

Endocrine Disruptors: Improving Regulatory Science Policy

Dose-Response:
An International Journal
2015:1-14
© The Author(s) 2015
DOI: 10.1177/1559325815611903
dos.sagepub.com



Paolo F. Ricci¹

Abstract

Law and science combine in the estimation of risks from endocrine disruptors (EDs) and actions for their regulation. For both, dose–response models are the causal link between exposure and probability (or percentage change) of adverse response. The evidence that leads to either regulations or judicial decrees is affected by uncertainty and limited knowledge, raising difficult policy issues that we enumerate and discuss. In the United States, some courts have dealt with EDs, but causation based on animal studies has been a stumbling block for plaintiffs seeking compensation, principally because those courts opt for epidemiological evidence. The European Union (EU) has several regulatory tools and ongoing research on the risks associated with bisphenol A, under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation and other regulations or directives. The integration of a vast (in kind and in scope) number of research papers into a statement of causation for either policy or to satisfy legal requirements, in both the United States and the EU, relies on experts. We outline the *discursive dilemma* and issues that may affect consensus-based results and a Bayesian causal approach that accounts for the evolution of information, yielding both value of information and flexibility associated with public choices.

Keywords

endocrine disruptors, bisphenol A, uncertainty, causation, science-policy, consensus, risk management

Introduction

Causation and its scientific evidence are necessary, but not sufficient, components of risk–cost–benefit analysis in public health. Causation is a key to understand and justify alternative choices of allowable exposure (eg, concentrations, doses, and dose rates). Causal analyses rely on dose–response models that relate exposure (eg, mass/volume) to specific adverse health effects and monotonic dose–response models that connect dose (eg, mg/kg body weight/d) using either cumulative probability of response or percentage effect. Some models are nonlinear (eg, nonmonotonic) at low doses. Causation is part of the legal evidence that justifies the often costly choices to reduce exposure to hazardous agents. An agency or other authority's standard setting may trigger judicial review by those who believe that the regulation went too far (and thus caused them costly actions) or not far enough (and thus endanger the public). This article deals with the legal–regulatory aspects of exposure to chemical endocrine disruptors (EDs) in the United States and European Union (EU) law by developing its following aspects:

1. judicial acceptance of scientific expert testimony,
2. policy science acceptance of conflicting scientific evidence,

3. aggregation of informed beliefs,
4. choices justified by theoretical principles, and
5. probabilistic causation based on Bayesian reasoning.

The effects of EDs at very low doses are unlike those predicted by the linear no-threshold model (LNT) at low doses (a monotonic function) and its main alternatives, the J-shaped biphasic (hormetic) models for cancer and the U-shaped dose response (both of which are nonmonotonic) for other end points. The reason is that it is increasingly clear that the correct dose–response model for those disruptors is nonmonotonic, due to biological mechanisms that are present at very low levels of exposure, but not at higher levels, as is commonly assumed for carcinogenic and other toxic end points. A class of models¹ that allows for all of these alternative qualitative behaviors may be useful to describe possible counterintuitive properties.

¹ PhD, LLM, Public Health Consultant

Corresponding Author:

Paolo F. Ricci, 685 Hilldale Ave, Berkeley, CA 94708, USA.
Email apricci@earthlink.net



Causal models are essential components of risk–cost–benefit analyses generally required under US environmental and health legislation and European constitutional law, under the Precautionary Principle. Specifically, the EU’s Consolidated Treaties² deal with the protection of the environment and public health and “the prudent and rational utilization of natural resources.” It states that:

Union policy on the environment shall aim at a high level of protection taking into account the diversity of situations in the various regions of the Union. It shall be based on the precautionary principle and on the principles that preventive action should be taken, that environmental damage should as a priority be rectified at source and that the polluter should pay.²

This command is bounded by the EU Commission as follows³:

The Commission stresses that the precautionary principle may only be invoked in the event of a potential risk and that it can never justify arbitrary decisions. The precautionary principle may only be invoked when the 3 preliminary conditions are met:

- identification of potentially adverse effects,
- evaluation of the scientific data available, and
- the extent of scientific uncertainty.

The European Commission guidance interprets this command by stating that³:

... the general principles of risk management ... are:
 proportionality between the measures taken and the chosen level of protection,
 nondiscrimination in application of the measures,
 consistency of the measures with similar measures already taken in similar situations or using similar approaches,
 examination of the benefits and costs of action or lack of action, and
 review of the measures in the light of scientific developments.

Roughly, public health risk analysis relies on 2 approaches to regulate hazardous exposures. One uses a specific model of dose response (response [probability, percentage] = f [exposure or dose]) developed from theoretical and empirical evidence, applies a policy risk level (response is measured by probabilities), and solves for its corresponding tolerable dose. The other relies on a dose that is demonstrably below the dose that causes no adverse health effects in an animal test system. This dose is calculated by dividing the dose known to cause no adverse health effects by a factor of safety (ie, the experimental no-effect dose is divided by a unit-less denominator that can be $\gg 10$). Specifically, hazardous exposures are regulated (eg, via numerical standards) at acceptable or tolerable levels. One such level is risk based: for instance, one in a million lifetime chance of cancer death. The other is deterministic: the tolerable daily intake (TDI) is obtained by applying the factors of safety discussed. For example, the US Environmental Protection Agency (EPA) developed some “reference doses” (in $\mu\text{g}/\text{kg}$ body

weight/d) using a safety factor of 10^3 , which is the result of multiplying 3 uncertainty factors each of which equals 10. These account for scaling from animals to humans (interspecies conversion uncertainty), human heterogeneity relative to the experimental animals, and converting from subchronic to chronic exposures.⁴ This reference dose was calculated from an experimental lowest observed adverse effect level (LOAEL).⁵

The EU and its Member States have different views on the importance of causal evidence and on the dose–response models used to assess risks associated with exposure. For example, if a carcinogen acts directly on a gene, then the assumption is that there is no threshold for that carcinogen. The LNT is used. When the regulation involves a tumorigenic dose ($\text{TD}_{x\%}$) that has been determined to cause cancer in 25% (or less) of the animals in a study, the tolerable exposure level for humans is 1 of 1000 times lower than the $\text{TD}_{x\%}$. If a carcinogen’s mode of action is epigenetic, it may be characterized by an experimental threshold (and thus the cognizant public agency may use factors of safety to yield a tolerable dose for humans). Some Member States do not use these approaches but rely on the scientific consensus about the danger from exposure.⁶ The difference in setting acceptable or tolerable doses is that some Member States use an LNT model, whereas others opt for the no observed adverse effect level or the LOAEL, and thus obtain thresholds. That is, the experimental exposure is decreased, through factors of safety, to establish a legally justified acceptable exposure. For example, under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation, discussed later, the derived no-effect level (DNEL) “represents a level of exposure above which humans should not be exposed.”⁷ When no DNEL can be derived, “REACH requires a qualitative assessment to be performed.”⁷ Also, “for non-threshold endpoints, if data allow, the development of a (semi) quantitative reference value (the DMEL, derived minimal effect level) may be useful.”⁷

Endocrine disruptors are ubiquitous and can be introduced in the body through various routes; a key mechanism of action is preventing a natural hormone to bind with its receptor. A vexing issue with these disruptors is their higher potency at low doses rather than at higher doses.⁸ This would prevent the inclusion of either a threshold or a J-shaped biphasic mechanism in which the descending arm of the J-shaped curve implies some reduction in the percentage response. Its ascending arm depicts increasing percentage adverse response. In the EU, these and many other toxicological findings have led to calls for a strategy that goes beyond REACH as well as bans on bisphenol A (BPA) by a number of Member States of the EU. For instance, European Food Safety Authority (EFSA)⁹ reconfirmed a TDI of 0.05 mg/kg body weight, although, in the United States, the US Food and Drug Administration (FDA) considers BPA to be safe. The EFSA’s reevaluation of BPA states that (emphasis omitted)¹⁰:

EFSA’s comprehensive re-evaluation of ... (BPA) exposure and toxicity concludes that BPA poses no health risk to

consumers of any age group (including unborn children, infants and adolescents) at current exposure levels. Exposure from the diet or from a combination of sources (diet, dust, cosmetics and thermal paper) is considerably under the safe level (the “tolerable daily intake” or TDI).

The EFSA’s experts have concluded that exposure should be set at 0.004 mg/kg body weight/d, from the existing 0.05 mg/kg body weight/d.¹⁰ The EFSA states that “uncertainties surrounding potential health effects of BPA on the mammary gland, reproductive, metabolic, neurobehavioral and immune systems have been quantified and factored in to the calculation of the TDI”.¹⁰

Studies conducted with low and high dose of BPA show effects at the low dose that are not apparent after exposure to the high dose, unlike the great majority of toxic chemicals.¹¹ Very low levels of BPA, through several pathways, can stimulate cellular response. Hence, this compound can be much more potent than previously thought, thus pointing to a revised TDI, which is what the EU is doing. Regarding dose–response mechanisms for EDs, different pathways lead to different non-linear dose–response curves (for the same outcome). Others believe that the descending and ascending arms of a U-shaped dose response can be due to different processes, and hence, the curve itself consists of 2 different processes, each of which is monotonically increasing and decreasing^{12,13}: a proliferative mechanism in one region and an inhibitory one in the other.¹⁴

Perspective on Public Regulatory Actions: Judicial Review

One of the clearest enunciations of practical risk-based policy analysis is that of the US EPA⁴:

conducts risk assessment to provide the best possible scientific characterization of risks . . . on a rigorous analysis of available information and knowledge, . . . a summary of the confidence or reliability of the information available to describe the risk, . . .

. . . can help guide risk managers to decisions that mitigate environmental risks at the lowest possible cost and which will stand up if challenged in the courts.

Understanding the overall importance of these citations requires a review of the US federal law, which we briefly explore. Standard setting by an agency of a government combines scientific and technical evidence with legal commands. The final test is the legal acceptability of the results from this combination, as exemplified by their review by the US Supreme Court (the Court). As a general rule, in the United States, judicial reversal of an agency’s decision—such as setting an environmental or occupational health standard—requires an egregious error by that agency to be struck down. For example, the US Supreme Court, in *Motor Vehicle Manufacturers’ Association v. State Farm Automobile Insurance Co*

(463 US 29, 1983), held that an agency rule can be reversed if an agency “relied on factors which Congress had not intended . . . entirely failed to consider an important aspect of the problem, offered explanations that run counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in point in view or the product of expertise.”

In *Industrial Union Department. AFL-CIO v. American Petroleum Institute* (448 US 607, 1980), the Court held that the Occupational Health and Safety Administration (OSHA) had the burden of showing that it is at least more likely than not (greater than 50% evidence) that long-term exposure to 10 ppm of benzene presents a significant risk of material health impairment. The Court required OSHA to develop better evidence of leukemia from occupational exposure to airborne benzene, concluding that “safe” is “not equivalent to risk-free.” Yet, the Court also held that the significance of risk is not a mathematical straitjacket, and OSHA’s findings of risk need not approach anything like scientific certainty. Moreover, the Court stated that “the reviewing court must take into account contradictory evidence in the record . . . , but the possibility of drawing two inconsistent conclusions from the evidence does not prevent an administrative agency’s findings from being supported by substantial evidence.”

However, the Court did not provide the risk acceptance criterion of “risk significance” (which is a number that US federal agencies provide, such as 1 in a million lifetime probability of cancer from environmental exposure).¹⁵ The Court held that when statutes contain terms such as “substantial release and significant amounts, an agency must establish a rational connection between the facts . . . and the choices.” The “rational connection” implies a weak (constitutional law) standard of review that ends up being deferential to agency rulemaking.^{15,16} In another case, *Baltimore Gas and Electric, Co v. Natural Resources Defense Council*, the Court unanimously reversed the DC Court of Appeals judgment that the US Nuclear Regulatory Commission (US NRC) had acted “arbitrarily and capriciously” (note 1). The Court held that a “most deferential” approach should be given an agency engaged in making legitimate predictions of risks that fell “within its area of expertise at the frontiers of knowledge, and when the resolution of . . . fundamental policy questions lies . . . with . . . the agency to which Congress has delegated authority.” The US NRC won. And, a US Court of Appeals held that the US NRC could adopt conservative assumptions “risking error on the side of over-protection rather than under-protection . . . when those assumptions have scientific credibility” (note 2). This result was a “satisfactory basis” for finding an “unreasonable risk” (statutorily defined in 5 USC §2058(f)(3)(A)). This court accepted the Commission’s bounding that a risk “somewhere between one in two thousand and one in fifty million, is appropriately left to the Commission’s discretion, so long as it was reasonable.” “Reasonable” is clarified in *Chevron v. Natural Resources Defense Fund* (note 3), where the Court stated that:

The Administrator's interpretation represents a reasonable accommodation of manifestly competing interests and is entitled to deference: the regulatory scheme is technical and complex, the agency considered that matter in detailed and reasoned fashion, and the decision involves reconciling conflicting policies . . . Judges are not expert in the field, and are not part of either political branch of the Government . . . When a challenge to an agency construction of a statutory provision, fairly conceptualized, really centers on the wisdom of the agency's policy, rather than whether it is a reasonable choice within a gap left open by Congress, the challenge must fail.

Of course, "agency wisdom" may have to be ascertained by having to go through several rounds of costly legal proceedings.¹⁷

European Union: Aspects of EDs Regulatory Law

In the EU, "all human health risk assessments of chemicals include hazard identification, dose-response assessment, exposure assessment and risk estimation/characterization."¹⁸ What is the flexibility inherent to using alternative dose-response models? Dose-response assessment consists of the mechanistic formulation and estimation of the parameters of the function or relationship between dose, or level of exposure, to a substance and the incidence of diseases. Hence, a choice is implicit, but how is it achieved? An indication of how the EU might deal with EDs, and thus BPA, is suggested by the EU Water Framework Directive (2000/60/EC), which looks at them as substances of equivalent concern to substances of relevance to the REACH Regulation.⁷ Endocrine disruptors are referred to in several EU regulations, for instance, (1) REACH (EC No 1097/2006), (2) the Plant Protection Products Regulation (EU No 1107/2009), (3) Biocidal Products Regulation (EU No 528/2012), and (4) the Cosmetics Regulation (EU No 1223/2009). Specifically, chemical agents should be identified as EDs through well-enunciated criteria for identification and approved testing methods/methods of detection and any other aspect of best practice and consistency with the state of the science (eg, Regulation EEC No 2377/90 [EU]). The REACH (EC No 1097/2006; EC 2006) illustrates key aspects of EU's regulatory law (through its secondary legislation). Because REACH is a Regulation, every Member State of the EU has to integrate it in their national legislation exactly as it is, unlike a Directive. The REACH is explicit⁷: "This Regulation should ensure a high level of protection of human health and the environment as well as the free movement of substances, on their own, in preparations and in articles, while enhancing competitiveness and innovation. This Regulation should also promote the development of alternative methods for the assessment of hazards of substances." The REACH also states that⁷: "it is based on the principle that it is for manufacturers, importers and downstream users to ensure that they manufacture, place on the market or use such substances that do not

adversely affect human health or the environment. Its provisions are underpinned by the precautionary principle."

This approach arises from a very different command than of US laws because the Precautionary Principle is a fundamental constitutional principle to justify environmental choices. This Principle may have the effect of trumping the choice of not acting because of scarce information on potential public exposure to agents that could cause serious or irreversible harm. For example, if the EU were to use the LNT because of lack of certainty about a mechanism of action associated with an ED, that choice would *prima facie* appear to be conservative and thus protective. Yet, this causal model is being increasingly demonstrated to be less conservative for EDs, due to fact that they cause increasing harm at very low doses, larger than that predicted by the LNT. The REACH⁷ also defines the role of the European Commission in assessing the evidence about EDs in Article 138(7): "By 1 June 2013 the Commission shall carry out a review to assess whether or not, taking into account latest developments in scientific knowledge, to extend the scope of Article 60(3) to substances identified under Article 57(f) as having endocrine disrupting properties. On the basis of that review the Commission may, if appropriate, present legislative proposals." For our work and suggestions contained in later section of this article, that basis includes⁷:

. . . hazard identification for the effect based on all available non-human information;

—the establishment of the quantitative dose (concentration)—response (effect) relationship.

. . .

When it is not possible to establish the quantitative dose (concentration)—response (effect) relationship, then this should be justified and a semiquantitative or qualitative analysis shall be included . . .

Unfortunately, the devil is in the details, and these do not appear to be specified. For example, under REACH⁷:

If one study is available, then a robust study summary should be prepared for that study. If there are several studies addressing the same effect, then, having taken into account possible variables (eg conduct, adequacy, relevance of test species, quality of results, etc.), normally the study or studies giving rise to the highest concern shall be used to establish the DNELs, and a robust study summary shall be prepared for that study or studies and included as part of the technical dossier. Robust summaries will be required for all key data used in the hazard assessment. If the study or studies giving rise to the highest concern are not used, then this shall be fully justified and included as part of the technical dossier, not only for the study being used but also for all studies demonstrating a higher concern than the study being used. It is important irrespective of whether hazards have been identified or not that the validity of the study be considered.

These qualitative clarifications can be expressed formally and thus limit possible ambiguities. This reasoning should be extended to empirical results that—for some EDs—confirm a region of probable supralinearity of the dose response,

provided that the scientific evidence on point produces sound theoretical understanding of the biological pathways and empirically validated experimental results applicable to humans (if that is the species of concern). Later sections of this article explain and exemplify how these aspects comply with the full extent of the Precautionary Principle, as the Treaty on the Functioning of the European Union (TFEU) commands.

Endocrine Disruptors: Admissibility of Scientific Expert Opinions in Judicial Proceedings

Endocrine disruptors in the US State Courts

Scientific experts deal with causation in both regulatory law and in litigation designed to provide monetary or other relief to individuals who believe that they were harmed by exposure to EDs. How to deal with uncertain evidence of cause and effect for diethylstilbestrol (DES) was assessed in a California case, *Sindell v. Abbott Laboratories* (26 Cal.3d 588, 163 Cal. Rptr. 132, 607 P.2d 924, cert. denied, 449 U.S. 912, 101S.Ct. 285, 66L.Ed.2d 140, 1980). There, the California Supreme Court developed a then novel legal theory: *market share* liability. The California Supreme Court held that a plaintiff could recover by showing that her injuries had been caused by DES and by joining as defendants “the manufacturers of a substantial share of the DES which her mother might have taken.” Each DES manufacturer becomes “liable for the proportion of the judgment represented by its share of [the] market unless it demonstrates that it could not have made the product which caused plaintiff’s injuries.” The New York Court of Appeals adopted *Sindell* in *Hymowitz v. Eli Lilly and Co.*, 73N.Y.2d 487, cert. denied, 110S.Ct. 350 (1989). Causation specific to EDs was addressed in *Beck v. Koppers Inc.*, (2006 US Dist. Lexis 2551 (N. D. Miss. 2006)), where it was claimed that dioxins cause an increased risk of breast cancer.

Epidemiological evidence is critical to a finding of cause and effect that is accepted as sound scientific evidence by the courts (eg, *Brock v. Merrel Dow Pharm.* 884 F.2d 166 [5th Cir. 1989]; *Chambers v. Exxon Corp.* 81 F. Supp. 2d 661 (M. D. La. 2000, affirmed 247 F.3d 240 [5th Cir. 2001]) seeking to resolve civil litigation in toxic torts (while animal studies may suffice for regulatory law choices). Although we cannot discuss how animal studies (as well as other studies) can establish causation, absent epidemiological evidence, if the exposure to which animals are subject is different from its human equivalent, US courts will generally not accept that evidence as being analogue of human exposure and response, as held by the US Supreme Court in *General Electric Co. v. Joiner*, 522 US 136 (1997). It found that expert testimony relying on that evidence was not reliable. Moreover, interspecies conversions are also deemed to be unreliable (*In re Human Tissue Product Liability Litigation*, 582 F. Supp. 2d 644 [D.N.J. 2008]), as are comparisons based on structure activity (*McClain v. Metabolife*, 401 F. 3d 1233 [11th Cir. 2005]). Clearly, scientific evidence is introduced by experts who justify their choices and conclusions in

administrative proceedings, such as Science Advisory Boards or in Congressional testimonies, as well as in trials. Hence, it is also important to understand how expert opinions are evaluated before being introduced in a trial, where that evidence undergoes further scrutiny and rebuttals through testimony by expert witnesses.

US Supreme Court: Admitting Expert Testimony in Judicial Proceedings

In 1993, the US Supreme Court, in *Daubert* (note 4), addressed which scientific results could be allowed in court, before trial (admissibility hearings precede trial and are under the exclusive jurisdiction of judges). Before *Daubert*, epidemiological evidence of the effect of Bendectin (note 5) based on chemical structure activity, in vitro, animal tests, and recalculations that had not been peer reviewed did not meet the legal test that controlled the admissibility of scientific evidence: the *Frye* test (note 6). It is a test of “general admissibility,” stating that:

(j)ust when a scientific principle or discovery crosses the line between the experimental and the demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while the courts will go a long way in admitting expert testimony deduced from well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.

The *Daubert* Court unanimously rejected *Frye*, holding that the general acceptance of a scientific finding is not required to be admissible into evidence, stating that:

... the “general acceptance” test is not a necessary precondition to the admissibility of scientific evidence under the Federal Rules of Evidence, but the Rules of Evidence—especially Rule 702—do assign to the trial judge the task of ensuring that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand. Pertinent evidence based on scientifically valid principles will satisfy those demands.

The *Daubert* Court added that “of course, it would be unreasonable to conclude that the subject of scientific testimony must be ‘known’ to a certainty; arguably there is no certainty in science.” The Court explained the role of peer review as a “relevant, though not dispositive, consideration in assessing the scientific validity of a particular technique or methodology.” The *Daubert* standard requires that admissible expert testimony, before trial, must be both relevant and reliable. Judges apply this standard to each study on which the expert relies, as well as the expert’s overall conclusions, where those studies have gone through vetting by both parties to the dispute, under oath, through depositions, at great cost to the parties to the dispute. It has been suggested that *Daubert* would have each study stand on its own merit through a corpuscular approach to expert testimony¹⁹ such that “if the plaintiff fails

to establish the relevance and scientific reliability of a sufficient number of individual studies, the trial judge will exclude the expert's testimony and (in the absence of other relevant and reliable expert testimony on causation) grant the defendant's motion for summary judgment before the jury ever enters the picture." Accordingly,²⁰ if the "corpuscular view,"¹⁹ then "we will begin to witness a divergence between judicial and regulatory approaches to evidence. In regulation, the strands of evidence are not assumed to stand by themselves. Rather, they are seen as pieces of a puzzle."¹⁶ Regarding our work, the implications are that the "corpuscular approach effectively prevents the expert in toxic tort cases from applying the 'weight-of-evidence' approach that regulatory agencies universally employ in addressing the risks that toxic substances pose to human beings."¹⁶

Neither scientific nor legal causation can hope for complete certainty. Statistical evidence alone, even when based on well-conducted studies, can be rebutted at trial using contradictory scientific theories and data. The strength of the legal causation combines scientific evidence and the legal standard of proof appropriate to the case—if and only if that evidence has passed the burden of admissibility. In the United States, when scientific uncertainty about causation was stated at trial as a "possibility," it was held to be insufficient (note 7) to demonstrate legal causation by the "preponderance of the evidence" test (the "more likely than not," >50%, test) (note 8). Of course, probabilistic reasoning alone will not be adequate for correctly integrating heterogeneous evidence and complex causal models of disease. As can be inferred from our discussion of US law, the acceptability of new science by the courts is a serious concern due to the inherent resistance of the judicial to accepting new methods (eg, the doctrine of *stare decisis*) and the fact that judges are the "gate-keepers of science."

Policy Science Controversies for BPA

We summarize aspects of the reasoning affecting policy debates concerning the effects of BPA¹⁴ in Table 1. To clarify, we use a citation¹⁴ to a US FDA statement that this agency believes that "there is a large body of evidence that indicates that FDA-regulated products containing BPA currently on the market are safe and that exposure levels to BPA from food contact materials, including for infants and children, are below those that may cause health effects." To reach this conclusion, the FDA had relied on 2 animal studies. In the first, its authors concluded that there was no effect from BPA exposure.²¹ However, because the Sprague Dawley rat is insensitive to estrogens and the study design lacked positive controls, whether BPA had no effect on the multiple generations of animals exposed or whether the rats were simply insensitive to estrogens is not decidable.²² The other study did not find low-dose effects of BPA.²³ However, this study had questionable positive controls.²⁴ There is disagreement at the consensus level, as a subcommittee of the US FDA in 2008 (Science Board Subcommittee on Bisphenol A) concluded that "coupling together the available qualitative and quantitative information

(including application of uncertainty factors) provides a sufficient scientific basis to conclude that the margins of safety defined by FDA as 'adequate' are, in fact, inadequate."²⁵ A summary discussion of the effects of EDCs states that it can be estimated "with confidence" that²⁰:

Effects will be due to "multiple hits" of environmental exposures and may occur only after a latent period of months to decades, requiring a lifespan research approach, including prospective human studies. There are multiple specific windows of enhanced susceptibility to metabolic disruptors across the lifespan, including paternal, in utero, early childhood, prepuberty, pregnancy (for the mother) menopause, and aging. Development, in utero and during the first few years of life, is the most sensitive window of susceptibility for metabolic disruption. The 2 sexes show differential susceptibility to metabolic disruption as well as different critical windows for, and different effects of, exposure. Understanding environmental effects on these diseases requires sensitive measures of personal exposures and sensitive end points to identify phenotypes. Effects of EDC exposure will vary depending on cooccurrence of other environmental stressors such as prescription drugs, sleep, hypercaloric diet, activity, stress, socioeconomic status, infections, microbiome, anxiety–depression, and so on, requiring a detailed analysis of potential interacting and confounding factors.

In judicial processes, the probative value of the evidence provided by both parties is assessed individually and then heuristically aggregated by decision makers (a jury or one or more judges). In agency rulemaking, that evidence is aggregated by agency experts so that a subset of that evidence is used numerically to establish a standard. We will turn our attention to this aspect next, focusing on regulatory law, rather than judicial aggregation, by outlining a quantitative approach to how to aggregate experts' opinions that allows a formal justification for any asserted consensus or, more precisely, aggregate opinions because it makes them transparent. We suggest that this aspect has been neglected but is needed. For example, in the EU, there are several initiatives (not yet completed) regarding the robustness, openness, and transparency of scientific assessments that directly affect the regulation of EDs. A critical aspect is coordination and consistency of the assessments conducted within a set of initiatives collectively falling under the PROMETHEUS project. The EFSA suggests a thorough emphasis on probabilistic reasoning (frequentistic and subjective).¹⁰ Additionally, EFSA discussions suggest a differentiation between concepts such as "weight of evidence"¹⁰ and the US EPA's concept of "evidence integration" under its Integrated Risk Information System (IRIS) program.²⁶ The EFSA is concerned with the balance of the overall scientific evidence.¹⁰ In particular, for EDs, there are different sources of uncertainty (eg, diet is more certain than nondietary exposures). It considers the strength of that evidence (eg, quality and confidence) and its dynamics (eg, how changes can affect a decision). Although these concepts aid the choice of the evidence to be used in setting standard by providing well-documented knowledge, they can be strengthened by additional quantitative analyses.

Table 1. Implications of BPA-Related Controversies Regarding Evidentiary Issues of Cause and Effect^a on Regulatory and Tort Law in the United States and the EU.

Controversy (for Extensive Detail, See Vandenberg et al) and This Special Issue	Basis for the Controversy	Probable Effect on Causation for Regulatory Law	Probable Effect on Causation for the Law of Toxic Torts
Incomplete knowledge of mechanisms at low dose rates.	Low-dose effects of BPA are larger than that occurring at higher doses.	Points to mechanisms of probable effect and is necessary but insufficient to show causation.	Evidence is wholly insufficient.
Human exposure and consequent risk: Large and well-controlled study of the possible health effects of BPA exposure on humans has been conducted, revealing positive correlations between urinary BPA concentrations and the prevalence of diabetes, heart disease, and liver toxicity (in Vandenberg et al [actual citations omitted]).	Duration, peaks, steady state, cumulative dose, and effects on response by those at risk.	Epidemiological studies are sufficient to show cause and effect but, as for other (smaller) epidemiological studies reported to have several limitations, including small sample sizes, limited details on subject selection criteria, and cross-sectional designs, they require additional evidence.	Epidemiological studies are necessary but currently insufficient to show cause and effect. The studies' designs prevent accurate assessments of the probable health risks of BPA.
Routes of exposure are incompletely characterized.	Mainly through ingestion but need data on all possible pathways of exposure, such as inhalation.	Not a limiting issue for regulation, depending on the statute or other regulatory instrument used.	Applicable to show confounding factors, shows incomplete understanding of exposure pathways, and study design limitations.
BPA's inactivation via conjugation and other biochemical processes.	This assumption is increasingly demonstrated to be weakened by new studies. Require in-depth assessments via PBPK and other models.	Key risk factors; inconsistent results between different species persist. Necessary but not sufficient condition for regulating BPA	Impact causal explanations of differences between mice and rat relative to human response, thus increasing epistemic uncertainty.
Patterns of exposure in animals should be consistent.	Relevance to humans: Taken together, these data indicate that animals exposed to BPA during gestation or the perinatal period show a wide variety of endocrine-related pathologies.	May attempt to mimic human exposure and indicates specific risks in subpopulations. Nontraditional organs may be at risk. Increases the understanding of necessary conditions for causation.	Shows that a variety of effects may be caused by exposure and that epidemiological studies may not be able to confirm effects in animals due to the cost of developing adequate sample sizes. Necessary and possibly sufficient condition for legal causation.
Carcinogenic effects at low doses in animals and humans.	Changes in cancer incidence in rats and mice.	Evidence of carcinogenicity has been ambiguous but is increasingly pointing to probable BPA-specific effects in rats and mice, as well as correlating with DES in human females. Can provide the necessary and sufficient conditions for regulatory action.	Applicable both as evidence and for causation. Requires mechanistic understanding of the cancer-specific disease process, statistical assessment, and account for PBPK and other biochemical events affecting BPA.

Abbreviations: BPA, bisphenol A; DES, diethylstilbestrol; EU, European Union; PBPK, physiologically based pharmacokinetic.

^a Developed from Vandenberg et al¹⁴; citations from her work.

Consensus-Based Aggregation of Evidence

A limited search of the several government Web sites does not disclose the details needed to understand how consensus is actually reached. Thus, it is difficult to discern voting criteria (other than majority-type criteria), how votes are counted and weighted (eg, Borda counts), and the details of aggregating expert opinions. Table 2 contains the typical elements of a

situation in which 3 experts are assessing the evidence before them. The experts are independent and should vote on all of the elements of a causal argument: antecedent, logical connection, and consequent. For simplicity, we use Boolean (true [T] or False[F]) states, do not include probabilities, use a logical connections (AND), and a single If . . . Then statement. The aggregation of the experts' judgments rule is by simple majority over

Table 2. Individual Opinions and Aggregate Results—the Discursive Dilemma.

Expert's Number	Proposition, p	If (p AND q) ↔ r	Therefore, Proposition r is	Comments
1	T	T	T	Logic
2	T	F	F	Logic
3	F	T	F	Logic
Aggregate judgment	T	T	F	Majority rule
NA	T: true; F: false	↔: if and only if (joint necessary and sufficient conditions) for r	NA	Kornhauser and Sager ²⁷ doctrinal paradox

all of the 3 components of the choice (using 3 rather than more experts is a minor simplification). The important issue regarding aggregation is the “discursive dilemma”: results are false, under the majority rule, whereas the same rule makes the process true and the premises true. All of the information summarized in Table 2 are essential for understanding consensus-based choices. The solution to this dilemma requires considering alternatives that are beyond our discussions.^{27,28} Moreover, when there are 3 or more voters, it is well known that only a dictatorial rule is satisfactory if voting (which is an aggregation of individual preferences) has to meet principles such as unanimity, anonymity, monotonicity, and systematicity.²⁸

Combining data through different mathematical operators include intersection (the min), various types of averages, and union (the max; note 9). Averages (note 10), from the arithmetic mean to the weighted generalized average, fall between the results obtained using intersections and unions:

1. Conjunctive (union; logical AND; t-norms; Min): given partial scores, the aggregate score will be high iff the partial scores are high.
2. Disjunctive (intersection; logical OR; t-conorms; Max): given partial scores, the aggregate score will be low iff partial scores are low. Limitations similar to conjunctive aggregation operators apply.
3. Compensatory (eg, averaging operators): low and high scores compensate each other. Averaging operators are monotonic and idempotent. Order statistics apply to cardinal information, and these operators are idempotent, continuous, monotonic, neutral, and compensative.
4. Noncompensatory (symmetric sums): scale reversal does not affect the results. These operators are continuous, nondecreasing, and commutative.

Computational aggregation has formal properties such as closure, continuity, idempotency, commutation, and others. Closure means that an uncertain number results from the aggregation of uncertain numbers. Idempotency means that an operation on uncertain quantities yields the same uncertainty. Finally, the commutative property implies order independence, and continuity implies that a small change in a quantity does not affect the final result.

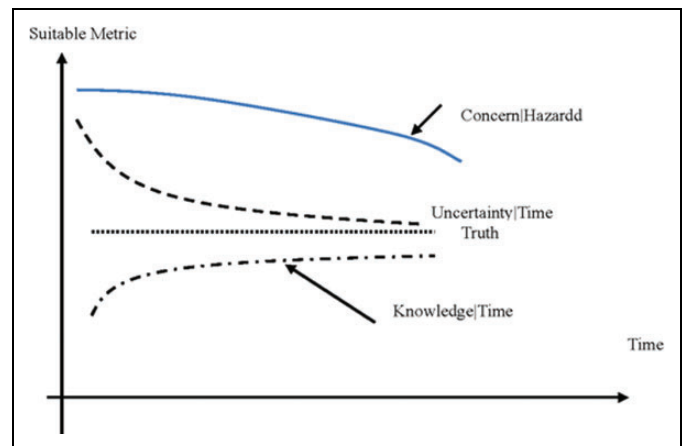


Figure 1. Plot of uncertainty and knowledge over the time horizon of the concern.

Discussion

Ideally, Aumann's²⁹ proofs that “like-minded” decision makers “cannot agree to disagree,” provided they have the same information knowledge (encoded by priors and likelihoods), should be the practical terminal target of the trajectories shown in Figure 1. However, several researchers and practitioners have shown that Aumann's convergence does not occur and that rational behaviors (by 2 or more stakeholders with independent of beliefs) do not converge because of different strategies taken by those actors to achieve their goals. We assume that a public decision has a common basis of scientific facts and knowledge and that these are available to all parties. Figure 1 is a hypothetical description of asymmetries regarding the eventual convergence of the trajectories of concern, knowledge, and an aggregate measure of epistemic uncertainty about that knowledge, conditioned on the nature of the hazard and time. The terminal state is labeled “truth.” The time horizon is the interval of time characterized by an initial policy concern about a hazard and its eventual regulation through any legally based process. We omit, for brevity, the fact that standard setting processes involve changes subsequent to the enactment of the standard, some of which are induced by litigation or by legislatively mandated revision that account for new scientific or technological developments. It follows that each trajectory is not deterministic.

Finally an aspect of these discussion has to do with (1) agreement/disagreement and (2) asymmetric information may be available to stakeholders, contrary to Aumann's ideas. Flexibility is the ability to consider alternative theories and decide on one to be optimal even though others disagree. This implies an explicit, formal analysis of the amount of information and knowledge. Often, depending on who the stakeholders are and their level of access to research funds, the information/knowledge/processing triplet can be asymmetric: some have more than others, and thus, uncertainty is higher. The higher the uncertainty, the greater the value of flexibility: "it is a question of the optimal course of action, with insufficient time or objective data for beliefs to converge."³⁰ Flexibility, however, has nothing to do with asymmetry of information because, even when there is symmetry of information and knowledge, beliefs may not converge. Therefore, commonality of interests does not guarantee that beliefs converge.

Informing the Public

Both the EU and the United States have legislation that commands public involvement and information in regulatory choices. The Office of Management and Budget (OMB, an agency of the White House) established *Agency Information Quality Guidelines* that control the collection, processing, and dissemination of information that has to do with risk assessment.³¹ The OMB refers to the Safe Drinking Water Act (SDWA) as the gold standard for justifying public decision making based on the minimization of risk (note 11). There, "influential information" is defined to be that scientific, financial, or statistical information, which will have or does have a clear and substantial impact on important public policies or important private sector decisions. Moreover, the OMB Guidelines also allow an individual to bring civil law suits to challenge the value of the influential information, including risk assessments. The SDWA (administered by the EPA) applies to the OMB's influential information, as follows³²:

1. Use of science in decision making. In carrying out this section, and to the degree that an Agency Action is based on science, the Administrator shall use:
 - i. the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices and
 - ii. data collected by accepted methods or best available methods (if the reliability of the method and nature of the decision justify use of the data).
2. Public information. . . the Administrator shall ensure that the presentation of the information on public health effects is comprehensive, informative, and understandable. The Administrator shall, in a document made available to the public in support of a regulation promulgated under this section, specify, to the extent practicable:
 - i. each population addressed by any estimate of public health effects,

- ii. the expected risk or central estimate of risk for the specific populations,
- iii. each appropriate upper-bound or lower-bound estimate of risk,
- iv. each significant uncertainty identified in the process of the assessment of public health effects and studies that would assist in resolving the uncertainty, and
- v. peer-reviewed studies known to the Administrator to support, are directly relevant to, or fail to support any estimate of public health effects and the methodology used to reconcile inconsistencies in the scientific data.

Regulation Should Follow Decision Theoretic Principles

Causal arguments directed to justify EDs policy combine qualitative and quantitative descriptions because:

1. causal explanations must be given to the stakeholders,
2. laws and regulations are based on scientific and legal causation,
3. scientific reasoning and explanations are or attempt to be causal,
4. stakeholders need to know how and why risky events can generate adverse consequences,
5. risk factors are used in apportioning liability to the sources of the hazard or hazards, and
6. risk reduction and minimization actions require causal knowledge to be fair and equitable.

The EFSA apparently used methods such that uncertainties are analyzed one by one and combined with expert judgments to yield a final TDI value. We suggest that the criterion selected to justify the results of any final (in the regulatory sense) reevaluation of BPA might be based on a specific decision theoretic criterion, which would support informed judgments that yield the final regulatory number sought for BPA. The reason is that different criteria yield different solutions, but each has formal properties that may be preferable on a case-by-case basis.

We disassociate "choices" (the items under analysis from which an optimal or preferred one is demonstrably superior to other choices being analyzed) from "decisions." Decisions are public actions undertaken by public decision makers who are legally bound to make them and who may be prosecuted for failure to act. Those decisions may or may not conform with the results of the analysis of choices and their optimality or preferential rankings and can be scrutinized by a court. Some elementary analytical criteria for justifying a choice, from a set of possible choices, properly to inform decision makers³³⁻³⁵ are briefly discussed next, as means to provide an initial answer to policy science issues developed throughout this article.

Several steps for modeling causation include³⁶:

1. identify a consistent, nonrandom association between exposure and response,

2. identify, explain, and include, in the causal system, the physics of the relation,
3. eliminate, or at least explicitly account for, the effect of confounding factors,
4. eliminate, or at least explicitly account for, sampling, information, and modeling biases,
5. test and confirm temporal precedence and conditional independence,
6. develop, explain, and confirm the effect of policy interventions through changes in the value of the variables affected (for a baseline causal model) on response.

For simplicity, uncertainty is handled through probabilities and their calculus. It suggests the following protocol:

1. define and identify the boundaries of the risky choice or problem,
2. define working hypotheses and conjectures and give a qualitative description of the processes leading to all relevant outcomes,
3. determine the state of knowledge about the decision process being studied,
4. assess the need for additional information and the stopping rules to limit additional costly information, based on value of information and value of flexibility calculations,
5. use experimental data to provide likelihoods (conditional probabilities),
6. choose the criterion that is consistent with risk reduction (eg, minimax cost, minimax regrets, and maximize expected monetary value or utility) and explain the rationale for the choice of criterion for selecting an alternative,³⁷
7. identify the optimal choice,
8. communicate it to the decision maker,
9. resolve outstanding issues, and
10. conclusions consisting of joint or marginal probability distributions and their (appropriate) moments, as needed by the stakeholders.

The combination of these 2 sets leads to an analysis of choices that can have different implications for informing decision makers. An analyst may use criteria such as the maximization or minimization of an objective function (eg, min [expected number of malformations in a population at risk]). Often used alternative criteria are as follows.

Max (expected utility) or max (expected net monetary values) criterion. The theory is expected utility (EUT) which is rational. That is, if its axioms (note 12) are accepted, then the best choice that the single decision maker can make is the one that has the highest expected utility (or positive expected monetary value). Choices are described as probabilistic gambles (in which $[0 \leq \text{probability} \leq 1]$). A theorem demonstrates that the optimal choice based on EUT is guaranteed by the decision maker consistency with those axioms.³⁸ If

monetary values are used instead of utilities, the criterion is the maximization of the net expected benefits from each action, over all possible actions. Because empirical evidence indicates that individuals often do not seek to maximize their expected utility, the criterion of maximization of the expected utility is descriptive but weakly predictive. The empirical findings that demonstrate violations of the axioms and assumptions that characterize EUT theory have led to several new theoretical variants, such as prospect theory. Those weaken one or more of the axioms (eg, the independence axiom in weighted utility theory and in rank-dependent utility theory) and are more consistent with human behavior.

Maximin, maximax, and minimax criteria. The maximin criterion consists of selecting the choice that is the maximum of the minima payoff of the choices available. The maximax criterion selects the maximum of the maxima. An alternative criterion is to minimize the maximum loss, this is the minimax criterion. Although these criteria are deterministic, they may be used in situations where there is considerable uncertainty and the decision maker feels uncomfortable in assigning probability numbers to outcomes. Probabilities are not used in these calculations, and the choice is said to be made under uncertainty.

The pessimism–optimism criterion. This criterion uses a coefficient (bounded between 0 and 1, included but it is not a probability measure) to capture attitudes that fall between pessimism and optimism. It includes the maximin and maximax criteria and is neither deterministic nor probabilistic.³⁹

In the next section, we suggest how a Bayesian approach meets these desiderata, particularly when the model consists of a Bayesian network (BN) that allows feedbacks. We also note that Bayesian methods combine expert opinions (as prior probabilities or distributions) with empirical results (as likelihoods). Decisions taken by duly empowered decision makers are informed by probabilistic analysis that develops the optimal or preferred choice, selected on criteria such as the EUT, which is then provided to the decision maker for his or her consideration but may not be used because other factors (eg, geopolitical) may affect his or her final decision.

Probabilistic Causation: Outline of a Bayesian Solution

Probabilities or probability distributions represent uncertain knowledge and beliefs. Specifically, prior probability distributions (such as density functions for continuous data and probability mass functions for discrete distributions) represent prior knowledge, information, judgments, and beliefs for the independent variables.^{40,41} Likelihood functions represent empirical evidence. All empirical data and modeling information are summarized by likelihood functions. Given a probability model $pr(y; x, b)$, the corresponding likelihood function for the

parameter \mathbf{b} (possibly a vector, hence the bold letter) is simply $\text{pr}(y; \mathbf{x}, \mathbf{b})$. It is a function of \mathbf{b} , instead of a function of X and Y . The pairs of observed x and corresponding y values are data from the X and Y random variables. Thus, representing data as vectors $(\mathbf{x}, \mathbf{y}) = (x_i, y_i), i = 1, 2, \dots, N$, the likelihood for a value \mathbf{b} of the parameter vector is the conditional probability of observing the data \mathbf{y} , conditioned on the data \mathbf{x} . The quantity $L(\mathbf{b}; \mathbf{x}, \mathbf{y}, \mathbf{pr})$ denotes the likelihood function for \mathbf{b} based on observed data \mathbf{x} and \mathbf{y} of n individual (x, y) pairs, assuming that the probability model is correct.

Updating prior beliefs with new evidence uses likelihoods. Given the prior beliefs $F(\mathbf{b})$ about the uncertain quantities \mathbf{b} and the evidence, measurements, model, and data summarized in L , posterior beliefs (meaning the combination of the prior and the likelihood) about \mathbf{b} are computed using Bayes' rule. An important step is the calculation of:

1. the posterior beliefs $F|L$ from prior beliefs that are encoded in F and
2. the assumptions and evidence encoded in L , which depend on the probability model, pr , short for $\text{pr}(y; \mathbf{x}, \mathbf{b})$, and on the observed data.

In practice, uncertainty about the correct model, pr_i , out of several alternative models, is often the largest affecting the analysis. To account for it, let $\{\text{pr}_1, \dots, \text{pr}_n\}$ denote the set of alternative models that are known to be (or are considered) mutually exclusive and collectively to exhaust all possible probability models. Let L_1, \dots, L_n denote the likelihood functions for alternative models, and let w_1, \dots, w_n be the corresponding judgmental probabilities, also called weights of evidence, that each model is correct. If the models are mutually exclusive and collectively exhaustive, these weights must sum to 1. The posterior probability distribution that is obtained from the prior F , data (\mathbf{x}, \mathbf{y}) , model weights is the weighted sum: $w_1(F|L_1) + \dots + w_n(F|L_n)$. The advantages of the probabilistic reasoning as just set out include the following:

1. Showing the distribution of values for the true parameter \mathbf{b} (or vector \mathbf{b}) based on available information and scientific judgments. The probability distributions show how much statistical variability can be reduced by further research (which would narrow the distributions).
2. Distinguishing the contributions of different sources of evidence and identifying specific areas where additional research is most likely to make a significant difference in reducing final uncertainty. This deals with the uncertainty of model building. Despite the advantages, probabilistic methods have limits to their ability to represent uncertainty. When more than one decision maker is involved, the unified probabilistic presentation of different types of uncertainties (eg, observer-independent stochastic characteristics with theoretical assumptions, subjective judgments, and speculations about unknown parameter and variable values) is as much a liability as it is an asset.⁴²

There are disadvantages. Two individuals with identical but incomplete objective information might express their beliefs with different prior probability distributions. Moreover, probability models cannot adequately express ambiguities about probabilities. For example, an estimated probability of .50 that a coin will come up heads on the next toss based on lack of information may not be distinguished from an estimated .50 that is based on 10 000 observations. The Bayesian view is that an analyst normatively should use either his or her own knowledge and beliefs to generate a probability model when objective knowledge is either incomplete or even inadequate. The opposing view is that the analyst has no justification, and should not be expected or required, to provide numbers in the absence of substantial and relevant knowledge. When the correct model is unknown and multiple models and weights are used, or when multiple sources of evidence giving partially conflicting posterior probabilities are combined, the resulting aggregate posterior probability distribution is ambiguous. An infinite variety of alternative models and weights are mapped by the formula $[w_1(F|L_1) + \dots + w_n(F|L_n)]$ onto the same aggregate posterior probability distribution for the risk estimate. A partial solution to these problems is to present posterior distributions and corresponding weights for each model separately. Some of the knowledge used to draw practical conclusions about risks can be abstract or nonquantitative. Other aspects of qualitative knowledge can be used to constrain probabilistic calculations but cannot be represented by probabilities. A set of mutually exclusive, collectively exhaustive hypotheses about the correct risk model is seldom known, making the use of probabilistic weights of evidence for different possible models inexact. Probability models inherently make the "closed-world assumption" that all the possible outcomes of a random experiment are known and can be described (and, in Bayesian analysis, assigned prior probabilities). This assumption can often be unrealistic because the true mechanisms may later turn out to be something entirely unforeseen. Conditioning on alternative assumptions about mechanisms only gives an illusion of completeness when the true mechanism is not among those considered.

Overall, the usefulness of Bayesian analysis is that it provides a formal method for updating scientific knowledge by requiring the researcher to think in probabilistically about events and the causal structure linking them. The analyst must disclose the reasons for his or her choice of prior distribution, shown them through the form of the prior distribution, and clarify them by giving the reasons for adopting that distribution over its alternatives. This requires the risk assessor having to specify how past information can be folded into a distribution function and assess the impact on the likelihood, which links the sample design to the structure of the model most likely to be determined by that sample. Finally, the analysis is transparent and can be discussed for lack of completeness, arbitrariness of assumptions, and adequacy of experimental results and links directly to value of information and flexibility.^{43,44} The analysis is independently replicable, using alternative models, and the difference between the results can be studied under a common and axiomatically correct methodology.

The last step provides further integration through probabilistic networks at various biological levels and their combination across those levels. It begins by developing the network that plausibly represents the disease of concern, choose probabilities to characterize uncertainty, conditioning them on current information and knowledge, and infer accordingly. In short, a Bayesian directed acyclic graph (DAG), and a modification to account for feedbacks, describes the key subprocesses causing a disease. Bayesian networks rely on DAGs to represent relationships between random variables. Acyclic implies no feedbacks,⁴⁰ a condition that can be relaxed (note 13). The structure of the graph is based on known or hypothesized relations (eg, a node, an arc, and another node). Bayesian networks describe probabilistic conditioning and thus the dependence–independence structure among the variables of the network. There must be “meaningful directionality” between the variables in the network. Testing “whether a proposed set of causal relationships is consistent with the available temporal probabilistic information” can be obtained with BNs.⁴⁰ An advantage of BNs arises from the causal meaning of the directed arcs. If S^1 represents a BN, S^2 another BN, and if $\text{pr}(S^1|D)$ is larger than $\text{pr}(S^2|D)$, then the change in probability makes the causal link stronger. A further advantage of a BN is the concise and comprehensive representation of the relationship between the risk factors. However, the “close-world” assumption, assuming knowledge of the distribution of each random variable, can be too demanding. This assumption involves knowing all relevant factors and separating them into the causally relevant, background factors and irrelevant factors. An aspect of this analysis that is useful in deciding between competing and alternative causal structures uses the principles of “stability” (meaning that there is an isomorphism between 2 structures) and “minimality” (meaning that the least complex of 2 structures, T_1 and T_2 , given the same data, is preferred). Minimality relates to the mathematical form and number of variables of the causal network. Stability relates to the probability of events in T_1 that make events in T_2 more probable; it refers to lack of extraneous probabilistic conditional independences in a BN.⁴⁰ This principle states that it is most improbable that the 2 competing structures overlap: a causal structure is minimally relative to a larger set of potential structure if no element of the minimal structure is preferred over those of any other structure in the class of structures considered: minimality and stability assess the uniqueness of the network. This is one of 2 important aspects of obtaining a solution, the other being its existence.

Conclusions

We believe that society can be better served when the full protocol used to decide on what is asserted through consensus is independently replicable and that its key axioms are tested. For instance, to the extent that the focus of this EFSA’s efforts is both for internal use as well as a means to inform scientific advisory group, the discussions we provide can be useful toward formulating rigorous science-policy choices. We have linked legal and scientific aspects of causation using the United

States and the EU as paradigms to help public decision making to limit chemical exposures, in particular EDs (BPA being prototypical). We conclude that US federal administrative law has developed to the point that it has the necessary and sufficient qualitative aspects to deal with human health risks associated with exposure to EDs. This is not yet the case for the EU as its efforts are still under development. We suggest that the EU should take note of the ways in which the United States has dealt with the admissibility of scientific evidence—in terms of relevance and reliability—and how it has balanced that evidence, a major concern of EFSA and US agencies. To advance this discourse, we conclude that a plausible way to deal with uncertain causation is provided by Bayesian methods, which we briefly outline. The reason for focusing on those methods is that they are understood by the courts, are used by agencies of the government and are prevalent in causal reasoning about uncertain exposure and response, conditioned on one or more biological or other mechanisms. Hence, these methods are established even though they may not be the best under all uncertain situations.

We find that consensus guides public policy choices—obtained though their formal analysis—to inform decision makers and stakeholders. The protocols for assessing evidence for or against a particular effect of exposure canonically consist of premises, rules, and conclusions. We offer initial guidance as how to formally frame the discourse that leads to an assertion of effect from exposure. This is a critical aspect in supporting a policy because the aggregation of experts’ opinion results in several paradoxes. One of those discussed, simple majority, can be robust under very specific circumstances. However, most alternative voting criteria involve a number of complexities that, paradoxically as well, generally lead to a dictatorial solution.

As both case and regulatory law indicate, the preferable form of evidence is human exposures and responses through a well-specified probabilistic model that can be used for prediction. Hence, regarding both factual and theoretical evidence, we conclude that epidemiology—although it has several problems—would be the best evidence for supporting regulatory law choices. However, we also conclude that linking EDs to the outcomes observed in mammalian species, or other testing protocols, is necessary to buttress epidemiological results, both at the ultimate end point level and at intermediate end points leading to it. The latter is essential for formulating the mechanistic aspects of epidemiological causal models. We conclude that a properly designed multilevel BN can approximate what is needed to produce replicable causal analyses and thus correctly inform standard setting. The final choice of standard value is a political choice by duly empowered officials. It is therefore outside the purview of the analysts. This is, of course, has a parallel in what happens in judicial decisions: the assigned of culpability is not for the lawyers to make. It is the exclusive province of judges.

Notes

1. 426 US 87 (1983). This is a more stringent legal standard than the tort law *more likely than not*, but less stringent than the *beyond a reasonable doubt* of criminal law.

2. *Kerr-McGee Nuclear Corp. v. U. S. Nuclear Regulatory Commission*, 17 ERC 1537 (1982).
3. 467 U. S. 837 (1984).
4. *Daubert et al., v. Merrell Dow Pharmaceuticals, Inc., L. Week* 61: 4805. In *Christophersen v. Allied-Signal*, No. 89-1995 (5th Cir. 1991), cert. den'd, 112 S. Ct. 1280 (1992), the scientific testimony about the causal association between exposure to nickel and cadmium and colorectal cancer was excluded because it was not mainstream science. Yet, some state courts have allowed into evidence scientific explanations that would fail *Frye*-type tests if those explanations are *sound, adequately founded ... and ... reasonably relied upon by experts*, (*Rubanick v. Witco Chemical*, 593 A. 2d 733 (NJ 1991)).
5. In *Brock v. Merrell Dow Pharmaceuticals*, 874 F.2d 307 (5th Cir.), 884 F.2d 166 (5th Cir.), reh'g en banc den'd, 884 F. 2d 167 (5th Cir. 1989).
6. *Frye v. U. S.*, 293 F. 1014 (D. C. 1923).
7. In US practice *sufficient evidence* is evidence that can proceed to be heard by the jury or by a judge acting as the fact-finder.
8. *Green v. American Tobacco Co.*, 391 F.2d 97 (5th Cir. 1968), 409 F. 2d 1166 (5th Cir. 1969). *Bowman v. Twin Falls*, 581 P.2d 770 (Id. 1978) held that to *require certainty when causation itself is defined in terms of statistical probability is to ask for too much*.
9. For instance, the arithmetic mean has meaning on intervals but not for ordinal data: for an aggregating formula to be *meaningful*, order statistics should be used instead. On the other hand, the median is order-preserving.
10. We do not discuss expectations for brevity.
11. 42 USC Sect. 300g-1(b)(3)(A), (B).
12. The axioms, based on binary preferences, are complete ordering, continuity, substitution and probability ordering between lotteries, and indifference between equally preferred lotteries (probabilistic gambles). Uncertainty is measured by probabilities that attach to outcomes, thus generating a gamble. We remind the reader that no known decision-theoretic criterion can meet all of the associated axioms. We also note that monetary values suffer from diminishing marginal utility, but gloss over this aspect for brevity.
13. An empirical deductive causal network M is defined by a set of dependent and independent variables forming the set V , each of the variables in V representing a distinct node. This representation includes intermediate variables that are explained by their predecessors and help to explain successors, unlike regression models. The arrows connect nodes with nodes such that each node is a function of the variables pointing to it. A BN can be succinctly stated as $M = (U, V, F)$, where U represents background variables, V the variables that are internal to the models and thus explicitly accounted for, and F represents functions. The probabilistic form of M , $(M, pr[U])$, allows updating as new evidence becomes available over one or more variables.⁴⁰

Acknowledgments

PFR thanks Louis Anthony (Tony) Cox jr and Huaxia Sheng for discussions and joint work that has led to some of the ideas contained in this paper, as well as 2 anonymous referees for their comments. A special debt of gratitude goes to Elena Fabbri for all her efforts in steering this special issue and review this paper: her contribution goes beyond collegial duty.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Ricci PF, Straja SR, Cox LA Jr. Changing the risk paradigms can be good for our health: J-shaped, linear and threshold dose-response models. *Dose Response*. 2012;10:177-189. Web site. http://scholarworks.umass.edu/dose_response/vol10/iss2/5 (accessed August 15, 2015).
2. European Commission, EU Comm2000. 2011. EUR-Lex, 12-0404-2011.
3. European Union Consolidated Treaties, Treaty on the Functioning of the European Union (TFEU), Lisbon Treaty, Article 191, paragraph 2.
4. United States Environmental Protection Agency. *Risk Assessment Guidance for Superfund Volume 1: Human Health Evaluation Manual*, EPA/540/R/99/005. Washington DC: United States Environmental Protection Agency; 2004.
5. Morrissey RE, George JD, Price CJ, Tyl RW, Marr MC, Kimmel CA. The developmental toxicity of bisphenol-A in rats and mice. *Fundam Appl Toxicol*. 1987;8(4):571-582.
6. Vogel D. The politics of risk regulation in Europe and the United States. *Yearbook of European Environ. Law*. 2002;3:1-42.
7. European Commission. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). 2006. Web site. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02006R1907-20150601&from=EN> (accessed August 15, 2007).
8. European Parliament. Library Briefing, Health Threats from Endocrine Disruptors: A scientific and regulatory challenge (Brussels, 12-06-2012). Presidency. 2012. Web site. http://www.mim.dk/en/EU2012/EU201204202012/Presidencyconclusions_htm (accessed August 15, 2015).
9. European Food Safety Authority. EFSA Updates Advice on Bisphenol A. 2010. Web site. <http://www.efsa.europa.eu/en/press/news/cef100930.htm> (accessed August 15, 2015).
10. European Food Safety Authority. Outcome of the Targeted Consultation of the EFSA Journal Editorial on Increasing Openness, Robustness and Transparency of Scientific Assessments, Technical Report. 2015.
11. Richter C, Birnbaum LS, Farabollini F, et al. In vivo effects of bisphenol A in laboratory rodent studies. *Reprod Toxicol*. 2007; 24(2):199-224.
12. Medlock KL, Lyttle CR, Kalepouris N, Newman ED, Sheehan DM. Estradiol down-regulation of the rat uterine estrogen receptor. *Proc Soc Exp Biol Med*. 1991;196(3):230-300.
13. Soto AM, Lin T-M, Sakabe K, Olea N, Damassa DA, Sonnenschein C. Variants of the human prostate LNCaP cell line as a tool to study discrete components of the androgen-mediated proliferative response. *Oncol Res*. 1995;49(10-11):3474-3481.

14. Vandenberg LN, Maffini MV, Sonnenschein C, Rubin BS, Soto AM. Bisphenol-A and the great divide: a review of controversies in the field of endocrine disruptors. *Endo Rev.* 2009;30(1):75-95.
15. Ricci PF, Gray NJ. Toxic Torts and Causation: Towards and Equitable Solution in Australian Law. Part II: Means-Ends Analysis, University of New South Wales Law Journal 44: 155-175.
16. Krinsky S. The weight of scientific evidence in policy and law. *Am J Public Health Suppl.* 2005;95(suppl 1):S129-S136.
17. Gold SC. When Certainty Dissolves into Probability: A Legal Vision of Toxic Causation for the Post-Genomic Era. 70 Washington & Lee Law Rev. 237. 2010. Web site. <http://scholarlycommons.law.wlu.edu/wlulr/vol70/iss1/5/>; 2013 (accessed August 15, 2015).
18. European Environment Agency. The European Environment—state and outlook. Chapter 5: Health risk assessment. 2011. Web site. <http://www.eea.europa.eu/publications/GH-07-97-595-EN-C2/chapter5h.html> (accessed November 18, 2015).
19. McGarity TO. Proposal for Linking Culpability and Causation to Ensure Corporate Accountability for Toxic Risks. In: William and Mary Environ. Law and Policy Rev. Web site. <http://scholarship.law.wm.edu/wmelpr/vol26/iss1/3>.
20. Heindel JJ, vom Saal FS, Blumberg B, et al. Parma Consensus Statement on Metabolic Disruptors. *Environ Health.* 2015; 14:54-61. Web site. <http://www.ehjournal.net/content/14/1/54/>. DOI 10.1186/s12940-015-0042-7.
21. Tyl RW, Myers CB, Marr MC, et al. Three-generation reproductive toxicity study of dietary Bisphenol A in CD Sprague-Dawley rats. *Toxicol Sci.* 2002;68(1):121-146.
22. Tyl RW, Myers CB, Marr MC, et al. Two-generation Reproductive Toxicity Study of Dietary Bisphenol A in CD-1 (Swiss) mice. *Toxicol Sci.* 2008;104(2):362-384.
23. vom Saal FS, Welshons WV. Large effects from small exposures. II. The importance of positive controls in low-dose research on bisphenol A. *Environ Res.* 2006;100(1):50-76.
24. Myers JP, vom Saal FS, Akingbemi BT, et al. Why public health agencies cannot depend upon 'good laboratory practices' as a criterion for selecting data: the case of bisphenol-A. *Environ Health Perspect.* 2009;117(3):309-315. doi:10.1289/ehp.0800173.
25. United States Food and Drug Administration. FDA Science Board Subcommittee on Bisphenol A. Scientific Peer-Review of the Draft Assessment of Bisphenol A for Use in Food Contact Applications. 2008. Web site. <http://www.fda.gov/ohrms/dockets/ac/08/briefing/2008-4386b1-05.pdf>.
26. US National Academies of Science. *Report Highlights, Review of EPA's Integrated Risk Information System IIRIS Process.* Washington, DC: NAS Press; 2014.
27. Kornhauser LA, Sager LG. The many as one: integrity and group choice in paradoxical cases. *Philosophy Public Aff.* 2004;32(3): 249-276.
28. List C, Pettit P. On the many as one: a reply to Kornhauser and Sager. *Philosophy Public Aff.* 2005;33(4):377-390.
29. Aumann RJ. Agreeing to disagree. *Ann Stat.* 1976;4(6): 1236-1239.
30. Boot A, Thakor AV. The many faces of disclosure. *Rev Financial Stud.* 2001;14(4):1021-1058.
31. Office of Management and Budget (the White House). Agency Information Guidelines, P.L. 106-554. 2002. Web site. <http://www.access-board.gov/the-board/policies/information-quality> (accessed November 18, 2015).
32. Safe Drinking Water Act, 42 U.S.C. Sect. 300f et seq. 1974, Amended 1974, 2002.
33. Institute of Medicine. *Environmental Decisions in the Face of Uncertainty.* Washington DC: National Academy for Science Press; 2013.
34. Mahnovski S. *Robust Decisions and Deep Uncertainty: An Application of Real Options to Public and Private Investment in Hydrogen and Fuel Cell Technologies.* Santa Monica, CA: RAND, Pardee Graduate School; 2007.
35. Lempert R, Collins M. Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches. *Risk Analysis.* 2007;27(4):1009-1026.
36. Cox LA Jr. *Risk Analysis Foundations, Models and Methods.* New York: Springer Science; 2002.
37. Gilboa I. A combination of expected utility and maxmin decision criteria. *J Math Psychol.* 1988;32:405-420.
38. Clemen RT, Reilly T. *Making Hard Decisions with Decision Tools.* Mason, OH: South-Western; 2001.
39. Jaffray JY, Jeleva M. Information Processing Under Imprecise Risk with the Hurwicz Criterion. In: 5th International Symposium on Imprecise Probability: Theories and Applications. 2007:233-242.
40. Pearl J. *Causality: Models, Reasoning and Inference.* Cambridge, MA: Cambridge University Press; 2000.
41. Howson C, Urbach P. *Scientific Reasoning: The Bayesian Approach.* Chicago, IL: Open Court; 2006.
42. Jeffrey R. *Subjective Probability: The Real Thing.* Cambridge, MA: Cambridge University Press; 2004.
43. Hallegatte S, Shah A, Lempert R, Brown C, Gill S. *Investment Decision Making Under Deep Uncertainty, Policy Research Working Paper 6193.* Washington, DC: The World Bank; 2012.
44. Chow BG. *Portfolio Optimization by Means of Multiple Tandem Certainty-Uncertainty Searches A Technical Description RR-270-A/OSD.* Santa Monica, CA: RAND Corporation; 2013.