Session BS10: Current and evolving concepts for the validation of safety assessment methods

Validation of Innovative Technologies and Strategies for Regulatory Safety Assessment Methods: Challenges and Opportunities

William S. Stokes[1] and Marilyn Wind[2]

[1]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA; [2]US Consumer Product Safety Commission, Bethesda, MD, USA NC, USA

Summary

Advances in science and innovative technologies are providing new opportunities to develop test methods and strategies that may improve safety assessments and reduce animal use for safety testing. These include high throughput screening and other approaches that can rapidly measure or predict various molecular, genetic, and cellular perturbations caused by test substances. Integrated testing and decision strategies that consider multiple types of information and data are also being developed. Prior to their use for regulatory decision-making, new methods and strategies must undergo appropriate validation studies to determine the extent that their use can provide equivalent or improved protection compared to existing methods and to determine the extent that reproducible results can be obtained in different laboratories. Comprehensive and optimal validation study designs are expected to expedite the validation and regulatory acceptance of new test methods and strategies that will support improved safety assessments and reduced animal use for regulatory testing.

Keywords: validation, safety testing, integrated testing strategies, integrated decision strategies

1 Introduction

Safety assessment methods are necessary to determine if new chemicals and products are safe or if they may adversely affect the health of people, animals, and the environment. Advances in science and innovative technologies are providing new opportunities to develop test methods and strategies that may improve safety assessments and reduce animal use for safety testing. Research continues to improve our understanding of the molecular and cellular alterations by which chemical exposures can cause or contribute to injury or disease. High throughput screening, toxicogenomics, and other approaches can now be used to rapidly measure many of the molecular, genetic, and cellular perturbations caused by chemicals. Robot operated laboratories can rapidly generate vast amounts of *in vitro* data for thousands of chemicals (Michael et al., 2008). Analysis of this data is ex-

pected to help identify panels of *in vitro* biomarkers that can be used to help assess chemical toxicity. Integrated testing strategies that consider information and data from such assays and various test methods are also being developed (Stokes, 2007).

Prior to their use for regulatory decision-making, new methods and strategies must undergo appropriate validation studies to determine if they can provide equivalent or improved protection compared to existing methods and to determine if reproducible results can be obtained in different laboratories (ICCVAM, 1997, 2003; OECD, 2005). Validation studies must be carefully designed to optimize test methods and to ensure that they generate adequate data for decisions on their regulatory acceptability (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Adequate validation will expedite the acceptance and use of new test methods and strategies that support improved safety

assessments and contribute to reduced animal use for regulatory testing. This paper will discuss emerging innovative technologies, concepts, and approaches applicable to regulatory safety assessments, and opportunities and challenges for their scientific validation.

2 Changing the paradigm of toxicity testing

Two recent reports have proposed using advances in science and technology to change the current paradigm of toxicity testing. These include the 2004 National Toxicology Program Roadmap, and the 2007 National Research Council (NCR) publication, *Toxicology in the 21st Century, A Vision and a Strategy* (NTP, 2004; NRC, 2007a). The NTP Roadmap envisions moving from toxicology studies that depend on observing the actual adverse outcome from chemical exposures, such as cancer and birth defects in animal models, to one based on understanding and detecting cellular and molecular perturbations in simpler models such as cell cultures and lower organisms that are predictive of these eventual adverse outcomes. To implement this vision, the NTP plan is to develop and validate improved testing methods and to ensure, where feasible, that such methods provide for the reduction, refinement, and replacement of animals.

The NTP report emphasizes that activities and assays developed under the NTP Roadmap will be done in cooperation and consultation with the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to maximize their value to regulatory agencies (NTP, 2004). ICCVAM is a U.S. interagency committee composed of 15 research and regulatory agencies that is charged with evaluating the scientific validity of new, revised, and alternative test methods proposed for regulatory testing (ICCVAM, 2003; Stokes and Schechtman, 2007).

The 2007 NRC report similarly envisions future testing based on an understanding of key toxicity pathways at the cellular and molecular levels and using predictive high throughput assays to detect the potential for chemicals to sufficiently alter these pathways to cause injuries or disease. The report states that the use of a comprehensive array of *in vitro* tests to identify relevant biological perturbations based on human biology could eventually eliminate the need for whole-animal testing and provide a stronger mechanistically based approach for environmental decision-making. However, a 2009 NRC report states that the realization of the promise of this vision is at least a decade away (NRC, 2009).

3 Emerging science and technology

New scientific advances and innovative technologies are now available to help develop future testing methods and strategies outlined in the NRC and NTP reports. These include high throughput screening, toxicogenomics, and computational modeling approaches.

High throughput screening involves the use of computerized robots to conduct the laboratory procedures necessary to study hundreds of compounds per day in multiple *in vitro* assays. The National Chemical Genomics Center at the National Human Genome Research Institute has a laboratory where such studies are conducted (Michael et al., 2008). In collaboration with the NTP and EPA, the lab is now conducting quantitative HTS using fifteen concentrations of each chemical (Collins et al., 2008). The lab uses 1536-well plates, which have a net testing capability of 1504 individual chemicals. Over 100,000 concentration response profiles can be generated per week. These profiles are then evaluated to determine if *in vitro* biomarker alterations are associated with known adverse health effects. Bioinformatics techniques will be used to identify complex relationships between different types of biological responses that may provide insights into critical toxic pathways (Schmidt, 2009).

Another NRC report published in 2007 addressed the application of toxicogenomic technologies to predictive toxicology and risk assessment (NRC, 2007b). Toxicogenomics is defined as the application of genomic technologies to study the adverse effects of environmental and pharmaceutical chemicals on human health and the environment. These technologies include genetics, genome sequence analysis, gene expression profiling, proteomics, metabolomics, and other related approaches. These are used to measure chemical-specific perturbations on expression patterns of genes, proteins, and metabolites in cells, tissues, and organisms. Such technologies are being investigated for their potential to improve the prediction of safety or potential hazards of chemicals to human health.

Computational modeling is being applied to estimate the absorption, distribution, metabolism, and excretion of chemicals (ADME) (NRC, 2007a). These models seek to estimate the relationship between the dose or amount of chemical exposure via oral, dermal, or inhalation routes, and the concentration of chemical that reaches individual cell types in various critical organs and tissues. These estimates will be essential for non-animal estimates of exposure levels that are safe and those that are likely to be associated with toxic effects. It is also important that data used to construct computational models is of high quality and derived from adequately designed studies.

4 Application of new science and technology to regulatory decision-making

As emerging scientific advances provide insights into the pathways and mechanisms of chemical toxicity, the National Toxicology Program and other public health agencies seek to apply this information so that it can be used to improve public health. Several recent and planned activities and initiatives have and will continue to investigate potential applications for public health decision-making. For example, at the request of the National Institute of Environmental Health Sciences, the National Academies recently formed a Standing Committee on the Use of Emerging Science for Environmental Health Decisions (NAS, 2009). The committee is charged with facilitating communication among government, industry, environmental groups, and the academic community about scientific advances that may be used in the identification, quantification, and

control of environmental impacts on human health. The topics covered will build on recent NRC reports on toxicity testing and toxicogenomics and will explore new developments in toxicology, molecular biology, bioinformatics, and related fields (NRC, 2007a, 2007b). Three workshops have been or will be held in the near future. (Fig. 1)

Mechanistic toxicity data from animal studies and humans are necessary to link *in vitro* pathway data to adverse health effects. To address this need, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Toxicological Methods (ICCVAM) recently convened an *International Workshop on Acute Chemical Safety Testing – Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations* (NICEATM, 2008). The primary goals of the workshop were to identify approaches for collecting additional mechanistic data from current *in vivo* testing that would support the development of predictive mechanism-based *in vitro* alternative models and that could also be used to identify earlier more humane endpoints.

5 Validation and acceptance of test methods based on new science and technology

In the United States, Federal laws require that new safety assessment methods proposed for regulatory safety assessment decisions must be determined to be sufficiently valid and acceptable for their proposed use (USC, 2000). National and internationally harmonized principles for validation and regulatory acceptance are available (ICCVAM, 1997; OECD, 2005). Determination of validity involves assessing the accuracy and reliability of the test method for a specific proposed purpose (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Accuracy assessments typically characterize sensitivity, specificity, and false positive and negative rates compared to existing reference data. Regulatory acceptance decisions involve reviewing the validation database to determine if the proposed use of the method

for decision-making will provide equivalent or improved protection compared to existing methods (USC, 2000). Reliability assessments determine if reproducible results can be obtained in different laboratories when using the proposed standardized test method protocol.

National and International authorities have agreed on validation and regulatory acceptance criteria for new, revised, and alternative test methods (ICCVAM, 1997; OECD, 2005). These are general criteria that should be appropriately addressed when considering the validity of test methods. The published criteria emphasize that flexibility is essential in interpreting and applying the criteria and that the extent that they will need to be addressed will depend on the intended purpose and nature of the proposed test (ICCVAM, 1997; OECD, 2005).

6 Validation of new science and technologies: challenges

New test methods based on scientific advances and technologies are likely to initially have limitations. Early definition of a test method's limitations can contribute to more efficient validation for the initial proposed uses and aid in identifying directed research to discover ways to address defined limitations. In some cases, test methods may be limited in terms of the physical and chemical properties of substances that can be tested. For example, the current NCGC HTS protocol is only capable of testing substances soluble in DMSO, so those that are not soluble cannot be adequately evaluated in this test system. The highest concentration that can be achieved in a test system may be limited by solubility in the required vehicle, which may not be sufficient for regulatory testing purposes. A significant limitation of most current *in vitro* testing methods is their inability to determine if there is metabolic activation of the substance to a more toxic or less toxic moiety. Additionally, there are still challenges in accurately estimating the toxicokinetics associated with specific exposures by various routes and the concentrations that will result in various critical target tissues. These limitations present challenges that will need to be

---

2009 Workshops:
National Academies' Standing Committee on Use of Emerging Science
for Environmental Health Decisions

- Use of Emerging Science and Technologies to Explore Epigenetic Mechanisms Underlying the Developmental Basis for Disease, July 30-31, 2009.
  http://dels.nas.edu/envirohealth/epigenetic.shtml

- Computational Toxicology: From Data to Analyses to Applications, September 21-22, 2009.
  http://dels.nas.edu/envirohealth/comptox.shtml

- The Exposome: A Powerful Approach for Evaluating Environmental Effects on Chronic Diseases, February 25-26. 2010. http://dels-old.nas.edu/envirohealth/exposome.shtml

Fig. 1: 2009 Workshops: National academies' standing committee on use of emerging science for environmental health decisions

addressed in order to fully move away from the use of intact living organisms for safety assessments.

Another significant challenge for evaluating the validity of new testing methods and strategies for human health safety assessments is the availability of high quality reference data from humans. For ethical reasons, most existing reference data is from animal studies. However, for some toxicity endpoints such as allergic contact dermatitis (ACD), there is considerable human testing data and experience from occupational and consumer exposures (ICCVAM, 1999; Basketter et al., 2007). These human data supported the validity of a new animal model for ACD testing that has many scientific and animal welfare advantages compared to the traditional animal tests for ACD. Improved ways of obtaining data regarding the health effects from human exposures and ways to more accurately extrapolate exposures and effects from animal models to humans are needed to help validate new test methods.

7 Validation of new science and technologies: opportunities

Early consideration of the potential application of new technologies for regulatory testing during research and development stages provides an important opportunity to incorporate efforts that will support the validation of eventual test methods. Early standardization and use of harmonized technology platforms for approaches such as toxicogenomics and HTS will allow for data from different studies to be compared and combined for data analyses. This will also help minimize experimental variables, aid in achieving more reproducible results across labs, and contribute to achieving a high signal to noise ratio. For example, a recent workshop developed recommendations for the standardization and validation of toxicogenomic-based platforms that will be evaluated for their potential use for safety assessments (Corvi et al., 2006).

There is also an opportunity to develop data during research and development that may contribute to the validation database supporting the validity of proposed test methods and approaches. Several critical factors should be considered during research, development, translation, and validation stages for new technologies. These include selection of reference substances, dose/concentration selection procedures, defining the test method purpose and potential regulatory use, and phased validation studies to develop an optimized test method protocol.

Reference Substances: Reference substances selected for evaluation of the new technology should have high quality data available from existing reference test methods or the species of interest for the toxicity endpoint under evaluation (ICCVAM, 2003; Stokes and Schechtman, 2007). Selection of reference data should generally address established selection criteria for reference substances (Stokes and Schechtman, 2007), which include:
– Represent the dynamic range of responses possible for the toxicity endpoint of interest and the range of potential responses that can be measured in the test system

– Represent the range of physical and chemical properties of substances for which the test system is proposed to be capable of testing (e.g., physical form, water solubility, pH, volatility)
– Represent the range of relevant biologic properties, as appropriate (e.g. peptide reactivity, mutagenicity)
– Represent the range of chemistry of substances proposed for evaluation in the new test method (i.e., chemical classes)
– Represent the range of known or suspected modes or mechanisms of action for the toxicity measured or predicted by the test method
– Supported by existing high quality data from the currently accepted test method, and where possible, data and/or experience in the species of interest (e.g. for humans, ethical test data or accidental exposures information)
– Readily available from commercial sources
– Avoidance of chemicals with excessive occupational or environmental hazard, if feasible.

Dose or concentration-setting procedures: The basis and procedures for determining the highest dose or concentration that will be tested should be clearly stated. For animal-based tests this is normally based on the highest minimally toxic dose (MTD) or a defined upper limit dose. For *in vitro* tests, this is normally the highest soluble concentration, the highest non-cytotoxic concentration, or a defined upper concentration based on the highest potential exposure that might occur (Stokes, 2006).

Test method purpose and regulatory applicability: The specific proposed purpose of the test method and the proposed or potential use for regulatory decision-making in the context of current or anticipated regulatory requirements should be clearly defined (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Proposed uses may range from serving as a complete replacement for a current existing test method to providing adjunct mechanistic data for weight-of-evidence decisions. For test methods proposed for use in chemical screening, the specific decisions that can be made with each possible test result must be clearly defined. For example, a positive result in a screening method might be used as the basis for hazard classification and labeling, while negative results associated with sufficient uncertainty may require further testing. Screening tests may also be proposed for prioritization decisions on whether further testing will or should be conducted. In such casest the uncertainty of the prediction of potential hazard or safety for a specific toxic endpoint should be characterized and transparent for the prioritization decision.

Phased validation studies: optimizing the test method protocol: Recent *in vitro* validation studies managed by NICEATM have shown that a validation study design consisting of several sequential progressive phases with coded chemicals was an efficient means of optimizing the test method protocol and minimizing intra- and inter-laboratory variation (Fig. 2) (ICCVAM, 1997; OECD, 2005; Paris et al., 2006; Stokes et al., 2007; Stokes and Schechtman, 2007). The initial laboratory evaluation phase involves a series of multiple testing with positive and vehicle

controls, with cycles of protocol modifications until all labs are able to obtain sufficiently reproducible results. Two stages of the second phase each test a small number of chemicals representative of the potential range of responses and vehicle solubility. After each phase, excessive experimental variation and discordance are stage and appropriate modifications made to the protocol. Retesting is conducted where substantive modifications are deemed necessary to confirm the effectiveness of these changes for obtaining consistent results. The last phase uses the final optimized protocol to generate data to assess accuracy and reliability.

8 Validation of integrated testing strategies

Integrated testing strategies involve considering all available information and data to determine if decisions can be made about the safety or hazard of substances in a stepwise or tiered manner. These are usually designed to minimize or avoid the use of animals. If there is not sufficient information and data for a decision at the initial level or tier, then testing proceeds to the next tier where a decision is made as to what is the most appropriate additional testing to conduct that might provide sufficient information for hazard classification decisions. Generally the stepwise testing proceeds from existing information and data to *in vitro* tests, followed by limited *in vivo* testing, and then to a full traditional *in vivo* test as the final tier, if necessary.

Normally, validation of testing strategies can be made using existing data, provided that there is sufficient data on the same substances for all of the test methods proposed for the test strategy. In designing prospective studies for testing strategies, it is important to ensure that all test substances are tested in all of the proposed test components proposed for the testing strategy. Each test method is assessed individually to determine which results can be useful for a hazard classification decision either alone or in combination with the various potential outcomes of each of the other test methods in the strategy. This involves determining the sensitivity and specificity for each of these possible combinations of test outcomes and assessing which ones can provide equivalent or improved predictions compared to the current existing test method.
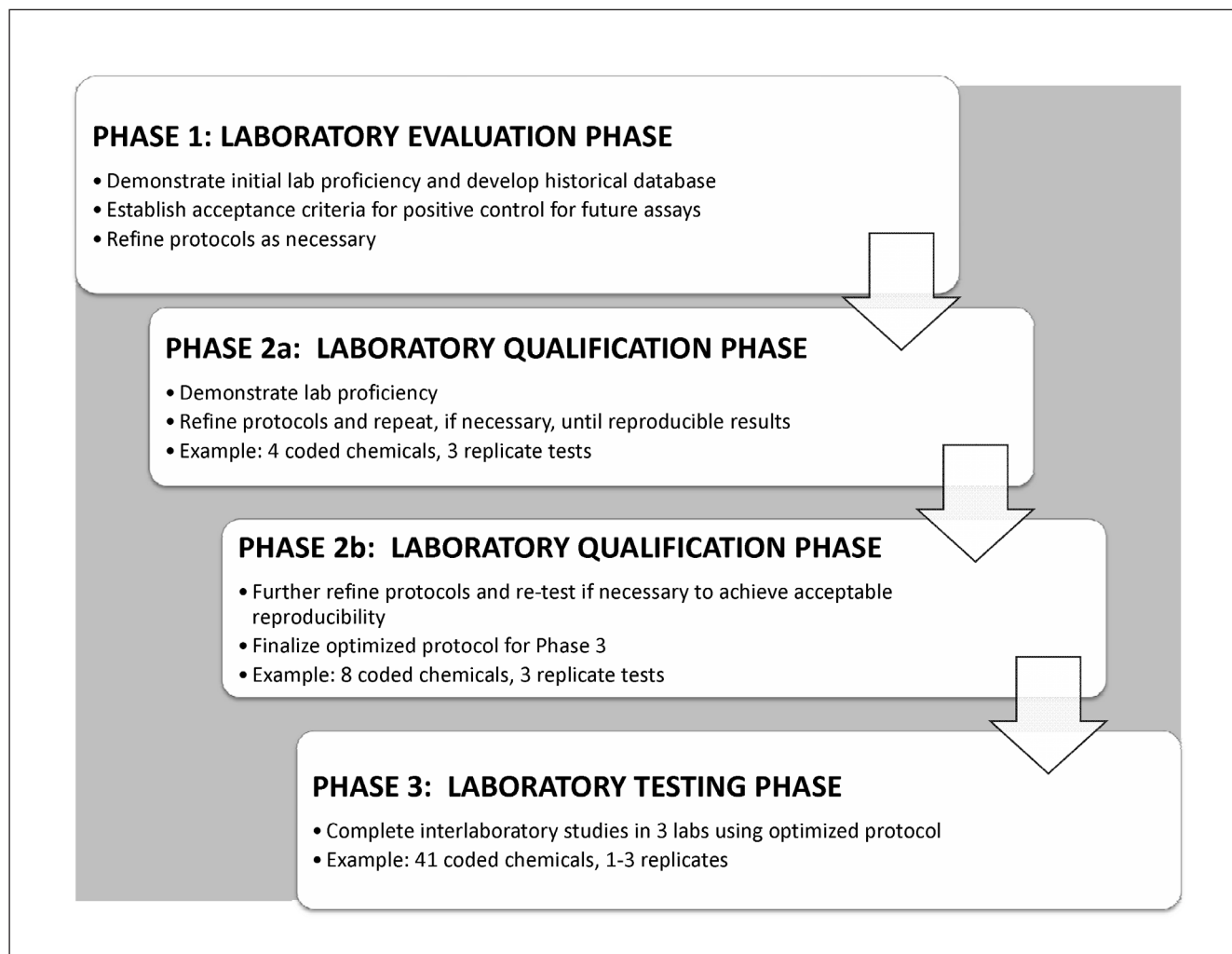

PHASE 1: LABORATORY EVALUATION PHASE
- Demonstrate initial lab proficiency and develop historical database
- Establish acceptance criteria for positive control for future assays
- Refine protocols as necessary

PHASE 2a: LABORATORY QUALIFICATION PHASE
- Demonstrate lab proficiency
- Refine protocols and repeat, if necessary, until reproducible results
- Example: 4 coded chemicals, 3 replicate tests

PHASE 2b: LABORATORY QUALIFICATION PHASE
- Further refine protocols and re-test if necessary to achieve acceptable reproducibility
- Finalize optimized protocol for Phase 3
- Example: 8 coded chemicals, 3 replicate tests

PHASE 3: LABORATORY TESTING PHASE
- Complete interlaboratory studies in 3 labs using optimized protocol
- Example: 41 coded chemicals, 1-3 replicates

Fig. 2: Validation study approach using sequential progressive phases

9 Integrated decision strategies

With some test methods, initially proposed single decision criteria may not provide sufficient certainty with regard to the predicted outcome for some specific results, while the remaining results may have sufficient certainty in terms of sensitivity and/or specificity. For example, a test method may have a false negative rate for a certain range of responses that is not considered adequately protective compared to the reference test method. Conversely, a test method may have a false positive rate for a certain range of responses that is sufficiently high so as to not be considered acceptable. In these situations, multiple decision criteria may be necessary, where each individual decision criteria provide sufficient certainty for responses within a specified range of test results. There may also be one or more decision criteria that identify a range of responses that are associated with an unacceptable level of uncertainty, and therefore should not be used for hazard or safety decisions. In this later
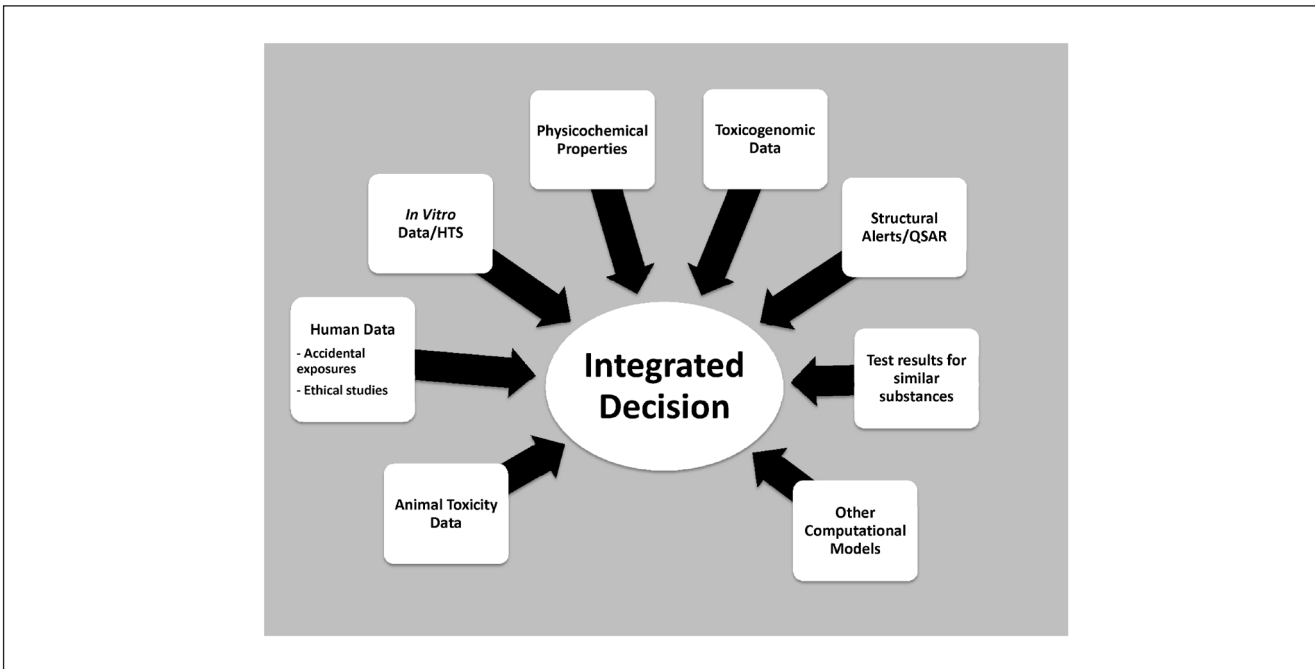

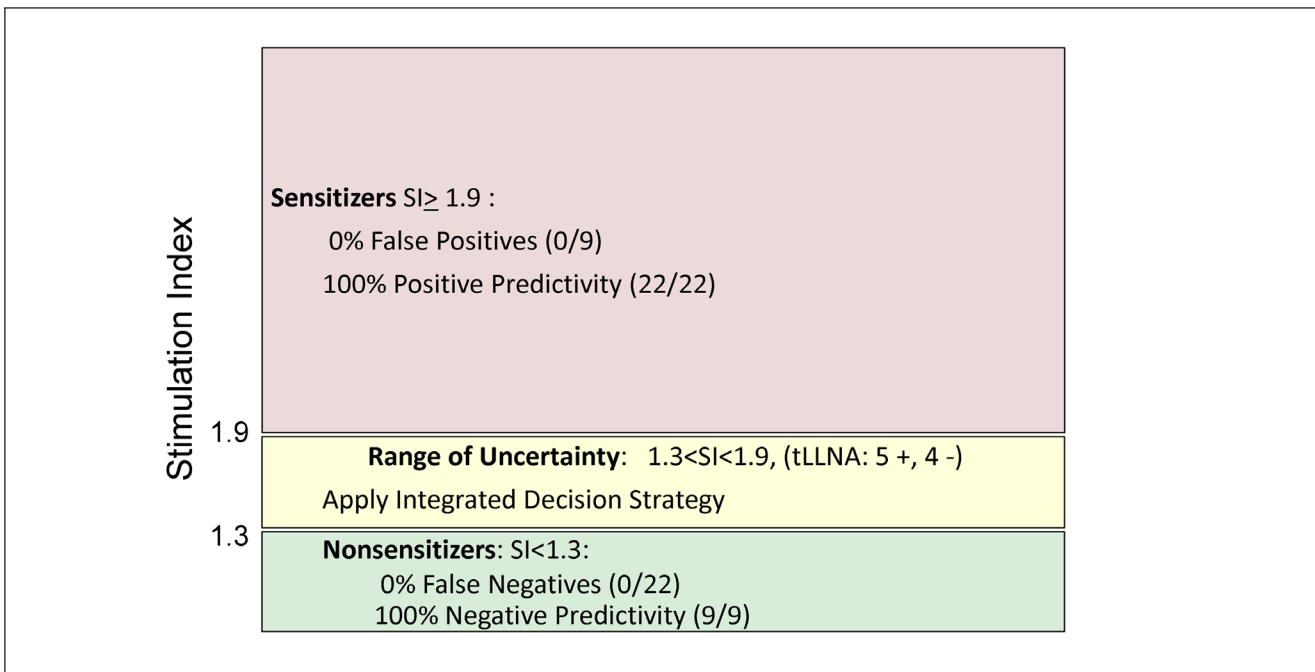**Fig. 3: Potential sources of data and information for integrated decision strategies**

Fig. 4: Integrated decision strategy – LLNA: BrdU-ELISA

situation, additional information or data could be used to reduce the uncertainty associated with these results using an integrated decision strategy to reach a hazard or safety decision. Integrated decision strategies using multiple sources of data and information can increase the certainty of hazard and safety predictions beyond the certainty associated with only a single source of data or information (Fig. 3).

Two examples of the need for integrated decision strategies are provided by the results of a recent ICCVAM test method peer review evaluation of two non-radioactive versions of the LLNA, the LLNA: DA and LLNA: BrdU-ELISA (ICCVAM, 2009a, 2009b, 2009c). For both test methods, a single decision criteria for whether a substance was a sensitizer or non-sensitizer could not be identified that would provide the same sensitivity and specificity as the traditional LLNA for the chemicals evaluated. However, in the LLNA: BrdU-ELISA, a decision criterion using a stimulation index (SI) ≥ 1.9 to classify substances as sensitizers was found to produce a false positive rate compared to the traditional LLNA of 0% [0/9] and a positive predictivity of 100% (22/22), which was obviously considered acceptable (Fig. 4)[1]. A second decision criterion of SI ≤ 1.3 to classify substances as non-sensitizers was found to produce a false negative rate compared to the traditional LLNA of 0% [0/22] and a negative predictivity of 100% (9/9) which also was considered acceptable (Fig. 4).

However, there were five sensitizers and four non-sensitizers among the reference test substances that produced an SI greater than 1.3 and less than 1.9 in the LLNA: BrdU-ELISA. Accordingly, SI results of greater than 1.3 and less than 1.9 were not considered sufficiently predictive to be used for hazard or safety decisions. To reduce the uncertainty associated with SI results in this range, additional information and data were considered necessary for evaluation in an integrated decision strategy to determine if this combined data would support a hazard decision. Additional information that could be considered included dose-response information, statistical comparison of treated vs. vehicle control groups, peptide reactivity, molecular weight, results from related substances, presence or absence of structural alerts, and *in vitro* and other testing data (Fig. 5).

The sensitivity and specificity associated with each of these other types of data or information must be available. For the LLNA: BrdU-ELISA, when an SI result occurs in the range of uncertainty, negative results for peptide reactivity and negative results in one or more *in vitro* assays for ACD were found to provide sufficient additional data to support a hazard decision as a non-sensitizer without resulting false negatives. This approach allowed an overall specificity of 100% for the validation database.

NICEATM and ICCVAM are currently assessing the other types of test method data and information that might be used in an integrated decision strategy for these two test methods. The sensitivity and specificity associate with the outcomes in each of these other types of data will need to be carefully assessed and incorporated into classification decisions. Successful applica-

---

[1] These results were obtained using the most prevalent outcome for substances that were tested multiple times.
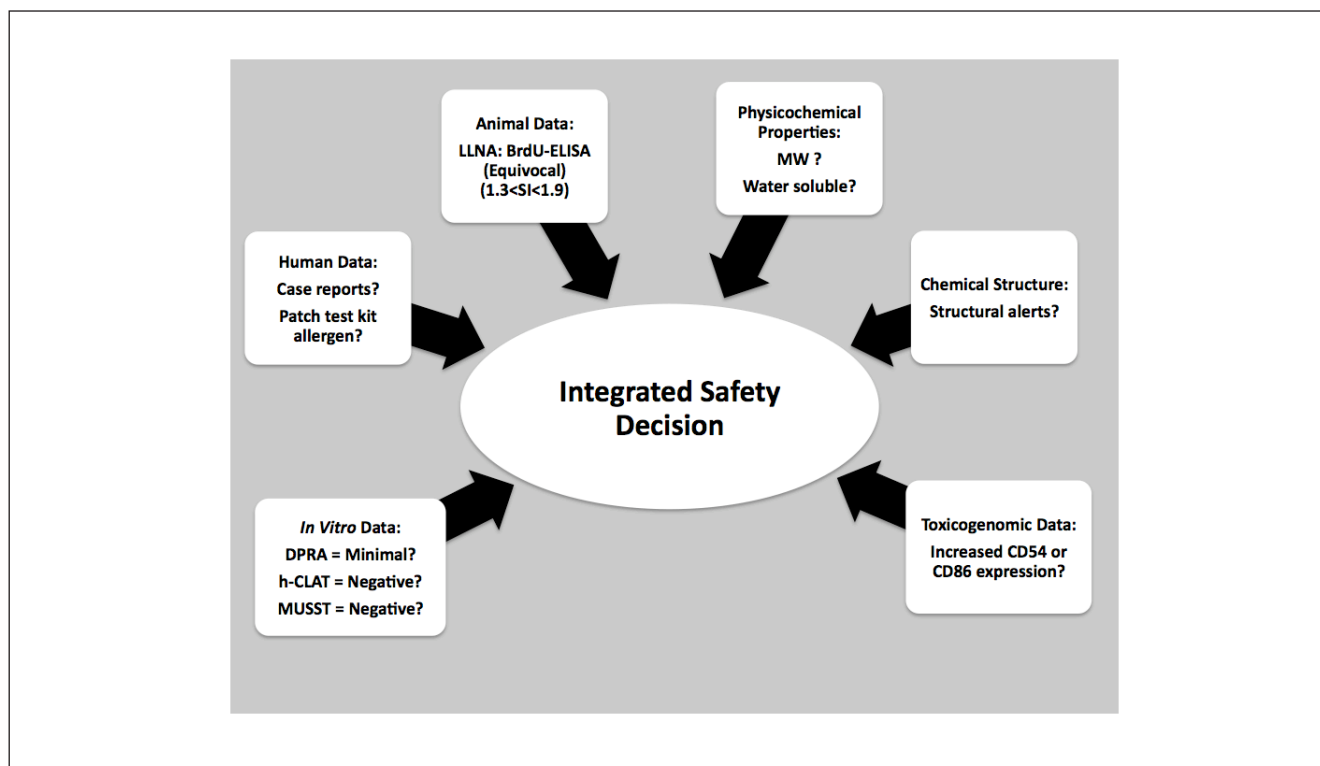

Fig. 5: Sources of potentially relevant data and information for an integrated decision strategy for uncertain results in the LLNA: BrdU-ELISA

tion of the integrated decision strategy approach is expected to produce acceptable classification decisions and avoid the need for additional testing.

10 Conclusions

Advances in science and innovative technologies are providing new opportunities to develop improved safety testing methods and strategies. Consideration of validation principles and potential application to regulatory decision-making during early stages of research, development, and validation will help expedite the scientific validation of these new methods and strategies. Validation databases will need to adequately characterize the usefulness and limitations of new proposed test methods and strategies, and support determinations of whether the new method or approach can provide equivalent or improved protection compared to existing test methods. New methods and integrated strategies should be developed and validated in consultation with relevant stakeholders and national validation centers in order to ensure adequate and appropriate studies. Comprehensive and optimal validation study designs are expected to expedite the validation and regulatory acceptance of new test methods and strategies that support improved safety assessments and contribute to reduced animal use for regulatory testing.

References

Basketter, D. A., Gerberick, G. F. et al. (2007). The local lymph node assay: current position in the regulatory classification of skin sensitizing chemicals. *Cutan. Ocul. Toxicol. 26(4),* 293-301.

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

Corvi, R., Ahr, H. J. et al. (2006). Meeting report: Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. *Environ. Health Perspect. 114(3),* 420-429.

ICCVAM (2009a). Independent Scientific Peer Review Panel Report: Updated Validation Status of New Versions and Applications of the Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemical and Products. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf (Accessed: September 17, 2009).

ICCVAM (2009b). Revised Draft Background Review Document: Non-Radioactive Murine Local Lymph Node Assay: Modified by Daicel Chemical Industries, Ltd. Based on ATP Content (LLNA: DA). Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/immunotox/llna-DA/BRDcomplete.pdf (Accessed: September 17, 2009).

ICCVAM (2009c). Revised Draft Background Review Document: Non-radioactive Murine Local Lymph Node Assay:

BrdU-ELISA Test Method Protocol. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/immunotox/llna-ELISA/BRDcomplete.pdf (Accessed: September 17, 2009).

ICCVAM (2003). ICCVAM Guidelines for the Nomination and Submission of New, Revised and Alternative Test Methods. NIH Publication Number 03-4508. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.pdf

ICCVAM (1999). The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 99-4494. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf (Accessed: September 17, 2009).

ICCVAM (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 Number 97-3981. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf (Accessed: September 17, 2009).

Michael, S., Auld, D. et al. (2008). A Robotic Platform for Quantitative High-Throughput Screening. *ASSAY and Drug Development Technologies 6(5),* 637-638.

NAS (2009). Emerging Science for Environmental Health Decisions. Available at: http://dels.nas.edu/envirohealth/index.shtml (Accessed September 17, 2009).

NICEATM (2008). Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Available at: http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm (Accessed September 17, 2009).

NRC (2009). *Science and Decisions: Advancing Risk Assessment. Washington*, DC: National Academy Press.

NRC (2007a). Toxicity testing in the 21st century: A vision and a strategy. Washington, DC: National Academies Press.

NRC (2007b). *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*. Washington, DC: National Academies Press.

NTP (2004). Toxicology in the 21st Century: The Role of the National Toxicology Program. Available at: http://ntp.niehs.nih.gov/ntp/main_pages/NTPVision.pdf (Accessed September 17, 2009).

OECD (2005). OECD Series on Testing and Assessment. Number 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: http://appli1.oecd.org/

olis/2005doc.nsf/linkto/env-jm-mono%282005%2914 (Accessed: September 17, 2009).

Paris, M., Strickland, J. et al. (2006). Reproducibility analyses for in vitro neutral red uptake methods from a validation study to evaluate in vitro cytotoxicity assays for estimating rodent acute systemic toxicity. *The Toxicologist 90, Abstract 1970*. Available at: http://iccvam.niehs.nih.gov/meetings/SOT06/sot06paris.pdf (Accessed September 17, 2009).

Schmidt, C. W. (2009). TOX 21: new dimensions of toxicity testing. *Environ Health Perspect 117(8)*, A348-353. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19672388

Stokes, W., Bremer, S. et al. (2007). NICEATM/ECVAM/JaCVAM Multi-phased International Validation Study of a Stably-Transfected Estrogen Receptor (ER) Transcriptional Activation (TA) Test Method. *The Toxicologist*. Available at: http://iccvam.niehs.nih.gov/meetings/SOT07/sot07stokes_pos.pdf

Stokes, W. S. (2007). Alternative and integrated approaches to skin sensitization: recent progress and the way forward (Abstract, 46th Annual Meeting of the Society of Toxicology). *Toxicologist 96*. Available at: Available: http://iccvam.niehs.nih.gov/meetings/SOT07/sot07stokes_pre.pdf. (Accessed September 17, 2009).

Stokes, W. S. and Schechtman, L. M. (2007). Validation and regulatory acceptance of new, revised, and alternative toxicological methods. *Principles and Methods of Toxicology* (1103-1128). A. W. Hayes. Philadelphia, Pennsylvania: Taylor and Francis.

Stokes, W. S. (2006). Validation of In Vitro Methods for Toxicology Studies. *Biological Concepts and Techniques in Toxicology: An Integrated Approach* (161-176). J. E. Rieviere. New York, NY: Taylor & Francis Group.

USC (2000). ICCVAM Authorization Act of 2000. Public Law 106-545. Available at: http://frwebgate.access.gpo.gov/cgi-bin/useftp.cgi?IPaddress=162.140.64.183&filename=publ545.pdf&directory=/diska/wais/data/106_cong_public_laws (Accessed: September 17, 2009).

Correspondence to

William S. Stokes, D.V.M., DACLAM
Director, NICEATM
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park
NC 27709, USA
e-mail: stokes@niehs.nih.gov