

Spheres of Influence



Assessing Assays

The mathematician George Box, recognized worldwide as a leading developer of statistical methods, once said that “all models are wrong, but some are useful.” Box’s comment could be said to apply to the models used in human cancer risk assessment. Rodent models used to test potential carcinogens are by their nature “wrong” because they merely simulate the response of the real target species—humans. James MacDonald, senior vice president for drug safety and metabolism at the Schering Plough Research Institute in Kenilworth, New Jersey, estimates that conventional rodent models correctly predict human cancer about 50% of the time. “It’s a flip of the coin,” he says. “A lot of the chemicals that test positive in rodent assays turn out not to be of concern for humans.”

PhotoDisc, Arnold Greenwell/EHP, Matt Ray/EHP

Improving animal models used in cancer risk assessment is therefore an important goal for public health. In 1996, an international group of scientists formed a research consortium dedicated to this task under the auspices of the International Life Sciences Institute (ILSI), a Washington, DC-based non-profit research foundation. Comprising experts from industry, government, and academia, the Committee on the Evaluation of Alternative Models for Carcinogenicity Testing has just completed its data collection and review. The results of the review, reflecting a six-year commitment and nearly \$35 million in collaborative funding, were published as a monograph in September 2001 in *Toxicologic Pathology*. Toxicologists see the publication of the monograph not only as a milestone for the project itself, but also as a victory for collaborative research. "This is a precedent setting effort," says Ray Tennant, director of the NIEHS's National Center for Toxicogenomics (NCT), who served on the committee. "A broad range of stakeholders came together to tackle a very complex issue in public health. I see this as a paradigm for how other, difficult public health problems might be resolved in the future."

The Problem Assay

To a large extent, the ILSI committee focused on *in vivo* alternatives to the "gold standard" used in cancer risk assessment for over 30 years: the chronic two-year rodent bioassay, performed in mice and rats of both sexes. Briefly, the test involves exposing animals to varying levels of a given chemical (beginning with the maximum tolerated dose, or MTD), sacrificing the animals once the exposure period has ended, and then checking their organ systems for numbers and types of tumors. Two implicit assumptions govern the test: that the animal and human responses will be the same, and that high-dose effects observed in the laboratory will also be relevant for ordinary human exposures.

With advancements in risk assessment, both of these assumptions have come under scrutiny. Today, toxicologists acknowledge that, depending on the chemical, the effect of tumor induction in animals may be species specific, with little relevance to humans. In perhaps the best-known example, the B6C3F1 mouse used in the two-year

bioassay is highly susceptible to chemically induced liver tumors, which are thought to arise from pathways that are unique to that strain. Scientists also question the relevance to humans of rodent tumors obtained only with the MTD. "If we're dealing with a situation in which the likely human exposure is in the same ballpark, then these [dosing regimens] may be applicable," says Jay Goodman, professor of pharmacology and toxicology at Michigan State University in Lansing. "But doses that are hundreds to thousands of times higher than normal exposures [such as those often given during animal testing] might be carcinogenic simply because they overwhelm detoxification pathways. In these cases, we see tumors along with gross histopathologic evidence of tissue damage."

In the face of these shortcomings, many experts believe the scientific value of the two-year bioassay is highly limited—barely worth the investments in personnel, animals, money, and time. "All we're doing with the [two-year test] is counting tumors," laments Samuel Cohen, chair of the Department of Pathology and Microbiology at the University of Nebraska Medical Center in Omaha. "We need tests that tell us more about a chemical's biological mode of action, so that we can be more confident about extrapolating to humans."

The Transgenic Alternatives

Transgenic rodent models provided by advances in molecular biology are widely seen as the most promising *in vivo* alternatives to the two-year rodent assay. These models are bred with genetic predispositions that increase their susceptibility to insult from carcinogens and to the rapid development of cancer. Scientists believe their use will accelerate cancer testing and provide important mechanistic insights into how carcinogens produce tumors.

Five of the six alternative *in vivo* models evaluated by the ILSI committee involve the use of transgenic species: the rasH2 transgenic mouse, the TgAC transgenic mouse, the p53 knockout mouse, the Xpa knockout mouse, and the p53/Xpa double knockout mouse. The review also lists one nontransgenic *in vivo* alternative, the neonatal mouse assay, and one *in vitro* test, the Syrian hamster embryo assay.

According to Ronald E. Cannon, a staff scientist at the NCT, each model

expresses a particular cancer-inducing genetic anomaly. The p53 knockout mouse is perhaps the best-known example. p53 is a tumor suppressor gene found in all mammalian cells. Under normal conditions, the gene responds to certain kinds of DNA damage by initiating cell death, thereby taking defective cells out of circulation. Defects in p53 are a serious health threat, implicated in as many as 50% of all human cancers. In the transgenic model, scientists intentionally mutate the gene to "knock out" p53. Without this normal suppressor, the cells quickly form tumors under carcinogen exposure.

The committee's selection of the six *in vivo* models was not arbitrary. In 1996, all these alternatives had just been recommended for cancer testing in the pharmaceutical industry by the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH), an independent group of industry and agency stakeholders from the United States, Europe, and Japan. That same year, the ICH recommendation was also recognized as guidance by the U.S. Food and Drug Administration (FDA), which since that time has accepted transgenic data from the models as valid in considering new drug applications.

But despite a willingness to consider such models by FDA regulators, most researchers had little experience with the models and were unsure about how to use them. It was from this sense of uncertainty that the ILSI project was born, says Robinson. "For us to feel more comfortable with the models, we felt we needed a broader, more standardized data set," she explains. "The goal of the [review] was to increase our confidence in the models and also to identify any special cross-model nuances that we should be aware of."

In addition to model selection, the choice of appropriate compounds for testing the models was also crucially important. From an initial list of 60 candidates, 21 nonproprietary agents, including both pharmaceutical and industrial chemicals, such as cyclosporin A, dieldrin, and diethylstilbestrol were ultimately chosen. Each compound had available mouse and rat data from the two-year bioassay, an established toxicology database of *in vitro* and *in vivo* modes of action, and data related to human exposure and effect. According to Denise Robinson, executive director

of ILSI's Health and Environmental Sciences Institute, particular emphasis was placed on suspected nongenotoxic carcinogens. These are also sometimes referred to as "rodent-only" carcinogens because tumors are elicited solely at the MTD, presumably via strain-specific mechanisms considered irrelevant to humans. The selection process also emphasized a range of biological effects, including immunosuppression, enzyme induction, and proliferation of peroxisomes (a precursor to tumors).

A key issue considered by the review committee was what additional mechanistic insights the alternative models might provide compared with the two-year assay. For example, suppose a chemical tests positive in one of the two-year bioassays and negative in a transgenic model. John E. French, a group leader at the NIEHS Laboratory of Environmental Carcinogenesis and Mutagenesis, says this discrepancy would suggest a "rodent-only" carcinogen acting through a species-specific pathway. Conversely, a positive result in both models would suggest a potential human carcinogen, possibly acting via a specific genetic pathway.

The Results

To reach conclusions with confidence, scientists must have faith that the transgenic models are reliable and accurate. However, the ILSI data indicate this is not always possible. According to Cohen, under the conditions employed during the tests, the models could not always distinguish between genotoxic and nongenotoxic carcinogens, nor did they identify human carcinogens with a 100% specificity. For example, estradiol—a known human breast and endometrium carcinogen—tested negative in the rasH2, Xpa, and TgAC models. Based on his review, Cohen says, "My personal feeling is that it's not appropriate to use any of these models as proof [or negation] of genotoxicity; the models provided exceptions in either direction."

The chief consensus among the review committee members is that four of the alternative models might substitute adequately for the two-year mouse bioassay, but only if the data are considered as part of an overall "weight-of-the-evidence" approach that takes information from multiple sources into account. These four models are the p53, Xpa, TgAC, and rasH2 models, which are considered genetically stable

(meaning the genetic specificity holds up over successive generations of breeding), flexible regarding dosing options (oral, gavage, or dietary), and able to produce a wide spectrum of tumors depending upon the type of chemical exposure. According to French, summary results from all studies published to date show that the accuracy of these models ranged from a low of 76% in TgAC to 82% in rasH2, reaching approximately 90% when results from p53 (for genotoxic carcinogens) and rasH2 (for both genotoxic and nongenotoxic carcinogens) were combined.

The researchers emphasize that alternative carcinogenicity testing paradigms should continue to rely on the rodent bioassay in combination with a single, targeted transgenic model. This line of reasoning is supported by the results of a meta-review of all the available transgenic carcinogenicity testing data (including that reviewed by ILSI) by John Pritchard, chief of the NIEHS Laboratory of Pharmacology and Chemistry. The results of this review indicate that, with such a combination, it was possible to correctly identify all the known human carcinogens contained in the data set. Results of the review are being finalized and, according to French, will likely be published in the fall.

Based on its review, the NIEHS has proposed that in some cases, even the rat test may be unnecessary. As part of an alternative testing paradigm for chemicals at its National Toxicology Program, the NIEHS will routinely perform abbreviated studies with transgenic mice concurrent with its standard three-month toxicity screening studies. The selection of the appropriate model, Cannon says, will be based on existing information about the chemical and its likely genetic target. According to Cannon, a positive result in the transgenic model may be sufficient to identify a chemical as a human carcinogen during the screening process. This is because human carcinogens are considered to be genotoxic, meaning that cancer develops from an interaction between potentially low levels of the chemical and DNA, as opposed to cancers identified in the two-year bioassay, which may result from toxic damage to tissues. "If we get a positive result without any clear-cut evidence of tissue damage, why would we want to even run the rat assay?" he asks. Cohen is

more cautious in this respect. "My difficulty with that is that we are not yet to the level where a positive result in a transgenic assay clearly indicates a human carcinogen," he says.

Both researchers agree, however, that substituting alternative models for the two-year rodent assay will save time (transgenic models typically require only 6–9-month exposures) and expense. "The alternative tests use far fewer animals," Cohen explains. "Although the appropriate numbers are still debated, it's currently 15 animals per dose group versus 50 animals in the two-year assay. That's a lot less pathology to be concerned with."

Even this reduction continues to be unacceptable to animal rights activists. "It's just more of the same," says Troy Seidle, research associate with People for the Ethical Treatment of Animals in Toronto, Canada, who insists (contrary to the beliefs of the ILSI committee members) that *in vitro* models like the ones evaluated by the ILSI committee are sufficiently predictive of human carcinogenicity to be useful in a regulatory context. "Nothing short of a nonanimal system will satisfy us, and we do not believe that reliance on transgenics provides a step in the right direction," he says.

Industry groups, on the other hand, are encouraged by the movement toward accepting transgenics in the cancer testing process. Gerald Long, senior research scientist with Eli Lilly and Company, says the results of the ILSI tests have allayed initial fears that transgenic models may be overly sensitive. "In fact, they appear not to be," he says. "The ILSI data combined with the results of the NIEHS review suggest that if you pick the appropriate alternative model, the results are equally as valid as the old two-species bioassays."

Cohen concludes that the adequacy of the transgenic alternatives marks an important divergence from the two-year assay, an archaic test that he believes has been ingrained in the regulatory and toxicology community for too long. "My hope is that this is the first step toward getting rid of the two-year bioassay altogether," he says. "I think this is what's going to happen eventually. The emerging use of these alternatives is the first step along a path that needs to start somewhere."

Charles W. Schmidt