Heavy dependence on animal testing data is not supported by science.

Regulating Unknown Risks

By Gio B. Gori
Health Policy Center

Many natural and manmade entities could conceivably pose health risks to humans. Those risks cannot be evaluated and quantified with objective scientific tests because of ethical concerns about human testing. Instead, we typically rely on the results of animal tests to surmise risk and to regulate such entities. The problem with this approach is that the use of animal tests in defining human risk is unwarranted.

Animal tests offer plausible human approximations only for acute toxicities that result from the rapid derailments of generic life processes that are conserved in many species, including humans. Approximations essentially vanish when testing for specific endpoints, such as cancer, endocrine disruptions, aberrations of reproduction and immunity, and other anomalies that usually take time to evolve and are triggered by complex causes that vary dramatically in different animals. In testing for such endpoints, the genetic and somatic differences of different species coalesce into causal mechanisms that are specific to each species. As a result, the outcomes of such tests become inconsistent across species or even among strains of the same species. Although this prevents reliable animal-to-human translations, animal tests have come to play a decisive role in the regulation of potential human risks.

The practice began with the requirement for life-long cancer tests in animals, harking back to the 1958 Delaney Clause of the Federal Food, Drug, and Cosmetics Act, prohibiting food additives that cause cancer in animals. In time, the scope of the clause expanded to include exposures at-large, hence extending the mistaken assumption that lifetime cancer tests in animals provide scientifically valid proxies of human conditions and risks. Although this assumption is scientifically untenable, it has nonetheless become ingrained in

Gio B. Gori is with the Health Policy Center in Bethesda, MD. In the 1970s, he was deputy director of the Division of Cancer Cause and Prevention at the National Cancer Institute, where he participated in the planning and conduct of the chemical carcinogenesis program. He was instrumental in the creation of the National Clearinghouse for Carcinogens, which later became the National Toxicology Program.

regulatory policymaking over the past half-century.

This regulatory artifice has been abetted and endorsed since early on by a series of colluding reports from committees of the National Research Council of the National Academy of Sciences. The most recent report in the series, the 2009 *Science and Decisions: Advancing Risk Assessment*, reads like an affirmation of faith, not a serious analysis from the top scientific institution in the country. It struggles to defend arbitrary assumptions as enablers of factual and enforceable evidence. Funded by the Environmental Protection Agency, the report is designed to buttress the EPA's 2005 *Guidelines for Cancer Risk Assessment*. Unmistakable hints in the guidelines and the NAS report indicate that the agency is aware that science and plain common sense do not allow the quantification of human cancer risks from animal tests. Yet, the thrust of the guidelines is to arrogate scientific justification from arbitrary default assumptions. This is attained, remarkably, through the subjective judgment of advisory committees beholden to the agency and charged with selecting, weighing, and interpreting animal test results.


MORGAN BALLARD

What is wrong with such an arrangement? In the first place, it places the imprimatur of science on the intolerable pretension of overcoming the absence of human data with the judgmental appraisal of animal tests. This imprimatur is then used to camouflage the arbitrary setting of regulatory targets, which are fashioned not through the factual consideration of human risks, but by manipulating the process to yield the most imponderable exposures that barely allow restricted uses. In effect the process manages solutions in the right direction, but uses irrational and irrelevant means.

So why not look for the sensible alternative of doing away with the expensive, biased, and fruitless rhetoric of irrelevant animal testing and instead concentrate on the simple, rational task of defining minimal operational exposures? A shift of this sort would restore reason and sanity to a now-contentious set of regulations and achieve better levels of public safety and welfare. As a bonus, it would redirect vast material and human resources toward research and developments more creative than the squandering and pretentious humdrum of animal bioassays. If this scenario sounds overly drastic, a closer look will reinforce its rationality and advantages.

Standards of Evidence Science's contributions to technology and to social and individual benefits have progressed because of validated evidence grounded on simple rules of the scientific method. Provisional validation depends on authenticable and relevant measurements with narrow error rates, coupled with warrants that confounding interferences have been controlled and that results are replicable in different hands. Conclusive proofs lean on counterfactual evidence — for instance, when adverse effects disappear once a presumed cause is removed or neutralized. No valid scientific evidence is possible if any of such criteria fails. Herein is the crux of the continuing frustration with unscientific and wanton justifications for regulations and policies that cannot be grounded on testable and relevant measures of human risk.

PRETENDING TO MEASURE UNMEASURABLE RISKS

Consider the archetypal example of carcinogen tests. They are officially required in mice and rats, even though using animal results to extrapolate the risks to humans is unwarranted. The lifespan of these rodents, their lifecycles, physiology, reproduction, anatomy, and ecology are not comparable to human conditions. Cancer tests in rats do not predict cancer tests in mice better than tossing a coin, despite the two species being so much closer to each other than to humans.

It has been long known that animal tests cannot predict human carcinogenic risks on a scientific basis. The International Agency for Research on Cancer conceded 30 years ago "at the present time a correlation between carcinogenicity in animals and possible human risk cannot be made on a scientific basis." The late Dr. David Rall, then director

of the National Toxicology Program, testified before Congress in 1981 that science could not assist in deriving human cancer risk from animal tests, though Rall insisted that Congress and the public ought to have "faith" in experts who examine animal entrails. The year before, the Occupational Safety and Health Administration identified more than 30 questions that would be crucial in forecasting human cancer risks from animal tests — questions that remain unanswered today. In 1983, a National Research Council committee reviewing human cancer risks from animal tests concluded "many components lack definitive scientific answers ... [because] the dominant analytic difficulty is the pervasive uncertainty."

Since objective formulas for animal-to-human translations are not available, regulators have resorted to inferring human risks by conditioning animal data through a series of invalid default assumptions. To wit, the regulators have simply ruled that mice and rats are proxies for humans, that tests must be conducted at maximum tolerated doses designed to be overtly toxic, that challenging animals at maximum doses during their life spans is equivalent to quite smaller doses that humans may experience, that the route of exposure is irrelevant, that the metabolism of animals at maximum tolerated doses is equivalent to the metabolism of humans at very low doses, that benign lesions are considered on par with malignant ones, that extrapolations from high doses in animals to low doses in humans must be linear, and more kindred assumptions. In the words of the 2005 EPA carcinogen testing guidelines, these justifications have extended to the astonishing official absurdity that, "[a]nimal studies are conducted at high doses in order to provide statistical power."

These are grotesque deviations from honest scientific thinking, and teaching such egregious nonsense at our colleges and universities would be cause for dismissal. Still, those regulatory aberrations have been repeatedly endorsed at the pinnacle of scientific authority. The 2009 NAS report baldly asserts, "Defaults need to be maintained for the steps in risk assessment that require inferences beyond those that can be clearly drawn from the available data or to otherwise fill common data gaps." It further claims that "the defaults involving science and policy judgments, such as the relevance of a rodent cancer finding in predicting low-dose-human risk, are used to draw inferences 'beyond the data,' that is, beyond what may be directly observable through scientific study."

Such assertions also appeared in earlier reports from the NAS that were also funded by regulatory agencies. The contemporary continuation of such double-speak tells of how resistant to criticism and to change the official risk-assessment apparatus has been. That apparatus is sustained by the colluding financial and ideological interests of regulators, politicians, advocacy groups, and of the regulated themselves who have learned to navigate a corrupt system, and who would regard change as costly.

If there is no objective relevant evidence, should it be permissible to burden a nation and the world with regulations justified by arbitrary guesses of risk falsely disguised as science? Apparently not, if one is to follow the White House "scientific integrity" presidential memorandum of March 2009, which compels all agencies of the Obama administration to the utmost respect of science in the formulation of public policies and regulations. The memorandum clearly speaks of the same reliable science that has advanced civilization with a clearer frame of mind, and with a host of unprecedented and functional technologies. President Obama could hardly advocate dialectical bureaucratic guesses, and obviously insists on validated scientific evidence: the sole enabler of applications that can be expected to work in the real time of real people. Thus we must ask, why are referees at national-arbiter scientific institutions endorsing questionable regulations, which they help masquerade as science?

Parallel default assumptions apply to epidemiologic surveys. Although relevant to humans, epidemiology faces daunting odds in measuring exposures and in accounting for multiple possible causes and biases that confound the results. Epidemiology lacks experimental safeguards, and mostly leads to conjectures of causality.

Among the founding fathers of contemporary epidemiology, Richard Doll and Richard Peto warned in 1981:

> [E]pidemiological observations ... have serious disadvantages.... [T]hey can seldom be made according to the strict requirements of experimental science and therefore may be open to a variety of interpretations. A particular factor may be associated with some disease merely because of its association with some other factor that causes the disease, or the association may be an artifact due to some systematic bias in the information collection. ...[T]hese disadvantages limit the value of observations in humans, but ... until we know exactly how cancer is caused and how some factors are able to modify the effects of others, the need to observe imaginatively what actually happens to various different categories of people will remain.

Of course, imagination is not factual evidence, and at best it provides incentive to new research. In fact, parallel warnings are found in the United States Federal Judicial Center's *Reference Manual on Scientific Evidence*, the official guidance to U.S. federal courts. According to the manual, "epidemiology cannot objectively prove causation; rather, causation is a judgment for epidemiologists and others interpreting the epidemiologic data." The manual goes on to note, "The existence of some [associated] factors does not ensure that a causal relationship exists."

Given those difficulties, epidemiologists have adopted default assumptions of their own in determining causality. These criteria, formulated early on by the British epidemiologist and statistician Sir Austin Bradford Hill, consider data attributes such as strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experimental evidence, and analogy.

Although adopted by the EPA as official procedures for risk assessments, authorities in epidemiology have been scornful of the validity and utility of these criteria in determining causation and risk. Kenneth Rothman and Sander Greenland, in their prominent textbook *Modern Epidemiology*, 2nd ed., summarize Hill's criteria as follows:

> [T]he standards of epidemiologic evidence offered by Hill are saddled with reservations and exceptions. Hill himself was ambivalent about the utility of these "standards" (he did not use the word criteria in the paper). On the one hand, he asked, "In what circumstances can we pass from this observed association to a verdict of causation?" Yet, despite speaking of causation, he disagreed that any "hard-and-fast rules of evidence" existed by which to judge causation: "None of my nine viewpoints [criteria] can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non.

In fact, epidemiologic conjectures can only be validated indirectly, as is the case for infectious diseases controlled by sanitation, vaccinations, and antibiotics, and for certain occupational hazards whose control or elimination followed the limitation of exposures. Thus, while default assumptions for animal tests and weak epidemiologic reports have become the legally enforceable currency of policies and regulations, they evidently negate a claim of factual validity. Hence the artifice to filter, weigh, assess, and interpret animal and epidemiologic inputs through a laborious process of default assumptions, gamed to give the impression of scientific validity.

To complete the disguise, default assumptions have been dressed up with procedures of Byzantine complexity and imbued with quantitative numerical pretensions. True to form, those pretensions are also trampled upon, with the application of arbitrary safety factors in the final setting of regulatory targets. This deceptive regulatory approach has proliferated worldwide with local variations, and is jealously guarded by an international bureaucratic alliance of truly Leviathan proportions. This alliance is very much loath to recant and rationalize the process, and is determined to further solidify its intrusive arbitrariness, as with the advancing regulation of putative endocrine disruptors masterminded by the EPA, and the "Registration, Evaluation, Authorization and Restriction of Chemicals" directive (REACH) of the European Union.

Seeking an Alibi Roughly, the battle lines have regulation and its cheerleaders on one side and market providers on the other. Given the enforcement powers of regulation, the fate of the contest might seem predictable, except that regulation is wise to proceed "without building so much instability that industries cannot plan," to borrow language from the Bipartisan Policy Center, warning that regulation would not exist without geese and golden eggs.

Fearful of being taken to task for proceeding regardless of "uncertainties result[ing] from a combination of lack of data and lack of conceptual understanding," regulatory agencies have sought the endorsement of committees that provide "expert" judgment in navigating through default assumptions. In reality, such committees are known to provide scant substantial advice. It is impossible to combine "incompatible judgments or models and the technical issue of training and calibration when there is a fundamental lack of knowledge and no opportunity for direct observation of the phenomenon being estimated." In the end, "[g]iven all of those limitations, there are few settings in which expert elicitation is likely to provide information necessary for discriminating among risk-management options."

Admittedly incapable of resolving evidence, the true function of these committees is to provide political cover for bureaucratic arbitrariness. There could be a problem, though, if their advice were to sway from the interests of regulators.

When appointing advisory committee members, regulators are careful to screen out anyone who might bring an unwanted point of view.

Hence, again in the words of the EPA report, the regulatory agencies strive to ensure their scientific advisory committees "fairly represent the mainstream" and avoid "so much constant debate that the rulemaking apparatus simply seizes up entirely." What this means is that when appointing advisory committee members, regulators are careful to screen out anyone who might bring an unwanted point of view to the committee's work. Currently excluded are all those with conceivable financial interests, including holding stock in enterprises whose products are under consideration, even if stock was held and sold several years back, or if it is or was held by wives, children, grandchildren, and close relatives. With dubious legality, agencies execute the screening through private contractors that keep extended dossiers of who may or may not qualify for advisory committees.

Agency claims that such screening avoids appointing advisors with conflicts of interest ring disingenuous, for it obviously avoids intellectual and technical criticism as well. Besides, three main sets of interests come to mind: those of the regulated, those of regulators, and the public interest. The last is difficult to define, other than thinking that the public expects reasoned safety, but also innovation and useful products. How public interest translates into regulation is left by law to regulators, who are bound to align public interest with interests of their own: interests that include healthy agency budgets, agency expansion, prestige, and professional security, as Nobelist James Buchanan has argued. The point is that a suspicious concern for financial interests

allows regulators to be confrontational and authoritarian, by concealing their weaknesses behind the consensus of hand-picked committees.

HISTORICAL CONTEXT

How could we regulate in the absence of objective and testable science to quantify the hypothetical risk posed by certain substances? The most reasonable approach would be to regulate for the least exposures conducive to useful applications. Unfortunately, ill-informed statutes gave the regulatory establishment a perfect excuse to create a monstrous bureaucratic construct, in keeping with a historical bureaucratic disdain for simplicity.

Instead of pausing to enlighten legislators about factual injunctions against using animals to assess complex human risks, regulators have chosen to fulfill their mandate by fabricating definitions of risks unconnected to the realities of harm or safety. The result is a health and safety regime incapable of a measurable effect on public health. In an uncharacteristic fit of earlier candor, the EPA admitted in its 1986 *Guidelines for Carcinogen Risk Assessment*, "It should be emphasized that the linearized multistage model leads to an upper limit to the risk that ... does not necessarily give a realistic prediction of risk. The true value of the risk is unknown and may be as low as zero." Similarly, the National Research Council concluded in its 1983 report mentioned above, "In the absence of completed risk assessments, risk management decisions continue to be made by state and federal agencies; however it is not known whether the decisions being made are health protective."

Despite their arbitrary grounding, the ensuing regulations are rigorously enforced. People are jailed and fined for acting contrary to official articles of faith, much as Galileo suffered. And just as in Galileo's day, regulatory prosecutors today also serve as judges and appoint jury members. All of this is considered legitimate because regulation is presumed to uphold the public interest. Yet, in a wager about human nature, it is a fair bet that public interests will be regularly trumped by institutional ones.

A Scientific Remedy? Could science offer a way out of this? Could it set some bright-line, objective standard for when regulation is, and is not, called for? Alas, current scientific research does not anticipate technical remedies anytime soon, despite fanciful forecasts that genomics, proteomics, and other "omics" sciences stand ready to help. Such approaches still would require new sets of default assumptions leading to judgmental risk speculations.

Taking cancer as the typical example, a true scientific test for human carcinogens devoid of default assumptions should provide a testable human-relevant mechanism of action at every step: from genomic or epigenetic events, to cell and organ targets, metabolism, repair and immune defenses, in situ establishment, angiogenesis, metastasis — to mention just some names. Similar considerations would hold for possible endocrine disruptors and other complex anomalies whose human risks remain untestable and putative. Verifiable mode-of-action approaches would have to integrate the probability of each step into an overall quantifiable probability of hazard both at personal and epidemiologic levels, which then could lead to reference doses, permissible exposures, and other regulatory targets. Nothing of the sort is remotely possible at the current stage of science, and we must accept that at any time science cannot have all desirable answers.

In the end, it should be apparent that the current approach to the regulation of incommensurable risks is incapable of a rational resuscitation, and should collapse under the weight of its absurd arrogance. The core philosophy of a remedy is that no bureaucratic, advocacy, or academic *nomenklatura*, and no appointee or expert, could have the wisdom and savvy to burden people with factually unsupported arbitrary dictates.

In a wager about human nature, it is fair to bet that public interests are regularly trumped by institutional ones.

The tyranny of arbitrariness should be the prime target of any significant improvement, for arbitrariness begets arbitrariness, and displaces health and environmental protections from being the principal goals of regulation. Indeed, arbitrariness makes it easy for special interests of all stripes to hijack regulation, as is the case today. What is more, arbitrariness becomes the main tool of regulatory proliferation, for it allows the limitless generation of public anxieties out of unsustainable conjecture, thus forcing new legislation and new regulatory assaults.

A clear example of such advancing degeneration is the current massive EPA program to regulate putative endocrine disruptors, in the absence of any epidemiologic evidence that such hazards affect public health. This is regulation against hypotheticals. Until now, the search for unknown hazards only occupied more or less serendipitous research, but how could the regulation of phantom hazards unknown to human health be justified as being science-based? The arbitrary imposition of imaginary hazards opens limitless horizons to invasive regulation, which should be checked before it metastasizes.

Arguably, even the arbitrary discretion of elected repre-

sentatives should be similarly constrained, but surely, bureaucratic temptations to adopt self-serving assumptions of evidence must be addressed. To this end, new laws ought to spell out the criteria for valid scientific evidence, and to demand regulations grounded on data that are independently testable and relevant to humans. This way, the mischief of "expert" opinions not supported by testable data would also be prevented. Legislators could find precedent and guidance in the Supreme Court's 1993 *Daubert v. Merrell Dow* opinion, which defines criteria that validate evidence that is admissible in federal courts. More comprehensive criteria would have to be found along the lines described earlier: namely, requirements for authenticable and relevant measurements with narrow error rates, coupled with warrants that spurious confounding interferences have been controlled, that results are consistently replicable, and that conclusive proofs are supported by counterfactual verifications.

A REMEDIAL PROPOSAL

Change would necessitate the repeal of the Delaney Clause and its doublets in various laws, and therefore the dismissal of long-term animal tests that are irrelevant to humans. At the present state of the art, regulation could only look for potential adverse effects that can be tested by acute and short-term tests in human volunteers and in animals, provided that animal tests offer testable translation to human conditions. Along with traditional tests for acute toxicities, short-term tests could be tailored to the nature of the matters considered and their proposed uses. Only epidemiologic studies could define long-term or chronic human hazards, if the evidence is validated by the objectivity criteria just mentioned. Take for instance the requirement of measurements with narrow margins of error: burdened with countless and often intractable uncertainties, epidemiologists have chosen a 5 percent margin of error as their threshold for statistical significance, thus accepting being wrong one time out of 20. Yet, would anyone drive a car whose brakes fail one out of 20 applications? Epidemiology and clinical trials ought to aspire to better evidentiary confidence before they could set regulations to inconvenience millions.

The intuitive, proven, and essential rule of toxics regulation is that the dose makes the poison. Therefore, in the absence of objective evidence of human risk, the regulatory recourse is to reduce exposures to the lowest levels compatible with useful applications, guided by the results of acute and short-term tests.

Issues of duplicative products could also raise interesting questions. Duplicate products may pose no issues if multiple formulations that are chemically and physically equivalent target a specific application, because it is unlikely to affect overall exposures. Considerations could vary when a singular application is the target of different formulations, which might be regulated based on their effectiveness, much as it is done for drugs on the basis of therapeutic indexes.

Public protection would be completed with an organized, detailed, and standing program of post-marketing epidemiologic surveillance for consumer and industrial exposures, similar to — but more pervasive and efficient than — the current post-marketing surveillance of medical products. Such a program should represent a major improvement over the current state of affairs, where general epidemiologic surveillance is not systematically organized and is largely left to the whim of investigators, with some exception for industrial exposures.

Necessary as reform might be, it would represent a significant departure from ingrained practices, and it would be resisted. Yet reform is a must, for the present arrangement is a shameful and intolerable abuse of public confidence. The amending of statutes is the first necessary step, and while fierce resistance could be expected from incumbent interests, the necessity of removing current default assumptions ought to win the day. Once that happens, reform will occur in a domino — if not painless — sequence: funds would be reallocated to imaginative productivities, careers would shift toward more research, new regulation procedures would arise, past regulations would be revisited, and industry would reassess how to interact with regulation.

Overall, the philosophy of an improvement allows that inevitable and necessary risk taking is better managed by real-time reason than by the costly and deceitful magic of default incantations. Reduced to transparent rational terms, the approach I have suggested would not fully erase intractable natural uncertainties, but it would remove the illusory claims that now defile the good name of science, and would redirect billion-dollar agency budgets toward more welfare-enhancing operations, while saving millions of animals. It would also close the deplorably farcical show of advisory committees that cannot objectively validate their own advice. Today's conflict of interest problems would be resolved or contained. More appealing still, many bright minds could return to productive research and developments, having rejected a regulatory paradigm mired in nonsensical rhetoric, unworthy of the intellectual honesty of science, and of the pledges of fairness and reason that the social contract of free people should warrant. R

Readings

- "An Evaluation of Chemical and Industrial Processes Associated with Cancer in Humans Based on Human and Animal Data," prepared by Ralph Althouse et al. *Cancer Research*, Vol. 43 (1980).
- "Cancer Risk Assessment: The Science That Is Not," by Gio B. Gori. *Regulatory Toxicology and Pharmacology*, Vol. 16 (1992).
- "Carcinogenesis Bioassay Data: Correlation by Species and Sex," by F. J. DiCarlo. *Drug Metabolism Reviews*, Vol. 15 (1984).
- "Results for 86 Two-Year Carcinogenicity Studies Conducted by the National Toxicology Program," by K. J. Haseman and D. D. Crawford. *Journal of Toxicology and Environmental Health*, Vol. 14 (1984).
- "Some Tautological Aspects of the Comparison of Carcinogenic Potency in Rats and Mice," by L. Bernstein et al. *Fundamental and Applied Toxicology*, Vol. 5 (1984).
- "The Causes of Cancer," by Richard Doll and Richard Peto. *Journal of the National Cancer Institute*, Vol. 66 (1981).
- "The Regulation of Carcinogenic Hazards," by Gio B. Gori. *Science*, Vol. 208 (1980).