Toxicology 371 (2016) 12–16

Contents lists available at ScienceDirect

Toxicology

journal homepage: www.elsevier.com/locate/toxicol

Upholding science in health, safety and environmental risk assessments and regulations

Michael Aschner[a], Herman N. Autrup[b], Sir Colin L. Berry[c], Alan R. Boobis[d], Samuel M. Cohen[e], Edmond E. Creppy[f], Wolfgang Dekant[g], John Doull[h], Corrado L. Galli[i], Jay I. Goodman[j], Gio B. Gori[k,*], Helmut A. Greim[l], Philippe Joudrier[m], Norbert E. Kaminski[n], Curtis D. Klaassen[o], James E. Klaunig[p], Marcello Lotti[q], Hans W.J. Marquardt[r], Olavi Pelkonen[s], I. Glenn Sipes[t], Kendall B. Wallace[u], Hiroshi Yamazaki[v]

[a] Professor, Molecular Pharmacology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, United States
[b] Emeritus Professor, Institute of Public Health, University of Aarhus, Aarhus, Denmark
[c] Professor Emeritus of Pathology, Queen Mary, London, UK
[d] Professor of Biochemical Pharmacology,Department of Medicine, Imperial College, London, UK
[e] Professor, Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA
[f] Head of Toxicology Department, Faculty of Pharmaceutical Sciences, Université Bordeaux Segalen, Bordeaux, France
[g] Professor of Toxicology, Department of Toxicology, University of Wuerzburg, Wuerzburg, Germany
[h] Emeritus Professor, Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, Kansas City, KS, USA
[i] Professor, Toxicology and Risk Assessment, Pharmacological and Biomolecular Sciences, University of Milan, Milan, Italy
[j] Professor of Pharmacology and Toxicology, Michigan State University, East Lansing, USA
[k] Emeritus, The Health Policy Center, Editor, Regulatory Toxicology and Pharmacology, Bethesda, MD, USA
[l] Emeritus Professor, Technical University of Munich, Munich, Germany
[m] Emeritus Director of Research INRA, Montpellier, France
[n] Professor, Pharmacology & Toxicology, Director, Institute for Integrative Toxicology, Michigan State University, East Lansing, MI, USA
[o] Affiliate Professor, Dept. of Environmental and Occupational Health, University of Washington, Seattle, WA, USA
[p] Professor, Department of Environmental Health, University of Indiana, Ellettsville, IN, USA
[q] Professor, Department of Cardiology, Thoracic and Vascular Sciences, School of Medicine, University of Padua, Padua, Italy
[r] Professor Emeritus, Experimental and Clinical Toxicology, University of Hamburg Medical School, Hambug, Germany
[s] Professor and Chair Emeritus, Department of Pharmacology and Toxicology, Faculty of Medicine, University of Oulu, Oulu, Finland
[t] Professor Emeritus, University of Arizona, Tucson, AZ, USA
[u] Professor & Associate Dean for Faculty Affairs, University of Minnesota Medical School Duluth, Duluth, MN, USA
[v] Dean of Graduate School & Professor, Drug Metabolism and Pharmacokinetics, Showa Pharmaceutical University, Machida, Tokyo, Japan

ARTICLE INFO

Article history:
Available online 14 September 2016

Keywords:
Scientific evidence
Regulation
Regulatory policy
Regulatory ethics
Animal bioassays
Hazard assessment
Risk assessment

ABSTRACT

A public appeal has been advanced by a large group of scientists, concerned that science has been misused in attempting to quantify and regulate unmeasurable hazards and risks.[1] The appeal recalls that science is unable to evaluate hazards that cannot be measured, and that science in such cases should not be invoked to justify risk assessments in health, safety and environmental regulations.

The appeal also notes that most national and international statutes delineating the discretion of regulators are ambiguous about what rules of evidence ought to apply. Those statutes should be revised to ensure that the evidence for regulatory action is grounded on the standards of the scientific method, whenever feasible. When independent scientific evidence is not possible, policies and regulations should be informed by publicly debated trade-offs between socially desirable uses and social perceptions of affordable precaution. This article explores the premises, implications and actions supporting the appeal and its objectives.

* Corresponding author.
E-mail addresses: Michael.Aschner@einstein.yu.edu (M. Aschner), ha@ph.au.dk (H.N. Autrup), colin@sircolinberry.co.uk (S.C.L. Berry), a.boobis@imperial.ac.uk (A.R. Boobis), scohen@unmc.edu (S.M. Cohen), edmond.creppy@u-bordeaux.fr (E.E. Creppy), dekant@toxi.uni-wuerzburg.de (W. Dekant), jdoull@kumc.edu (J. Doull), corrado.galli@unimi.it (C.L. Galli), goodman3@msu.edu (J.I. Goodman), gorigb@msn.com (G.B. Gori), helmut.greim@lrz.tum.de (H.A. Greim), philippe.joudrier@neuf.fr (P. Joudrier), kamins11@msu.edu (N.E. Kaminski), curtisklaassenphd@gmail.com (C.D. Klaassen), jklauni@indiana.edu (J.E. Klaunig), marcello.lotti@unipd.it (M. Lotti), marquardt@uke.uni-hamburg.de (H.W.J. Marquardt), olavi.pelkonen@oulu.fi (O. Pelkonen), Sipes@email.arizona.edu (I. G. Sipes), kwallace@d.umn.edu (K.B. Wallace), hyamazak@ac.shoyaku.ac.jp (H. Yamazaki).

[1] An Appeal for the Integrity of Science and Public Policy. Toxicology, September 4, 2016. doi:10.1016/j.tox.2016.08.015.

http://dx.doi.org/10.1016/j.tox.2016.09.005
0300-483X/

The scientific context

As a central premise of the appeal, the raison d'être of health and safety regulations is to control hazards and prevent risks, based on testable evidence provided by toxicology and epidemiology, or informed by sensible considerations of precaution. Hazards of interest entail interactions of atoms, molecules, radiations or other physical forces that can be observed and measured directly, or through instrumentation. Hazards may cause adverse effects in exposed populations when hazard potency and the intensity and duration of exposure combine to exceed no-effect thresholds. In turn, the risks of adverse effects are assessed by measuring the frequency of such effects against the intensity of hazard exposure in differently exposed and in non-exposed populations.

Epidemiology has the singular advantage of dealing with humans, and has been successful in identifying and preventing infectious diseases tied to single necessary causes. However, most chronic diseases of current interest – cancer, cardiovascular disorders and more – are linked to multiple and simultaneous hazards, which generally raise dominant barriers to unambiguous causal determinations for most retrospective epidemiologic studies. For the same reasons, prospective intervention studies have not been more successful despite subject matching and randomization efforts, which may mitigate the influence of a few variables of interest but not the bulk of multifactorial confounders of causal interpretations. Essentially, prevailing difficulties exist in measuring individual or group exposures reproducibly, and in measuring and controlling multiple externalities capable of confounding observations and results.[2]

With the exception of cigarette smoking, certain infections, and some occupational and medical exposures that can be reasonably measured, the contributions of multifactorial epidemiology to public health and policy have been precautionary—a conclusion especially true in the wake of numerous and massive intervention trials designed to test initial epidemiologic hypotheses, trials that have regularly disappointed.[3] The effective role of epidemiology is to continue investigations of occupational and other restricted settings where exposures and externalities are amenable to measurement, and to provide tentative causal clues that toxicology could investigate.

As an experimental science, toxicology is expected to follow the standards of the scientific method in attaining quantifiable evidence of physical hazards. Much has been written about the philosophical underpinnings of the method,[4] although the method and science itself would be meaningless without a few evident and essential operational standards. These ask for numerical measurement with explicit and suitably small error rates, for authentic representations of what is being measured, and for measurements that are relevant to the issues being considered, i.e. relevant to humans when testing for human hazards. They also ask for the control of externalities that may confound observations and conclusions, for detailed procedural descriptions, and for results that are reproducible by independent investigators. Ground controls should also be included to allow counterfactual inferences. Precise, authentic, relevant and reproducible measurements are the foundations of reliable scientific evidence.[5]

Unlike the absolute truths of purely intellectual disciplines, such as mathematics, geometry and formal logic, natural sciences postulate empirical truths in a probabilistic context, because of the inherent approximations of measurements, the multitude of potentially confounding variables and the natural vagaries of atoms, molecules and overall matter. Although philosophically provisional, such truths – or natural laws – are generally verified counterfactually by reliable controls and applications: airplanes fly, radio waves convey signals, and therapies cure. Indeed, it is the operational standards of the scientific method as just described, which have allowed science to accumulate a body of empirical knowledge sufficiently certain to enable all successful technologies and applications that sustain advanced societies.

Empirical science also includes a research activity dealing with knowledge-in-the-making, aiming at validating emerging hypotheses to virtual certainty. Yet, hypotheses are not theorems, and research conjectures and preliminary findings are scientific in the sense of being part of scientific research, but are not part of the validated and operational knowledge of science.[6] Thus, as the appeal implies, it should be unethical to use untested research presumptions in justifying policies and regulations that substantially interfere with national economies, that influence the anxieties, choices and behavior of billions of citizens, and that can impose massive penalties and even detention on transgressors. In this light, the appeal maintains that hazards characterized by the scientific method can justify regulation on their own account—a justification that is not permissible when the significance of putative hazards cannot be assessed empirically.

Testing for and measuring human hazards

For ethical and practical reasons, tests for the regulation of potential human hazards are conducted in animals, mostly rats and mice. Although animals are not Man, short-term animal tests offer experimentally verifiable insights on short-term adverse effects in animals and humans, and on the threshold exposure conditions

[2] Green, M.D., Friedman, D.M., Gordis, L. *Reference Guide on Epidemiology*. Reference Manual on Scientific Evidence, Third Edition. Federal Judicial Center, National Research Council. National Academy Press, Washington DC. 2011. http://www.fjc.gov/public/pdf.nsf/lookup/SciMan3D12.pdf/$file/SciMan3D12.pdf

[3] Werkö, L. *The enigma of coronary heart disease and its prevention*. Acta Med Scand. 1987;221:323–333. Werkö, L. *Analysis of the MRFIT screens: A methodological study*. J. Int. Med. 1995;237:507–518. Hakama M., Beral V., Cullen J., Parkin M. *UICC workshop on evaluating interventions to reduce cancer risk*. Int. J. Cancer. 1989;43:967–969. Strandberg T.E., Salomaa V.V., Vanhanen H.T., Neukkarinen V. A.; Sarna S.J., Miettinen T.A. *Mortality in participants and non-participantsof a multifactorial prevention study of cardiovascular diseases: A 28 year follow up of the Helsinki Businessman Study*. Br. Heart J. 1995;74:449–454. Luepker R.V., Murray D.M., Jacobs D.R., Mittelmark, M.B., Bracht N., Carlaw R., et al.; *Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Health Program*. Am. J. Publ. Health. 1994;84:1383–1393. The Multiple Risk Factor Intervention Trial Research Group. *Mortality after 16 years for participants randomized to the Multiple Risk factor Intervention Trial*. Circulation. 1996;94:946–951. Feinlieb M.; *New directions for community intervention studies*. Am. J. Publ. Health. 1997;86:1696–1698. Shaten J.B., Kuller L.H., Kjeselberg M.O., Stamler J., Ockene J.K., Cutler J.A., et al.; *Lung Cancer Mortality after 16 years in MRFIT Participants in Intervention and Usual-Care Groups*. Ann. Epidemiol. 1997;7:125–136. Taubes G.; *The soft science of dietary fat*. Science. 2001;291:2536–2545. Taubes G.; *Epidemiology faces its limits*. Science 1995;269:164–169. Editorial; *Do epidemiologists cause epidemics?* Lancet 1993;341:993–994.

[4] Gauch H.G., Jr.; *Scientific Method in Practice*. Cambridge University Press, Cambridge, UK, 2003.

[5] Hand D.J.; *Measurement theory and practice. The world through quantification*. Arnold, London, UK, 2004.

[6] Berry, Sir Colin; *Relativism, Regulation and the Dangers of Indifferent Science*. Toxicology 2010; 267; 7–13. Gori G.B.; *Science, Imaginable Risks, and Public Policy: Anatomy of a Mirage*. Regulatory Toxicology and Pharmacology. 1996;23:304–311.

that enable or prevent those effects.[7] Assisted by pharmacokinetic experiments and modeling, such tests can determine or predict the absorption, internal distribution, metabolism and excretion rates of natural or synthetic agents in animals and humans. They can also determine the physiologic concentrations and timing needed for the occurrence and observation of short-term adverse effects in different organs and body compartments. Matching such evidence with biomarker and field or simulated exposure data, and with benchmarks for thresholds of toxicologic concern (TTC),[8] it is possible to assess objectively and to regulate safe levels of short-term exposure. This is feasible because adverse effects occurring in less than 90 days generally concern basic and stable physiologic mechanisms that are conserved in many species including humans, and are not significantly overcome by transient random disturbances.

Animals, however, are not credible human surrogates in testing for hazards that depend on chronic exposures, with effects that evolve stochastically over protracted times, and through multiple and mostly unpredictable, unknown, or incidental modes of action. Examples are cancer, cardiovascular and neurological deficits, endocrine disruptions, aberrations of reproduction and immunity, and other anomalies triggered by complex events over a long time. In testing for such endpoints, the genetic, somatic, behavioral, life history, disease sensitivities, environmental and dietary adaptations, and other disparities in different species diverge into different causal opportunities often species-specific and most likely unique to individual pathologies.[9] Hence, the outcomes of chronic tests are inconsistent across species, and among strains and individuals of the same species.

Clearly, the primary responsibility of regulatory toxicology is to ensure that animal results are true proxies for human effects, but methods for a valid objective translation of chronic test results are not presently available. Research to resolve this problem should be a major priority in toxicology, and the foundation of a central framework for testing potential chronic human hazards. The appeal deplores that standard chronic animal tests are prescribed to justify massively costly regulations according to a set of mandatory default assumptions, without concern of whether the results obtained are defensible proxies for chronic human responses.[10]

Are animal test results valid proxies for human hazards?

The archetypal example of chronic animal tests is the bioassay for carcinogens. It officially entered in regulatory practice with the 1958 Delaney Clause of the Federal Food, Drug, and Cosmetics Act, which prohibits food additives that cause cancer in animals.[11] In time, the intent of the Clause migrated beyond foods, thus imposing the assumption that lifetime animal tests adequately assess human risks due to other hazards and exposures.

Cancer bioassays, however, yield different results depending on how they are set up and run, and it would have been sensible at the start to identify test conditions relevant to the forecast of human experiences. Instead, regulatory guidelines have imposed as a matter of law that human hazards be derived from bioassays run and interpreted according to a standard set of default assumptions, all in contrast with human experiences. Of these, the most significant are that rats and mice must be accepted as valid human surrogates; that bioassay must be run at overtly toxic maximum tolerated doses for the lifetime of the animals; that bioassay interpretation must assume that such lifetime animal exposures represent humans exposed to much smaller doses and shorter durations; that the metabolism of animals at maximum tolerated doses parallels the metabolism of humans at low doses; that benign lesions are on par with malignant ones; that the route of exposure is irrelevant; that extrapolations from high doses in animals to low doses in humans must be linear, and more. Included is the perplexing statement that"*[a]nimal studies are conducted at high doses in order to provide statistical power . . .*."[12]

Such prescribed assumptions have reduced much of chronic regulatory bioassays into a formulaic paradigm that has inhibited intellectual and experimental inquiry, while knowing and admitting all along that objective science is effectively removed from the process.

The International Agency for Research on Cancer, an arm of the World Health Organization conceded 35 years ago "*. . . a correlation between carcinogenicity in animals and possible human risk cannot be made on a scientific basis*."[13] The late Dr. David Rall, then director of the National Toxicology Program, testified in 1981 before an astonished Congressman Albert Gore, that science could not derive human cancer risk from animal tests. He insisted that Congress and the public ought to have *faith* (sic) in experts who examine animal entrails searching for human cancer risks.[14] A National Research Council committee reviewing human cancer risks and animal tests, concluded "*. . . many components lack definitive scientific answers . . . [because] . . . the dominant analytic difficulty is the pervasive uncertainty*."[15]

Indeed, as the appeal implies, cancer bioassay results in rats and

[7] Cohen, S.M.; *Human carcinogenic risk evaluation: an alternative approach to the two-year rodent bioassay*." Toxicol. Sci. 2004;80(2): 225–229. Cohen, S.M.; *Evaluation of Possible Carcinogenic Risk to Humans Based on Liver Tumors in Rodent Assays: The Two-Year Bioassay Is No Longer Necessary*. Toxicologic Pathology 2010;38 (3): 487–501. Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W.; *IPCS framework for analyzing the relevance of a cancer mode of action for humans*. Crit. Rev. Toxicol. 2006;36(10): 781–792. Cohen, S. M., Arnold L.L.; *Chemical carcinogenesis*. Toxicol. Sci. 2011;120 Suppl. 1: S76–92. Boobis, A.R., et al.; *IPCS framework for analyzing the relevance of a noncancer mode of action for humans*. Crit. Rev. Toxicol. 2008;38(2): 87–96.

[8] Cramer, G. et al. *Estimation of toxic hazard —a decision tree approach*. Food Cosmet. Toxicol. 1978;16:255–76. Kroes, R. et al. *Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet*. Food Chem. Toxicol. 2004;42:65–83. EFSA. *Scientific Opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC)*. EFSA Journal 2012;10(7) 2750. U.S. Food and Drug Administration. *Threshold of Regulation (TOR) for indirect food additives*. 1995;21C.F.R. §170.39.

[9] Shibata D.; Heterogeneity and tumor history. Science 2012;336:304–305.

[10] U.S. Environmental Protection Agency. *Guidelines for Carcinogen Risk Assessment*. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. March 2005. http://www2.epa.gov/sites/production/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf

[11] FEDERAL FOOD, DRUG, AND COSMETIC ACT; §348. *Food additives*, c,3,A. http://www.gpo.gov/fdsys/pkg/USCODE-2011-title21/pdf/USCODE-2011-title21-chap9.pdf

[12] Ref. 10, page A4.

[13] IARC; Working Group, International Agency for Research on Cancer. An evaluation of chemicals and industrial processes associated with cancer in humans based on human and animal data. Cancer Res. 1980;43:1 52. http://cancerres.aacrjournals.org/content/40/1/1.full-text.pdf

[14] Rall D.P.; Hearing before the Subcommittee on Investigations and Oversight. Committee on Science and Technology, U.S. House of Representatives, July 15, 1981, pp. 52, 53, 57. U.S. Government Printing Office, Washington, DC. pp 52–57. http://babel.hathitrust.org/cgi/pt?view=image;size=100;id=mdp.39015008433222;page=root;seq=1

[15] National Research Council; *Risk assessment in the federal government: Managing the* process. National Academy Press, Washington DC, 1983. pp. 36 and 11. http://www.nap.edu/catalog/366.html

mice cannot be relevant to humans because the lifespan and histories of these rodents, their genetics, physiology, metabolism, reproduction, anatomy, ecology and more are not comparable to human conditions.[16] Not surprisingly, it was early noted that the results of official cancer tests in rats and mice do not match better than tossing a coin.[17]

The current regulation of putative hazards

Concerned about adding legitimacy to default assumptions, regulatory agencies have funded a series of committees of the National Academy of Sciences (the Academy) over the last decades, which ended up endorsing those assumptions. A 2009 report of the Academy reads *"[d]efaults 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."* And further " . . . *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."*[18] Bereft of testable evidence, the Academy's committee and the regulatory agency appear to be fully aware that science and commonsense do not allow predictive testing for human cancer hazards in animals. All the same, they endorse and prescribe such tests under the illusory mantle of "regulatory science".

The appeal deplores the enforcement of unwarranted animal tests as instruments of human risk assessment, with the approval of top scientific institutions.[10,15,18,19,20] Far reaching and very costly regulations have been enacted, assisted by gratuitous protestations of scientific validity, and covered by elaborate complexities of jargon and quantitative illusions. Corollary disciplines have been constructed to add artifices in human cancer hazard and risk assessment from bioassay data. They include complex procedures for micro-extrapolations of oral slope factors and inhalation risk units for each tumor type observed in animals, from which claims of carcinogen potencies in humans are supposedly derived. Debates continue on whether mutagenicity and genotoxicity can predict carcinogenicity in animals and in humans. Ever more complex mathematical and statistical algorithms are devised to figure whether similarities of chemical structure may predict carcinogenicity in animals and humans. Spurred by Academy's declarations, new generations of toxicologists believe that carcinogen potency in humans could be predicted by quick sketchy signals from in vitro micro cell cultures, genomic mappings, and other extremely reductionist assays.[19] Although these models may have a research role, they are fraught with major uncertainties: crucial ones stemming from default assumptions tied to chronic bioassay and similarly precarious benchmarks. Such satellite exercises have been forced into regulation without testable relation to human hazards and risks.

Still, the results of costly cancer bioassays are not the final justification of regulatory action, which is further extended with the application of offhand safety factors and the judgmental appraisals of advisory committees. Safety factors are arbitrary multipliers applied to augment risk derivations from bioassay results, summing up to notional adjustments often exceeding several orders of magnitude. The functions of advisory committees are more nuanced. In the words of a 2009 Academy report: *"[w]hen large uncertainties result from a combination of lack of data and lack of conceptual understanding (for example, a mechanism of action at low dose), some regulatory agencies have relied on expert judgment to fill the gaps or establish default assumptions. Expert judgment involves asking a set of carefully selected experts a series of questions . . . "*[21] But the same Academy report highlights how such committees are confronted with major difficulties in attempting 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 . . . "* And therefore *"[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."*[22]

If expert committees are admittedly incapable of resolving evidence that does not exist, why use committees in the first place? A contextual answer is provided by a report of the US Office of Government Ethics, to the effect that advisory committees are to offer technical support to offset an appearance of official arbitrariness. To this end, the Ethics Office report cites the statutory requirement that the Academy and government agencies at large must recruit advisory committee members who are free of conflicts of interest.[23] Admittedly, the obvious interests in regulation are those of the regulated, of the regulators and the public interest, but the statute directs that only financial interest potentially favoring the regulated are to be scrutinized and suppressed. To formalize enforcement, agencies investigate individual financial conflicts of interest, keep dossiers of people who do not qualify, and have sole discretion on appointments.

All said, it remains unclear whether bioassay results, safety factors and advisory committees have resulted in regulations with testable public health outcomes. Earlier, the US-EPA 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."*[20] Similarly, the Academy concluded" . . . *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."*[24]

Clearly, the open use of illusory evidence and rhetorical arguments clashes with the fundamental public health motives and responsibilities of regulation. This being reality, the appeal implies that a first step toward a remedy is to admit there is no objective way to regulate unknowable and putative long-term hazards and risks, in ways that ensure explicit public health protection. Facing the unknown, regulatory intervention could only be precautionary, which raises further questions.

[16] Gori GB; Long-term animal bioassays: is the end near? Toxicol Pathol. 2013 Jul;41(5) 805–807. doi: 10.1177/0192623312467524.

[17] Purchase, I.F.H. (1980). *Inter-species comparison of carcinogenicity*. Br. J. Cancer 41, 454–468. Haseman, K.J., and Crawford, D.D. (1984). *Results for 86 two year carcinogenicity studies conducted by the National Toxicology Program*. J. Toxicol. Environ. Health 14, 621–639. Di Carlo, F.J. (1984). *Carcinogenesis bioassay data: Correlation by species and sex*. Drug Metab. Rev. 15, 409–413. Bernstein, L., et al. (1985). *Some tautologous aspects of the comparison of carcinogenic potency in rats and mice*. Fundam. Appl. Toxicol. 5, 79–88. Ashby J, Tennant RW; *Definitive relationships between chemical structure, carcinogenicity, and mutagenicity for 301 chemicals tested by the U.S. NTP*. Mutat. Res. 1991;257:229–306.

[18] National Research Council; *Science and Decisions. Advancing Risk Assessment*. National Academy Press. 2009. p.192 http://www.nap.edu/catalog.php?record_id=12209

[19] National Research Council; *Toxicity Testing in the 21 st Century: A Vision and a Strategy*. National Academy Press, Washington, DC. (2007). http://www.nap.edu/catalog.php?record_id=11970

[20] EPA; US Environmental Protection Agency. *Guidelines for Carcinogen Risk Assessment*. Publication EPA/630/R-00/004. Page 13. September 1986, Washington DC. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=30004TZX.txt

[21] Ref. 18, p. 101.

[22] Ref. 18, p. 102.

[23] Office of Government Ethics. *Acts affecting a personal financial interest*. 18 U.S.C. § 208. https://www.oge.gov/Web/oge.nsf/Resources/18+U.S.C.+%C2%A7+208:+Acts+affecting+a+personal+financial+interest

[24] Ref 18. P. 17.

A call for action

Is it reasonable, ethical and defensible to justify public policy sanctions based on blank assertions of precaution? In a civics context, such justification is not sustainable, for executive decisions grounded on imagined precaution are by necessity arbitrary.[25] Precaution is costly, with a price that can only be assessed against the cost of what is being forfeited, inclusive of regulation's costs. Hence, absent independently verifiable evidence, the appeal argues that precautionary regulations should be defined by publicly debated tradeoffs between uses socially perceived as desirable and socially perceived degrees of affordable precaution.

Ways to define these social perceptions would be a key to such tradeoffs, and clearly beyond the insular compass of regulatory agencies. Yet, worldwide legislators have delegated the development and enforcement of policies and regulations to unelected administrators. Reflecting a perplexing tradition of wide executive discretion, statutes that discipline such delegated authority are relaxed on how evidence is to be obtained and assimilated in policies and regulations. A prominent instance is how the US Administrative Practices Act addresses this task:

> "*Except as otherwise provided by statute, the proponent of a rule or order has the burden of proof. Any oral or documentary evidence may be received, but the agency as a matter of policy shall provide for the exclusion of irrelevant, immaterial, or unduly repetitious evidence. A sanction may not be imposed or rule or order issued except on consideration of the whole record or those parts thereof cited by a party and supported by and in accordance with the reliable, probative, and substantial evidence*."[26]

Accordingly, a regulatory agency has the burden of proof, but also unlimited discretion in selecting preferred information. Moreover, the language does not define what may constitute "reliable, probative, and substantial evidence", nor does it instruct how to proceed when such evidence is unavailable and cannot be obtained. As a remedy, the appeal asks for statutory modifications to prescribe the application of the scientific method whenever feasible, in obtaining independent evidence to enable regulation. Modifications should not dwell on the well-known philosophical framework of the scientific method, but rather on the nuts and bolts of its implementation. Accordingly, a statutory modification should ask that evidence be grounded on empirical measurements with explicit and suitably small error rates, with authentic representations of what is being measured, and with testable relevance to the issues being considered, i.e. relevant to humans when testing for human hazards. It also should ask for the measurement and control of externalities that may confound observations and conclusions, for detailed procedural descriptions, and for results that are reproducible by independent investigators.

In the absence of scientific hazard assessments, a modified statute should detail how to define precautionary regulations by conducting public tradeoff debates between socially desirable uses and socially affordable precaution. To complement and reinforce precaution, an epidemiologic surveillance program could be prescribed to monitor early signs of possible adverse effects in most heavily exposed cohorts, similar to what is being done in the pharmacovigilance of medicines and medical devices.

Statutory modifications would be achieved by political pressure from scientific societies and academies, and by securing media coverage of the compelling ethical, logical and political reasons underlying this effort in free societies. It would be expedient to focus first on dominant countries and country blocks, aiming to set in motion a cascade of events in other countries. The United States offers unique opportunities, assisted by the 1993 Daubert opinion of the US Supreme Court, which mandates scientific method standards as the litmus test for evidence admissible in federal courts.[27]

Under the proposed modifications, the regulation of short-term hazards based on short-term animal and human tests may not change substantially. Long-term hazards would be objectively regulated if verifiable evidence were available, as it happens with successful occupational health interventions. In the absence of such evidence, precautionary regulations of putative long-term hazards would rely on projected exposure simulations and short-term pharmacokinetic determinations, linked to exposure benchmarks and thresholds of toxicologic concern (TTC). Different for various exposures and exposed populations, TTC benchmarks have been extensively discussed and often used in regulation,[6] but owing to their precautionary component they would need to be socially redefined and approved, and not be entrusted to narrow administrative discretion. Exposures to putative chronic hazards that may exceed TTC benchmarks would require public debates to define what minimal exposures would allow desirable uses. Essential to those debates is a public acceptance that real or putative hazards cannot be fully avoided, especially in complex societies.

Conclusion

Change is overdue: especially so in the US, where statutory warrants for reform have been available for a long time. The US Administrative Practice Act details a list of official administrative misbehaviors, identified as sufficiently unlawful to "*. . . set aside agency action, findings, and conclusions found to be (1) arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law; (2) contrary to constitutional right, power, privilege, or immunity; (3) in excess of statutory jurisdiction, authority, or limitations, or short of statutory right; (4) without observance of procedure required by law; (5) unsupported by substantial evidence . . .; or (6) unwarranted by the facts . . .*."[28] The list sounds as a verbatim indictment of what is wrong with the present regulatory arrangement. Could it become an earlier incentive for reform?

In the end, moving from illusory assessments of risk to socially balanced assessments of utility and precaution would not erase intractable uncertainties about putative long term hazards and risks. Yet, it would reaffirm the integrity of science and restore public trust in government regulations and policies. Billion dollars could shift to fund more research grants and other creative operations, while saving untold numbers of animals. Many bright minds could return to fruitful research, no longer tied to a regulatory paradigm unworthy of the intellectual honesty of science. Indeed, vast regulatory resources bereft of verifiable social and public health benefits ought to be made productive, under the pledge of fairness and reason that the social contract of free people should warrant.

Acknowledgments

The authors declare no conflicts of interest. No external funding was sought or obtained for this article. Originally drafted by Gio Batta Gori and Sir Colin Berry the article was enriched by the contributions of all authors. Co-authors Kendall B. Wallace and Hans W.J. Marquardt are co-editors of this journal.

[25] Hamburger P.; *Is Administrative Law Unlawful?* University of Chicago Press. Chicago 2014. DeMuth C.; Can the administrative state be tamed? J. Legal Analysis February 29, 2016:1–70.

[26] Administrative Procedure Act. 5 U.S.C. Subchapter II; § 556 (d). http://www.archives.gov/federal-register/laws/administrative-procedure/556.html

[27] Daubert v. Merrell Dow Pharms., Inc., (1993) 509 U.S. 579, 590, 593–94.

[28] US Administrative Procedures Act. Section 10, (e). http://www.justice.gov/sites/default/files/jmd/legacy/2014/05/01/act-pl79-404.pdf