FORUM SERIES - PART IX

Toxicity Testing in the 21st Century: Using the New Toxicity Testing Paradigm to Create a Taxonomy of Adverse Effects

Kim Boekelheide[1] and Sarah N. Campion

Department of Pathology and Laboratory Medicine, Brown University, Providence, Rhode Island 02912

[1] To whom correspondence should be addressed at Department of Pathology and Laboratory Medicine, Brown University, Box G-E5, Providence, RI 02901. Fax: (401) 863-9008. E-mail: kim_boekelheide@brown.edu.

Distinguishing adaptive from adverse responses is fundamental to understanding toxicity and to implementing regulatory guidelines that are protective of human health. However, what we consider to be an adverse effect may change over time as the cultural acceptance of risk alters and new knowledge and insight accumulate. The fact that the identification of an adverse effect is subject to change is obvious, necessary, and uncomfortable. In this commentary, a framework for defining adverse effects is proposed for the emerging paradigm of toxicity testing in the 21st century—a paradigm that focuses on human cells, *in vitro* approaches, toxicity pathways, and high-throughput techniques. The traditional meaning of an adverse effect as a change at the organismal level is not compatible with this new system of toxicity testing. Instead, based on the experience of accident investigators, we propose that a Toxicological Factors Analysis and Classification System will use the database resulting from the high-throughput toxicity testing of the future to develop a *Taxonomy of Adverse Effects*. Similar to an accident, predisposing "latent failures" identified within categories of the toxicant response database will be associated with the "active failure" of an adverse effect.

Key Words: toxicity testing; *in vitro*; high throughput; adverse; adaptive.

In his 1990 book *Human Error*, Reason (1990) developed the "Swiss cheese" model of human error, also called the "cumulative act effect." This model, originally designed to explain the sequence of events that could lead to a nuclear power plant accident, was subsequently adopted by the Department of Defense for use in aircraft accident investigation, becoming formalized as the Human Factors Analysis and Classification System (HFACS). Using HFACS, the Navy analyzed its extensive database of more than 300 aircraft accidents and created a *Taxonomy of Unsafe Operations*. The most interesting aspect of this framework is the separation of predisposing events into distinct sequential categories, only the

last of which, called *Unsafe Acts*, was the act of the operator that resulted in the accident. This last event, termed an "active failure," was conditional on previous "latent failures" of the system grouped into various categories (*Organizational Influences*, *Unsafe Supervision*, and *Preconditions for Unsafe Acts*). The Swiss cheese analogy comes into play because of the presence of "holes," due to failed or absent defenses, in these underlying categories. When the latent failures in the system are sufficiently associated (i.e., when the holes are aligned), the active failure of an accident can take place.

Here, we explore this conceptual model as a framework for understanding adverse effects, and for planning on how we will incorporate new knowledge into defining adverse effects, in the context of toxicity testing in the 21st century.

CREATING A TAXONOMY OF ADVERSE EFFECTS

As defined by Lewis *et al*. (2002), an adverse effect is "A biochemical, morphological or physiological change... that ... adversely affects the performance of the whole organism or reduces the organism's ability to respond to an additional environmental challenge." By this definition, adverse effects are limited in scope to apical events or active failures of the whole organism. On the other hand, the approaches proposed for toxicity testing in the 21st century are nonapical, focused on human cells, *in vitro* approaches, toxicity pathways, and high-throughput techniques. Is the traditional meaning of an adverse effect even compatible with this new system of toxicity testing?

Before addressing this important question, let us backfill and introduce the new paradigm of toxicity testing. The National Research Council (NRC, 2007) of the National Academies produced a report entitled *Toxicity Testing in the 21st Century: A Vision and a Strategy* that provides a roadmap for the future of toxicity testing. This Forum article is one of a series over the past year that comments on the report and the follow-up paper

"Toxicity Testing in the 21st Century: Bringing the Vision to Life" by Andersen and Krewski (2009). The basic proposal is to reorient testing to the molecular level rather than observing apical responses at the level of whole organisms. The NRC report's vision identified a sequence of steps for evaluating toxicants, including (1) chemical characterization, (2) assessment of toxicity pathway responses and targeted testing, (3) dose-response and extrapolation modeling, (4) benchmarking to population and exposure data, and (5) decision making within risk contexts. Each of these steps will require new tools and thought processes that together create a coherent system of toxicity testing.

So back to our important question: How do we rationalize the traditional definition of adversity as an apical event with the proposed new testing paradigm that is based on a detailed analysis of the building blocks of chemistry and biology? In parallel with the growth of a coherent system of accident investigation, the answer may lie in the development of a systematic approach to the analysis of the data generated by the new testing paradigm—a Toxicological Factors Analysis and Classification System (TFACS), using the parlance of accident investigators. Applying TFACS to the enormous database expected to result from the high-throughput toxicity testing of the future will naturally lead to the creation of a *Taxonomy of Adverse Effects*. Similar to an accident, predisposing latent failures identified within categories of the toxicant response database will be associated with the active failure of an adverse effect. A presentation of possible categories and patterns within the toxicant response database that can inform us about adversity follows a discussion of the nature of adversity itself and the analytical approaches available to us.

ADVERSITY IS DEFINED BY WHAT WE KNOW

A real advantage of the new toxicity-testing paradigm is its comprehensive approach. Using high-throughput techniques, every toxicant that is tested will be assessed for a multiplicity of fundamental subcellular responses. Our past inability to acquire such a comprehensive understanding of the toxicity response space has been very limiting.

Lead toxicity is a great example of how such limitations have changed our perception of adversity over time. The ancient Romans knew that lead was toxic, but their adverse effects of concern were saturnine gout and colic. Up until the mid-20th century, lead-induced inhibition of hemoglobin synthesis leading to anemia was the most sensitive adverse effect identified for regulatory purposes. We now know that developmental neurotoxicity is an adverse effect of lead that occurs at low levels of exposure. Indeed, it took much of the 20th century to understand that lead toxicity occurs in children (Gibson, 2005) and produces persistent neurodevelopmental effects (Byers and Lord, 1943) with long-lasting consequences for behavior and intellectual function (Lanphear *et al.*, 2005;

Needleman, 1995). Only in the last quarter century has this new understanding of children as a susceptible subpopulation been translated into regulatory action that limits exposure of this vulnerable group. While toxicologists focused on verifying the importance of each apical end point or active failure in turn, the biological implications of the underlying molecular mimicry shared by lead, zinc, and calcium were underappreciated. This is just the type of situation that TFACS can address, finding the latent failures inherent to shared chemical structure and inadvertent biological pathway activation that predispose to an adverse outcome.

As new scientific tools and approaches have been developed, a greater appreciation of the underlying processes of nature has been revealed. An important lesson from the history of lead toxicity is that our understanding of what constitutes an adverse effect is ever changing. For lead toxicity, as the known adverse effects became increasingly subtle apical end points, new tools were required for their definitive identification and understanding. There are many current examples of debate regarding what constitutes an adverse effect in an area of scientific controversy. For an endocrine disruptor, is an induced change in a hormone level sufficient evidence of adversity, or does there need to be an end-organ alteration in histopathology, weight, or function? How about a change in gene expression associated with hormone production or action—is this an adverse effect? For many, uncertainty about what is adaptive or adverse is unsettling, like moving the goal posts after the game has started. Unfortunately, the future portends the need for a more nuanced view of adversity as part of a continuum of complex subcellular responses.

THE AMES TEST—BACK TO THE FUTURE

Enhanced sensitivity at the cost of lowered specificity is undoubtedly a major concern raised by the proposal to create a Taxonomy of Adverse Effects from the largely *in vitro* high-throughput toxicity testing of the future. This concern was articulated by the title of an earlier Forum article in this series (Pragmatic challenges for the vision of toxicity testing in the 21st century in a regulatory context: another Ames test? . . . or a new edition of "the Red Book"? [Meek and Doull, 2009]). The Ames test and its derivatives are great mutagenicity tests, but as carcinogenicity tests, they are often derided because of the large numbers of false positives generated.

Carcinogenicity is clearly a complex adverse effect, so it is useful to examine how the *in vitro* Ames test approach has been employed to address it. In this regard, a report of the European Centre for the Validation of Alternative Methods workshop entitled "How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests" (Kirkland *et al.*, 2007) is illuminating. In a gathering of more than 20 genotoxicity experts, this workshop identified a research agenda relevant not only to genotoxicity testing but also to the toxicity testing

vision developed by the NRC report. The workshop participants recognized the very poor discrimination of noncarcinogens from carcinogens using *in vitro* genotoxicity tests and proposed an ambitious plan that was specific and lengthy, detailing areas for further research and development, including (1) culture medium composition and toxicant oxidation, (2) deficiencies in the cell lines, (3) lack of normal metabolism, (4) cytotoxicity as a confounder, and (5) the need for new cell systems, such as three-dimensional culture models. As a bottom line, the workshop participants concluded "... that better guidance on the likely mechanisms resulting in positive results that are not biologically relevant for human health ... is needed" (Kirkland *et al.*, 2007).

As an example of how better mechanistic insight can be informative, the incorporation of gene expression profiling into *in vitro* genotoxicity assessment provides increased discriminatory power regarding carcinogenicity (Staal *et al.*, 2006). A comparison of the effects of polycyclic aromatic hydrocarbons (PAHs) on gene expression with other end points indicative of carcinogenicity (e.g., DNA-adduct formation, Ah-receptor binding) provides insight into how we can discriminate better between carcinogenic and noncarcinogenic compounds. PAHs were found to generally induce a compound-specific response on gene expression, but when responses were examined at the pathway level, carcinogenic and noncarcinogenic compounds could be successfully distinguished (Staal *et al.*, 2006). Pathways likely to be activated by carcinogens include DNA repair and oxidative stress response pathways among others. Pathway responses at the gene or protein level provide better guidance because they can not only serve as a discriminatory tool but also reveal mechanistic information related to the carcinogenicity. Understanding the mechanism behind a positive hit will provide greater confidence in whether it is a true or false positive, a vast improvement over the traditional Ames test. This confidence will extend beyond genotoxicity testing, with the mechanistic insight derived from gene and protein expression profiling providing increased discriminatory power regarding adverse effects in the new toxicity testing vision.

The focus on using the new tools that are available to reveal toxicant mode of action is fundamental to the success of the new testing paradigm. The ongoing refinement of the Ames test approach presages the future of toxicity testing and illustrates the need for time, investment, and dedicated research to succeed. Already very valuable for understanding genotoxic potential, a mechanistically defined and thoroughly understood Ames test approach will one day be highly predictive of carcinogenicity for certain classes of compounds.

BEYOND VIRCHOW—PRINCIPLES OF ADVERSITY AT THE MOLECULAR LEVEL

Virchow (1860), the famous German physician, said "... the cell is really the ultimate morphological element in which there is any manifestation of life ..." in articulating his vision of the cellular basis of disease. The implementation of the microscope as a diagnostic tool and the development of a taxonomy of disease organized around cellular dysfunction contributed greatly to the advances in medicine that took place during the first half of the 20th century. The determination of the double-helix structure of DNA in 1953 by Watson and Crick heralded the replacement of the cellular basis of disease by the modern era of molecular pathogenesis. The tools now at our disposal are phenomenal, rapidly evolving, and supported by remarkable computational power.

We are now in a post-Virchow era, and defining an adverse effect purely by its apical manifestations is no longer enough. The toxicity testing of tomorrow will use new molecular techniques to generate enormous amounts of data describing fundamental subcellular responses to toxicants. A TFACS framework, yet to be designed, will provide a path forward to mining this data, generating a deep understanding of toxicity and adversity at the molecular level.

The NRC report provides some guidance for creation of a TFACS framework. Roughly speaking, one can imagine that the TFACS categories might correspond to the initial sequential steps of toxicity testing as proposed in the NRC report: chemical characterization, assessment of toxicity pathway responses and targeted testing, and dose-response and extrapolation modeling. The testing itself will identify response patterns consistent with vulnerabilities in each of these categories (latent failures) that when sufficiently aligned predispose to an adverse effect (active failure). Analysis of the response patterns of numerous toxicants within the TFACS framework will, over time, develop and refine a Taxonomy of Adverse Effects. This conceptual approach is illustrated in Figure 1.

Looking into the future, one imagines that the features that predispose to adversity within the TFACS categories (Fig. 2) are similar to those already associated with an adverse effect. At the level of the first TFACS category, chemical characterization, response patterns will be identified by evaluating quantitative structure-activity relationships, physical and chemical properties, environmental concentrations, and possible metabolites and toxic properties of the test chemicals. Vulnerabilities predisposing to an adverse effect will likely include a biologically reactive chemical or metabolic product of a chemical, predicted molecular interactions with critical cellular macromolecules, and chemicals that have a high probability of reaching and persisting in the environment.

Assessment of toxicity pathway responses, the next category within the TFACS framework, will provide another level of predictive power for active failures. Chemical exposure may result in activation of adaptive response pathways (e.g., heat shock protein response pathways) and/or pathways that produce adverse consequences (e.g., proapoptotic pathways). Activation of adaptive pathways, while protective, may be a red flag for eventual adverse outcomes should the pathway
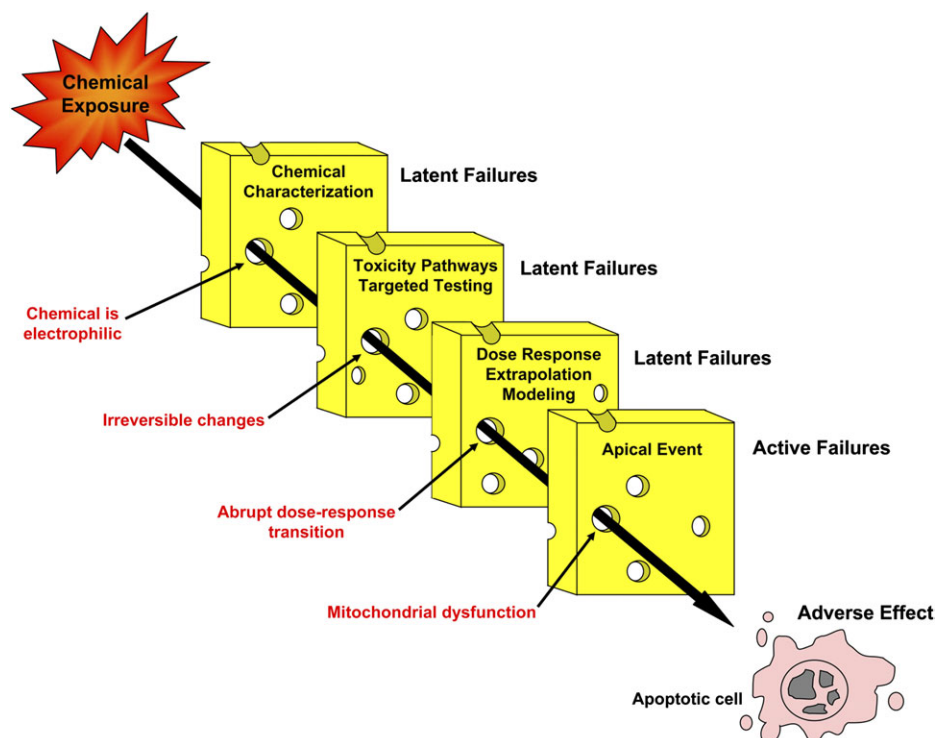
FIG. 1. The Swiss cheese model of adverse effects. Chemical exposure may result in vulnerabilities, or latent failures, in TFACS categories, represented by the pieces of cheese. Examples of failures for each of these categories are listed in red. When the response patterns are sufficiently aligned, they can predispose to an adverse effect or active failure, illustrated by the apoptotic cell.

perturbation remain for a sufficient period of time or intensity. In the context of toxicity pathway alterations, adaptive responses to toxicant exposure will likely be characterized by reversible changes, a limited scope of toxicity pathway–

FIG. 2. The TFACS framework. Within the three TFACS categories illustrated in this example, chemical characterization, toxicity pathways and dose–response, and extrapolation modeling, there are various features that will be used to identify latent failures or characteristics associated with adverse effects.

induced alterations, and gradual changes in dose-response. On the other hand, latent failures predisposing to an adverse effect will likely be characterized by irreversible changes, a spreading of toxicity pathway responses, and abrupt dose-response transitions. As we develop this new framework, we will likely identify selected pathways of concern that may be more indicative of an active failure than other pathways. Therefore, the specific toxicity pathways affected and the biological relevance of the pathway alteration are important considerations when characterizing a response as adaptive versus adverse. In a previous forward-thinking report from the NRC entitled *Scientific Frontiers in Developmental Toxicology and Risk Assessment* (Committee on Developmental Toxicology, Board on Environmental Studies and Toxicology, National Research Council, 2000), signaling pathways important during development were characterized and the importance of understanding how molecular perturbations in these pathways can result in adverse outcomes was stressed. This report provides an excellent example of how adverse effects will be classified in the future based on a more complete understanding of toxicity pathways.
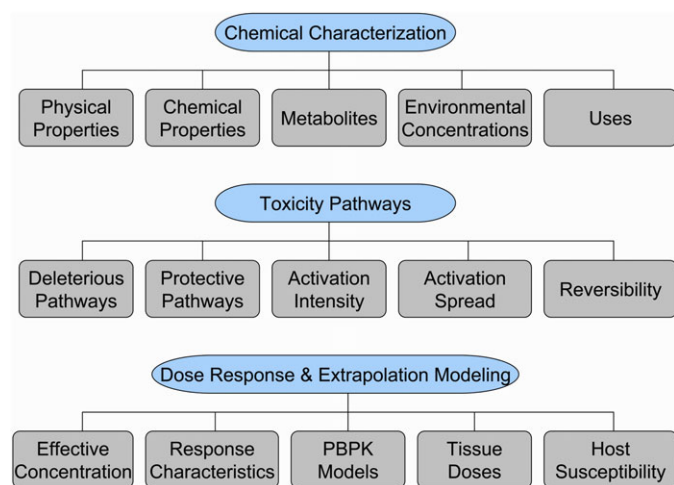
Within the third TFACS category, dose-response and extrapolation modeling, there are additional predisposing manifestations of adversity. The characterization and interpretation of the dose-dependent changes in protective and adverse toxicity pathways are at the core of the new toxicity testing

paradigm. At some low dose, a pathway may begin to be disrupted by a toxicant exposure, but the pathway will continue to function due to a homeostatic response (an "adaptive" behavior). At a higher dose, the adaptive response is overwhelmed, and an adverse effect takes place. Dose-response modeling is critical to identifying adverse effects, especially since there can be dose-dependent transitions in the principal mechanism of toxicity (Slikker *et al.*, 2004). As discussed earlier, an abrupt dose-response transition and a transition in mechanism of toxicity at different doses will likely be indicative of an adverse effect. Physiologically based pharmacokinetic models will be needed to determine if the doses that cause toxicity pathway alterations *in vitro* are comparable to the human blood/ tissue concentrations that would result from environmental exposure levels. A response will be classified as adverse if the dose required to elicit the effect is environmentally relevant.

The development of the Taxonomy of Adverse Effects can only be accomplished through national and international collaboration among laboratories. This will require the establishment of standardized data collection and integration methods. Publicly available databases must be the norm, facilitating collaboration and comparison of data among laboratories and maximizing the utility of the resource. The large amounts of data that will be generated and stored in these databases must be analyzed to identify adverse and adaptive effects of chemical exposure. Therefore, development of standardized bioinformatic techniques for data analysis will also be necessary. These are just a few of the steps that will be essential to the creation of a robust and comprehensive Taxonomy of Adverse Effects.

A fully fleshed out Taxonomy of Adverse Effects is the holy grail of the new toxicity testing paradigm. Implementing a systems-oriented approach to toxicity testing and developing a TFACS framework of data analysis will inevitably lead to a Taxonomy of Adverse Effects as the output of the next 10–20 years of work (or however long it takes). We all seek a mode of action–based molecular understanding of how the initiating events arising from the interactions of a toxicant with a living system produce adverse effects. One advantage of this new approach is a deeper and coherent appreciation of the contributing components that ultimately manifest as an adverse effect. For commercial aviation, the systematic use of such a coherent approach to accident investigation has markedly decreased the frequency of active failures in the past quarter century. The goal for the future of toxicity testing is exactly the same—to build a testing system that is very robust in its identification and understanding of the predisposing manifestations of adversity.

REFERENCES

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

Byers, R. K., and Lord, E. E. (1943). Late effects of lead poisoning on mental development. *Am. J. Dis. Child.* **66,** 471–494.

Committee on Developmental Toxicology, Board on Environmental Studies and Toxicology, National Research Council. (2000). *Scientific Frontiers in Developmental Toxicology and Risk Assessment.* National Academy Press, Washington, DC.

Gibson, J. L. (2005). A plea for painted railings and painted walls of rooms as the source of lead poisoning amongst Queensland children. 1904. *Public Health Rep.* **120,** 301–304.

Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., Elhajouji, A., Glatt, H., Hastwell, P., Hayashi, M., *et al.* (2007). How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: report of an ECVAM workshop. *Mutat. Res.* **628,** 31–55.

Lanphear, B. P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D. C., Canfield, R. L., Dietrich, K. N., Bornschein, R., Greene, T., *et al.* (2005). Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environ. Health Perspect.* **113,** 894–899.

Lewis, R. W., Billington, R., Debryune, E., Gamer, A., Lang, B., and Carpanini, F. (2002). Recognition of adverse and nonadverse effects in toxicity studies. *Toxicol. Pathol.* **30,** 66–74.

Meek, B., and Doull, J. (2009). Pragmatic challenges for the vision of toxicity testing in the 21st century in a regulatory context: another Ames test? or a new edition of "the Red Book"? *Toxicol. Sci.* **108,** 19–21.

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

Needleman, H. L. (1995). Behavioral toxicology. *Environ. Health Perspect.* **103**(Suppl. 6), 77–79.

Reason, J. (1990). *Human Error.* Cambridge University Press, New York.

Slikker, W., Jr, Andersen, M. E., Bogdanffy, M. S., Bus, J. S., Cohen, S. D., Conolly, R. B., David, R. M., Doerrer, N. G., Dorman, D. C., Gaylor, D. W., *et al.* (2004). Dose-dependent transitions in mechanisms of toxicity: case studies. *Toxicol. Appl. Pharmacol.* **201,** 226–294.

Staal, Y. C., van Herwijnen, M. H., van Schooten, F. J., and van Delft, J. H. (2006). Modulation of gene expression and DNA adduct formation in HepG2 cells by polycyclic aromatic hydrocarbons with different carcinogenic potencies. *Carcinogenesis* **27,** 646–655.

Virchow, R. L. K. (1860). *Cellular Pathology.* R.M. De Witt, London.