessment that combines large-scale ambient monitoring of pollutants with advanced geographic information system (GIS) applications, dispersion models, and land-use regression (LUR) models to estimate exposures of large populations. Those methods—and decades of investment in nationwide air-pollution surveillance networks—have allowed researchers to establish long-term exposure models for large prospective cohort studies and to investigate long-term consequences of air pollution, such as cancer and cardiovascular disease, while controlling for major potential confounders. Studies based on those advances—exemplified by recent publications from the European Study of Cohorts for Air Pollution Effects (ESCAPE) consortium (Beelen et al. 2014)—have led a working group of the International Agency for Research on Cancer (IARC) to conclude that there is "sufficient" evidence to conclude that ambient air pollution is carcinogenic to humans and that the evidence is "sufficient" to conclude that airborne PM is carcinogenic to humans (IARC 2015).

The evidence on the causal relationship of air pollution with lung cancer (IARC 2015) is strong, and hazard identification is not at issue with regard to regulatory decision-making, at least in high-income countries with well-established evidence-based air-quality standards. However, there are a number of unanswered scientific questions concerning air pollution and cancer that are still relevant to regulatory decision-making; for these questions, 21st century science has the potential to reduce uncertainty around key issues relevant to tightening and targeting air-quality regulation. This particular case study illustrates how new and emerging science can be used to address lingering questions about well-defined hazards or data-rich chemicals and considers the following key issues:

- *Identifying critical air-pollution sources and components*. (1) Air pollution is a mixture that reflects its many sources; its composition varies by time and space. (2) The composition of the pollutant mix is not fully characterized, and research suffers from the "lamp-post syndrome" (that is, it has focused on a few target or indicator pollutants, such as EPA's criteria pollutants, including PM and nitrogen dioxide). (3) There is potential for interaction and synergy among different components of the air pollution mixture with implications for overall mixture toxicity.
- *Characterizing the exposure–response relationship*. (1) On the basis of available epidemiological evidence, there is no apparent threshold for the long-term effects of air pollution at current levels in the United States and elsewhere, particularly on total mortality and on cancer (Raaschou-Nielsen et al. 2013; Beelen et al. 2014; Hamra et al. 2014). (2) The power to detect effects and characterize risks precisely at low exposures is difficult even in large cohorts, such as the ESCAPE and American Cancer Society cohorts. (3) There are various hypotheses about the possible mechanisms by which air pollution causes long-term adverse effects at current exposures, and the mechanisms are likely to vary by outcome and pollutant mixture. (4) Specific groups might be at greater risk because of particular characteristics, such as genetics, life stage, disease status, or co-exposure to other agents.
- *Addressing emerging concerns*. There is an expanding list of possible adverse health effects of long-term exposure to air pollution. For example, some evidence indicates possible adverse effects on neurodevelopment

in children and decline of cognitive function in adults (Calderon-Garciduenas et al. 2014; Chen et al. 2015).

This case study develops two parallel examples. One is based on lung cancer, and the main concerns are estimating the exposure–response relationship, especially at low exposures, as experienced in the United States and much of Europe and identifying mechanisms involved and key mixture components that might drive cancer risk. The second example, neurodevelopment in children, has been chosen for different reasons. The questions concern mainly hazard identification because causal associations with air pollution for any specific neurodevelopmental outcome are far from well-established. The uncertainties in a number of neurodevelopmental outcomes reflect the challenges in investigating rare but severe outcomes, such as autism, that require large pregnancy cohorts that have detailed air-pollution assessments and the difficulties in comparing results among studies that evaluate a large array of neuropsychological effects and cognitive function at different developmental ages in children exposed to various pollutant mixtures.

## Lung Cancer: Characterizing the Exposure–Response Relationship and Identifying Key Mixture Components

Current epidemiological tools are unlikely to offer direct answers to the related problems of characterizing risk precisely at low doses and determining the shape of the exposure–response curve partly because there are limits to the size of cohorts that can be assembled and because exposure-measurement error is unavoidable with the available tools. However, those problems can be addressed with new and emerging approaches and tools described below that help to characterize exposure more precisely and to probe mechanisms more deeply.

### External Exposome

One critical issue in characterizing the exposure–response relationship is defining exposures more precisely, particularly at low levels of exposure. New exposure-assessment approaches centered around the concept of the exposome can help to address that issue. As defined in Chapter 1, the term *exposome* refers to the totality of a person's exposure. It is discussed here because of the emergence of new tools that provide time-integrated measurements of multiple pollutants at the individual level with greater spatial and temporal resolution than could be achieved previously (see Chapter 2). Such measurements potentially will help to characterize the exposure–response relationship better by reducing exposure-measurement error and by providing the needed inputs for measurement-error correction models.

The new exposure approaches contrast sharply with those used in past studies. Originally, epidemiological studies of air pollution relied on exposure classifications that were based on a few measurements in a few locations. Even the well-known Harvard Six Cities Study (Dockery et al. 1993), initiated in 1974, relied on central site measurements in the six selected cities. The wave of time-series studies that began about 3 decades ago fully incorporated the temporal detail of exposure measures but still used monitoring data that were limited spatially, such as central site monitors. Later cohort studies also incorporated more temporally refined measures, such as hourly or daily ambient monitoring station values, but again were spatially limited, often taken at one or a few stations per city. Citywide average exposures during specified periods were then applied to all residents in a design that would now be recognized as ecological or semiecological (that is, population-level assignment of exposure but with individual-level covariate information) (Künzli and Tager 1997). That approach, reflected in the Six Cities Study, ignores within-city variation and implicitly assumes that there is little spatial heterogeneity of air pollutants or that residents moved around cities enough to be similarly exposed to various pollutant sources. Neither assumption is correct in practice. Thus, measurement error was implicit in those studies, which nonetheless found associations with indicators of PM exposure, most likely because it was possible to exploit the high temporal resolution and fluctuations in air pollutants, especially in assessing short-term effects, such as in the time-series studies of mortality.

New tools are being developed to capture spatial variation in effects better (Coker et al. 2015). Early 21st century advances—such as GIS applications, dispersion models, and LUR models—have added a major refinement of capturing spatial variation in exposure assessment. Before those advances, exposures were generally assigned on the basis of residential location, and that practice accounted for some of the within-city variation. Reliance on residential location, however, did not fully capture or integrate exposures from multiple sources on larger geographic scales. For example, in Europe and the United States, investigators used tailored measurements of $PM_{2.5}$ in a number of cities with multiple land-use characteristics of each area (traffic, ports, population density, and factories) to predict concentrations at individual addresses with reasonably good performance by using LUR models and sometimes adding a temporal component to the estimates with data from routine ambient monitoring (Raaschou-Nielsen et al. 2013). However, those measurements were affected by measurement error as suggested by comparisons with, for example, personal-exposure monitoring campaigns. The latter are based on the use of backpacks or similar devices containing instruments that measure exposure at the individual level with great temporal and spatial resolution; such campaigns are gen-

erally conducted for shorter periods, such as 2–4 weeks, for feasibility. The external exposome measures showed the complexity of capturing the entirety of personal exposure to PM. For example, cooking was shown to be an important source of exposure to ultrafine particles. Such studies added to earlier understanding that personal exposure to air pollution can vary widely in time and space and be driven by specific time–activity patterns, such as time spent at home, in traffic, at work, and in restaurants. Without an understanding of such variation, exposure estimates can be quite inaccurate and bias risk estimates (Nieuwenhuijsen et al. 2015). New personal devices that will measure a large variety of pollutants are under development, as reported in Chapter 2.

However, none of the new sensor technologies is likely to be feasibly implemented (in terms of data handling and security) at the individual level in large cohorts over the extended periods (decades) necessary to investigate risks of chronic disease outcomes. Studies that are sufficiently large and have the detailed exposure information needed to address the key questions related to lung cancer and air pollution are not likely to be undertaken. Crowd sourcing or anonymous data collection using sensors might be a feasible alternative if implemented within existing or new general cohorts. The resulting data would then be used to refine exposure models and estimates associated with participants in such cohorts. Possible limitations of such data-collection methods include sampling bias and measurement error in the devices that might feasibly be deployed (see NRC 2012 for a more detailed discussion of possible limitations). The committee anticipates further refinements of exposure estimates within cohort studies. The refinements might be achieved by including extensive time–activity data in sophisticated spatiotemporally refined pollution models and by controlling measurement error better, which would reduce one major contributor to uncertainty in the burden of lung cancer attributable to air pollution.

New and emerging approaches also will be helpful for addressing the other challenge noted above that is related to characterizing the specific mixture components and the corresponding sources that drive lung-cancer risk. Most evidence on the health effects of PM air pollution from epidemiological studies—for example, on lung cancer—is based on estimated PM mass as the indicator of exposure. But PM is a complex mixture, and particles of different size and compositions might differ in toxicity and carcinogenic potential. Furthermore, PM exists within the broader air-pollution mixture.

New modeling approaches can provide estimates of concentrations of various PM components and characteristics and facilitate the exploration of the relationships between specific PM components and health risk. Recent studies have comprehensively characterized sources of outdoor air pollution and incorporated LUR models for estimating ambient $PM_{10}$, $PM_{2.5}$, and nitrogen dioxide (Raaschou-Nielsen et al. 2016). Models have then been developed for elemental composition (x-ray fluorescence), elemental and organic carbon, polycyclic aromatic hydrocarbons (PAHs), benzene, and ultrafine particles, which have been studied little because of difficulties in exposure assessment (Chang et al. 2015). Exposure estimation for ultrafine PM is now possible with, for example, an innovative mobile monitoring design that has been shown to be reliable and cost-effective (Hudda et al. 2014).

There are opportunties to use new in vitro and in vivo assays to evaluate and compare toxicity of PM samples. One of the properties of particles likely to reflect toxicity is oxidative potential, a property for which novel assays have been developed that measure the reduction of antioxidants in lung-lining fluid (Kelly and Fussell 2015). By analyzing the spatial and temporal variability of the oxidative potential of PM in filters, one can characterize the determinants of that variation and develop new spatially resolved air-pollution models for oxidative potential (Yang et al. 2015).

The air-pollution models alone, however, provide information only on ambient outdoor-pollutant concentrations and do not incorporate data on locations of members of the population needed for an exposure-assessment approach that would integrate data on various spaces. The models do not specifically take into account indoor exposure sources or indoor exposures to outdoor pollutants that have penetrated indoors. Recent advances in GIS (for example, route modeling) and microenvironmental models (for example, indoor-to-outdoor exposures) have led to the development of more detailed personal-exposure models that can be fed by rich sources of detailed data on population time–activity patterns, which should reflect time spent indoors. Regarding outdoor exposures, many cities hold information on origin and destination travel details from prepaid card systems or survey data on travel. Combined with regional or national surveys on time-use, those data constitute a rich additional source for personalized exposure models. Detailed data on personal and population-wide air-pollution exposures and space–time activity patterns from monitoring campaigns are required to evaluate new exposure models and thus support their use in providing improved exposure estimates for epidemiological studies and risk assessment.

**Internal Exposome**

The internal exposome can be investigated with two broad approaches: directly with analytical chemistry (as described in Chapter 2) and indirectly with several -omics technologies. Direct measurement focuses on the exogenous chemicals that can be found in internal fluids and measured with great sensitivity given current analytical-chemistry methods. Indirect measurements are based on

changes in DNA, RNA, proteins, or metabolites from which exposure to particular exogenous chemicals can be inferred. Genomics, transcriptomics, epigenetics, and proteomics allow only indirect inferences on exposures, and metabolomics and adductomics might allow direct measurements.

The use of -omics technologies described in this appendix allows the study of changes—for example, in blood or urine—that can help to characterize adverse effects of air pollutants, to refine exposure, to identify mechanisms, and to identify groups at risk. Here, the committee describes the potential contributions of the different -omics technologies in relation to the regulatory issues raised above and provides a few examples intended to show the potential of the rapidly developing science. A systematic review on the topic was not possible, given the scope of the relevant literature and the rapid development of this field. See Chapter 1 for definitions of -omics technologies.

*Genomics*

Carcinogenesis is understood to be a multistep process to which genetic and nongenetic changes contribute (see Smith et al. 2016). For lung cancer and air pollution, information on genetic determinants of risk would be useful for public-health protection. Genomics can be based on the systematic investigation of genetic (inherited) variants that lead to or increase susceptibility to air-pollution–related disease or can be based on the study of somatic mutations induced by air pollution in cells. Concerning inherited susceptibility, several genetic variants (such as GSTM1) have been investigated in the candidate gene era; more recently, variants have been identified thanks to genome-wide association studies (see, for example, Kachuri et al. 2016). The associations of genetic variants with lung cancer are mostly weak, but the findings of some variants associated with lung-cancer risk have identified groups in the population that are potentially more susceptible to carcinogens.

A potentially fruitful approach for identifying susceptible groups is to develop profiles of susceptibility that are based on genetic pathways. For example, Bind et al. (2014) used a pathway-analysis approach to investigate whether gene variants that are associated with such pathways as oxidative stress, endothelial function, and metal processing modified the association of PM exposure and fibrinogen, C-reactive protein, intercellular adhesion molecule-1, or vascular-cell adhesion molecule-1.

Concerning somatic (acquired) mutations, the sequencing of several types of cancer tissues has shown that mutational patterns can reflect environmental mutagens (Nik-Zainal et al. 2015). For example, lung cancer has a mutational pattern that strongly resembles that induced by benzo[a]pyrene (B[a]P) in in vitro assays that use immortalized mouse embryo fibroblasts (Nik-Zainal et al. 2015). The results revealed that B[a]P induces a characteristic mutation signature: predominantly G→T mutations for B[a]P as opposed to C→T and CC→TT for ultraviolet radiation and A→T for aristolochic acid, a carcinogenic and mutagenic compound. Thus, the study suggests that the carcinogenicity caused by smoking (and possibly air pollution) could be due to the PAH component in smoke (or ambient air). Mechanistically, that information is of great importance.

Genomics could thus prove useful in two ways. First, genetic (inherited) variants that contribute to modulating the cancer risk associated with air-pollution exposure could be identified. Identification of populations at greater (or less) risk would refine understanding of the exposure–response relationship and point to a susceptible population. Second, if a molecular signature in tumor tissue (somatic mutations) were linked specifically to air-pollution exposure, burden could be more effectively quantified and exposure–response models developed for particular phenotypes defined by etiology. The committee notes that substantial research indicates differences in mutational spectra of lung cancers between smokers and never smokers, although markers that are definitive for any specific type of environmental exposure have not yet been identified. Third, even if signatures are not identified, mechanistic insights that support biological plausibility further and perhaps provide insights concerning mixture components could be gained.

*Epigenomics*

Environmental exposures are able to change epigenetic signatures, for example, the methylation pattern of DNA or chromatin. DNA methylation and the associated repressed or activated transcription of genes might affect carcinogenesis (Vineis et al. 2010). Changes in methylation of the aryl-hydrocarbon receptor (AHR) repressor gene show that methylation can be used as a marker of exposure to smoking (Shenker et al. 2013) and to monitor the effect of cessation of exposure (Guida et al. 2015). Some authors have used AHR repressor methylation as a marker for in utero exposure of the fetus to tobacco-smoke components from maternal smoking (Joubert et al. 2012). Epigenetic markers in cord blood and placental tissue could also be used to detect possible effects of air-pollution exposure on the fetus and might be useful in addressing the question of whether maternal exposure to air pollution leads to developmental effects (Novakovic et al. 2014). And epigenetic markers might provide information on exposure to air pollution and even particular components.

How informative epigenetics is in studying risks of disease or health outcomes depends on whether the markers are permanent, whether they develop during a critical age window, and whether the right tissue can be investigated; methylation markers are tissue-specific. A few studies have investigated the effects of air-pollution exposure on DNA-methylation patterns (see, for example, Baccarelli et al. 2009) and focused on methylation of long interspersed element-1 (LINE-1) and Alu elements as measures of whole-genome methylation in blood cells. LINE-1 and Alu elements are retrotransposons, that is, repetitive and mobile sequences in the genome. LINEs comprise a substantial proportion of the genome, and LINE-1 and Alu methylation correlates with overall cellular levels of DNA methylation. Air pollution was found to alter LINE-1 methylation (Baccarelli et al. 2009; Demetriou et al. 2012).

Epigenetic changes might also be integral to carcinogenesis, perhaps to the same extent as genetic mutations. Fasanelli et al. (2015) showed that the same genes (including the AHR respressor gene) for which methylation changes are associated with smoking predict lung-cancer risk. Similar studies are not available for air pollution and lung cancer.

Given the substantial current emphasis on the epigenome and the environment, the committee anticipates that the utility of epigenetics in risk assessment will be determined over the next decade. Studies that span the life course are in progress, and there is opportunity for marker validation over longer times, although this research would require multiple biological samples from well-characterized large cohorts.

*Transcriptomics*

Transcriptomics can lead to the identification of perturbations in gene expression relevant to lung carcinogenesis due to environmental exposures, including exposure to air pollution. Thus, transcriptomics is expected to be a key tool in research, for example, for identifying which specific components of an air-pollution mixture are biologically active and might have a role in causing lung cancer. Transcriptomics might also help to reveal interactions of mixture components by showing that the overall effect of a mixture on gene expression is greater than the sum of gene expression of the individual components.

Gene-expression changes have been linked to air-pollution exposures in in vitro and animal experiments. Specifically, exposure to air pollution leads to increased or decreased expression of genes that are relevant to immune or inflammatory actions. Although few observations have been made in humans, Wittkopp et al. (2016) performed an exploratory analysis and tested whether gene expression was associated with air-pollution exposures in a Los Angeles area cohort of elderly subjects who were exposed to $PM_{2.5}$ at an average of 10-12 $\mu g/m^3$. The authors found positive associations of traffic-related pollutants (including nitrogen oxides and PAH content in $PM_{0.25–2.5}$ or $PM_{0.25}$) with the expression of several candidate genes, particularly Nrf2-mediated genes, which indicated involvement of oxidative stress pathways. A number of genes have been found to be dysregulated by using transcriptomics tools in studying lung cancer (see, for example, Amelung et al. 2010).

*Proteomics*

As noted in Chapter 1, proteomics refers to the measurement of the whole compartment of proteins in a biological sample with high-throughput techniques. Like transcriptomics, it might be useful in characterizing toxicity of individual air-pollution components, identifying interactions of air-pollution components, and identifying pathways that might be involved in a response to air pollution and possibly related to lung carcinogenesis For example, the association between long-term exposure to air pollution and inflammatory markers was investigated with a proteomic approach (Mostafavi et al. 2015), and immune–inflammatory perturbations were observed at high exposures. Little work has been conducted on the proteome in relationship to air pollution.

*Adductomics*

DNA and protein adducts have long been measured in relation to air-pollution exposure (Demetriou et al. 2012; Demetriou and Vineis 2015). Specific adducts, such as PAH–DNA adducts, have been measured. Adductomics is a new approach to identifying exposure biomarkers with a systematic, high-throughput search of all potential adducts resulting from external exposures or internally generated compounds. As part of the exposome concept, adductomics typically involves an untargeted investigation that analyzes hydrolysis products of albumin by using mass spectrometry. Electrophilic chemicals or their metabolites that bind to albumin are also likely to bind to DNA. Thus, protein-based adductomics can potentially be used to identify genotoxic, electrophilic components in a mixture. Adductomics might also be helpful in refining exposure–response relationships, including the shape of the exposure–response curve for lung cancer, because the high sensitivity of adductomics reduces misclassification and uncertainty. That research would require repeat samples from prospective cohorts, and one of the pillars of modern epidemiology is the availability of large prospective cohorts with multiple samples that create an opportunity to study the stability of signals. Some of the markers integrate exposures over relatively long periods and would thus be useful for exposure estimation.

*Metabolomics*

Metabolomics can be performed on plasma, serum, or urine samples by several methods, including high-resolution mass spectrometry coupled to ultra-high-performance liquid chromatography for untargeted analyses. Metabolic features that characterize exposed groups are identified by multivariate statistics with appropriate correction for false discovery rate. Metabolites unique to exposed groups are then identified with more targeted investigations. However, metabolomics data are subject to high intraindividual variability, and many metabolites have short lives, which might limit their utility in estimating longer-term exposures. Annotation is another limiting factor; researchers are unable to characterize features detected with, for example, mass spectrometry without additional chemical analyses. In principle and with likely future technical developments, however, metabolomics could become a useful tool for achieving several goals, as suggested in Table B-3: the identification of specific metabolites related to mixture components and their interactions, better characterization of exposure by linking metabolites to external measurements, and reconstruction of molecular and biochemical pathways, which would contribute to mechanistic knowledge and identification of pathways.

## Concluding Remarks

Early and still evolving findings from epidemiological research that uses -omics techniques are starting to suggest that air pollutants might act via pathways that involve inflammation and oxidative stress. In addition, there might be mutational signatures that are characteristic of air-pollution exposure vs, for example, smoking, although air pollution and cigarette smoke have several common components, such as PAHs. The small samples of early studies, however, do not allow sound quantitative estimation of pathway perturbations at low doses. Although the evidence is limited, some consistency is emerging among different -omics platforms, such as transcriptomics, epigenomics, and proteomics. The consistency among platforms can be investigated by using statistical techniques known as cross-omics (Vineis et al. 2013). The long-term goal is to couple external exposome approaches to reduce measurement error at the individual level with a suite of -omics investigations that characterize the various steps involved in carcinogenesis by investigating, for example, mutational spectra, epigenetic changes, inflammation, and cell proliferation in human samples. That research is expected to lead to more accurate quantitative risk assessment.

Overall, -omics technologies will facilitate exploration of all the characteristics of carcinogens and the pathways that lead from exposure to diseases. The main challenges are related to the variability of measures due to technical reasons and biological intraindividual variation, the long latency of cancer with decades between exposure and disease onset and the multiple steps involved, and the lack of access to precursor lesions—there is access only to surrogate tissues, such as blood—to study molecular changes that take place in target cells. Regardless of the

**TABLE B-3** Relevant Regulatory Questions and How -Omics Technologies Might Help to Answer Them in the Case of Lung Cancer[a]

| | -Omics Technologies | | | | | |
|---|---|---|---|---|---|---|
| Regulatory Question | Genomics | Epigenomics | Transcriptomics | Proteomics | Adductomics | Metabolomics |
| Identifying Critical Air-Pollutions Sources and Components | | | | | | |
| Characterize toxicity and long-term effects of mixture components | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Investigate interaction potential of mixture components | | | ✓ | ✓ | | ✓ |
| Characterizing the Exposure–Response Relationship | | | | | | |
| Characterize exposure better | | | | | ✓ | ✓ |
| Identify mechanisms | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Identify groups at greater risk | ✓ | | | | | |

[a]This table is related to the current knowledge and uses of -omics in the field of lung carcinogenesis. Assignment of checkmarks in the table is likely to change with advances in the science of -omics and in the understanding of lung carcinogenesis.

challenges, the -omics technologies offer opportunities to identify critical components of air-pollution mixtures and to refine the exposure–response relationship as illustrated in Table B-3.

### Neurodevelopmental Effects and Particulate Air Pollution: Determining Whether a Causal Relationship Exists

Determining whether there is a causal relationship between neurodevelopmental effects and PM is potentially of great public-health importance. It has long been known that fetuses, infants, and young children are more sensitive than adults to diverse environmental toxicants because of the vulnerability accompanying developmental, growth, and maturation processes (WHO 1986; NRC 1993; Anderson et al. 2000; Perera et al. 2004; Grandjean and Landrigan 2006). One topic of particular concern is neural development. A large body of research has addressed the influences of air pollution on fetal growth, including head circumference (Vrijheid et al. 2011; Stieb et al. 2012; van den Hooven et al. 2012; Backes et al. 2013; Proietti et al. 2013; Smarr et al. 2013). More recently, epidemiologists have become interested in potential effects of PM air pollutants because some combustion components of PM, such as PAHs and their derivatives, have shown neurodevelopmental toxicity in some experimental and small pathology studies (Calderon-Garciduenas et al. 2002; Takeda et al. 2004). In this section, the committee briefly discusses the epidemiological studies that have linked air-pollution exposures to neurodevelopmental effects and offers some suggestions on how ES21 and Tox21 tools and methods could be used to strengthen or improve the epidemiological studies. The committee notes that epidemiological studies that address neuropsychological effects of air pollution have been summarized by Guxens and Sunyer (2012) and Suades-González et al. (2015) and are not discussed here. The section concludes with some general considerations related to developmental neurotoxicity (DNT) and possible approaches for studying DNT.

### Epidemiological Evidence of Associations Between Air Pollution and Neurodevelopment in Children

Epidemiological studies have begun to investigate the association between various air pollutants and neurodevelopmental effects in children. The characteristics and designs of the key studies are summarized in Table B-4. Several small cohort studies in the United States, Poland, and China have shown adverse neurodevelopmental effects in children exposed in utero to PAHs (Perera et al. 2006, 2009; Tang et al. 2008, 2014; Edwards et al. 2010; Lovasi et al. 2014). PAH exposure in the studies was measured through short-term (48-hour) personal-exposure measurements during pregnancy or as PAH–DNA adducts in cord blood. The adverse effects reported were decreases in mental function or IQ and motor developmental delays early in childhood, but these effects were not observed consistently at all ages at which the children were examined. An additional cohort study in the United States linked adverse neuro-developmental effects (IQ and attention disorders) in children with increases in children's lifetime exposure to black carbon, which is related to traffic (Suglia et al. 2008; Chiu et al. 2013); however, only in boys was black-carbon exposure associated with attention disorders, and this suggests possible sex-specific vulnerability. A large European study combined six birth cohorts (Guxens et al. 2014) and reported that nitrogen dioxide, but not other air pollutants, was associated with delayed psychomotor development in children 4 years old and younger; no associations with cognitive or language development were seen. In addition, several Asian studies and a Polish study reported associations of different types of air pollutants and exposure periods with various developmental outcomes (see Table B-4 below). Most of the studies were small, tested children at different developmental ages and for different functions or disorders, and measured exposures prenatally or postnatally, focusing on different pollutants and sources. Thus, additional studies are needed to replicate or confirm some of the reported findings before conclusions about associations of air pollution with adverse neurodevelopment outcomes can be drawn from epidemiological data.

The limitations of the epidemiological studies might be addressed by ES21 and Tox21 approaches. The following paragraphs summarize the challenges and possible approaches to addressing them.

- Studies testing children's neuropsychological function at different ages are time-consuming and expensive, and researchers have to balance various factors, such as the extent and variety of functional assessments, cohort size, and length of follow-up. Feasibility and costs are major concerns. Those problems are exemplified in the most recent review of epidemiological studies (Suades-González et al. 2015), which still did not identify sufficient data to conduct quantitative meta-analyses because of heterogeneity in the methods used to assess exposures and outcomes. With respect to cognitive and psychomotor development, Suades-González et al. (2015) decided that for only one exposure (PAHs) were there enough high-quality studies available to conclude that there was "sufficient evidence" of an association but not a causal relationship. For other air pollutants, modern exposure assessment and modeling—GIS or dispersion modeling supported by satellite data and ground-level monitoring networks—might facilitate adding comparable air-pollution exposure measures to those completed or current expensive human studies of neurodevelopment (for example, studies using neuroimaging or extensive functional testing). Eventually, the research conducted might

**TABLE B-4** Study Design of Epidemiological Studies That Have Investigated Neurodevelopmental Effects of Air Pollution

| Study Characteristics | Exposure Details | Principal Outcomes Investigated | Selected Findings | Reference |
|---|---|---|---|---|
| N = 46,039 singleton births in Japan on January 10–17 or July 10–17, 2001 | Evaluated maternal exposure to air pollution related to municipality-level traffic, including PM, $NO_2$, CO, and $SO_2$ in the 9 months before birth. Air-pollution measurements were taken from general and roadside stations nationally. | Milestone delays were measured through a series of questions administered at ages 2.5 and 5.5 years. Questions were not validated or selected from an established scale, but have been used in previous studies. | Estimated air-pollution exposure during gestation was positively associated with some risk of several developmental milestone delays at both ages—verbal and fine motor development at age 2.5 years and behaviors related to inhibition and impulsivity at 5.5 years. | Yorifuji et al. 2016 |
| N = 183 children, 3 years old, born to black and Dominican women in New York, NY, mother–child pairs recruited in 1998-2003 | Evaluated prenatal exposure to airborne PAHs, secondhand tobacco smoke, and pesticides; PAHs were monitored during pregnancy with personal air sampling.<br><br>Umbilical cord blood was taken at delivery, and maternal blood within 2 days postpartum was analyzed for cotinine, heavy metals, and pesticides. | The Bayley Scales of Infant Development- Revised were used to assess cognitive and psychomotor development at ages 12, 24, and 36 months to generate an MDI and corresponding PDI. Behavioral problems were measured on the Child Behavior Checklist. | Prenatal exposure to PAHs of the mothers was not associated with PDI or behavioral problems. However, high prenatal exposure to PAHs (the upper quartile of the distribution) was associated with lower MDI at the age of 3 years, but not 1 or 2 years. | Perera et al. 2006 |
| N = 249 children, 5 years old, born to black and Dominican women in New York, NY, mother–child pairs recruited 1998–2003.<br><br>Note: This cohort is the same as Perera et al. 2006. | PAHs were measured in women in their third trimester with a personal monitoring device during the daytime hours for 2 consecutive days; monitor was placed near the bed at night. Pumps operated continuously during this period, collecting vapors and particles ≤ 2.5 μm in diameter. | The WPPSI-R was used to determine verbal, performance, and full-scale IQ scores. | Women who had higher exposure to PAHs during pregnancy were significantly more likely to have infants with lower full-scale and verbal IQ scores tested at the age of 5 years. After adjustment for maternal intelligence, quality of the home caretaking environment, environmental tobacco-smoke exposure, and other potential confounding factors, high PAH levels (above the median of 2.26 $ng/m^3$) were significantly and inversely associated with full-scale and verbal IQ scores but not with performance IQ scores. | Perera et al. 2009 |
| N = 326 children, born to black and Dominican women in New York, NY in 1998–2006.<br><br>Note: This cohort is the same as Perera et al. 2006. | PAH exposures were measured with personal ambient air monitors worn for 2 consecutive days and placed at the bedside at night during the third trimester of pregnancy. Housing disrepair was self-reported by mothers, and neighborhood characteristics were estimated within a 1-km network from the prenatal address overlaid with data from the 2000 US Census. Indicators measured included number of residents below the federal poverty line, high-school diploma or equivalent degree attained, and low neighborhood English-language proficiency. | The WPPSI-R was used to assess intelligence and neurodevelopment at of age 5 years. Spanish scores were excluded because of difference in the Spanish- and English-language versions. | Prenatal PAH exposure above the median was significantly associated with lower total WPPSI-R and verbal scores. The mean differences were 3.5 total points and 3.9 verbal points between high and low PAH exposure groups, respectively. | Lovasi et al. 2014 |
| N = 214 children born to women in Krakow, Poland | Exposure to eight PAHs was measured with personal air monitors carried over a 48-hour period during the second or third trimester of pregnancy; monitors were kept at the bedside at night during this period. | At age 5 years, RCPM were used to assess a child's nonverbal reasoning ability. | A higher prenatal exposure (above the median of 17.96 $ng/m^3$) to airborne PAHs (range, 1.8–272.2 $ng/m^3$) was significantly associated with decreased RCPM scores at the age of 5 years, after adjustment for potential confounding variables. This corresponds to an estimated average decrease of 3.8 IQ points. | Edwards et al. 2010 |

| | | | | |
|---|---|---|---|---|
| N = 1,257 US children, 6–15 years old; data collected from 2001–2004 cycles of NHANES. | PAH exposure was based on urinary metabolite concentrations measured in the 2001–2002 and 2003–2004 cycles. | Outcomes were measured by parental reporting of (1) ever doctor-diagnosed ADHD (2) ever doctor- or school representative-identified LD and (3) receipt of SE or early intervention services. | Higher concentrations of fluorine PAH metabolites in children were associated with 2-fold increased odds of needing SE, somewhat more in males than in females. | Abid et al. 2014 |
| N = 202 children in Boston, MA, participating in a prospective birth cohort study (1986–2001) | Exposure to BC was estimated with a model on the basis of child's residence during study follow-up. Data collected from more than 80 locations in the greater Boston area were used to complete a spatiotemporal LUR model to predict 24-hour measures of traffic exposure. | Cognitive tests were administered at ages 8–11 years and included the K-BIT (assesses verbal and nonverbal intelligence) and the WRAML (evaluates a child's ability to actively learn and memorize a variety of information). | With adjustment for sociodemographic factors, birth weight, blood lead concentration, and tobacco smoke, BC exposure was associated with decreases in the vocabulary (-2.2), matrices (-4.0), and composite intelligence quotient (-3.4) scores of the K-BIT and visual subscale (-5.4) and general index (-3.9) of the WRAML. | Suglia et al. 2008 |
| N = 174 children, 7–14 years old in Boston, MA.<br><br>Note: This cohort is the same cohort as Suglia et al. 2008 | Traffic-related black carbon (BC) concentrations were estimated over child's lifetime using a spatiotemporal model for 24-hour measures of BC based on 6,021 observations from >2,079 unique exposure days at 82 locations in greater Boston area. Models took into consideration warm (May–October) and cold (November–April) seasons. | The Conners' CPT was used to assess attention disorders and neurological functioning at ages 7–14 years. | In this population of urban school-aged children, there was a positive association between higher BC and increased commission errors and lower HRT, even after adjustment for child IQ, age, sex, and other variables. Sex-stratified analysis showed statistically significant associations between BC and both commission errors and HRT in boys, but BC was not significantly associated with any outcomes in girls. | Chiu et al. 2013 |
| N = 9,482 children in six European population-based birth cohorts: the Netherlands, Germany, France, Italy, Greece, and Spain; mother–infant pairs recruited in 1997-2008. | LUR models were used to estimate $NO_x$ in all study regions and PM with diameter <2.5, <10, and 2.5–10 μm, and $PM_{2.5}$ absorbance in subregions. Monitoring campaigns took place primarily from October 2008 to January 2011.<br><br>$NO_x$ was measured at least three times per week for 2 weeks within 1 year. $PM_{2.5}$ absorbance was measured in a subgroup of regions by reflectance of $PM_{2.5}$ filters. To obtain final analyses, a back-extraction procedure was used to estimate the concentrations during each pregnancy of each woman. | Cognitive and psychomotor development was assessed at ages 1–6 years. Different neuropsychological tests for cognitive and psychomotor development were administered, including McArthur Communicative Development Inventory, Bayley Scales of Infant Development I–III editions, Denver Developmental Screening Test II, McCarthy Scales of General Cognition, and Ages and Stages Questionnaire. | Air-pollution exposure during pregnancy, particularly $NO_2$ (of which traffic is a major source) and $PM_{2.5}$, was associated with delayed psychomotor development in children (-0.68 points in the global development score) for each 10 $\mu g/m^3$ increase in $NO_2$). Cognitive development measured at similar ages was not related to air-pollution exposure during pregnancy. | Guxens et al. 2014 |
| N = 520 mother–child pairs in three regional centers in South Korea studied in January 1, 2006–December 31, 2008 | Exposure to $PM_{10}$ and $NO_2$ during pregnancy was estimated with inverse distance-weighting method. A mini-volume air sampler was used to measure outdoor ambient $PM_{10}$; a passive sampler was used to measure outdoor ambient $NO_2$; sampling was performed over 24 hours. | The Korean Bayley Scale of Infant Development II was used to measure neurodevelopment progress. Results were expressed as MDI and PDI at 6, 12, and 24 months. | There was a negative association between maternal exposure to $PM_{10}$ and MDI and PDI throughout the first 24 months of life. Maternal $NO_2$ exposure was associated with impairment of PDI but not with cognitive function. A multiple-linear-regression model showed significant effects of prenatal air-pollution exposure ($PM_{10}$ and $NO_2$) on MDI and PDI at 6 months, but no significant associations were found at 12 and 24 months. | Kim et al. 2014 |

*(Continued)*

**TABLE B-4** Continued

| Study Characteristics | Exposure Details | Principal Outcomes Investigated | Selected Findings | Reference |
|---|---|---|---|---|
| N = 533 mother–infant pairs in 29 villages or cities in Taiwan selected in October 2003–January 2004; followed up at 6 and 18 months. | Hourly ambient concentrations of CO, $O_3$, $PM_{10}$, $SO_2$, $NO_2$, THCs, and NMHCs were measured at the Taiwan Air Quality Monitoring Network. Participant exposure was considered to be the average taken during the period 7 am to 7 pm. Air-pollutant exposure for each child was measured by linking data from the air-quality monitoring stations of the town to the exposure period from the beginning of gestation to 18 months after birth. The gestational period was divided into 3 trimesters, and the postpartum ranges were birth–6 months, 7–12 months, and 13–18 months. | Neurodevelopmental performance was measured by parent responses to a screening instrument, the TBCS. The scale consists of four developmental divisions: gross motor, fine motor, language/communication, and social/self-care abilities. Parents completed two neurobehavioral development scales at each interview; responses consisted of never, sometimes, and all the time. Scales have good predictive validity, and dimensions correlate with the Bayley Scales of Infant Development. | Various indexes of ambient air pollution, even low $SO_2$ exposure, during pregnancy and up to the age of 12 months were associated with poor subclinical neurodevelopment (neurobehavioral effects and poor gross motor development) in early childhood. | Lin et al. 2014 |
| N = 133 children born March 4, 2002–June 19, 2002, in three Tongliang, China county hospitals; followed for 2 years | Study carried out in an area in China with a seasonally operated coal-fired power plant. PAH–DNA adducts, Pb, and Hg were measured in umbilical-cord blood samples collected at delivery. HPLC was used to analyze B[a]P–DNA adducts in extracted white blood cell DNA. A PE-800 Zeeman atomic absorption spectrometer with background correction system was used to measure Pb in samples. | Physical, emotional, and behavioral development of 2-year-old children was measured with the GDS. Children received DQs for each of motor behavior, language behavior, personal behavior, and social behavior. | Increased cord adduct concentration was inversely associated with decreases in the motor area DQ, language area DQ, and average DQ after adjustment for cord lead concentration, environmental tobacco smoke, sex, gestational age, and maternal education level. High cord blood lead was also significantly associated with decreased social area DQ and average DQ. The frequency of developmental delay ranged from 9.1% (social) to 13.6% (motor), with an average score of 6.4%. | Tang et al. 2008 |
| N = 150 children born March 4, 2002–June 19, 2002, compared with a cohort of 158 children born March 2, 2005–May 23, 2005; both cohorts consisted of children born in Tongliang, China.<br><br>Note: This cohort is the same as Tang et al. 2008 | Two mini-volume samplers were used at three sites in Tongliang in March 2002–February 2003 and in March 2005–February 2006 to collect 72-hour PAH samples. Overall PAH concentrations were measured by analyzing B[a]P–DNA adducts in extracted white blood cells collected from the umbilical cord at delivery and from the mother within 1 day postpartum. | Birth weight, length, and head circumference were measured at birth or more than once after birth if the child was delivered by cesarean section. Neurodevelopment was measured with the GDS at the age of 2 years. As above, DQs were developed for motor, adaptive, language, and social behavior. | The power plant was closed between the recruitment of the two cohorts. Patterns of developmental delay in all DQ areas except language were improved in the 2005 post-shutdown cohort compared with the 2002 cohort. | Tang et al. 2014 |

Abbreviations: ADHD, attention deficit hyperactivity disorder; BC, black carbon; CO, carbon monoxide; CPT, Continuous Performance Test; DQ, developmental quotient; GDS, Gesell Developmental Schedules; Hg, mercury; HPLC, high-performance liquid chromatography; HRT, hit reaction time; K-BIT, Kaufman Brief Intelligence Test; LD, learning disability; LUR, land-use regression; MDI, mental-development index; NHANES, National Health and Nutrition Examination Survey; NMHC, nonmethane hydrocarbon; $NO_x$, nitrogen oxides; $NO_2$, nitrogen dioxide; $O_3$, ozone; PAH, polyaromatic hydrocarbon; Pb, lead; PDI, psychomotor-development index; PM, particulate matter; RCPM, Raven Coloured Progressive Matrices; SE, special education; $SO_2$, sulfur dioxide; TBCS, Birth Cohort Study Scale; THC, total hydrocarbon; WPPSI-R, Wechsler Preschool and Primary Scale of Intelligence-Revised; WRAML, Wide Range Assessment of Memory and Learning.

provide sufficient sample size, appropriate exposure gradients, and possibly information about source-specific or chemical-specific pollution components to generate results that allow quantitative or causal evaluation of air pollutants and neurodevelopment.

- Key limitations in many DNT studies of air pollution are that they cannot address multiple air-pollutant exposures (mixtures) and most likely can ascertain potential confounders only incompletely, given the limited knowledge of social and cultural determinants of neurodevelopment and the strong association of neurodevelopment with socioeconomic status (SES). GIS could help to disentangle the role of SES by allowing, for example, area-level adjustment for correlates of SES. Computer-resource–intensive multilevel spatial modeling in a Bayesian framework might also allow addressing spatially correlated confounders and pollutant mixtures (Coker et al. 2015, 2016).
- In future studies with smaller samples, it might be possible to use personal air monitoring or biomarker approaches that include new sensor technologies if instruments are small and lightweight and if measurements are less expensive and thus feasible. The new approaches would allow monitoring over extended periods in pregnancy or early life. With the exception of PAH adducts, there are no good biomarkers for toxic PM components. Monitoring only particles does not allow assessment of the toxicity of their components, and particle composition probably depends on the sources that generate the particles. However, combining continuous particle monitoring with repeated collection of relevant biosamples (such as maternal and infant blood, urine, and placenta) would also allow the use of -omics tools to find new exposure biomarkers in human samples and possibly some biomarkers predictive of outcomes (see, for example, Janssen et al. 2015; Saenen et al. 2015). Nontargeted approaches might be useful for identifying new biomarkers.

### General Considerations Related to Developmental Neurotoxicity and Possible Assessment Approaches

Historically, establishing causal linkages between neurodevelopmental disorders and environmental exposures, such as exposure to air pollution, has been difficult for a variety of reasons, including the need for large populations in epidemiological studies, the complexity of capturing the full array of relevant exposures before and during pregnancy, the long latency between exposure and effect (particularly for neurodegenerative disorders), the lack of defining pathology of some disorders (such as schizophrenia or autism spectrum disorder), and inherent limitations of animal models and in vitro assays. Perspectives and strategies for assessing DNT more comprehensively have been published by various stakeholders and will not be recapitulated here (Aschner et al. 2010; Bal-Price et al. 2015; Felter et al. 2015). This discussion highlights the unique challenges associated with trying to assess DNT and provides some possible approaches to doing so.

The most notable challenge unique to brain and behavioral targets is the dynamic complexity of the developing brain and a fundamental lack of understanding of the etiology of complex behavioral disorders, such as intellectual disability and emotional impairment. A disease-centric approach to DNT risk assessment is particularly challenging and unlikely to be feasible because many neural disorders, especially neuropsychiatric disorders, are syndromes with a spectrum of hallmark features and lack defining neuropathology or clear etiology. Thus, it is not plausible or rational to use a framework that attempts to make clear linkages between exposure, DNT mechanisms, and neural disease. Only a few such models have been proposed for DNT, and they are all too general (for example, oxidative stress) and do not explain the pathology well. Furthermore, the evidence does not support their acceptance with confidence, particularly in the neuroscience community. Instead, risk assessment of and chemical screening for DNT will have to be conducted in recognition that in the absence of an extraordinary situation (major accident or industrial exposure) clear linkages between exposure and a clinically diagnosed neural disease will be challenging.

Although perspectives on how to improve DNT risk assessment in a regulatory context differ, there is general agreement that testing for DNT should focus on evolutionarily conserved, fundamental events in neurodevelopment. Those events include neural induction, cell migration, axonal guidance, synapse formation and pruning, and apoptosis. Perturbation of the critical events underlies the primary deficits in neural disorders. Given that perspective, developmental neurotoxicants would be identified by their capacity to alter the fundamental events, regardless of their specific cellular or molecular mechanisms. Examples in which that perspective has yielded critical insight in connection with air pollution include evidence that $PM_{2.5}$ induces oxidative stress in homogenates of rat brain regions and disrupts blood–brain barrier integrity, thereby enhancing neurotoxicity by activated macrophages and microglia (Fagundes et al. 2015; Liu et al. 2015). In mice, developmental exposure to ultrafine particles induced sex-specific neurotoxicity (including excitotoxicity and glial activation) and behavioral changes indicative of heightened impulsivity and hyperactivity—behavioral changes also associated with exposure of children to air pollution (Allen et al. 2014). Furthermore, in utero exposure to B[a]P during peak periods of neurogenesis in mice leads to behavioral learning deficits (McCallister et al. 2016).

Rapidly evolving experimental, epidemiological, computational, and toxicity-screening strategies are poised

to assess neurotoxicity and neuroendocrine disruption better and to fill critical testing gaps. Thus, DNT is a topic in which the application of Tox21 approaches would be particularly opportune and advantageous. For example, neuroinflammatory responses to air pollution have now been observed in human, animal, and in vitro studies (Costa et al. 2014); the results suggest the potential for contributions of Tox21 approaches that include the use of animal models and human tissues to assess DNT risks posed by air pollution and other exposures.

Tox21 approaches, including DNT assays, could also be used to address the challenges of identifying the air-pollution components that are contributing to neural disease. They could allow rapid testing of specific particle neurotoxicity and could help to identify markers of particle sources responsible for greater toxicity. For example, little is known about what PAHs are present in exposure mixtures; environmental samples can contain hundreds of individual parent or substituted PAHs, and bioactivity and toxicity of PAHs depend heavily on chemical structure (Wang et al. 2011). New methods could increase our understanding of the structure and toxicity relationships of neurobehavioral deficits if the full suite of chemicals present in samples could be identified and their individual or composite activities understood. Specifically, a suite of in vitro and high-throughput integrated systems could be used to classify PAHs by identifying their biological targets or pathways. Those systems could initially use untargeted global assessments—such as proteomics, metabolomics, transcriptomics, and epigenomics—to identify activity signatures for chemical classification and modeling. Recent studies in zebrafish, for example, evaluated and compared the developmental toxicity of 38 oxy-PAHs and revealed patterns of responses associated with PAH structural features (Knecht et al. 2013). In addition, full-genome RNA-sequence studies in zebrafish revealed that even for PAHs that produce toxicity through binding and activation of the AHR, subtle differences in PAH structure yield different overall developmental gene-expression changes and indicate that measuring P450 induction as a measure of AHR activation might be problematic (Goodale et al. 2015). Once targets of individual PAHs are identified, Tox21 approaches might be exploited further to predict how mixtures of PAH interact to produce neurotoxicity. In vitro functional assays of nervous-system development and function could be implemented to identify chemicals and mixtures that alter end points relevant to the nervous system. High-throughput integrated systems, such as zebrafish, might play a pivotal role in connecting identified molecular-response data with neurobehavioral measures (Truong et al. 2014; Reif et al. 2016). Optimization and scale up of assays that probe more complex behaviors in adult zebrafish (discussed in Chapter 3) should provide new avenues to link chemical exposures to functionally relevant neurobehavioral end points.

Despite enthusiasm for improving testing approaches and the emergence of new assays for DNT, implementation has been slow. For example, lack of assay coverage in EPA's ToxCast for neurotoxicity end points or neuronal targets is a well-recognized limitation. An initial attempt to use the ToxCast data to rank tested chemicals in terms of neurotoxicity failed because of poor assay coverage of suitable end points and low reliability of existing assays (Filer et al. 2014). Stakeholder meetings and workshops have helped to identify better ways to integrate emerging tools and approaches for DNT but require the inclusion of more neuroscientists and developmental endocrinologists to ensure that fundamental pathways in neurophysiology are evaluated and that sexual dimorphisms, region-specific sensitivity, and dynamic critical windows of exposure are considered in assay development (Crofton et al. 2014; McPartland et al. 2015). A battery of assays that incorporates the most up-to-date neuroscience tools and principles and that provides data relevant for regulatory science and risk-based decision platforms will be needed. Identifying and leveraging the most promising approaches and technologies will require active engagement of experts in disciplines outside traditional toxicology, especially the neurosciences. Accomplishing a multidisciplinary approach and encouraging a multidisciplinary research program for assay development and evaluation can be achieved by coordinating with relevant scientific societies and groups that have the needed expertise and with relevant funding agencies, such as the National Institute of Environmental Health Sciences.

How the adult human brain accomplishes complex cognitive and social processing remains mysterious and is the focus of intense research that is using a broad array of tools. Even less is known about when key aspects of the complex systems are organized in development or about how sexual dimorphisms emerge (Reinius and Jazin 2009; Yang and Shah 2014; Hawrylycz et al. 2015; Loke et al. 2015). The role of glia is also gaining substantial attention because these cells, particularly astrocytes and microglia, appear to play a more fundamental role in neural development than previously thought (Schwarz and Bilbo 2012; Schitine et al. 2015). Thus, assessments of neurodevelopmental consequences of chemical exposures must be undertaken with the understanding and acceptance of the fact that fundamental understanding about how the brain develops remains to be achieved, let alone how it enables us to engage in uniquely human behaviors and what contributes to the cognitive and social capacities that define our species. More research is needed on DNT, particularly given its critical consequences and society's high level of concern about its adverse effects. Addressing the challenges associated with DNT will require collaborative engagement of a broad array of disciplines, from neuroscientists who can address fundamental questions about the vulnerability of the brain to exogenous chemi-

cal exposures to population scientists who can assess the effects of chemical exposures in human populations.

## REFERENCES


Abid, Z., A. Roy, J.B. Herbstman, and A.S. Ettinger. 2014. Urinary polycyclic aromatic hydrocarbon metabolites and attention/deficit hyperactivity disorder, learning disability, and special education in US children aged 6 to 15. J. Environ. Public Health 2014:628508.

Allen, J.L., X. Liu, D. Weston, L. Prince, G. Oberdörster, N.J. Finkelstein, C.J. Johnston, and A. Cory-Slechta. 2014. Developmental exposure to concentrated ambient ultrafine particulate matter air pollution in mice results in persistent and sex-dependent behavioral neurotoxicity and glial activation. Toxicol. Sci. 140(1):160-178.

Amelung, J.T., R. Bührens, M. Beshay, and M.A. Reymond. 2010. Key genes in lung cancer translational research: A meta-analysis. Pathobiology 77(2):53-63.

Anderson, L.M., B.A. Diwan, N.T. Fear, and E. Roman. 2000. Critical windows of exposure for children's health: Cancer in human epidemiological studies and neoplasms in experimental animal models. Environ. Health Perspect. 108(Suppl. 3):573-594.

Aschner, M., K.M. Crofton, and E.D. Levin. 2010. Emerging high throughput and complementary model screens for neurotoxicology. Neurotoxicol. Teratol. 32(1):1-3.

Baccarelli, A., R.O. Wright, V. Bollati, L. Tarantini, A.A. Litonjua, H.H. Suh, A. Zanobetti, D. Sparrow, P.S. Vokonas, and J. Schwartz . 2009. Rapid DNA methylation changes after exposure to traffic particles. Am. J. Respir. Crit. Care Med. 179(7):572-578.

Backes, C.H., T. Nelin, M.W. Gorr, and L.E. Wold. 2013. Early life exposure to air pollution: How bad is it? Toxicol. Lett. 216(1):47-53.

Bal-Price, A., K.M. Crofton, M. Sachana, T.J. Shafer, M. Behl, A. Forsby, A. Hargreaves, B. Landesmann, P.J. Lein, J. Louisse, F. Monnet-Tschudi, A. Paini, A. Rolaki, A. Schrattenholz, C. Suñol, C. van Thriel, M. Whelan, and E. Fritsche. 2015. Putative adverse outcome pathways relevant to neurotoxicity. Crit. Rev. Toxicol. 45(1):83-91.

Beelen, R., O. Raaschou-Nielsen, M. Stafoggia, Z.J. Andersen, G. Weinmayr, B. Hoffmann, K. Wolf, E. Samoli, P. Fischer, M. Nieuwenhuijsen, P. Vineis, W. Xun, K. Katsouyanni, K. Dimakopoulou, A. Oudin, B. Forsberg, L. Modig, A.S. Havulinna, T. Lanki, A. Turunen, B. Oftedal, W. Nystad, P. Nafstad, U De Faire, N. Pedersen, C.G. Östenson, L. Fratiglioni, J. Pennell, M. Korek, G. Pershagen, K.T. Eriksen, K. Overvad, T. Ellermann, M. Eeftens, P. H. Peeters, L. Meliefste, M. Wang, B. Bueno-de-Mesquita, D. Sugiri, U. Krämer, J. Heinrich, L. de Hoogh, T. Key, A. Peters, R. Hampel, H. Concin, G.Nagel, A. Ineichen, E. Schaffner, N. Probst-Hensch, N. Künzli, C. Schindler, T. Schikowski, M. Adam, H. Phuleria, A. Vilier, F. Clavel-Chapelon, C. Declercq, S. Grioni, V. Krogh, M. Tsai, F. Ricceri, C. Sacerdote, C. Galassi, E. Migliore, A. Ranzi, G. Cesaroni. C. Badaloni, F. Forastiere, I. Tamayo, P. Amiano, M. Dorronsoro, M. Katsoulis, A. Trichopoulou, B. Brunekreef, and G. Hoek. 2014. Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. Lancet 383(9919):785-795.

Bind, M.A., B. Coull, H. Suh, R. Wright, A. Baccarelli, P. Vokonas, and J. Schwartz. 2014. A novel genetic score approach using instruments to investigate interactions between pathways and environment: Application to air pollution. PLoS One 9(4):e96000.

Blackburn, K., and S.B. Stuard. 2014. A framework to facilitate consistent characterization of read across uncertainty. Regul. Toxicol. Pharmacol. 68(3):353-362.

Calderón-Garcidueñas, L., B. Azzarelli, H. Acuna, R. Garcia, T.M. Gambling, N. Osnaya, S. Monroy, M. del Rosario Tizapantzi, J.L. Carson, A. Villarreal-Calderon, and B. Rewcastle. 2002. Air pollution and brain damage. Toxicol. Pathol. 30(3):373-389.

Calderón-Garciduenas, L., R. Torres-Jardón, R.J. Kulesza, S. Park, and A. D'Angiulli. 2014. Air pollution and detrimental effects on children's brain. The need for a multidisciplinary approach to the issue complexity and challenges. Front Hum. Neurosci. 8:613.

Chang, S.Y., W. Vizuete, A. Valencia, B. Naess, V. Isakov, T. Palma, M. Breen, and S. Arunachalam. 2015. A modeling framework for characterizing near-road air pollutant concentration at community scales. Sci. Total Environ. 538:905-921.

Chapin, R.E., J. Delaney, Y. Wang, L. Lanning, B. Davis, B. Collins, N. Mintz, and G. Wolfe. 1999. The effects of 4-nonylphenol in rats: A multigeneration reproduction study. Toxicol. Sci. 52(1):80-91.

Chen, J.C.., X. Wang, G.A. Wellenius, M.L. Serre, I. Driscoll, R. Casanova, J.J. McArdle, J.E. Manson, H.C. Chui, and M.A. Espeland. 2015. Ambient air pollution and neurotoxicity on brain structure: Evidence from Women's Health Initiative Memory Study. Ann. Neurol. 78(3):466-476.

Chiu, Y.H., D.C. Bellinger, B.A. Coull, S. Anderson, R. Barber, R.O. Wright, and R.J. Wright. 2013. Associations between traffic-related black carbon exposure and attention in a prospective birth cohort of urban children. Environ Health Perspect. 121(7):859-864.

Coker, E., J. Ghosh, M. Jerrett, V. Gomez-Rubio, B. Beckerman, M. Cockburn, S. Liverani, J. Su, A. Li, M.L. Kile, B. Ritz, and J. Molitor. 2015. Modeling spatial effects of PM(2.5) on term low birth weight in Los Angeles County. Environ Res. 142:354-364.

Coker, E., S. Liverani, J.K. Ghosh, M. Jerrett, B. Beckerman, A. Li, B. Ritz, and J. Molitor. 2016. Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. Environ. Int. 91:1-13.

Costa, L.G., T.B. Cole, J. Coburn, Y.C. Chang, K. Dao, and P. Roque. 2014. Neurotoxicants are in the air: Convergence of human, animal, and in vitro studies on the effects of air pollution on the brain. Biomed. Res. Int. 2014:736385.

Crofton, K., E. Fritsche, T. Ylikomi, and A. Bal-Price. 2014. International Stakeholder NETwork (ISTNET) for creating a developmental neurotoxicity testing (DNT) roadmap for regulatory process. ALTEX 31(2):223-224.

Demetriou, C.A., and P. Vineis. 2015. Carcinogenicity of ambient air pollution: Use of biomarkers, lessons learnt and future directions. J. Thorac. Dis. 7(1):67-95.

Demetriou, C.A., O. Raaschou-Nielsen, S. Loft, P. Møller, R. Vermeulen, D. Palli, M. Chadeau-Hyam, W.W. Xun, and P. Vineis. 2012. Biomarkers of ambient air pollution and lung cancer: A systematic review. Occup. Environ. Med. 69(9):619-627.

Dockery, D.W., C.A. Pope, III, X. Xu, J.D. Spengler, J.H. Ware, M.E. Fay, B.G. Ferris, Jr., and F.E. Speizer. 1993. An association between air pollution and mortality in six US cities. N. Engl. J. Med. 329(24):1753-1759.

Edwards, S.C., W. Jedrychowski, M. Butscher, D. Camann, A. Kieltyka, E. Mroz, E. Flak, Z. Li, S. Wang, V. Rauh, and F. Perera. 2010. Prenatal exposure to airborne polycyclic aromatic hydrocarbons and children's intelligence at 5 years of age in a prospective cohort study in Poland. Environ. Health Perspect. 118(9):1326-1331.

EPA (US Environmental Protection Agency). 2009. Integrated Science Assessment for Particulate Matter (Final Report). EPA/600/R-08/139F. US Environmental Protection Agency, Washington, DC [online]. Available: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=216546 [accessed July 25, 2016].

EPA (US Environmental Protection Agency). 2011. Estimation Programs Interface (EPI) Suite for Microsoft® Windows, Version 4.1. U. S. Environmental Protection Agency, Washington, DC.

Fagundes, L.S., A.daS. Fleck, A.C. Zanchi, P.H. Saldiva, and C.R. Rhoden. 2015. Direct contact with particulate matter increases oxidative stress in different brain structures. Inhal. Toxicol. 27(10):462-467.

Fasanelli, F., L. Baglietto, E. Ponzi, F. Guida, G. Campanella, M. Johansson, K. Grankvist, M. Johansson, M.B. Assumma, A. Naccarati, M. Chadeau-Hyam, U. Ala, C. Faltus, R. Kaaks, A. Risch, B. De Stavola, A. Hodge, G.G. Giles, M.C. Southey, C.L. Relton, P.C. Haycock, E. Lund, S. Polidoro, T.M. Sandanger, G. Severi, and P. Vineis. 2015. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat. Commun. 6:10192.

Felter, S.P., G.P. Daston, S.Y. Euling, A.H. Piersma, and M.S. Tassinari. 2015. Assessment of health risks resulting from early-life exposures: Are current chemical toxicity testing protocols and risk assessment methods adequate? Crit. Rev. Toxicol. 45(3):219-244.

Filer, D., H.B. Patisaul, T. Schug, D. Reif, and K. Thayer. 2014. Test driving ToxCast: Endocrine profiling for 1858 chemicals included in phase II. Curr. Opin. Pharmacol. 19:145-152.

Goodale, B.C., J. La Du, S.C. Tilton, C.M. Sullivan, W.H. Bisson, K.M. Waters, and R.L. Tanguay. 2015. Ligand-specific transcriptional mechanisms underlie aryl hydrocarbon receptor-mediated developmental toxicity of oxygenated PAHs. Toxicol. Sci. 147(2):397-411

Grandjean, P. and P.J. Landrigan. 2006. Developmental neurotoxicity of industrial chemicals. Lancet 368(9553):2167-2178.

Guida, F., T.M. Sandanger, R. Castagné, G. Campanella, S. Polidoro, D. Palli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, G. Severi, S.A. Kyrtopoulos, P. Georgiadis, R.C. Vermeulen, E. Lund, P. Vineis, and M. Chadeau-Hyam. 2015. Dynamics of smoking-induced genome-wide methylation changes with time since smoking. Hum. Mol. Genet. 24(8):2349-2359.

Guxens, M., and J. Sunyer. 2012. A review of epidemiological studies on neuropsychological effects of air pollution. Swiss Med. Wkly. 141:w13322.

Guxens, M., R. Garcia-Esteban, L. Giorgis-Allemand, J. Forns, C. Badaloni, F. Ballester, R. Beelen, G. Cesaroni, L. Chatzi, M. de Agostini, A. de Nazelle, M. Eeftens, M.F. Fernandez, A. Fernández-Somoano, F. Forastiere, U. Gehring, A. Ghassabian, B. Heude, V.W. Jaddoe, C. Klümper, M. Kogevinas, U. Krämer, B. Larroque, A. Lertxundi, N. Lertxuni, M. Murcia, V. Navel, M. Nieuwenhuijsen, D. Porta, R. Ramos, T. Roumeliotaki, R. Slama, M. Sørensen, E.G. Stephanou, D. Sugiri, A. Tardón, H. Tiemeier, C.M. Tiesler, F.C. Verhulst, T. Vrijkotte, M. Wilhelm, B. Brunekreef, G. Pershagen, and J. Sunyer. 2014. Air pollution during pregnancy and childhood cognitive and psychomotor development: Six European birth cohorts. Epidemiology 25(5):636-647.

Hamra, G.B., N. Guha, A. Cohen, F. Laden, O. Raaschou-Nielsen, J. Samet, P. Vineis, F. Forastiere, P. Saldiva, T. Yorifuki, and D. Loomis. 2014. Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis. Environ. Health Perspect. 122(9):906-911.

Hawrylycz, M., J. Miller, V. Menon, D. Feng, T. Dolbeare, A.L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C. Lee, A. Bernard, M.F. Glasser, D.L. Dierker, J. Menche, A. Szafer, F. Collman, P. Grange, K.A. Berman, S. Mihalas, Z. Yao, L. Stewart, A. Barabási, J. Schulkin, J. Phillips, L. Ng, C. Dang, D.R. Haynor, A. Jones, D.C. Van Essen, C. Koch, and E. Lein. 2015. Canonical genetic signatures of the adult human brain. Nat. Neurosci. 18(12):1832-1844.

Hossaini, A., M. Dalgaard, A.M. Vinggaard, P. Pakarinen, and J.J. Larsen. 2003. Male reproductive effects of octylphenol and estradiol in Fischer and Wistar rats. Reprod. Toxicol. 17(5):607-615.

Hudda, N.T. Gould, K. Hartin, T.V. Larson, and S.A. Fruin. 2014. Emissions from an international airport increase

particle number concentrations 4-fold at 10 km downwind. Environ. Sci. Technol. 48(12):6628-6635.

IARC (International Agency for Research on Cancer). 2015. Outdoor Air Pollution. Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 109. Lyon: IARC [online]. Available: http://monographs.iarc.fr/ENG/Monographs/vol109/index.php [accessed July 25, 2016].

Janssen, B.G., H.M. Byun, W. Gyselaers, W. Lefebvre, A.A. Baccarelli, and T.S. Nawrot. 2015. Placental mitochondrial methylation and exposure to airborne particulate matter in the early life environment: An ENVIRONAGE birth cohort study. Epigenetics 10(6):536-544.

Joubert, B.R., S.E. Håberg, R.M. Nilsen, X. Wang, S.E Vollset, S.K. Murphy, Z. Huang, C. Hoyo, Ø. Midttun, L.A. Cupul-Uicab, P.M. Ueland, M.C. Wu, W. Nystad, D.A. Bell, S.D. Peddada, and S.J. London. 2012. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ. Health Perspect.120(10):1425-1431.

Kachuri, L., C.I. Amos, J.D. McKay, M. Johansson, P. Vineis, H.B. Bueno-de-Mesquita, M.C. Boutron-Ruault, M. Johansson, J.R. Quirós, S. Sieri, R.C. Travis, E. Weiderpass, L. Le Marchand, B.E. Henderson, L. Wilkens, G. Goodman, C. Chen, J.A. Doherty, D.C. Christiani, Y. Wei, L. Su, S. Tworoger, X. Zhang, P. Kraft, D. Zaridze, J.K. Field, M.W. Marcus, M.P. Davies, R. Hyde, N.E. Caporaso, M.T. Landi, G. Severi, G.G. Giles, G. Liu, J.R. McLaughlin, Y. Li, X. Xiao, G. Fehringer, X. Zong, R.E. Denroche, P.C. Zuzarte, J.D. McPherson, P. Brennan, and R.J. Hung. 2016. Fine-mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. Carcinogenesis 37(1):96-105.

Kelce, W.R., and L.E. Gray, Jr.1997. Endocrine disruptors: Effects on sex steroid hormone receptors and sex development. Pp. 435-474 in Drug Toxicity in Embryonic Development, Vol. 2, R.J. Kavlock, and G.P. Daston, eds. Berlin: Springer.

Kelly, F.J., and J.C. Fussell. 2015. Linking ambient particulate matter pollution effects with oxidative biology and immune responses. Ann. N.Y. Acad. Sci. 1340:84-94.

Kim, E., H. Park, Y.C. Hong, M. Ha, Y. Kim, B.N. Kim, Y. Kim, Y.M. Roh, B.E. Lee, J.M. Ryu, B.M. Kim, and E.H. Ha. 2014. Prenatal exposure to $PM_{2.5}$ and $NO_2$ and children's neurodevelopment from birth to 24 months of age: Mothers and Children's Environmental Health (MOCEH) study. Sci. Total Environ. 481:439-445.

Knecht, A.L., B.C. Goodale, L. Truong, M.T. Simonich, A.J. Swanson, M.M. Matzke, K.A. Anderson, K.M. Waters, and R.L. Tanguay. 2013. Comparative developmental toxicity of environmentally relevant oxygenated PAHs. Toxicol. Appl. Pharmacol. 271(2):266-275.

Künzli, N., and I.B. Tager. 1997. The semi-individual study in air pollution epidemiology: A valid design as compared to ecologic studies. Environ. Health Perspect. 105(10):1078-1083.

Laws, S.C., S.A. Carey, J.M. Ferrell, G.J. Bodman, and R.L. Cooper. 2000. Estrogenic activity of octylphenol, nonylphenol, bisphenol a and methoxychlor in rats. Toxicol. Sci. 54(1):154-167.

Lin, C.C., S.K. Yang, K.C. Lin, W.C. Ho, W.S. Hsieh, B.C. Shu, and P.C. Chen. 2014. Multilevel analysis of air pollution and early childhood neurobehavioral development. Int. J. Environ. Res. Public Health 11(7):6827-6841.

Liu, F., Y. Huang, F. Zhang, Q. Chen, B. Wu, W. Rui, J.C. Zheng, and W. Ding..2015. Macrophages treated with particulate matter $PM_{2.5}$ induce selective neurotoxicity through glutaminase-mediated glutamate generation. J. Neurochem. 134(2):315-326.

Loke, H., V. Harley, and J. Lee. 2015. Biological factors underlying sex differences in neurological disorders. Int. J. Biochem. Cell. Biol. 65:139-150.

Lovasi, G.S., N. Eldred-Skemp, J.W. Quinn, H.W. Chang, V.A. Rauh, A. Rundle, M.A. Orjuela, and F.P. Perera. 2014. Neighborhood social context and individual polycyclic aromatic hydrocarbon exposures associated with child cognitive test scores. J. Child Fam. Stud. 23(5):785-799.

McCallister, M.M., Z. Li, T. Zhang, A. Ramesh, R.S. Clark, M. Maguire, B. Hutsell, M.C. Newland, and D.B. Hood. 2016. Revealing behavioral learning deficit phenotypes subsequent to in utero exposure to benzo(a)pyrene. Toxicol Sci. 149(1):42-54.

McPartland, J., H.C. Dantzker, and C.J. Portier. 2015. Building a robust 21st century chemical testing program at the US Environmental Protection Agency: Recommendations for strengthening scientific engagement. Environ. Health Perspect. 123(1):1-5.

Mikkilä, T.F., J. Toppari, and J. Paranko. 2006. Effects of neonatal exposure to 4-tert-octylphenol, diethylstilbestrol, and flutamide on steroidogenesis in infantile rat testis. Toxicol. Sci. 91(2):456-466.

Mostafavi, N., J. Vlaanderen, M. Chadeau-Hyam, R. Beelen, L. Modig, D. Palli, I.A. Bergdahl, P. Vineis, G. Hoek, S.A. Kyrtopoulos, and R. Vermeulen. 2015. Inflammatory markers in relation to long-term air pollution. Environ. Int. 81:1-7.

Nieuwenhuijsen, M.J., D. Donaire-Gonzalez, I. Rivas, M. de Castro, M. Cirach, G. Hoek, E. Seto, M. Jerrett, and J. Sunyer. 2015. Variability in and agreement between modeled and personal continuously measured black carbon levels using novel smartphone and sensor technologies. Environ. Sci. Technol. 49(5):2977-2982.

Nik-Zainal, S., J.E. Kucab, S. Morganella, D. Glodzik, L.B. Alexandrov, V.M. Arlt, A. Weninger, M. Hollstein, M.R. Stratton, and D.H. Phillips. 2015. The genome as a record of environmental exposure. Mutagenesis 30(6):763-770.

Novakovic, B., J. Ryan, N. Pereira, B. Boughton, J.M. Craig, and R. Saffery. 2014. Postnatal stability, tissue, and time

specific effects of AHRR methylation change in response to maternal smoking in pregnancy. Epigenetics 9(3):377-386.

NRC (National Research Council). 1993. Pesticides in Diet of Infants and Children. Washington, DC: National Academy Press.

NRC (National Research Council). 2012. Exposure Science in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

Perera, F.P., D. Tang, Y.H. Tu, L.A. Cruz, M. Borjas, T. Bernert, and R.M. Whyatt. 2004. Biomarkers in maternal and newborn blood indicate heightened fetal susceptibility to procarcinogenic DNA damage. Environ. Health Perspect. 112(10):1133-1136.

Perera, F.P., V. Rauh, R.M. Whyatt, W.Y. Tsai, D. Tang, D. Diaz, L. Hoepner, D. Barr, Y.H. Tu, D. Camann, and P. Kinney. 2006. Effects of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. Environ. Health Perspect. 114(8):1287-1292.

Perera, F.P., Z. Li, R. Whyatt, L. Hoepner, S. Wang, D. Camann, and V. Rauh. 2009. Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. Pediatrics 124(2):e195-e202.

Proietti, E., M. Röösli, U. Frey, and P. Latzin. 2013. Air pollution during pregnancy and neonatal outcome: A review. J. Aerosol. Med. Pulm. Drug Deliv. 26(1):9-23.

Raaschou-Nielsen, O., Z.J. Andersen, R. Beelen, E. Samoli, M. Stafoggia, G. Weinmayr, B. Hoffman, P. Fischer, M.J. Nieuwenhuijsen, B. Brunekreef, W.W. Xun, K. Katsouyanni, L. Dimakopoulou, J. Sommar, B. Forsberg, L. Modig, A. Oudin, B. Oftedal, P.E. Schwarze, P. Nafstad, U. De Faire, N.L. Pedersen, C.G. Ostenson, L. Fratiglioni, J. Penell, M. Korek, G. Pershagen, K.T. Eriksen, M. Sørensen, A. Tjønneland, T. Ellerman, M. Eeftens, P.H. Peeters, K. Meliefste, M. Wang, B. Bueno-de-Mesquita, T.J. Key, K. de Hoogh, H. Concin, G. Nagel, A. Vilier, S. Grioni, V. Krogh, M.Y. Tsai, F. Ricceri, C. Sacerdote, C. Galassi, E. Migliore, A. Ranzi, G. Cesaroni, C. Badaloni, F. Forastiere, I. Tamayo, P. Amiano, M. Dorronsoro, A. Trichopoulou, C. Bamia, P. Vineis, and G. Hoek. 2013. Air pollution and lung cancer incidence in 17 European cohorts: Prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol. 14(9):813-822.

Raaschou-Nielsen, O., R. Beelen, M. Wang, G. Hoek, Z.J. Andersen, B. Hoffmann, M. Stafoggia, E. Samoli, G. Weinmayr, K. Dimakopoulou, M. Nieuwenhuijsen, W.W. Xun, P. Fischer, K.T. Eriksen, M. Sørensen, A. Tjønneland, F. Ricceri, K. de Hoogh, T. Key, M. Eeftens, P.H. Peeters, H.B. Bueno-de-Mesquita, K. Meliefste, B. Oftedal, P.E. Schwarze, P. Nafstad, C. Galassi, E. Migliore, A. Ranzi, G. Cesaroni, C. Badaloni, F. Forastiere, J. Penell, U. De Faire, M. Korek, N. Pedersen, C.G. Östenson, G. Pershagen, L. Fratiglioni, H. Concin, G. Nagel, A. Jaensch, A. Ineichen, A. Naccarati, M. Katsoulis, A. Trichpoulou, M. Keuken, A. Jedynska, I.M. Kooter, J. Kukkonen, B. Brunekreef, R.S. Sokhi, K. Katsouyanni, and P. Vineis. 2016. Particulate matter air pollution components and risk for lung cancer. Environ. Int. 87:66-73.

Reif, D.M., L. Truong, D. Mandrell, S. Marvel, G. Zhang, and R.L. Tanguay. 2016. High-throughput characterization of chemical-associated embryonic behavioral changes predicts teratogenic outcomes. Arch. Toxicol. 90(6):1459-1470.

Reinius, B., and E. Jazin. 2009. Prenatal sex differences in the human brain. Mol. Psychiatry 14(11):987-989.

Saenen, N.D., M. Plusquin, E. Bijnens, B.G. Janssen, W. Gyselaers, B. Cox, F. Fierens, G. Molenberghs, J. Penders, K. Vrijens, P. De Boever, and T.S. Nawrot. 2015. In utero fine particle air pollution and placental expression of genes in the brain-derived neurotrophic factor signaling pathway: An ENVIRONAGE Birth Cohort Study. Environ Health Perspect. 123(8):834-840.

Schitine, C., L. Nogaroli, M.R. Costa, and C. Hedin-Pereira. 2015. Astrocyte heterogeneity in the brain: From development to disease. Front. Cell. Neurosci. 9:76.

Schwarz, J.M., and S.D. Bilbo. 2012. Sex, glia, and development: Interactions in health and disease. Horm. Behav. 62(3):243-253.

Shenker, N.S., S. Polidoro, K. van Veldhoven, C. Sacerdote, F. Ricceri, M.A. Birrell, M.G. Belvisi, R. Brown, P. Vineis, and J.M. Flanagan. 2013. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum. Mol. Genet. 22 (5):843-851.

Smarr, M.M., F. Vadillo-Ortega, M. Castillo-Castrejon, and M.S. O'Neill. 2013. The use of ultrasound measurements in environmental epidemiological studies of air pollution and fetal growth. Curr. Opin. Pediatr. 25(2):240-246.

Smith, M.T., K.Z. Guyton, C.F. Gibbons, J.M. Fritz, C.J. Portier, I. Rusyn, D.M. DeMarini, J.C. Caldwell, R.J. Kavlock, P. Lambert, S.S. Hecht, J.R. Bucher, B.W. Stewart, R. Baan, V.J. Cogliano, and K. Straif. 2016. Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environ. Health Perspect. 124(6):713-721.

Stieb, D.M., L. Chen, M. Eshoul, and S. Judek. 2012. Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. Environ. Res. 117:100-111.

Suades-González, E., M. Gascon, M. Guxens, and J. Sunyer. 2015. Air pollution and neuropsychological development: A review of the latest evidence. Endocrinology 156(10):3473-3482.

Suglia, S.F., A. Gryparis, R.O. Wright, J. Schwartz, and R.J. Wright. 2008. Association of black carbon with cognition among children in a prospective birth cohort study. Am. J. Epidemiol. 167(3):280-286.

Takeda, K., N. Tsukue, and S. Yoshida. 2004. Endocrine-disrupting activity of chemicals in diesel exhaust and diesel exhaust particles. Environ. Sci. 11(1):33-45.

Tang, D., T. Li, J.J. Liu. Z. Zhou, T. Yuan, Y. Chen, V.A. Rauh, J. Xie, and F. Perera. 2008. Effects of prenatal exposure to coal-burning pollutants on children's development in China. Environ. Health Perspect. 116(5):674-679.

Tang, D., T.Y. Li, J.C. Chow, S.U. Kulkarni, J.G. Watson, S.S. Ho, Z.Y. Quan, L.R. Qu, and F. Perera. 2014. Air pollution effects on fetal and child development: A cohort comparison in China. Environ. Pollut. 185:90-96.

Truong, L., D. Reif, L. St. Mary, M. Geier, H.D. Truong, and R.L. Tanguay. 2014. Multidimensional in vivo hazard assessment using zebrafish. Toxicol. Sci. 137(1):212-233.

Tyl, R.W., C.B. Myers, M.C. Marr, D.R. Brine, P.A. Fail, J.C. Seely, and J.P. Van Miller. 1999. Two-generation reproduction study with para-tert-octylphenol in rats. Regul. Toxicol. Pharmacol. 30(2 Pt 1):81-95.

van den Hooven, E.H., F.H. Pierik, Y. de Kluizenaar, S.P. Willemsen, A. Hofman, S.W. van Ratingen, P.Y. Zandveld, J.P. Mackenbach, E.A. Steegers, H.M. Miedema, and V.W. Jaddoe. 2012. Air pollution exposure during pregnancy, ultrasound measures of fetal growth, and adverse birth outcomes: A prospective cohort study. Environ Health Perspect. 120(1):150-156.

Vineis, P., A. Schatzkin, and J.D. Potter. 2010. Models of carcinogenesis: An overview. Carcinogenesis 31(10):1703-1709.

Vineis, P., K. van Veldhoven, M. Chadeau-Hyam, and T.J. Athersuch. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology. Environ. Mol. Mutagen. 54(7):461-467.

Vrijheid, M., D. Martinez, S. Manzanares, P. Dadvand, A. Schembari, J. Rankin, and M. Nieuwenhuijsen. 2011. Ambient air pollution and risk of congenital anomalies: A systematic review and meta-analysis. Environ. Health Perspect. 119(5):598-606.

Wang, W., N. Jariyasopit, J. Schrlau, Y. Jia, S. Tao, T.W. Yu, R.H. Dashwood, W. Zhang, X. Wang, and S.L. Simonich. 2011. Concentration and photochemistry of PAHs, NPAHs, and OPAHs and toxicity of $PM_{2.5}$ during the Beijing Olympic Games. Environ. Sci. Technol. 45(16):6887-6895.

WHO (World Health Organization). 1986. Principles for Evaluating Health Risks from Chemicals during Infancy and Early Childhood: The Need for a Special Approach. Environmental Health Criteria 59. Geneva: World Health Organization.

Wittkopp, S., N. Staimer, T. Tjoa, T. Stinchcombe, N. Daher, J.J. Schauer, M.M. Shafer, C. Sioutas, D.L. Gillen, and R.J. Delfino. 2016. Nrf2-related gene expression and exposure to traffic-related air pollution in elderly subjects with cardiovascular disease: An exploratory panel study. J. Expo. Sci. Environ. Epidemiol. 16(2):141-149.

Wu, S., K. Blackburn, J. Amburgey, J. Jaworska, and T. Federle. 2010. A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. Regul. Toxicol. Pharmacol. 56(1):67-81.

Wu, S., J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, and K. Blackburn. 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem. Res. Toxicol. 26(12):1840-1861.

Yang, C.F., and N.M. Shah. 2014. Representing sex in the brain, one module at a time. Neuron. 82(2):261-278.

Yang, A., M. Wang, M. Eeftens, R. Beelen, E. Dons, D.L. Leseman, B. Brunekreef, F.R. Cassee, N.A. Janssen, and G. Hoek. 2015. Spatial variation and land use regression modeling of the oxidative potential of fine particles. Environ. Health Perspect. 123(11):1187-1192.

Yorifuji, T., S. Kashima, M. Higa Diez, Y. Kado, S. Sanada, and H. Doi. 2016. Prenatal exposure to traffic-related air pollution and child behavioral development milestone delays in Japan. Epidemiology 27(1):57-65.

# C

# Case Studies on Site-Specific Assessments

As discussed in Chapter 5, understanding the risk associated with a spill or a hazardous waste site requires identifying and quantifying the chemicals present, characterizing the chemical toxicity, and estimating the mixture toxicity and associated risk. This appendix provides a case study related to each element. The first case study describes approaches for refining exposure estimates for known chemicals at a hypothetical site and approaches for identifying the uncharacterized chemicals at the site. The second addresses the generation of toxicity data and exposure information on a data-poor chemical after its accidental release. The third explores a biological read-across approach for assessing mixtures at a hypothetical site.

## IDENTIFYING CHEMICALS AT A SITE

For this case study, the setting is a large, historically contaminated site that comprises land and surface water near a major population center (for example, Love Canal, the Portland Harbor, or the Houston Ship Channel). Recent site characterization has produced an extensive set of environmental monitoring data for air, water, and soil at the site. The data cover multiple times and are geographically distributed throughout the site. Biomonitoring data are available from serum, urine, and hair in a representative sample of people who live and work in the area surrounding the site. The biomonitoring data are geographically distributed but in some cases limited to single times.

Targeted analytical chemistry produced concentration data on about 50 toxicologically well-characterized chemicals in environmental media and human blood, urine, and hair (see Table C-1). The chemicals represent four major chemical classes: polycyclic aromatic hydrocarbons, industrial chemicals and solvents, plasticizers, and pesticides. Information on metabolism and pharmacokinetics of many of the chemicals in rodents and humans is available. Assessments of external exposure of the population around the site (children, adults, and senior adults) to each chemical by the oral, dermal, and inhalation routes, where appropriate, have been conducted. Nontargeted analyses of the same environmental and biomonitoring samples revealed 5,000 unidentified substances in the environmental media, 3,000 in serum, 2,000 in urine, and 800 in hair; 300 of the unidentified analytes are common to the environmental media and all biomonitoring samples (see Figure C-1).

For this case study, the tasks become refining exposure assessment of the known chemicals, translating the external-exposure predictions into internal-exposure predictions, and identifying the unknown chemicals at the site. The following sections explore those various tasks.

### Assessment of Known Chemicals and Chemical Mixtures

The initial step in this case study would be to assemble existing exposure data on the identified (known) chemicals and refine their exposure estimates for testing. The relative composition, variability, and concentration ranges of the chemicals in the various media would be assessed and quantified, taking into account the exposure routes of interest. For example, testing designed to evaluate risks associated with dermal exposures might focus on concentrations of chemicals in soil, water, and air that would come into contact with skin. Similarly, mixtures that should be evaluated for inhalation toxicity in portal-of-entry tissues (lung tissue) might best be defined by air concentrations of mixture components. Alternatively, oral exposures for toxicity testing could initially be defined by the composition and concentrations of components of soil and water or other media that might be ingested and absorbed in sufficient amounts to influence total exposure.

Once exposures have been defined, the task is to translate exposures from external measures to internal predictions to appropriate concentrations for in vitro testing with pharmacokinetic models or measurements obtained from biomonitoring. The accuracy of the model estimates will be determined partly by the amount of information available on absorption, distribution, metabolism, and excretion (ADME) processes. Cheminformatic and high-throughput systems can provide information on, for example, metabolism by hepatocytes, absorption by caco-2 cells, and binding to plasma proteins that could be used to estimate pharmacokinetic parameters (Wetmore

**TABLE C-1** Site-Specific Chemicals Identified by Targeted Chemistry Analysis

| Class | Rank[a] | Chemical Name |
|---|---|---|
| Polycyclic aromatic hydrocarbons | 10 | Benzo(B)fluoranthene |
| | 38 | Benzo(A)anthracene |
| | 80 | Naphthalene |
| | 138 | Fluoranthene |
| | 168 | Acenaphthene |
| | 185 | Dibenzofuran |
| | 255 | Pyrene |
| High-production-volume industrial chemicals | 30 | Benzidine |
| | 54 | Pentachlorophenol |
| | 84 | 2,4,6-Trichlorophenol |
| | 98 | 2,4-Dinitrotoluene |
| | 101 | 4,6-Dinitro-*o*-cresol |
| | 137 | 1,2,3-Trichlorobenzene |
| | 142 | 2,4,5-Trichlorophenol |
| | 172 | Cresol, *para*- |
| | 181 | Phenol |
| | 195 | Cresol, *ortho*- |
| | 206 | *n*-Nitrosodiphenylamine |
| | 260 | 2,6-Dinitrotoluene |
| Plasticizers | 58 | Di-*n*-butyl phthalate |
| | 77 | Di(2-ethylhexyl)phthalate |
| | 266 | Bis(2-ethylhexyl)adipate |
| Pesticides | 13 | DDT, P,P'- |
| | 18 | Dieldrin |
| | 25 | Aldrin |
| | 26 | DDD, P,P'- |
| | 28 | Heptachlor |
| | 34 | γ-Hexachlorocyclohexane |
| | 37 | Disulfuron |
| | 40 | Endrin |
| | 41 | Diazinon |
| | 44 | Endosulfan |
| | 47 | Heptachlor epoxide |
| | 53 | DDT, O,P'- |
| | 55 | Methoxychlor |
| | 65 | Chlorpyriphos |
| | 89 | 2,4-Dinitrophenol |
| | 99 | Ethion |
| | 103 | Dimethylarsinic acid |
| | 131 | Azinphos-methyl |
| | 144 | Dicofol |
| | 148 | Parathion |
| | 155 | Trifluralin |
| | 166 | Phorate |
| | 200 | Ethoprop |
| | 232 | Dimethoate |
| | 244 | 2,4-D Acid |
| | 246 | Butylate |
| | 250 | Diuron |
| | 269 | Metolachlor |
| | 272 | Carbaryl |

[a]Rank is from the ATSDR 2015 Substance Priority List in which rank is based on frequency, toxicity, and potential for human exposure at Superfund sites.

et al. 2012, 2014). Genetic analysis of single-nucleotide polymorphisms related to human pharmacokinetics could provide information on variability in pharmacokinetic parameters in the population of concern. Ultimately, the pharmacokinetic parameters are helpful for evaluating the relationships between external and internal exposures and guiding selection of test concentrations. The data on individual chemical and mixture exposure and the related pharmacokinetic data would ideally be used to establish test concentrations or exposures for the appropriate in vivo or in vitro test systems that reflect the composition of real-world exposures at the site.

### Assessment of Chemicals of Unknown Identity

Nontargeted analyses of samples from the site revealed 5,000 unidentified chemicals in the environmental media, 3,000 in serum, 2,000 in urine, and 800 in hair (see Figure C-1). All sample types had 300 unidentified chemicals in common. One key challenge in nontargeted analysis of complex samples is to identify the unidentified chemicals accurately. Without chemical identifications, the ability to quantify exposure, conduct toxicity testing, and evaluate the plausibility of exposure–disease associations is extremely limited. To identify unknowns, standard reference materials for industrial and other chemicals and their metabolites are needed. Analytical features of the standard reference materials—such as elution time, exact mass, isotopic signature, and fragmentation pattern from, for example, gas chromatography (GC), liquid chromatography (LC), and tandem mass spectrometry (MS/MS)—can be matched to analytical features in the sample to identify the chemicals of interest. Chemical-identity libraries that contain the analytical spectra of reference standards are growing, particularly for endogenous metabolites (for example, the Human Metabolome Database, HMD), but more progress needs to be made before nontargeted analyses can become routine. The following discussion provides approaches for making progress in this field.

Two general approaches—an experimentally driven approach and another driven by cheminformatics (Horai et al. 2009; Neumann and Bocker 2010)—have been suggested to overcome the obstacles presented by the lack of chemical-identity libraries. In the experimentally driven approach, chemical-identity libraries similar to the HMD that include exact mass, elution times, isotopic signature, and mass fragmentation patterns (see Figure C-2) could be created for ToxCast and other chemicals. To support that effort, the US Environmental Protection Agency (EPA) has obtained authentic chemical standards for thousands of ToxCast chemicals and placed them in a chemical repository. Development of a complete chemical-identity library for the ToxCast chemicals (and addition of this information to such databases as the HMD) would enable measurements of these chemicals in environmental media and human biofluids. However, a major limitation in the experimental approach is the absence of standards for common environmental degradation products or metabolites that are likely to be found in biofluids. As chemical-identity libraries grow, archived GC, LC-MS, or MS/MS spectra can be searched to make new identifications.

Nuclear magnetic resonance (NMR) methods present another experimental approach to identification of unknown chemical features. The methods hold great promise because NMR analysis allows identification and quantitation of chemicals without an authentic standard. A noted limitation of the approach is its need for relatively high concentrations of target chemicals in the sample (1 μM; Bingol and Brüschweiler 2015) and its relatively low throughput. Advanced labeling techniques (Clendinen et al. 2015) and methods that involve combinations of NMR, MS, and other analytical techniques, however,

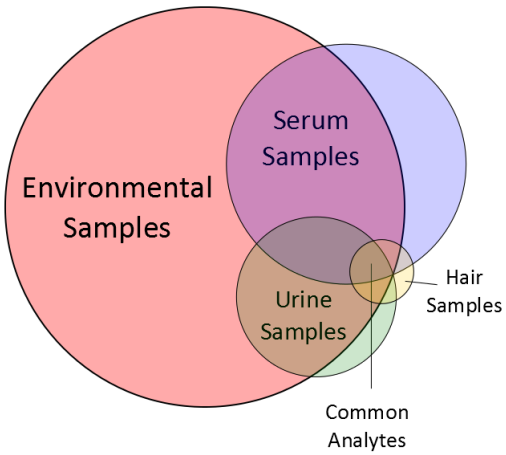



**FIGURE C-1** Hypothetical distribution of unidentified analytes in environmental media and biomonitoring samples. Analysis revealed a total of 5,000 analytes in the environmental samples, 3,000 in the serum samples, 2,000 in the urine samples, and 800 in the hair samples. The four sample types had 300 analytes in common.

show promise for future applications (Bingol and Brüschweiler 2015).

Ion-mobility spectrometry–mass spectrometry (IMS-MS) analysis is another promising experimental approach for library-building and rapid identification of chemical features of unknowns (Ewing et al. 2016; May et al. 2016). In IMS-MS analyses, chemicals separate on the basis of their collisional cross-sectional (CCS) area during flow through a nitrogen- or helium-filled tube with a charge separation. Separation times are in the milliseconds and allow the potential for very high-throughput sample analysis. One potential advantage of IMS-MS over other analytical approaches for chemical identification for which authentic standards do not exist is that the CCS area can be calculated in silico with good accuracy (2–5% error; Paglia et al. 2014). The high throughput of the IMS-MS techniques and the possibilities of in silico library-building could produce large libraries of known chemicals, metabolites, and degradation products even if the chemical standards are not available. Those libraries could then be used to assign provisional identifications or identifications with probability statements. Furthermore, IMS-MS chemical fragmentation patterns can be matched to those in existing databases, such as the HMD, for improved chemical identification.

The other general approach is based on cheminformatics and can circumvent the challenges associated with limited chemical-identity libraries and the lack of standard reference materials. Applied in concert with emerging analytical chemistry approaches and computational methods, cheminformatics holds great potential for rapid identification or classification of unknown analytes. For example, quantitative structure–activity relationship methods that compare chromatographic behaviors of unknown analytes could be combined with other data to provide predictions about select chemical properties of the analytes. Computational approaches based on physicochemical properties have been used to predict elution times (Shah et al. 2010; Kangas et al. 2012), MS-MS fragmentation patterns (Heinonen et al. 2008; Wolf et al. 2010; Perdivara et al. 2013), and CCS area (Paglia et al. 2014). Using one or more of the analytical approaches with other cheminformatic tools for predicting metabolism and environmental degradation products (Dimitrov et al. 2010) might help to create in silico libraries that grow in breadth and accuracy and can be used to transition from nontargeted to targeted analysis.

The approaches described here represent essential methods for making the rapid transition from nontargeted to targeted analysis. For site-specific assessments with many unidentified chemicals, the approaches would provide a means of identifying analytes progressively for later hazard or risk assessment. For this case study, the committee assumed that the approaches applied to the environmental media, serum, urine, and hair samples would yield a list of 300 chemicals that are found with greatest consistency and at the highest concentrations in all samples (see Figure C-1). Chemicals that are found in environmental media and biological samples will constitute a logical choice for targeted toxicity testing because they might have a higher exposure potential than chemicals found only in environmental media.

As the number of identified chemicals increases, the data could be used to identify signatures of exposure to chemicals and mixtures. Such efforts would help to strengthen the exposure narrative and identify real-world mixtures for toxicity testing. The approaches for ranking based on hazard and bioactivity reported by Rager et al. (2016) (see Figures 2-7 and 2-8) are potentially applicable in some context of complex exposures. Other ES21 tools would then be used as needed and as described in Chapter 2 and the above section to provide better exposure char-

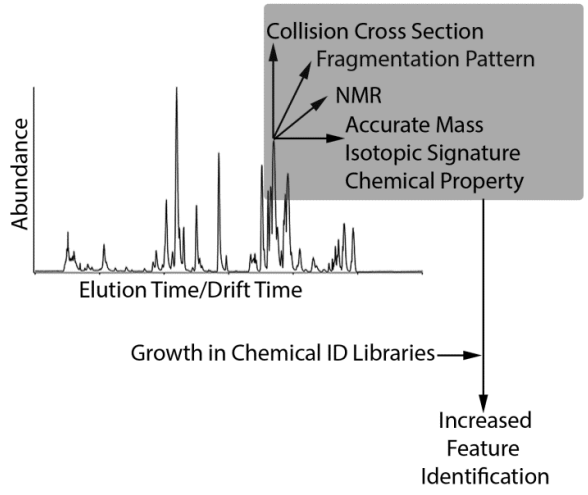



**FIGURE C-2** GC, LC, and ion-mobility spectrometry–MS/MS platforms allow the use of multiple types of data—including isotopic signature, elution time, fragmentation pattern, ionization source, collision cross-sectional area, and physicochemical properties—to identify unknown chemicals.

acterization through a more complete understanding of exposure pathways, fate and transport, and biokinetics.

## CHARACTERIZING TOXICITY AFTER A CHEMICAL RELEASE

This case study considers the environmental release of a chemical that has few toxicity data and approaches for characterizing toxicity rapidly to inform decision-making. 4-Methylcyclohexanemethanol (MCHM) was the major component of a chemical mixture that was spilled into the Elk River about 1 mile upstream of a water-intake facility for the city of Charleston, West Virginia, in 2014. The immediate public-health response was a "do not drink" order, but there was not enough information to provide guidance on what types of adverse health effects might be expected from MCHM or at what exposure levels. Primarily because hazard data were sparse, an acceptable concentration of MCHM in water being consumed by the local population and the potential risks associated with exposures to it could not be easily estimated. A few models and data streams that could be used in such situations are described below. The discussion provides general guidance but is not intended to be exhaustive. For example, only the exposure scenario related to drinking of tap water is presented here. In emergency scenarios, advice would also be given on whether people, including children and infants, could bathe in the water and whether the water could be used for cooking, washing clothes, and cleaning and provided to pets. Furthermore, although the focus is on MCHM, other chemicals at concentrations of at least 1% were present in the spilled material, including 4-(methoxymethyl)cyclo-hexanemethanol (4–22%), methyl 4-methylcyclohexane carboxylate (5%), 1,4-cyclohexanedimethanol (1–2%), and glycol phenyl ethers (propylene and dipropylene, whose concentrations were unknown).

Measured and model-predicted chemical-property information that is relevant for estimating MCHM environmental fate and toxicokinetics and for conducting an exposure assessment can be obtained from publicly available databases and software, including EPA's EPI Suite™ program (EPA 2011), which is primarily used here to obtain chemical property, fate, and bioaccumulation information. MCHM is a relatively small (128.2 g/mol) neutral organic chemical that has a solubility limit of about 2,000 mg/L and an octanol–water partition coefficient ($K_{OW}$) of about 350 (EPA 2011). It is a relatively volatile chemical (vapor pressure of about 8 Pa); however, its water solubility results in an air–water partition coefficient ($K_{AW}$) of about 0.0003 (EPA 2011). Screening-level evaluative mass-balance fate models that are included in EPI Suite (EPA 2011) indicate that once released to surface water, such as a river, MCHM is not distributed significantly from water to air or sediment. The biodegradation half-life in surface water is estimated to be about 15 days (EPA 2011). Predicted bioaccumulation factor for MCHM in fish is about 20 L/kg (relatively low), and the biotransformation half-life in fish is less than 1 day (relatively short). Those screening data indicate low persistence and bioaccumulation of MCHM in the environment; chemical concentrations in the river and in possible food sources from the river would be expected to decrease relatively quickly. Long-term, chronic exposures to local residents would not be expected. More sophisticated and resource-intensive models could be used to provide more refined situation-specific calculations for the expected change in environmental concentrations over time. For example, modeling tools could be used to estimate the time that it would take for concentrations in the river at the water-intake facility to decrease. Similar tools could be applied for the water-distribution system (after intake at the treatment facility).

Measured MCHM concentrations in drinking water in the first 2 days after the spill were about 1–4 mg/L (Foreman et al. 2015; Whelton et al. 2015). To determine the safety of the water for consumption, such sensitive populations as young infants and lactating and pregnant women would need to be considered. The 95th percentile drinking-water intakes by lactating women, pregnant women, and young infants are 0.055 L/kg-day (EPA 2011), 0.043 L/kg-day (EPA 2011), and 0.24 L/kg-day (EPA 2008), respectively. Given an MCHM concentration of 2 mg/L, the estimated acute (48-hour) intake in drinking water would be 0.48 mg/kg-day for the most exposed group, young infants. Lactating women would take in 0.11 mg/kg-day. Water concentrations in Charleston tap water declined to less than 1 mg/L 5 days after the spill and continued to decline to about 0.002 mg/L 3 weeks after the spill (Foreman et al. 2015). Thus, the MCHM intake 3 weeks after the spill by the 95th percentile drinking-water consumers would have declined to 0.48 μg/kg-day in young infants and 0.11 μg/kg-day in lactating women. The predicted half-life of MCHM in humans is about 2 hours (Arnot et al. 2014), so internal concentrations are expected to decrease relatively quickly after exposure events because MCHM is not persistent or bioaccumulative in humans.

A number of symptoms were reported in the community either through emergency-room visits or in follow-up surveillance by the Centers for Disease Control and Prevention and the Kanawha Charleston Health Department. Vomiting, nausea, diarrhea, and sore throat were most associated with reported drinking of the water, whereas skin irritation and rash were associated with bathing (Whelton et al. 2015). At the time of the spill, animal data were available on acute and subacute toxicity, site-of-contact irritation, skin sensitization, and genotoxicity, but there was no information on potential developmental toxicity or long-term health effects. The information generated af-

ter the spill primarily used Tox21 tools described in Chapter 3 and provide a good example of how these tools can be used qualitatively to provide support for public-health decisions. The following discussion provides several approaches for estimating or evaluating MCMH toxicity.

A rapid approach for estimating the potential for adverse effects is chemical structural comparison with known toxicants. Published methods can be used to determine whether there are reports in the literature on chemicals that have similar structural features. Wu et al. (2013) published a decision tree for developmental toxicity that was based on a chemical structural analysis of about 900 chemicals. The decision tree contained no precedents for developmental toxicity of chemicals that had the structural features of MCMH. Although that approach does not provide a definitive answer, it is a rapid means of determining whether a chemical has a signal for developmental toxicity. It is also possible to look for structurally similar chemicals in large toxicology databases, such as those amalgamated under EPA's Aggregated Computational Toxicology Resource program. In this case, no chemicals that had high structural similarity to MCHM were identified.

The National Toxicology Program (NTP) undertook a number of short-term assays intended to determine whether MCHM has activity against targets of concern (NTP 2016a). The testing included in vitro assays in 27 cell types, querying activity on signaling pathways relevant for development, rapid-turnaround assays in *Caenorhabditis elegans* and zebrafish embryos, and a 5-day toxicogenomics study in rats. No signals were generated from any in vitro assays up to relatively high concentrations (almost 100 μM) or in assays with *C. elegans* or zebrafish, although a minor contaminant of MCHM did have some activity in zebrafish embryos at about 100 μM. The toxicogenomics study was used to generate a biological no-observed-effect level (NOEL) for gene expression that is reported to be in the range of 6-99 mg/kg-day (the range is attributed to different methods used for data analysis). That screening-level study used six doses from 0.1 to 500 mg/kg-day (administered orally) for 5 days and evaluated gene expression in liver and kidney. A biological response was reported in liver at 6–99 mg/kg-day with no effect on kidney gene expression (NTP 2016b). The acute 95th percentile water-consumption exposure intake rates of 0.48 mg/kg-day for infants and 0.11 mg/kg-day for lactating women are lower than the NOEL for gene expression by factors of about 12–200 and 60–1,000, respectively. The committee notes that longer-term exposures were much lower. Because this example did not account for other exposure routes, which could add to ingestion exposure, the findings support the do-not-drink order issued for the entire service area (Whelton et al. 2015). Data gaps regarding other exposure routes could have been addressed by testing gene expression after administering the chemical by other relevant routes or by using physiologically based pharmacokinetic models to estimate the contribution of dermal and inhalation exposure to the total systemic concentration. Policy on interpreting the data streams will need to be created; this example is informative in describing the types of data that can be generated quickly to support risk-management decisions.

In summary, although the data differed from a standard toxicology evaluation, they were sufficient to indicate that MCHM was not structurally similar to known developmental toxicants or genotoxicants and that it did not have biological activity consistent with that of a potent developmental or systemic toxicant. A few animal studies that reported a sensitive readout (global gene expression) identified an MCHM concentration that was without biological effect, and that information supports a NOEL of about 100 mg/kg-day or somewhat lower, depending on the method of analysis. Exposure estimates derived from measurements of drinking water could be compared with the NOEL and other hazard data, and models could be used to provide initial indications of the time required for environmental concentrations to decrease to acceptable concentrations after a spill.

## PREDICTING TOXICITY OF REAL-WORLD CHEMICAL MIXTURES

Once chemicals at a site or part of a spill have been identified, the first question to address is whether toxicity data on them exist. For some chemicals, there are health assessments, such as those generated by the Integrated Risk Information System program, the International Agency for Research on Cancer monographs program, and the *Report on Carcinogens* program. Some assessments might be out of date or have notable limitations, so there might be some benefit of using Tox21 tools and approaches described in Chapter 3 to produce additional hazard and dose–response data, perhaps focused on previously identified end points of concern or to provide missing data on variability. Many chemicals, however, will not have been assessed or not have many toxicity data, such as MCHM (described in the case study above). For those chemicals, there would be clear cost and time advantages of using Tox21 tools and approaches described in Chapter 3. For example, the potential for identified substances to pose a human health hazard can be estimated quantitatively or qualitatively by chemical structure–activity modeling (Sutter et al. 2013), by combining structural information and bioactivity profiling (Low et al. 2011, 2013), by assessing bioactivity with in vitro assays that represent a wide array of tissues and biological targets (Judson et al. 2014), by establishing appropriate points of departure followed by in vitro to in vivo extrapolation (Judson et al. 2011), and by using population-based and other in vitro models to derive chemical-specific variability estimates (Abdo et al. 2015a,b). Although the toxicity evaluation

would initially be performed on an individual-chemical basis, real-world exposures are to the chemical mixtures that have been detected in environmental samples. Adding complexity to the situation is that many of the chemicals in a mixture will not have been identified. This case study provides an approach for investigating the potential hazard posed by such mixtures.

For the toxicity assessment of complex mixtures observed in environmental samples, tissues, and biofluids, such as in the first case study described in this appendix, a biological read-across approach (Low et al. 2013; Grimm et al. 2016) that relies on bioactivity-profiling data from various in vitro toxicity assays, high-content screening assays, and possibly high-throughput genomic analyses could be used to probe potential hazards. A biological read-across might be the most expedient approach for identifying potential human health hazards posed by the uncharacterized mixtures. Heterogeneity of tissue or organ toxicity, interindividual variability, and other factors can be addressed through bioactivity profiling of real-world mixtures by using human cell models in monoculture, co-cultures of various cell types, or more complex tissue-on-a-chip models.

Figure C-3 provides an overview of the biological read-across approach. Generally, chemical representatives of various toxicant classes, such as those listed in Table C-1, should be tested in a panel of in vitro assays that will also be used to test the environmental samples to establish the range of responses. Likewise, "designed" mixtures can be created—for example, on the basis of chemical-use patterns or other exposure-based data—and tested. The testing will yield a database of the biological effects of persistent environmental pollutants from a panel of diverse in vitro assays that can be used to move the unknown mixtures into classes of known chemicals or designed mixtures and to conduct the read-across to predict potential human health hazards posed by the real-world mixtures as described further below.

The database of bioactivity readouts from representative chemicals and designed mixtures can be used as a training set for the classification models that evaluate differences between chemicals or chemical classes. The results of that activity can then be used to compare (read-across) the environmental mixtures that have unknown chemical composition with representative chemicals or designed mixtures. For example, a series of machine-learning–based models could be constructed that define biological spaces that separate one class from all others (one-vs-all) or separate a single class from another class (one-vs-one). Ultimately, a real-world environmental mixture can be profiled in the same assay battery, and the resulting bioactivity readout can be used to obtain a quantitative estimate and qualitative response related to whether the mixture behaves like a particular toxicant or toxicant class in a specific assay or assay battery.

Ultimately, high-dimensional in vitro toxicity or transcriptomic data can be used to read-across a particular mixture of unknown chemical composition to known ref-

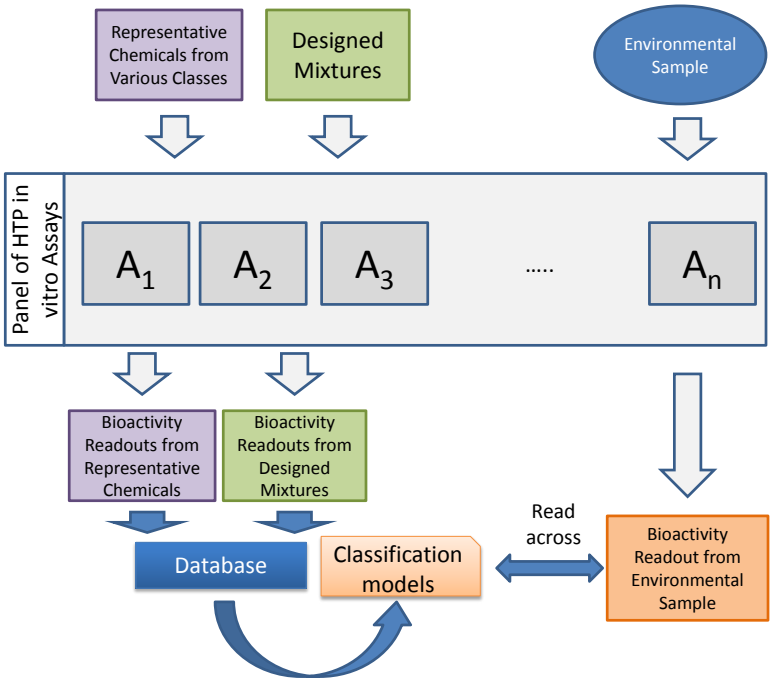



**FIGURE C-3** Biological read-across that provides an approach to assessing the hazard posed by complex mixtures.

erence chemicals or chemical combinations and establish a "biological analogue" that consists of a mixture of reference chemicals for which existing toxicity benchmarks are available. If the read-across–based mixture is used as a surrogate for the original mixture, standard methods for deriving cumulative risk estimates that are based on individual chemical exposure estimates and decision benchmark methods can be applied. Although the read-across mixture might have a different chemical composition from the real-world mixture, one can assume that their biological similarity based on the in vitro toxicity testing is adequate for informing environmental decisions.

Tox21 methods of evaluating mixtures can be used to establish dose–response relationships for various bioactivity by evaluating serial dilutions of the mixture or extracts. The resulting data can be compared with the bioactivity of the samples collected at different locations at the site or adjacent areas or with the bioactivity of historical samples from the same site. A challenge in this method is similar to the one that exists for extrapolating in vitro exposure to in vivo exposure. In vitro–in vivo extrapolation (IVIVE) methods are now used to estimate the daily human oral dose, called the oral equivalent dose, necessary to achieve steady-state in vivo blood concentrations equivalent to the points of departure derived from the in vitro assays (NRC 2014). IVIVE-adjusted data from in vitro assays can be directly compared with exposure information and improve chemical priority-setting by adding a risk context to the high-throughput in vitro screening (Wetmore et al. 2013). However, IVIVE research efforts have focused on individual chemicals, not on mixtures. A study of comparative analysis of in vitro cytotoxicity of pesticide mixtures with potential human exposures is an example of computing oral equivalent doses for mixtures by using the reverse-dosimetry approach (Abdo et al. 2015a). In that study, incorporation of dosimetry with in vitro data and conversion to an oral equivalent dose of each mixture allowed a risk-relevant ranking of the mixtures that considered chemical pharmacokinetic behavior; additional exposure data were used to adjust the potencies. However, additional experimental and methodological work is needed to bridge in vitro testing data on mixtures and exposure estimates.

## REFERENCES

Abdo, N., B.A. Wetmore, G.A. Chappell, D. Shea, F.A. Wright, and I. Rusyn. 2015a. In vitro screening for population variability in toxicity of pesticide-containing mixtures. Environ. Int. 85:147-155.

Abdo, N., M. Xia, C.C. Brown, O. Kosyk, R. Huang, S. Sakamuru, Y.H. Zhou, J.R. Jack, P. Gallins, K. Xia, Y. Li, W.A. Chiu, A.A. Motsinger-Reif, C.P. Austin, R.R. Tice, I. Rusyn, and F.A. Wright. 2015b. Population-based in vitro hazard and concentration-response assessment of chemicals: The 1000 genomes high-throughput screening study. Environ. Health Perspect. 123(5):458-466.

Arnot, J.A., T.N. Brown, and F. Wania. 2014. Estimating screening-level organic chemical half-lives in humans. Environ. Sci. Technol. 48(1):723-730.

Bingol, K., and R. Brüschweiler. 2015. Two elephants in the room: New hybrid nuclear magnetic resonance and mass spectrometry approaches for metabolomics. Curr. Opin. Clin. Nutr. Metab. Care 18(5):471-477.

Clendinen, C.S., G.S. Stupp, R. Ajredini, B. Lee-McMullen, C. Beecher, and A.S. Edison. 2015. An overview of methods using (13)C for improved compound identification in metabolomics and natural products. Front. Plant Sci. 6:611.

Dimitrov, S., D. Nedelcheva, N. Dimitrova, and O. Mekenyan. 2010. Development of a biodegradation model for the prediction of metabolites in soil. Sci. Total Environ. 408(18):3811-3816.

EPA (US Environmental Protection Agency). 2008. Child-Specific Exposure Factors Handbook (Final Report). EPA/600/R-06/096F. US Environmental Protection Agency, Washington, DC.

EPA (US Environmental Protection Agency). 2011. Estimation Programs Interface (EPI) Suite for Microsoft® Windows, Version 4.1. US Environmental Protection Agency, Washington, DC.

Ewing, M.A., M.S. Glover, and D.E. Clemmer. 2016. Hybrid ion mobility and mass spectrometry as a separation tool. J. Chromatogr. A. 1439:3-25.

Foreman, W.T., D.L. Rose, D.B. Chambers, A.S. Crain, L.K. Murtagh, H. Thakellapalli, and K.K. Wang. 2015. Determination of (4-methylcyclohexyl)methanol isomers by heated purge-and-trap gc/ms in water samples from the 2014 Elk River, West Virginia, chemical spill. Chemosphere 131:217-224.

Grimm, F.A., Y. Iwata, O. Sirenko, G.A. Chappell, F.A. Wright, D.M. Reif, J. Braisted, D.L. Gerhold, J.M. Yeakley, P. Shepard, B. Seligmann, T. Roy, P.J. Boogaard, H. Ketelslegers, A. Rohde, and I. Rusyn. 2016. A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. Green Chem. 18(16):4407-4419.

Heinonen, M., A. Rantanen, T. Mielikäinen, J. Kokkonene, J. Kiuru, R.A. Ketola, and J. Rousu. 2008. FiD: A software for ab initio structural identification of product ions from tandem mass spectrometric data. Rapid Commun. Mass Spectrom. 22(19):3043-3052.

Horai, H., M. Arita, Y. Ojima, Y. Nihei, S. Kanaya, and T. Nishioka. 2009. Traceable analysis of multiple-stage mass spectra through precursor-product annotations. Pp. 173-178 in Proceedings of German Conference on Bioinformatics, September 28-29, 2009, Wittenberg, Germany, I. Grosse, S. Neumann, S. Posch, F. Schreiber, and P. Stadler, eds. Lecture Notes in Informatics-Proceedings Vol 157. Bonn: Kollen Druck+Verlag [online]. Available: http://subs.emi

s.de/LNI/Proceedings/Proceedings157/173.pdf [accessed July 27, 2016].

Judson, R.S., R.J. Kavlock, R.W. Setzer, E.A. Hubal, M.T. Martin, T.B. Knudsen, K.A. Houck, R.S. Thomas, B.A. Wetmore, and D.J. Dix. 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chem. Res. Toxicol. 24(4):451-462.

Judson, R., K. Houck, M. Martin, T. Knudsen, R.S. Thomas, N. Sipes, I. Shah, J. Wambaugh, and K. Crofton. 2014. In vitro and modeling approaches to risk assessment from the US Environmental Protection Agency ToxCast programme. Basic Clin. Pharmacol. Toxicol. 115(1):69-76.

Kangas, L.J., T.O. Metz, G. Isaac, B.T. Schrom, B. Ginovska-Pangovska, L. Wang, L. Tan, R.R. Lewis, and J.H. Miller. 2012. In silico identification software (ISIS): A machine learning approach to tandem mass spectral identification of lipids. Bioinformatics 28(13):1705-1713.

Low, Y., T. Uehara, Y. Minowa, H. Yamada, Y. Ohno, T. Urushidani, A. Sedykh, E. Muratov, V. Kuz'min, D. Fourches, H. Zhu, I. Rusyn, and A. Tropsha. 2011. Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. Chem. Res. Toxicol. 24(8):1251-1262.

Low, Y., A. Sedykh, D. Fourches, A. Golbraikh, M. Whelan, I. Rusyn, and A. Tropsha. 2013. Integrative chemical-biological read-across approach for chemical hazard classification. Chem. Res. Toxicol. 26(8):1199-1208.

May, J.C., R.L. Gant-Branum, and J.A. McLean. 2016. Targeting the untargeted in molecular phenomics with structurally-selective ion mobility-mass spectrometry. Curr. Opin. Biotechnol. 39:192-197.

Neumann, S., and S. Bocker. 2010. Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. Anal. Bioanal. Chem. 398(7-8):2779-2788.

NRC (National Research Council). 2014. A Framework to Guide Selection of Chemical Alternatives. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2016a. West Virginia Chemical Spill: NTP Studies [online]. Available: http://ntp.niehs.nih.gov/results/areas/wvspill/studies/index.html [accessed July 27, 2016].

NTP (National Toxicology Program). 2016b. West Virginia Chemical Spill: 5-Day Rat Toxicogenomic Studies, June 2015 NTP Update [online]. Available: http://ntp.niehs.nih.gov/ntp/research/areas/wvspill/micronucleus_update_508.pdf [accessed July 27, 2016].

Paglia, G., J.P. Williams, L. Menikarachchi, J.W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, A. Moseley, D. Grant, J. Langridge, B.O. Palsson, and G. Astarita. 2014. Ion mobility derived collision cross sections to support metabolomics applications. Anal. Chem. 86(8):3985-3993.

Perdivara, I., L. Perera, M. Sricholpech, M. Terajima, N. Pleshko, M. Yamauchi, and K.B. Tomer. 2013. Unusual fragmentation pathways in collagen glycopeptides. J. Am. Soc. Mass. Spectrom. 24(7):1072-1081.

Rager, J.E., M.J. Strynar, S. Liang, R.L. McMahen, A.M. Richard, C.M. Grulke, J.F. Wambaugh, K.K. Isaacs, R. Judson, A.J. Williams, and J.R. Sobus. 2016. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environ. Int. 88:269-280.

Shah, A.R., K. Agarwal, E.S. Baker, M. Singhal, A.M. Mayampurath, Y.M. Ibrahim, L.J. Kangas, M.E. Monroe, R. Zhao, M.E. Belov, G.A. Anderson, and R.D. Smith. 2010. Machine learning based prediction for peptide drift times in ion mobility spectrometry. Bioinformatics 26(13):1601-1607.

Sutter, A., A. Amberg, S. Boyer, A. Brigo, J.F. Contrera, L.L. Custer, K.L. Dobo, V. Gervais, S. Glowienke, J. van Gompel. N. Greene, W. Muster, J. Nicolette, M.V. Reddy, V. Thybaud, E. Vock, A.T. White, and L Müller. 2013. Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regul. Toxicol. Pharmacol. 67(1):39-52.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell, III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thomas. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125(1):157-174.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, L. Li, H.J. Clewell, III, R.S. Judson, K. Freeman, W. Bao, M.A. Sochaski, T.M. Chu, M.B. Black, E. Healy, B. Allen, M.E. Andersen, R.D. Wolfinger, and R.S. Thomas. 2013. Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol. Sci. 132(2):327-346.

Wetmore, B.A., B. Allen, H.J. Clewell, III, T. Parker, J.F. Wambaugh, L.M. Almond, M.A. Sochaski, and R.S. Thomas. 2014. Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicol. Sci. 142(1):210-224.

Whelton, A.J., L. McMillan, M. Connell, K.M. Kelley, J.P. Gill, K.D. White, R. Gupta, R. Dey, and C. Novy. 2015. Residential tap water contamination following the Freedom Industries chemical spill: Perceptions, water quality, and health impacts. Environ Sci Technol. 49(2):813-823.

Wolf, S., S. Schmidt, M. Müller-Hannemann, and S. Neumann. 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 11:148.

Wu, S., J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, and K. Blackburn. 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem. Res. Toxicol. 26(12):1840-1861.

# D

# Case Study on Assessment of New Chemistries

The case study in this appendix describes a hypothetical scenario in which there are three choices of "new" chemicals for use in the manufacture of a product that will result in human exposure. Initial testing shows that the chemicals in question will most likely leach out of the product and possibly end up in food or water that will be ingested by people. In addition, contact with skin during the regular handling of the product is a possible route of human exposure. Finally, the chemical might become aerosolized and inhaled by workers in the manufacturing facility or as a result of indoor consumer use of the product. Therefore, chemical exposure is possible through inhalation, ingestion, and dermal pathways, and the chemical could pose a threat to human health.

For illustrative purposes, the committee chose to use three related drugs (weak acids)—ibuprofen, ibufenac and diclofenac—on which various amounts of in vitro data are publicly available. Table D-1 provides the chemical structures and selected physicochemical properties. To reflect a possible real-world scenario, a key assumption of this case study is that only the in silico and in vitro data presented here are available for the screening assessment; in vivo and clinical data are presumed to be "not yet available." However, because the adverse effects of the chemicals on people have been studied, one can compare the results of the approach with actual human-safety outcomes. The example is intended to illustrate how available and emerging screening-level tools and data (read-across, screening-level models, and available high-throughput in vitro data) could be applied to inform decision-making and to identify some of the key data gaps and sources of uncertainty that are relevant to risk assessments. The committee notes that most practical approaches for assessing chemical similarity would exclude diclofenac from this comparison because of the chlorine and amine moieties that are not present in the other two chemicals. The committee includes it here for the sake of illustration, but it should be noted that there are limits to how dissimilar chemicals can be used in a read-across scenario.

## STRUCTURAL ALERTS

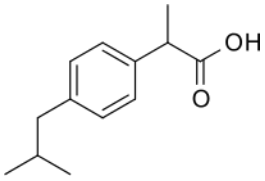
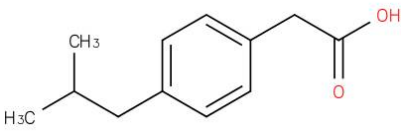
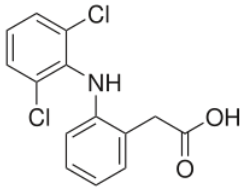
All molecules that contain an arylacetic acid group can undergo acyl glucuronidation, a major metabolic conjugation pathway in mammals for chemicals that contain these groups. Acyl glucuronides have been implicated—although it is not definitively proved—as a cause of adverse effects in humans because they form protein adducts (Shipkova et al. 2003) (see Figure D-1) . Common risk concerns are liver injury and hypersensitivity reactions (Regan et al. 2010). The relative reactivity and half-life of the acyl glucuronide has been suggested as a differentiating factor between chemicals that cause adverse events and ones that are of less concern. Other researchers suggest that arylacetic acids can undergo coenzyme A (CoA) conjugation, and interference of the CoA conjugates with lipid metabolism and other cellular processes can lead to the observed toxicity (Darnell and Weidolf 2013). The metabolic scheme might need to be confirmed experimentally to reduce uncertainty (Patlewicz et al. 2015).

## IN VITRO DATA

To ensure data consistency among the chemicals in question, in vitro data were gathered only from the ToxCast website, and they are summarized in Table D-2 (EPA 2016). Only assays that yielded activity below a 10 μM threshold are considered because they constitute 20% of the observed assay activity for diclofenac and would most likely be the cause of the toxicity used to set assay doses. It is important to note that the assays used in the ToxCast program do not represent the entire spectrum of biological processes that might be relevant to human health (that is, all possible adverse effects of exposure to chemicals); therefore, there are likely to be gaps in knowledge of how the three chemicals would interact in a biological system. To give some context to the values in Table D-2, diclofenac was tested in a zebrafish toxicity screen and had a lowest effect level of 64 μM (Truong et al. 2014).

Data on ibufenac are not available, but given its structural similarity to ibuprofen and comparable physicochemical properties, one would expect ibufenac to have an in vitro activity profile similar to that of ibuprofen.

**TABLE D-1** Chemical Structures and Selected Measured and Predicted Properties[a]

| Chemical structure and name |  Ibuprofen |  Ibufenac |  Diclofenac |
|---|---|---|---|
| Molar mass (g/mol) | 206.3 | 192.3 | 296.2 |
| log $K_{OW}$[b] | 3.97 | 3.35 | 4.51 |
| log $K_{AW}$[c] | -5.21 | -5.33 | -9.71 |
| pKa[d] | 4.4 | 4.4 | 4.2 |
| logD (pH 7.4)[e] | 0.45 | 0.22 | 1.37 |
| Air half-life (h) | 10.8 | 12.7 | 0.78 |
| Predicted whole-body biotransformation half-life (h) (chemical similarity score) | 3.6 (0.36, low similarity) | 2.1 (0.24, low similarity) | 14.9 (0.36, low similarity) |

[a]Physicochemical properties are from EPA's EPI Suite™ (EPA 2011) and ACD Labs (ACD 2015). The whole-body biotransformation half-lives shown here are predicted from structure by using a screening-level quantitative structure–activity relationship (QSAR) model (Arnot et al. 2014a). Various methods can be used to determine the applicability domain of a QSAR prediction. Here, the chemical similarity score is a measure of the similarity, in structure and properties, of a predicted chemical to chemicals in the training dataset on the basis of a nearest-neighbors approach (Brown et al. 2012). The three chemicals have similar molar mass and partitioning and dissociation properties, and absorption efficiencies are expected to be similar in the chemicals but different for each exposure pathway.
[b]log $K_{OW}$ (or logP) is the $\log_{10}$ of the octanol–water partition coefficient of the neutral species.
[c]log $K_{AW}$ is $\log_{10}$ of the air–water partition coefficient of the neutral species.
[d]pKa is the $\log_{10}$ of the acid dissociation constant.
[e]logD is the $\log_{10}$ of the distribution coefficient of neutral and ionic species between octanol and water at pH 7.4.

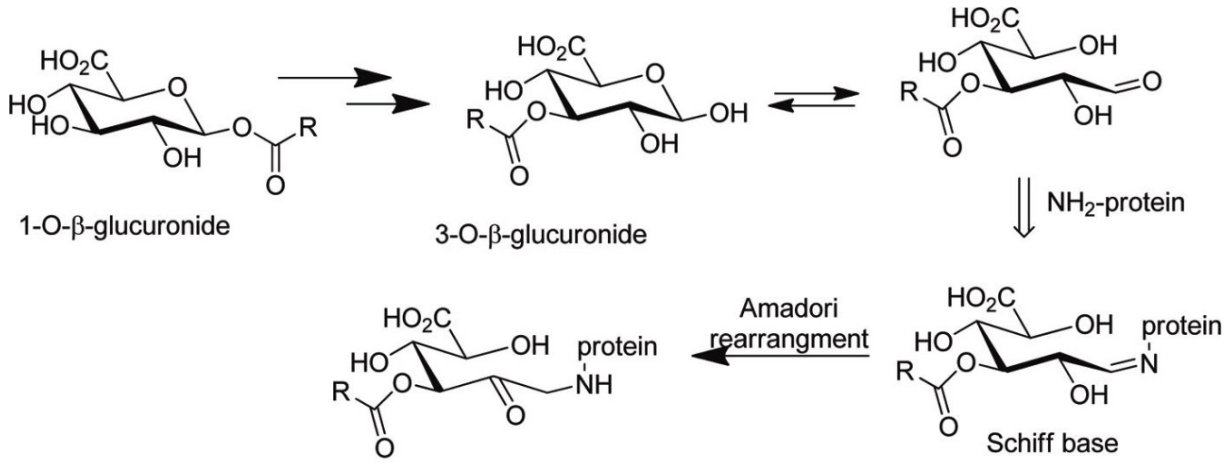



**FIGURE D-1** Metabolism of 1-O-β-glucuronide. Source: Stepan et al. 2011.

**TABLE D-2** Data from In Vitro Assay in Which Chemicals Had Activity Below 10 μM

| Assay Activity | Platform | Diclofenac $AC_{50}$ (μM) | Ibuprofen $AC_{50}$ (μM) |
|---|---|---|---|
| Decrease in interleukin 8 (IL-8) | BioSeek | – | 0.002 |
| Decrease in matrix metalloproteinase-1 (MMP-1) | BioSeek | – | 0.003 |
| Suppression of prostaglandin $E_2$ secretion ($PGE_2$) | BioSeek | 0.010 | 1.203 |
| Inhibition of cyclooxygenase 1 (COX1) | NovaScreen | 0.163 | 3.0 |
| Inhibition of cyclooxygenase 2 (COX2) | NovaScreen | 0.215 | 30.0 |
| Increase in cell proliferation | BioSeek | – | 0.251 |
| Binding of peroxisome proliferator-activated receptor gamma (PPAR-γ) | NovaScreen | 0.523 | – |
| Decrease in collagen III | BioSeek | 26.108 | 3.509 |
| Decrease in interleukin 6 (IL-6) | BioSeek | – | 3.977 |
| Increase in thrombomodulin | BioSeek | 4.742 | 17.674 |
| Activation of pregnane X receptor (PXR) | Attagene | 7.438 | – |
| Decrease in low-density lipoprotein (LDL) receptor | BioSeek | – | 7.637 |
| Increase in macrophage colony-stimulating factor (M-CSF) | BioSeek | – | 7.639 |
| Decrease in monocyte chemotactic protein 1 (MCP1) | BioSeek | 7.704 | – |
| Activation of PPAR-γ | Attagene | 8.256 | 39.710 |
| Activation of glucocorticoid receptor (GR) | NovaScreen | 8.671 | – |
| Activation of estrogen receptor element (ERE) | Attagene | – | 9.566 |

Source: Data from PubChem. Available at https://pubchem.ncbi.nlm.nih.gov/.

### Suppression of Prostaglandin Synthesis

Diclofenac is a potent inhibitor of cyclooxgenase 1 and 2 (COX1 and COX2), and inhibition of these enzymes can decrease prostaglandin biosynthesis (Vane 1971). Decreased secretion of prostaglandin E2 ($PGE_2$) was observed in the Bioseek platform. Ibuprofen is a weak nonspecific inhibitor of COX1 ($IC_{50}$, about 18 μM) and COX2 ($IC_{50}$, about 370 μM) (Noreen et al. 1998) but also showed a similar suppression of $PGE_2$ in the BioSeek platform. $PGE_2$ is linked to suppression of T-cell receptor signaling and inflammation responses (Wiemer et al. 2011). However, $PGE_2$ is also a vasodilator, so suppression of its secretion might lead to an increase in blood pressure and to cardiac toxicity (Strong and Bohr 1967).

Drugs that inhibit COX1 or COX2, such as celecoxib and rofecoxib, have been linked with causing cardiovascular events (Johnsen et al. 2005), and rofecoxib, a selective COX2 inhibitor, was withdrawn from the US market after being linked to heart attacks and strokes. Inhibitors of COX1 have been linked to causing ulceration and bleeding in the gastrointestinal tract as a result of suppressing the secretion of the protective prostaglandins $PGE_2$ and $PGI_2$ (Süleyman et al. 2007). Inhibitors of COX1 might also affect renal function by changing the role of prostaglandins on renal hemodynamics and glomerular filtration rate (GFR) (DuBois et al. 1998; Morita 2002).

Diclofenac and ibuprofen increase thrombomodulin (TM) in the BioSeek platform. TM is a cell-surface receptor for thrombin on endothelial cells that is involved in blood coagulation (Gerlitz et al. 1993). Increases in TM might increase clotting times but similarly reduce the risk of stroke and myocardial infarction (Esmon et al. 1982).

The low-density lipoprotein (LDL) receptor mediates the endocytosis of LDL. The accumulation of LDL in the blood is involved in the development of atherosclerosis, which is the process responsible for most cardiovascular diseases (Hobbs et al. 1992). A decrease in the LDL receptor might lead to an increased risk of cardiovascular events in people who are predisposed to atherosclerosis or who have cardiovascular conditions.

### Liver Effects

Diclofenac is shown to increase the activity of the pregnane X receptor (PXR). PXR is a nuclear receptor that has important roles in integrating pathways related to fatty acid, lipid, and glucose metabolism (Wada et al.

2009). It also senses the presence of foreign substances and responds by upregulating proteins involved in their oxidation and others involved in their clearance (Kliewer 2003), and it is a transcriptional regulator of the cytochrome P450 gene CYP3A4, a major metabolizing enzyme for many drugs that is highly expressed in the liver.

Both diclofenac and ibuprofen activate the peroxisome proliferator–activated receptor gamma (PPAR-γ) that regulates fatty acid storage and glucose metabolism, although only at relatively high concentrations in the case of ibuprofen. The genes activated by PPAR-γ increase lipid uptake and adipogenesis by fat cells (Zou et al. 2016). PPAR-γ agonists have been used in the treatment of hyperlipidemia and hyperglycemia and therefore might induce hypoglycemia in healthy subjects (Spiegelman 1998; Rangwala and Lazar 2004). Some drugs that were designed to activate PPAR-γ have been linked with hepatotoxicity (troglitazone: Watkins 2005), cardiovascular events (rosiglitazone: Singh et al. 2007), and an increased incidence of in bladder cancer (pioglitazone: Ferwana et al. 2013). However, no direct link has been established between the activation of PPAR-γ and those adverse events.

### Immune-Response Effects

Diclofenac and ibuprofen have effects on various cellular processes that are involved in inflammation and tissue repair. For example, diclofenac decreases the expression of monocyte chemotactic protein 1 (MCP1). MCP1 promotes movement of monocytes, memory T cells, and dendritic cells to sites of inflammation (Mukaida et al. 1998; Xue et al. 2015).

Similarly, diclofenac is an agonist of the glucocorticoid receptor (GR), which is expressed in almost every cell in the body and regulates genes that control development, metabolism, and immune response (Rhen and Cidlowski 2005; Lu et al. 2006). The activated GR complex prevents the movement of transcription factors from the cytosol into the nucleus, resulting in changes in expression of nuclear anti-inflammatory proteins and cytosolic proinflammatory proteins.

Ibuprofen decreases the secretion of Interleukin 8 (IL-8) as measured in the BioSeek platform. IL-8 is a chemokine that is produced by macrophages and other cell types, such as epithelial cells, airway smooth muscle cells (Hedges et al. 2000), and endothelial cells. IL-8 induces chemotaxis in neutrophils and causes them to migrate toward sites of infection and promotes phagocytosis at the infection site. It is also a potent promoter of angiogenesis and an important mediator of the immune reaction in the innate immune system response. Ibuprofen decreases the secretion of IL-6, which acts as a pro-inflammatory cytokine and an anti-inflammatory myokine (Schöbitz et al. 1994).

## HAZARD IDENTIFICATION

On the basis of the available in vitro data, structural comparisons, and knowledge of structural alerts, a key safety concern about all three chemicals would be liver injury through the formation of reactive acyl glucuronides or acyl coenzyme A conjugates that would cause tissue damage and impaired organ function. Relative reactivity of the acyl conjugates would play an important role in determining the risk of liver injury. Chemicals that have alkyl substitutions at the α-carbon atom have been shown to have lower reactivity with protein nucleophiles; this suggests that inherent electronic and steric effects affect the overall rate of acyl glucuronide rearrangement (Stepan et al. 2011) and so could have a profound effect on the reactivity of the conjugates in the case of ibuprofen (Wang et al. 2004; Walker et al. 2007; Baba and Yoshioka 2009). The risk of liver injury could be increased by induction of cytochrome P-450s through activation of PXR and by lipid dysfunction as a result of activation of PPAR-γ.

Cardiovascular toxicity in the form of increased blood pressure and increased clotting times and renal damage or gastrointestinal bleeding caused by the suppression of prostaglandin secretion are also of concern with diclofenac and ibuprofen and by inference, ibufenac.

As discussed in Chapter 3, for inhibitors of G-protein–coupled receptors, the anticipated pharmacological response is often observed in vivo at plasma concentrations up to 3 times the measured $IC_{50}$ of the chemical in question (McGinnity et al. 2007). As a general rule of thumb, a 100-fold difference between the measured $IC_{50}$ or the inhibition constant in a cell-free assay and the circulating plasma $C_{max}$ free[1] concentration could be considered to be adequate to pose minimal risk of toxicity from a pharmacological interaction. It is worth noting that for more phenotypic cellular responses, such as those measured by the BioSeek platform, more research is required to establish an appropriate translation from in vitro to in vivo.

## EXPOSURE ASSESSMENT

In this hypothetical case study, the three chemicals of interest have not been used in commercial products; therefore, there are no monitoring data, and there are no emissions and use data on which to formulate a typical risk-based evaluation. However, the available premarket toxicity or bioactivity data identified above can be used to develop parameters for exposure models that can "back-calculate" the rates of chemical use for various scenarios that correspond to specific hazard thresholds. The selected threshold for such simulations could be a concentration from a bioassay in the case of ibuprofen or diclofenac or a read-across value in the case of ibufenac determined from

[1] $C_{max}$ free is the maximum measured or observed concentration of the fraction of the chemical that is unbound to plasma proteins.

in vivo, in vitro, or computational methods or a no-effect threshold method, such as one that uses a threshold of toxicological concern.

The general exposure-assessment approach outlined here is analogous to the critical-emission-rate concept that has been applied in ecological assessments (Arnot et al. 2006) and to concepts applied in reverse toxicokinetics that is used to calculate intake rates expressed as oral equivalent doses (OEDs;[2] mg/kg-day) from the in vitro testing data (Rotroff et al. 2010). In the present case, toxicokinetic models are combined with indoor-fate models to back-calculate the rates of chemical use that correspond to illustrative exposure scenarios. The simulations can consider various assumptions and contexts for exposure and chemical or product scenarios. The exposure models used in such simulations can vary in complexity according to the amount of data needed to satisfy all the parameter requirements. In that regard, tiered modeling strategies might be helpful.

In this example, a one-compartment, whole-body toxicokinetic model that considers primary routes of exposure and intake (dermal, ingestion, and inhalation) and routes of elimination (for example, exhalation, renal excretion, biotransformation, egestion, and desquamation) is linked to a representative indoor environment (Arnot et al. 2014b) to back-calculate the rates of chemical use for three hypothetical exposure scenarios:

- *Scenario 1*. The chemical is released to air in a defined indoor environment. Exposure pathways include inhalation, dermal permeation (from passive diffusion in air), and nondietary ingestion (from hand-to-surface and surface-to-mouth contact).
- *Scenario 2*. The chemical is applied directly to skin and assumed to be left on indefinitely. Exposure pathways include dermal permeation and inhalation (from volatilization of the chemical from dermal application).
- *Scenario 3*. The chemical is ingested.

Simplifying assumptions are steady-state calculations and no charged species (that is, no explicit calculation for charged species; only the neutral form is simulated). The latter assumption is similar to recent hazard and risk-based calculations that used ToxCast data in which the potential for chemical dissociation was ignored; that is, acids and bases were treated as nondissociating neutral organics (Rotroff et al. 2010; Wetmore et al. 2012; Shin et al. 2015).

The first step is to translate the in vitro bioassay concentrations ($C_{in\ vitro}$) that correspond to the observed bioactivity to in vivo concentrations ($C_{in\ vivo}$). Here, the committee uses the same assumptions as in recent applications of ToxCast data for OED calculations: $C_{in\ vivo,\ blood} = C_{in\ vitro}$ (Rotroff et al. 2010; Wetmore et al. 2012; Shin et al. 2015). However, more explicit calculations should be used to account for differences in the in vitro and in vivo systems; for example, free dissolved concentrations rather than assumed nominal in vitro concentrations could be used (see discussion in Chapter 2). The steady-state volume of distribution is assumed to be 0.5 L/kg (35 L) for the three chemicals to relate blood concentrations to whole-body concentrations. Models for volume of distribution and other methods to address differential concentrations among and within tissues could be considered. The lowest $AC_{50}$ from the available ToxCast assays is selected as the hazard threshold on which to base parameter values for the exposure model. That value for ibuprofen and ibufenac is 0.002 μM, and the selected threshold for diclofenac is 0.01 μM (see Table D-2).

The second step is to select the parameters needed for the exposure models to calculate chemical fate in various environments. For the sake of illustration, the committee assumes an adult human in a single room, although infants and children have greater breathing rates relative to body weight; the evaluative model requires the following chemical-specific information: $K_{OW}$, $K_{AW}$, and degradation half-lives in air (see Table D-1). Quantitative structure–activity relationship (QSAR) models are used here to predict whole-body biotransformation half-life data. Half-lives could also be determined by scaling in vitro assay data derived from hepatocytes to liver (see, for example, Rotroff et al. 2010) or whole-body half-life estimates. In addition, hepatic, renal, or other compartment-specific QSAR models could be used to provide parameter values for pharmacokinetic models that are used for exposure assessment. Ideally, multiple lines of evidence (for example, various measured and predicted estimates) will show concordance in key information used in the model simulations (chemical partitioning properties and reaction half-lives), and this concordance will foster confidence in the assessment results. If chemicals are shown to have high environmental persistence, adding far-field human-exposure models to the assessment is warranted to account for possible far-field exposure pathways (chemical dispersed and diffused into food and water).

## RISK CHARACTERIZATION

The results of the back-calculation simulations are summarized in Table D-3. The calculations yield the indoor air release (Scenario 1), application (Scenario 2), and ingestion (Scenario 3) rates in milligrams per day corresponding to the selected in vitro bioactivity-assay data (assumed threshold values). The results could be used for interim guidance on use scenarios for each chemical and for comparative analyses between the three candidate chemicals. If one assumes that all three chemicals are

[2]The OED is the chemical intake rate that corresponds to an assumed steady-state blood concentration related to the in vitro bioactivity.

**TABLE D-3** Indoor Release (Scenario 1), Application (Scenario 2), and Ingestion (Scenario 3) Rates (mg/d) Corresponding to Selected In Vitro Bioactivity Data

| Chemical | Scenario 1: Release to Indoor Air | Scenario 2: Application Directly to Skin | Scenario 3: Ingestion |
|---|---|---|---|
| Diclofenac | 10 | 1.1 | 0.14 |
| Ibufenac | 1.8 | 0.15 | 0.13 |
| Ibuprofen | 1.9 | 0.16 | 0.08 |

used in the same quantity, the chemicals and scenarios with the lowest rates correspond to the greatest potential to achieve the in vitro bioactivity threshold. For example, for diffuse release to air in an indoor environment (Scenario 1), diclofenac shows the highest emission or release rate and, so could pose the lowest potential concern of the three chemicals. Of the three exposure scenarios, Scenario 3 results in the lowest use and application rates for all chemicals. Overall, the ranges of values are not large because the chemicals have similar properties for partitioning, reaction, and bioactivity (that is, the same bioactivity value is used for ibuprofen and ibufenac on the basis of structural read-across).

The values in Table D-3 do not show the uncertainty in the calculations and do not account for interindividual variability in the pharmacokinetics and pharmacodynamics that one would expect in a large diverse population. The results of this example are illustrative, but the general concept can be helpful in determining putative use and release scenarios for premarket chemicals. The application of exposure models to back-calculate emission and use rates corresponding to a toxic threshold or bioactivity can also be useful for evaluating commercial chemicals when emission and use rates are unknown or highly uncertain. Ultimately, confidence in the calculated emission and use rates depends on the confidence in and suitability of the toxicity (threshold) data and the exposure-model estimates. For the three chemicals in this example, measured volumes of distribution are about one-third to one-half the assumed values, and half-lives in adults are one-seventh to one-half the values used in this premarket assessment (Obach et al. 2008). For risk-based decision-making, additional analyses for various life stages and alternative use scenarios should be considered as warranted.

To put the exposure estimates in Table D-3 into context, the typical over-the-counter medicines that contain ibuprofen recommend an oral dose of 200–400 mg every 4 hours with a maximum dose of 1,200 mg in any 24-hour period for persons over 12 years old.[3] However, doctors can prescribe ibuprofen to be given orally at up to 3,200 mg/day in doses of up to 800 mg at any one time.[4] Similarly, ibuprofen has been approved in Europe for administration to children 3 to 6 months of age at a starting dose of 50 mg taken orally three times a day. Ibuprofen is contraindicated in pregnant women in their third trimester, and doctors recommend that women during the first 6 months of pregnancy not take it, if that is possible.

Diclofenac is approved for use by prescription, and the maximum recommended daily oral dose is 150 mg in adults; it is not recommended for use in children under 12 years old. It is also contraindicated for use by pregnant women. Ibufenac was withdrawn from the market because of severe hepatotoxicity and jaundice in patients taking the drug. At the time, the maximum recommended daily oral dose of ibufenac was 750 mg.

In Europe, both ibuprofen and diclofenac were approved for use as a topical gel (ibuprofen, 5% w/w gel; diclofenac, 2.32% w/w gel). A maximum daily application of the diclofenac gel was 8 g, which is equivalent to 160 mg of the active ingredient. Similarly, the recommended application of the ibuprofen gel was up to 125 mg, four times per day, which is equivalent to 25 mg of the active ingredient. However, only 22% of the dose is absorbed through the skin. Compared with the oral route of administration of ibuprofen, the plasma exposure is considered to be much lower and unlikely to cause systemic side effects.

The estimated oral and dermal exposures in the present example would be substantially below the therapeutic doses for most populations, including children. However, it should be noted that at therapeutic doses, some side effects and adverse events are observed with various, albeit relatively low, frequencies that might not necessarily be considered tolerable in an environmental or occupational risk assessment in which long-term, low-level exposures of a broad population demographic have to be considered. Similarly, conclusions cannot be drawn at this time about whether the estimated doses would ensure the protection of the most sensitive group—pregnant women.

[3]See https://www.medicines.org.uk/emc/medicine/15681.

[4]See http://www.accessdata.fda.gov/drugsatfda_docs/anda/2001/76-112_Ibuprofen.pdf.

## REFERENCES


ACD (Advanced Chemistry Development). 2015. ACD/Percepta Suite. ACD, Toronto, ON, Canada.

Arnot, J.A., D. Mackay, E. Webster, and J.M. Southwood. 2006. Screening level risk assessment model for chemical fate and effects in the environment. Environ. Sci. Technol. 40(7):2316-2323.

Arnot, J.A., T.N. Brown, and F. Wania. 2014a. Estimating screening-level organic chemical half-lives in humans. Environ. Sci. Technol. 48(1):723-730.

Arnot, J.A., X. Zhang, I. Kircanski, L. Hughes, and J. Armitage. 2014b. Develop Sub-module for Direct Human Exposures to Consumer Products. Technical report for the US Environmental Protection Agency, by ARC Arnot Research & Consulting Inc., Toronto, Canada.

Baba, A., and T. Yoshioka. 2009. Structure-activity relationships for the degradation reaction of 1-beta-O-acyl glucuronides. Part 3: Electronic and steric descriptors predicting the reactivity of aralkyl carboxylic acid 1-beta-O-acyl glucuronides. Chem. Res. Toxicol. 22(12):1998-2008.

Brown, T.N., J.A. Arnot, and F. Wania. 2012. Iterative fragment selection: A group contribution approach to predicting fish biotransformation half-lives. Environ. Sci. Technol. 46(15):8253-8260.

Darnell, M., and L. Weidolf. 2013. Metabolism of xenobiotic carboxylic acids: Focus on coenzyme a conjugation, reactivity, and interference with lipid metabolism. Chem. Res. Toxicol. 26(8):1139-1155.

DuBois, R.N., S.B. Abramson, L. Crofford, R.A. Gupta, L.S. Simon, L.B. Van De Putte, and P.E. Lipsky. 1998. Cyclooxygenase in biology and disease. FASEB J. 12(12):1063-1073.

EPA (US Environmental Protection Agency). 2011. Estimation Programs Interface (EPI) Suite for Microsoft® Windows, Version 4.1. US Environmental Protection Agency, Washington, DC.

EPA (US Environmental Protection Agency). 2016. Toxicity Forecasting: Advancing the Next Generation of Chemical Evaluations [online]. Available: https://www.epa.gov/chemical-research/toxicity-forecasting [accessed July 26, 2016].

Esmon, C.T., N.L. Esmon, and K.W. Harris. 1982. Complex formation between thrombin and thrombomodulin inhibits both thrombin-catalyzed fibrin formation and factor V activation. J. Biol. Chem. 257(14):7944-7947.

Ferwana, M., B. Firwana, R. Hasan, M.H. Al-Mallah, S. Kim, V.M. Montori, and M.H. Murad. 2013. Pioglitazone and risk of bladder cancer: A meta-analysis of controlled studies. Diabet. Med. 30(9):1026-1032.

Gerlitz, B., T. Hassell, C.J. Vlahos, J.F. Parkinson, N.U. Bang, and B.W. Grinell. 1993. Identification of the predominant glycosaminoglycan-attachment site in soluble recombinant human thrombomodulin: Potential regulation of functionality by glycosyltransferase competition for serine474. Biochem. J. 295(Pt 1):131-140.

Hedges, J.C., C.A. Singer, and W.T. Gerthoffer. 2000. Mitogen-activated protein kinases regulate cytokine gene expression in human airway myocytes. Am. J. Respir. Cell Mol. Biol. 23(1):86-94.

Hobbs, H.H., M.S. Brown, and J.L. Goldstein. 1992. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum. Mutat. 1(6):445-466.

Johnsen, S.P., H. Larsson, R.E. Tarone, J.K. McLaughlin, B. Nørgård, S. Friis, and H.T Sørensen. 2005. Risk of hospitalization for myocardial infarction among users of rofecoxib, celecoxib, and other NSAIDs: A population-based case-control study. Arch. Intern. Med. 165(9):978-984.

Kliewer, A. 2003. The nuclear pregnane X receptor regulates xenobiotic detoxification. J. Nutr. 133(7 Suppl.):2444S-2447S.

Lu, N.Z., S.E. Wardell, K.L. Burnstein, D. Defranco, P.J. Fuller, V. Giguere, R.B. Hochberg, L. McKay, J.M. Renoir, N.L. Weigel, E.M. Wilson, D.P. McDonnell, and J.A. Cidlowski. 2006. International union of pharmacology. LXV. The pharmacology and classification of the nuclear receptor superfamily: Glucocorticoid, mineralocorti-coid, progesterone, and androgen receptors. Pharmacol. Rev. 58(4):782-797.

McGinnity, D.F., J. Collington, R.P. Austin, and R.J. Riley. 2007. Evaluation of human pharmacokinetics, therapeutic dose and exposure predictions using marketed oral drugs. Curr. Drug Metab. 8(5):463-479.

Morita, I. 2002. Distinct functions of COX-1 and COX-2. Prostag. Oth. Lipid M. 68-69:165-175.

Mukaida, N., A. Harada, and K. Matsushima. 1998. Interleukin-8 (IL-8) and monocyte chemotactic and activating factor (MCAF/MCP-1), chemokines essentially involved in inflammatory and immune reactions. Cytokine Growth Factor Rev. 9(1):9-23.

Noreen, Y., T. Ringbom, P. Perera, H. Danielson, and L. Bohlin. 1998. Development of a radiochemical cyclooxygenase-1 and -2 in vitro assay for identification of natural products as inhibitors of prostaglandin biosynthesis. J. Nat. Prod. 61(1):2-7.

Obach, R.S., F. Lombardo, and N.J. Waters. 2008. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab. Dispos. 36(7):1385-1405.

Patlewicz, G., N. Ball, P.J. Boogaard, R.A. Becker, and B. Hubesch. 2015. Building scientific confidence in the development and evaluation of read-across. Regul. Toxicol. Pharmacol. 72(1):117-133.

Rangwala, S.M., and M.A. Lazar. 2004. Peroxisome proliferator-activated receptor gamma in diabetes and metabolism. Trends Pharmacol. Sci. 25(6):331-336.

Regan, S.L., J.L. Maggs, T.G. Hammond, C. Lambert, D.P. Williams, and B.K. Park. 2010. Acyl glucuronides: The good, the bad and the ugly. Biopharm. Drug Dispos. 31(7):367-395.

Rhen, T., and J.A. Cidlowski. 2005. Antiinflammatory action of glucocorticoids- new mechanisms for old drugs. N. Engl. J. Med. 353(16):1711-1723.

Rotroff, D.M., B.A. Wetmore, D.J. Dix, S.S. Ferguson, H.J. Clewell, K.A. Houck, E.L. Lecluyse, M.E. Andersen, R.S. Judson, C.M. Smith, M.A. Sochaski, R.J. Kavlock, F. Boellmann, M.T. Martin, D.M. Reif, J.F. Wambaugh, and R.S. Thomas. 2010. Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol. Sci. 117(2):348-358.

Schöbitz, B., E.R. de Kloet, and F. Holsboer. 1994. Gene expression and function of interleukin I, interleukin 6 and tumor necrosis factor in the brain. Prog. Neurobiol. 44(4):397-432.

Shin, H.M., A. Ernstoff, J.A. Arnot, B.A.Wetmore, S.A. Csiszar, P. Fantke, X. Zhang, T.E. McKone, O. Jolliet, and D.H. Bennett. 2015. Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays. Environ. Sci. Technol. 49(11):6760-6771.

Shipkova, M., V.W. Armstrong, M. Oellerich, and E. Wieland. 2003. Acyl glucuronide drug metabolites: Toxicological and analytical implications. Ther. Drug Monit. 25(1):1-16.

Singh, S., Y.K. Loke, and C.D. Furberg. 2007. Long-term risk of cardiovascular events with rosiglitazone: A meta-analysis. JAMA 298(10):1189-1195.

Spiegelman, B.M. 1998. PPAR-gamma: Adipogenic regulator and thiazolidinedione receptor. Diabetes 47(4):507-514.

Stepan, A.F., D.P. Walker, J. Bauman, D.A. Price, T.A. Baillie, A.S. Kalgutkar, and M. Aleo. 2011. Structural alert/reactive metabolite concepts as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicitiy: A perspectove based on the critical examination of trends in the top 200 drugs marketed in the United States. Chem. Res. Toxicol. 24(9):1345-1410.

Strong, C.G., and D.F. Bohr. 1967. Effects of prostaglandins E1, E2, A1, and F1-alpha on isolated vascular smooth muscle. Am. J. Physiol. 213(3):725-733.

Süleyman, H., B. Demircan, and Y. Karagöz. 2007. Anti-inflammatory and side effects of cyclooxygenase inhibitors. Pharmacol. Rep. 59(3):247-258.

Truong, L., D.M. Reif, L. St Mary, M.C. Geier, H.D. Truong, and R.L. Tanguay. 2014. Multidimensional in vivo hazard assessment using zebrafish. Toxicol. Sci. 137(1):212-233.

Vane, J.R. 1971. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. Nat. New Biol. 231(25):232-235.

Wada, T., J. Gao, and W. Xie. 2009. PXR and CAR in energy metabolism. Trends Endocrinol. Metab. 20(6):273-279.

Walker, G.S., J. Atherton, J.Bauman, C. Kohl, W. Lam, M. Reily, Z. Lou, and A. Mutlib. 2007. Determination of degradation pathways and kinetics of acyl glucuronides by NMR spectroscopy. Chem. Res. Toxicol. 20(6):876-886.

Wang, J., M. Davis, F. Li, F. Azam, J. Scatina, and R. Talaat. 2004. A novel approach for predicting acyl glucuronide reactivity via Schiff base formation: Development of rapidly formed peptide adducts for LC/MS/MS measurements. Chem. Res. Toxicol. 17(9):1206-1216.

Watkins, P.B. 2005. Insight into hepatotoxicity: The troglitazone experience. Hepatology 41(2):229-230.

Wetmore, B.A., J.F. Wambaugh, S.S. Ferguson, M.A. Sochaski, D.M. Rotroff, K. Freeman, H.J. Clewell, III, D.J. Dix, M.E. Andersen, K.A. Houck, B. Allen, R.S. Judson, R. Singh, R.J. Kavlock, A.M. Richard, and R.S. Thomas. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol. Sci. 125(1):157-174.

Wiemer, A.J., S. Hegde, J.E. Gumperz, and A. Huttenlocher. 2011. A live imaging cell motility screen identifies prostaglandin E2 as a T Cell stop signal antagonist. J. Immun. 187(7):3663-3670.

Xue, J., F. Chen, J. Wang, S. Wu, M. Zheng, H. Zhu, Y. Liu, J. He, and Z. Chen. 2015. Emodin protects against concanavalin A-induced hepatitis in mice through inhibiting activation of the p38 MAPK-NF-κB signaling pathway. Cell Physiol. Biochem. 35:1557-1570.

Zou, Q., W. Hong, Y. Zhou, Q. Ding, J. Wang, W. Jin, J. Gao, G. Hua, and X. Xu. 2016. Bone marrow stem cell dysfunction in radiation-induced abscopal bone loss. J. Orthop. Surg. Res. 11:3.

# E

# A Bayesian Example: Predicting Dose–Response Relationships from High-Throughput Data and Chemical Structure

This appendix illustrates the use of Bayesian methods to address a common problem in the analysis of high-throughput data that have relatively large measurement error for the purpose of characterizing dose–response relationships. Bayesian methods can be particularly useful for synthesizing data and quantifying uncertainty. To illustrate the utility of Bayesian methods for datasets that have diverse features, the committee provides an analysis that links two types of data that are captured in two distinct datasets. The first dataset contains measurements of dose–response relationships of 969 chemicals on one specific end point related to the activation of the nuclear pregnane X receptor (PXR) pathway. PXR is involved in the sensing of and initiation of metabolism in response to xenobiotics that enter the body and has a role in lipid homeostasis. Activation of the PXR pathway is associated with beneficial and injurious processes, and measurements of the activation of PXR provide information about the biological activity of a chemical. The data on PXR activation were taken from the US Environmental Protection Agency ToxCast Phase II data in the AttaGene test system, which uses a HepG2 human liver hepatoma cell line to measure transcription factor activity through gene expression (Judson et al. 2010a,b). The second dataset contains information about the structures of the tested chemicals. It characterizes each chemical structure according to 39 features, which are the major principal features extracted from 770 chemical descriptors produced by the Mold2 program (Hong et al. 2008). The features describe the structure of each of the 969 chemicals in the dataset. The exercise involves the quantitative structure–activity relationship (QSAR) task of relating chemical structure to a dose–response curve. The information can be used to reduce the uncertainty in the dose–response relationship for PXR activation measured for a chemical and to predict the dose–response relationship for an untested chemical.

The task of relating chemical structures to dose–response curves is challenging because of the large number of potentially relevant chemical features and the lack of prior knowledge relating the features to the dose–response curves for the outcome being studied (PXR activation). Simple statistical QSAR models that do not allow for interactions among the structural features are expected to have poor performance and to underestimate the uncertainty in the prediction. In contrast, more complex statistical approaches, such as flexible Bayesian models, allow relationships between different types of data to be unknown beforehand while borrowing information and allowing learning of lower-dimensional structure. By fitting a single Bayesian hierarchical model to the entire set of chemical-structure descriptors and dose–response curves, the model can adapt the width of the uncertainty bands accordingly and accurately reflect the scope of available information. This full Bayesian approach thus extends the standard QSAR concept of domain of applicability and provides flexible and adaptive measures of uncertainty.

Figure E-1 shows the raw dose–response data for PXR activation by the chemicals under consideration. As expected for so many chemicals that have broadly different chemical structures, the dose–response relationships are highly variable. To predict dose–response values of a new chemical only on the basis of information available on its chemical structure, it is important to predict the dose–response curve with a good appraisal of the uncertainty in the prediction. The accuracy of a prediction depends partly on whether a chemical in the training dataset is similar in structure to the new chemical under consideration.

To capture nonlinear relationships between dose and response and how the shapes of the relationships are associated with different chemical structures, two assumptions are made: that each dose–response curve is continuous (that is, no "jumps") and that when two chemicals

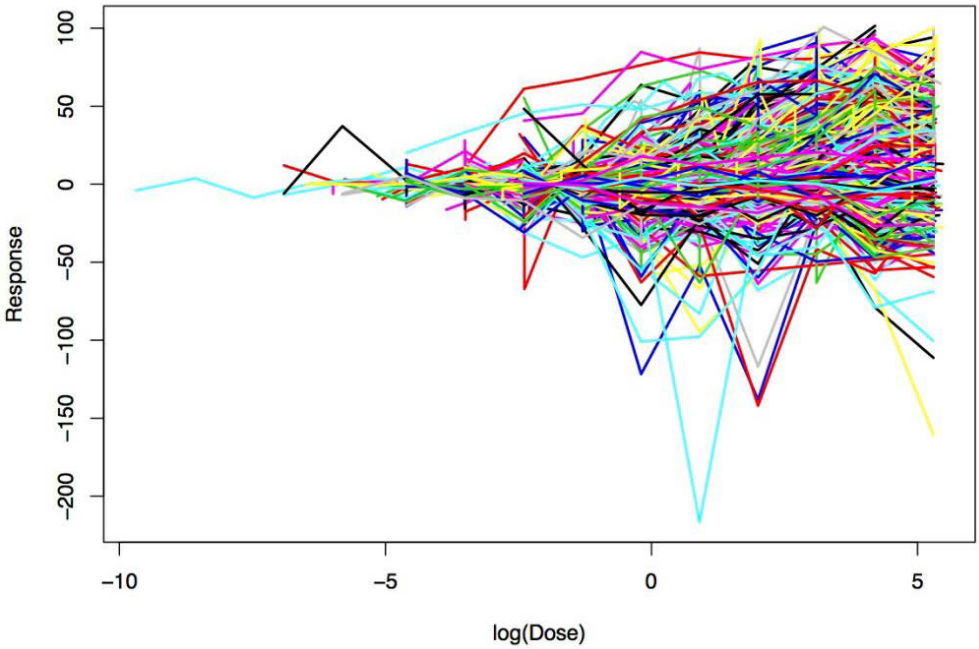



**FIGURE E-1** Dose–response records of PXR activation for 969 chemicals represented in the AttaGene ToxCast Phase II data. Dose is presented as concentration (μM) and response as fold increase or decrease in transcription.

are structurally alike (defined by a distance metric) their dose–response curves are similar.

Nonparametric Bayesian approaches provide a convenient framework for applying the two assumptions for curve estimation. Specifically, the dose–response curves are allowed to be completely unknown instead of our assuming that the curves follow a particular parametric form, such as a Hill function. That is accomplished by choosing a prior probability distribution for the entire curve. There is a rich literature on such priors; the Gaussian processes (GPs) provide a commonly used choice that is routinely used for many applications. For example, GPs are used routinely in epidemiological studies that collect information on spatial locations to incorporate "random effects" that characterize unmeasured spatially indexed covariates, which might act as confounders.

In the present setting, a GP prior is chosen that allows the dose–response curves to change flexibly according to chemical dose and chemical-structural features. Under the Bayesian nonparametric model used, two response measurements are assumed to be highly correlated a priori when the doses are similar and the chemical structures are similar, and the correlation gradually decays as doses and structural features move farther apart. The GP prior is chosen to allow wide uncertainty in the unknown curves before including information in the database. If one generated samples from the prior, the credible bands (Bayesian versions of confidence bands) would be wide. However, if the prior distribution is updated with information in the full dataset (not just for a single chemical but for all 969 chemicals), a much more accurate estimate of the curve and narrower credible bands are obtained.

Figure E-2 shows, after fitting of the model, the estimated dose–response curve and 95% credible bands for one chemical with the observed PXR dose–response data on that chemical. The figure shows that the estimated curve provides a good fit to the data with narrow uncertainty bands. The estimated curve differs somewhat from that obtained by estimating the dose–response curve nonparametrically on the basis of data only on that chemical (not shown); in particular, the uncertainty bands are narrower, and the curve is shifted slightly from a simple interpolation of the means at each dose. Those properties reflect the borrowing of information on chemicals that have related structures.

In addition to improving estimation of the dose–response curve for chemicals on which there are direct dose–response data, the approach can be used to predict dose–response curves for chemicals on which there is information only on structural features. For a chemical that has a known structure but lacks dose–response data, the actual experimental data can be replaced with a model-

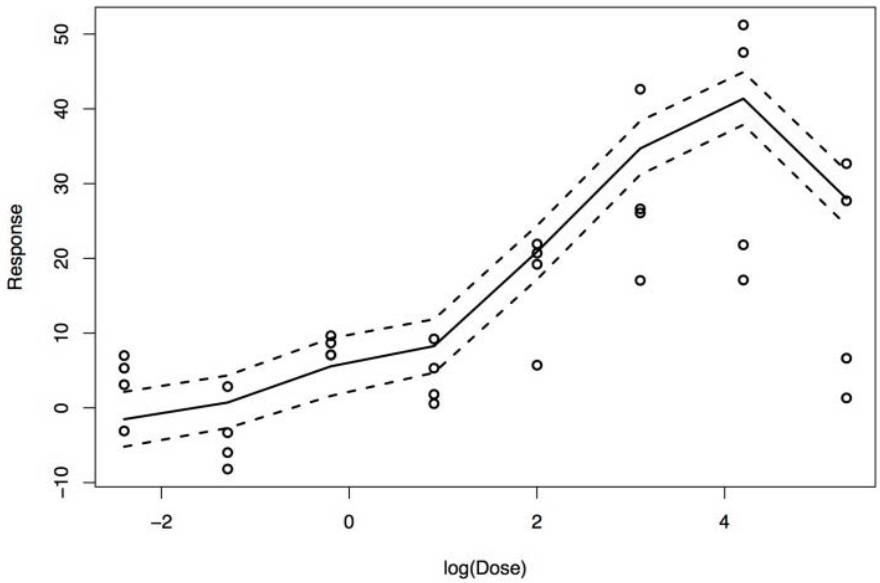



**FIGURE E-2** Estimated dose–response curve (solid line) for PXR activation and 95% credible interval (dashed lines) for one chemical. The credible interval is for the mean curve and so is not expected to enclose most of the data points (circles). The estimated dose–response curve is based on the full QSAR and PXR datasets for 969 chemicals in addition to the data points shown. Dose is presented as concentration (μM) and response as fold increase or decrease in transcription.

based statistical prediction. That prediction will be more accurate for chemicals that are structurally similar to chemicals in the database.

To illustrate the performance of the Bayesian modeling, the committee used data on 800 chemicals as training data on which to base the relationships between chemical structure and PXR dose–response relationship by fitting a Bayesian hierarchical model. The committee set aside the structure and PXR dose–response data on the remaining 169 chemicals. To illustrate predictive accuracy, the committee then compared the predicted curves and credible bands with the held-out data.

Figure E-3 shows predicted PXR dose–response relationships for two chemicals drawn from the 169 chemicals that were not used in the development of the Bayesian predictive model. Thus, the data points shown in the figure were not used in predicting the dose–response curve and estimating the uncertainty bands. Note also that the uncertainty bands are wider than those shown in Figure E-2, as expected because the bands in Figure E-2 include direct observations of the dose–response curve, and the dose–response prediction in Figure E-3 bases the estimated relationship only on chemical-structure information. For one chemical, shown first in Figure E-3, there is not a strong observed relationship between chemical dose and PXR activation, and the predicted dose–response relationship accordingly reflects a lack of clear dose-response, at least at lower doses. The dose-response relationship for the second chemical is more defined, as are the direct observations of the dose-response relationship that were not used to create the curve shown. The curve and confidence bands provide a relatively good fit to the observations.

Although Figure E-3 shows only two chemicals for illustration, good performance was observed across the 169 "test" chemicals. In cases in which the estimated dose–response curve had wide uncertainty bands indicating uncertainty in the prediction, the bands were wide enough to contain the curves providing a good fit to the observed data on the chemical.

This example illustrates the utility of Bayesian methods for data integration. Primary advantages are flexibility, the ability to borrow information from different data types, and uncertainty quantification. The committee used a nonparametric Bayesian approach with a GP prior; there is an increasing literature on applying similar approaches in a rich variety of applications, and there are many packages for routinely fitting GP-based models in practice (Vanhatalo et al. 2013). As illustrated in this example, flexible Bayesian hierarchical modeling avoids overly restrictive parametric assumptions that might not

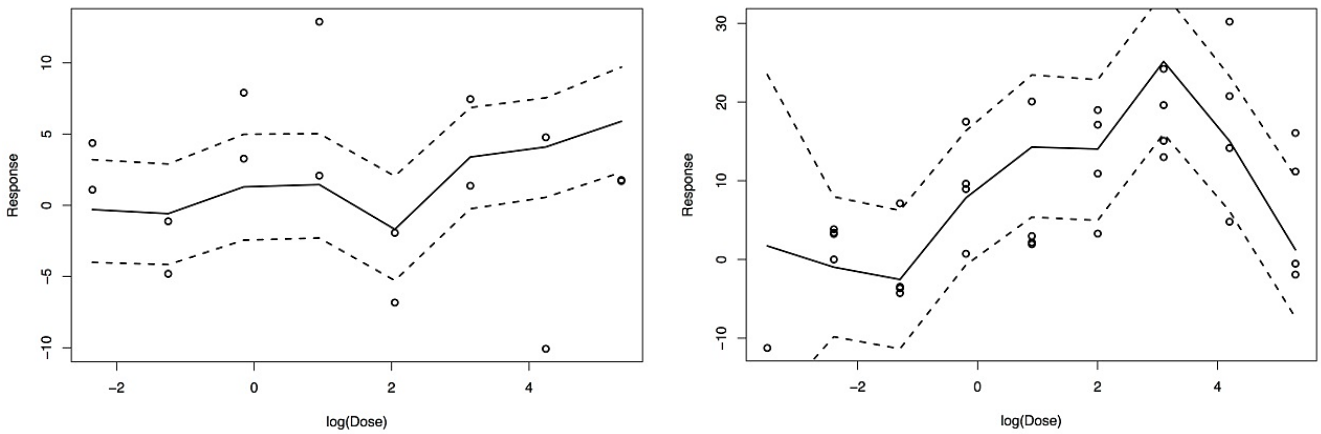



**FIGURE E-3** The predicted dose–response curves (solid line) and 95% credible intervals (dashed lines) for PXR activation for two chemicals. Dose is presented as concentration (μM) and response as fold increase or decrease in transcription. The predictions, based only on chemical structures, match the observed responses (circles) well. That is, data on the chemicals shown were not used to build the Bayesian model used to make the predictions.

be justifiable biologically while allowing incorporation of information from different data sources adaptively. In this context, *adaptively* means that one learns the similarities in the data sources and how much it makes sense to use the sources as reflected in the uncertainty bands. The increasingly large databases of results for a variety of assays and chemicals can thus be used to inform the current analysis and interpretation and eventually can support the collection of fewer data on future chemicals as the relationships among chemicals and disparate end points are increasingly understood and reflected in good predictive models.

## REFERENCES

Hong, H., Q. Xie, W. Ge, F. Qian, H. Fang, L. Shi, Z. Su, R. Perkins, and W. Tong. 2008. Mold(2), molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. J. Chem. Inf. Model. 48(7):1337-1344.

Judson, R.S., M.T. Martin, D.M. Reif, K.A. Houck, T.B. Knudsen, D.M. Rotroff, M. Xia, S. Sakamuru, R. Huang, P. Shinn, C.P. Austin, R.J. Kavlock, and D.J. Dixon. 2010a. Analysis of eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity. Environ. Sci. Technol. 44(15):5979-5985.

Judson, R.S., K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, D.M. Reif, D.M. Rotroff, I. Shah, A.M. Richard, and D.J. Dix. 2010b. In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project- Supplemental Information. Environmental Health Perspect. 118(4):485-492 [online]. Available: http://ehp.niehs.nih.gov/wp-content/uploads/118/4/ehp.0901392.s001.pdf [accessed November 15, 2016].

Vanhatalo, J., J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, and A. Vehtari. 2013. GPstuff: Bayesian modeling with Gaussian processes. J. Mach. Learn. Res. 14:1175-1179.