

The Report on Carcinogens

What Went Wrong and What
Can Be Done to Fix It

By Richard B. Belzer

JANUARY 2012



The Report on Carcinogens

What Went Wrong and What Can Be Done to Fix It

By Richard B. Belzer

Executive Summary

There is nothing wrong in principle with publishing periodic reports identifying substances that pose carcinogenic risks to humans. Cancer remains a serious disease even though advances in diagnosis and treatment have rendered most types much less often fatal than they were when President Richard Nixon declared the “war on cancer” in 1971. But it would be a mistake to continue basing these reports on scientific knowledge and primitive technology dating from the 1960s.

The National Toxicology Program’s (NTP) *Report on Carcinogens* (RoC) is one such periodic report. The NTP has interpreted its statutory charge in a way that never was consistent with the law authorizing its preparation, resulting in Reports that never could live up to Congress’ original intent. Though the law requires the NTP to estimate the number of Americans actually exposed, and to list substances only if a significant number of Americans are exposed to them, the NTP functionally ignores exposure. The law also requires the NTP to estimate the reduction in cancer incidence resulting from regulatory standards, but it does not perform that required task, either.

Problems with the RoC begin with the NTP’s listing criteria. A careful review of the text shows that they are mere tautologies. For example, a substance is deemed to be a known carcinogen if the NTP decides that the evidence from human studies is *sufficient*. The minimum threshold for designation as a known carcinogen is unknown to the public because the NTP never says what is required for evidence to be *sufficient*. Thus, a substance is a known human carcinogen if the NTP says the evidence is *sufficient*. Conversely, the evidence is sufficient if the NTP says the substance is a known carcinogen. Similarly circular logic pervades the definition of a *reasonably anticipated* human carcinogen.

Worse, the NTP appears to be institutionally incapable of incorporating decades of advancements in scientific knowledge into its listing decisions, and there is no transparent way to scientifically rebut or reverse a listing decision once it has been made.

Both of the statutory categories for assignment (*known* and *reasonably anticipated*) imply that a causal relationship has been demonstrated with near certainty in the first case, and with an unspecified but lesser confidence level in the second. But the NTP’s listing criteria do not require any demonstration of causality. Rather, the NTP assumes that causality is demonstrated when it decides to list. This is clear from the grammatical structure of the criteria, which treats causality as a merely parenthetical element.

This enables the NTP to reserve to itself the discretion to consider whatever information it wants, to exclude whatever information it wants, and to evaluate that information in accordance with whatever ad hoc criteria it wants to apply. The NTP does not constrain itself to scientific information, either. By withholding

from the public the weight of evidence scheme, the NTP preserves the policy discretion to give any weight it wants to policy goals and objectives, and to keep those weights hidden from public view.

This paper concludes with specific recommendations for statutory reforms that would improve the scientific quality of listing decisions and the practical utility of the RoC for screening-level risk-benefit decision-making. Each recommendation would help restore science to its intended role and end the NTP's science charade.

Background

Secretary of Health, Education and Welfare (HEW) Joseph Califano established the National Toxicology Program (NTP) in November 1978 and transferred to it the function of conducting laboratory tests of chemicals, previously performed by the National Cancer Institute (NCI). The NCI had designed these studies primarily for researching carcinogenicity, not to inform regulatory decision making.¹ Separately, the secretary assigned the NTP the responsibility of producing the *Report on Carcinogens (RoC)* to comply with a new congressional mandate.²

The law called for annual reports, but in 1993 reporting was made biennial.³ The NTP has been unable to adhere to either schedule. It has issued 12 reports, with the 11th and 12th biennial reports published seven years apart—in 2004 and 2011, respectively. That is partly the result of technical complexity to which the law is insensitive.

Mostly, however, delays occur because each RoC is highly controversial. Contrary to Congress' intent, RoC listings are not scientific determinations so much as policy decisions justified, where possible, by science. Responsibility for this failure is shared by Congress, which legislated ambiguously, and the NTP, which implemented ambiguous statutory language in ways that drained the RoC of scientific legitimacy.

The RoC process has its own major deficiencies. The law allows only federal agencies to nominate substances for listing or delisting, denying the public any reasonable opportunity for input into the process.⁴ The NTP decides which substances to advance to the review stage without any transparent and reproducible criteria. If the NTP decides to list a substance, it is assigned to one of two categories established in the law: *known human carcinogen* or *reasonably anticipated to be a human carcinogen*.

Having no statutory instruction concerning how to make these determinations, the NTP devised its own criteria, which are tautological and immune to reasoned dissent or scientific refutation. Peer review consists of ratifying the NTP's assignments in accordance with these criteria. The underlying science is not rigorously reviewed. The NTP permits public participation but does not credibly respond to critical views. With rare exception, substances that are reviewed for listing are reviewed exactly once, for the evidentiary hurdle for delisting a substance appears to be exceedingly high.⁵

This paper consists of three sections and a case study of styrene provided in an appendix. (A longer version of this paper that elaborates more fully on these points and includes case studies of naphthalene and formaldehyde is available from the author's website.)⁶

*RoC listings
are not scientific
determinations
so much as policy
decisions justified,
where possible,
by science.*

- **Section I** provides background on the NTP's program of laboratory testing and its practice of making strong policy inferences on the basis of limited information. The way the NTP interprets the results of laboratory provides useful insight into the way it classifies a substance's human carcinogenicity.
- **Section II** describes the statutory text the NTP is supposed to implement to prepare the RoC and how it has implemented this directive. The NTP's listing criteria are shown to be tautological, evading the implied statutory requirement to demonstrate causality.
- **Section III** concludes with an array of recommendations for reform. Each is a simple change in the statutory wording that would significantly improve the quality and reliability of the RoC.
- **The appendix** consists of a case study on NTP's recent listing of styrene. This listing has a remarkably weak epidemiological foundation, which is contradicted by state-of-the-art scientific data showing near certain proof of the absence of a human cancer risk.

How the NTP Interprets Results of Its Laboratory Experiments

The Department of Health and Human Services—along with its predecessor, HEW—has long had a mix of regulatory, public health, and scientific functions. Among its scientific functions has been an expansive program of laboratory testing of chemicals on animals,⁷ typically rodents, conducted under the auspices of the NCI beginning around 1961 and then the new NTP in 1978.⁸ When these tests are available, they appear to trump all other scientific information in RoC listing decisions.

The NTP classifies the results of its laboratory experiments for carcinogenicity as follows:

- **Clear Evidence of Carcinogenic Activity** is demonstrated by studies that are interpreted as showing a dose-related (i) increase of malignant neoplasms, (ii) increase of a combination of malignant and benign neoplasms, or (iii) marked increase of benign neoplasms if there is an indication from this or other studies of the ability of such tumors to progress to malignancy.
- **Some Evidence of Carcinogenic Activity** is demonstrated by studies that are interpreted as showing a chemical-related increased incidence of neoplasms (malignant, benign, or combined) in which the strength of the response is less than

that required for clear evidence.

- ***Equivocal Evidence of Carcinogenic Activity*** is demonstrated by studies that are interpreted as showing a marginal increase of neoplasms that may be chemically related.
- ***No Evidence of Carcinogenic Activity*** is demonstrated by studies that are interpreted as showing no chemical-related increases in malignant or benign neoplasms.
- ***Inadequate Study of Carcinogenic Activity*** is demonstrated by studies that cannot be interpreted as valid for showing either the presence or absence of carcinogenic activity because of major qualitative or quantitative limitations.⁹

Each category contains scientific information (such as, for example, “studies ... showing a dose-related increase of malignant neoplasms”), but the NTP couches each such fragment of scientific information with ambiguous, non-scientific caveats (for example, “studies *that are interpreted as* showing a dose-related increase of malignant neoplasms” [emphasis added]). These caveats are wholly subject to the personal predilections of NTP scientists and senior managers, or to the agency’s institutional interests.

The descriptors for these categories imply, but do not actually include, scientific content of a probabilistic nature. “Clear evidence” is superior to “some evidence,” which is superior to “equivocal evidence,” which is superior to “no evidence” in determining probability. But the lines dividing these categories are murky at best, and subject to personal bias, politics, and agency bureaucratic interests.

Early on, these judgments varied substantially across toxicologists. Over time, however, the NCI (and then NTP) built a record on how they made these judgments, and that record became the functional equivalent of legal precedent. For example, if the output of an experiment on Substance H looked more like Substances A, D, F, and G than like Substances B, C, and E, it was virtually certain NTP would assign it the same classification as the former. In this way, the NTP’s RoC program developed a similar pattern of consistency. Over time, this pattern of consistency has become confused with scientific accuracy.¹⁰

Dose, the key determinant of human health risk,¹¹ has a peculiar role in this process. A laboratory experiment that shows carcinogenic “activity” in a rodent at doses of a given substance thousands of times greater than a human would ever experience earns the descriptor “clear evidence of carcinogenicity,” even if its actual or likely contribution to human cancer incidence is negligible or zero.¹²

The lines dividing categories are murky at best, and subject to personal bias, politics, and agency bureaucratic interests.

The conduct of chronic two-year bioassays has become a cottage industry for the NTP and numerous private contractors, with several guaranteed markets, including the NTP's own RoC.

Laboratory experiments in rodents, most notably the chronic two-year bioassay,¹³ have been widely touted as the “gold standard” for toxicology. The two-year bioassay is a 104-week controlled laboratory experiment in which different groups of mice or rats of both sexes are exposed to two or three different doses of a test agent, with other groups unexposed to serve as controls.

These experiments have been widely criticized by biologists,¹⁴ toxicologists,¹⁵ and committees of the National Research Council (NRC) over several decades. In 1993, an NRC committee raised profound doubts about the scientific value of chronic two-year bioassays because they rely on the maximum tolerated dose (MTD),¹⁶ which yields many false positives.¹⁷ A more recent committee advocated a fundamental change in direction for toxicological research.¹⁸ Ironically, it appears that as far back as 1979, at least one senior government official expressed the hope that better tests would begin to replace the chronic two-year bioassay as soon as 1985.¹⁹ Since then, the number of chronic two-year bioassays performed by the NTP has risen from about 190 to about 600, with the number of experiments limited only by appropriations and regulatory requirements.²⁰ Thus, the “gold standard” endorsement appears to have more to do with tradition and rent seeking than objective evidence of the value of the scientific information these studies produce.

Changing direction will be difficult for several reasons. The conduct of chronic two-year bioassays has become a cottage industry for the NTP and numerous private contractors, with several guaranteed markets, including the NTP's own RoC.²¹ In addition, any substantial change would make comparisons within the historical record problematic because hundreds of substances have been subject to this ancient protocol. Better approaches likely would produce many fewer false positives, and it would be awkward for the NTP, and the toxicology world generally, to admit that the “gold standard” had produced so much error. For these reasons, the chronic two-year bioassay appears to be in no danger of being replaced by higher quality methods that predict human cancer risk with greater selectivity.

The Statutory Design of the Report on Carcinogens Contrasts with the NTP's Implementation

Many things can cause cancer, but the statutory text requires the NTP to adopt a narrow view of cancer etiology. Only “substances”—predominantly man-made chemicals—matter. Thus, under the NTP criteria, the extent to which cancer is primarily a byproduct of DNA mutations due to aging is not relevant.²²

The relevant statutory text consists of three clauses:

The Secretary shall publish a biennial report which contains—

(A) a list of all substances

*(i) which either are known to be carcinogens
or may reasonably be anticipated to be
carcinogens and*

*(ii) to which a significant number of persons
residing in the United States are exposed;*

*(B) information concerning the nature of such exposure
and the estimated number of persons exposed to such
substances;*

(C) a statement identifying

*(i) each substance contained in the list under
subparagraph (A) for which no effluent, ambient,
or exposure standard has been established by
a Federal agency, and*

*(ii) for each effluent, ambient, or exposure standard
established by a Federal agency with respect
to a substance contained in the list under
subparagraph (A), the extent to which, on the
basis of available medical, scientific, or other
data, such standard, and the implementation of
such standard by the agency, decreases the risk
to public health from exposure to the substance.²³*

The first part of Clause (A) sets forth a two-prong scientific threshold. A substance must be (i) either known or reasonably anticipated to be a carcinogen, *and* (ii) a significant number of persons residing in the United States must be exposed to it.²⁴ Each prong is difficult to implement scientifically because it contains crucial non-scientific language.

“Known” human carcinogens. The NTP’s definition of a *known* human carcinogen is based on *sufficient* evidence:

There is sufficient evidence of carcinogenicity from studies in humans, which indicates a causal relationship between exposure to the agent, substance, or mixture, and human cancer.²⁵*

“Reasonably anticipated” carcinogen. The NTP’s definition of a *reasonably anticipated* human carcinogen is based on three distinct paths, reproduced verbatim below. A careful review of the text shows that there are 11 different ways by which the NTP can deem a substance a *reasonably anticipated* human carcinogen. Path (A) has one, Path (B) has eight, and Path (C) has two:

Path A. There is limited evidence of carcinogenicity from studies in humans, which indicates that causal interpretation is credible, but that alternative explanations, such as chance, bias, or confounding factors, could not adequately be excluded;*

Path B. There is sufficient evidence of carcinogenicity from studies in experimental animals, which indicates there is an increased incidence of malignant and/or a combination of malignant and benign tumors

- 1. in multiple species or at multiple tissue sites, or*
- 2. by multiple routes of exposure, or*
- 3. to an unusual degree with regard to incidence, site, or type of tumor, or age at onset; or*

Path C. There is less than sufficient evidence of carcinogenicity in humans or laboratory animals; however, the agent, substance, or mixture belongs to a well-defined, structurally related class of substances whose members are listed in a previous Report on Carcinogens as either known to be a human carcinogen or reasonably anticipated to be a human carcinogen, or there is convincing relevant information that the agent acts through mechanisms indicating it would likely cause cancer in humans.²⁶

Path (A) is the only one involving human data. It consists of the same text as the known carcinogen criterion except that it requires only *limited* rather than *sufficient* evidence.²⁷ The minimum threshold to qualify as *limited* evidence is not defined.

Paths (B) and (C) are the NTP’s answer to the question, *What should be done when all positive data are from animal tests?* Science provides no answers.

The default practice has been to simply assume that substances that cause cancer in animals at high doses also cause cancer in humans at low doses.²⁸

Path (B) requires the existence of sufficient evidence of carcinogenicity in animals rather than humans. A careful reading reveals eight possible ways to get there:

- Increased incidence of malignant tumors at multiple tissue sites in a single species;
- A combination of malignant and benign tumors in multiple species;
- A combination of malignant and benign tumors at multiple tissue sites in the same species;
- A combination of malignant and benign tumors by multiple routes of exposure in the same species;
- A combination of malignant and benign tumors in one species to an unusual degree with regard to incidence;
- A combination of malignant and benign tumors in one species to an unusual degree with regard to site;
- A combination of malignant and benign tumors in one species to an unusual degree with regard to type of tumor; and
- A combination of malignant and benign tumors in one species to an unusual degree with regard to age at onset.

As before, the minimum threshold for evidence to be *sufficient* is not defined. Similarly, ambiguous words that are integral parts of some of these paths (e.g., *increased* incidence, *unusual degree*) are not defined.

Path (C) is a pair of catchall categories for situations in which the NTP apparently believes that a substance ought to be designated as a reasonably anticipated human carcinogen, but the data are too weak, too controversial, or too burdened by negative or equivocal data. Either the following circumstances is enough:

- The substance “belongs to a well-defined, structurally related class of substances” previously listed as a carcinogen;
- “There is convincing relevant information that the agent acts through mechanisms indicating it would likely cause cancer in humans.”²⁹

Both circumstances appear to be forms of guilt by association that obviate the usual need for evidence. Nothing needs to be known about the actual carcinogenicity of the substance.³⁰

The default practice has been to simply assume that substances that cause cancer in animals at high doses also cause cancer in humans at low doses.

The NTP has kept its definitions secret, so the public can only guess at what they mean.

When all 11 options are taken together, it becomes clear that the NTP’s definition of a *reasonably anticipated* human carcinogen is potentially quite capacious. Exactly how capacious depends on how the NTP interprets the many ambiguous nonscientific terms, most notably the thresholds for *sufficient*, *less than sufficient*, and *limited* evidence.

The inscrutability of “sufficiency” and similar descriptors. The NTP’s listing criteria depend on several crucial descriptors that it has chosen not to define. Thus, to be a *known* human carcinogen requires “sufficient evidence of carcinogenicity from studies in humans,” but the NTP never defines the minimum amount of evidence necessary to be *sufficient*. Similarly, to be a *reasonably anticipated* human carcinogen requires “limited evidence of carcinogenicity from studies in humans,” “sufficient evidence of carcinogenicity from studies in experimental animals,” or “less than sufficient evidence of carcinogenicity in humans or laboratory animals,” plus other information that does not even qualify as evidence. But the NTP does not define the minimum amount of evidence necessary to be *limited* or *less than sufficient*.

To be clear, the NTP could not have defined them in scientific terms because they do not have scientific meaning. They have meaning in other areas—in law, for example, where *sufficiency* defines the threshold burden of proof for criminal or civil litigation purposes, and it varies depending on context. Evidence may be *sufficient* in civil litigation if it meets a “preponderance of evidence” standard—typically greater than 50 percent likelihood. In criminal litigation, however, evidence is *sufficient* only if it meets the “beyond reasonable doubt” standard.³¹

Unfortunately, the NTP has kept its definitions secret, so the public can only guess at what they mean. We do not know if the NTP requires evidence to be “beyond a reasonable doubt” (≥ 95 percent), a “more likely than not” (> 50 percent), or maybe well below 50 percent. For all we know, the NTP might be applying a “beyond reasonable doubt” standard in which the null hypothesis is that a substance *is* a carcinogen, and thus it is the duty of negative evidence to show that there is less than 5 percent chance that it is not.³² Of course, the NTP could have been more revealing about its evidentiary threshold. For example, it could have explicitly used one of these established legal principles. It did not do so.

Both “known” and “reasonably anticipated” require causation, which the NTP assumes rather than assesses. To say a substance is a *known* human carcinogen means a causal relation is well established and beyond controversy. These are attributes one would expect from an implication of virtual scientific

certainty. There must be a strong, specific, and selective statistical association in epidemiologic studies supported by clear biological theory and evidence that is relevant to humans. Except in the extraordinary circumstance where high-quality human data are available from a well-controlled experiment,³³ there should be multiple high-quality human studies from different populations, with consistent, robust results that are insensitive to all reasonable assumptions about such matters as model specification.

The inference of a strong causal relationship implies that there are no credible grounds for doubt or skepticism, such as the presence of contradictory evidence or important limitations in study design or data quality. For example, there must not be serious weaknesses in positive studies, nor can there be studies of similar or greater quality showing that the observed cancer incidence has other identified or scientifically plausible origins or high-quality studies that show no carcinogenic effects.

In other words, to be scientifically deemed a *known* carcinogen, the database must convincingly refute the null hypothesis of no effect. This is the conventional scientific position that A is not said to cause B unless and until proven otherwise. Evidence marshaled to prove the case must meet very high internal quality standards and be compellingly consistent.³⁴

Presumably, the causation threshold for being a *reasonably anticipated* human carcinogen is lower than that for being a *known* human carcinogen. But how much lower? The NTP does not say. Interestingly, the legislative history provides useful insight concerning what Congress intended. An earlier version of the bill would have required the RoC to list *suspected* carcinogens, a much more expansive category, but in the bill enacted into law, *suspected* was replaced with *reasonably anticipated*. In the words of one of the sponsors of the earlier bill containing the expansive term *suspected* carcinogen, this was done “in order to make it absolutely clear in the statute that there must be reasonable ground for designating a substance as a putative carcinogen.”³⁵ The definition of “reasonableness,” of course, is eternally elusive and Congress did not give additional guidance concerning what it meant. While Congress might have thought this would ensure “absolute clarity” about the need for “reasonableness,” in practice the NTP has interpreted *reasonably anticipated* to mean about the same thing as *suspected*.

Instead of assessing causality, the NTP merely assumes it. Recall that to earn designation as a *known* human carcinogen, all that is required is *sufficient* evidence from studies in humans. If such evidence exists, it is assumed to “indicate a causal relationship.” In the definition of a *reasonably anticipated* human

The inference of a strong causal relationship implies that there are no credible grounds for doubt or skepticism, such as the presence of contradictory evidence or important limitations in study design or data quality.

A substance belongs in one of two categories if the NTP says it does, nothing more and nothing less.

condition, the test is somewhat more complex but just as tautological. A substance belongs in one of these two categories if the NTP says it does, nothing more and nothing less.

This becomes clear when the sentence construction is examined carefully. In the *known* human carcinogen criterion, the clause *which indicates a causal relationship* is preceded by a comma. Grammatically, this means the clause is a parenthetical element; it can be removed from the sentence without changing the sentence's meaning.³⁶ Thus, the full meaning of the *known* human carcinogen criterion can be obtained by excising everything after the comma: "There is sufficient evidence of carcinogenicity from studies in humans." But that, of course, is a mere tautology. A substance which the NTP defines as *known* human carcinogen must have *sufficient* supporting evidence, as that term is understood in common English usage. Were the evidence *not sufficient* it would have to be *insufficient*, but it would make no sense to say a substance is a *known* carcinogen based on insufficient evidence.

Causality is also grammatically implied in the *reasonably anticipated* criterion. For *known* carcinogens, conflicting evidence or doubts about the strength of positive evidence appear not to exist. For *reasonably anticipated* carcinogens, questions about conflicting evidence or doubts about the strength of positive evidence are acknowledged, but deemed unpersuasive. The full meaning of the *reasonably anticipated* criterion is similarly obtained by excising everything after the comma in Paths (A) and (B) but retaining the text following the semicolon in Path (C), as follows:

- There is *limited* evidence of carcinogenicity from studies in humans; or
- There is *sufficient* evidence of carcinogenicity from studies in animals; or
- There is *less than sufficient* evidence of carcinogenicity in humans or animals; however, the agent, substance, or mixture belongs to a well-defined, structurally related class of substances whose members are listed in a previous Report on Carcinogens as either known to be a human carcinogen or reasonably anticipated to be a human carcinogen, or there is convincing relevant information that the agent acts through mechanisms indicating it would likely cause cancer in humans.

The first and second bullets are unrevealing. Designation as a reasonably anticipated human carcinogen means that the requisite evidentiary threshold was met, but the identity of that threshold is not defined. The third bullet denies

the existence of any minimum evidentiary standard at all. It permits information that is not admissible as evidence—information that does not pertain to the substance itself—to be the sole basis for conviction.

Weight-of-evidence weightlessness. The NTP cannot make any determination without what is called a weight of evidence (WoE) scheme. This term of art has become popular in risk assessment circles, but it is borrowed from the expressly nonscientific field of law.³⁷ A *scientific* WoE scheme is one that is designed and implemented so as to exclude all nonscientific considerations, such as scientists' personal opinions and risk management preferences, agencies' bureaucratic interests, and the like. Transparency and reproducibility are crucial process requirements, for without them it cannot be verified that a WoE scheme in fact contains only scientific evidence or that the agency applied its scheme as written.

The NTP uses a WoE scheme but it has never made it public. For a *known* carcinogen, positive results from human studies are required, but how many such studies are needed? How strong must they be? How are negative or equivocal data taken into account? The NTP's definition answers none of these questions. Moreover, the NTP elucidates no transparent way whereby a presumption of causality could be rebutted with scientific evidence. In short, the NTP reserves to itself the discretion to consider whatever information it wants, to exclude whatever information it wants, and to evaluate that data in accordance with whatever ad hoc criteria it wants to apply.

The NTP does not constrain itself to use only scientific information. Rather, it permits itself the discretion to give any weight it wants to policy goals and objectives, and to keep those weights hidden from public view. A WoE scheme that is not transparent or reproducible might be legally adequate so long as the NTP acknowledged that its listing decisions were exercises in policy judgment and not scientific determinations. But the NTP professes that its listing determinations are scientific, and they do not include non-scientific considerations such as precautionary risk management policy judgment. Nowhere in the listing criteria does the NTP acknowledge that judgment of any kind is being exercised.³⁸

The NTP admits that its listing decisions have no practical utility for decision making. The preamble to the RoC contains several interesting disclaimers. For example, the NTP acknowledges that listing decisions capture only theoretical or hypothetical cancer risk: “[T]he listing of substances in the RoC only indicates a *potential hazard* and *does not establish the exposure conditions that would pose cancer risks* to individuals in their daily lives” [emphasis added].³⁹ In short, NTP listing conveys nothing about actual human cancer risk. It means

The NTP reserves to itself the discretion to consider whatever information it wants, to exclude whatever information it wants, and to evaluate that data in accordance with whatever ad hoc criteria it wants to apply.

The statutory categories known and reasonably anticipated are not scientific terms, so substances cannot be assigned to them through scientific means.

only that cancer risk is possible at some unspecified dose that may or may not be relevant to human experience. The NTP then attempts to shift to others the responsibility of demonstrating that listed substances cause cancer in humans: “[F]ormal risk assessments are the responsibility of the appropriate Federal, state, and local health regulatory and research agencies.”⁴⁰ This is an astounding admission, for it means the NTP knows that the RoC does not fulfill the charge given to it by Congress.

The NTP has implemented the law this way at least in part because the law invited it to do so. The statutory categories *known* and *reasonably anticipated* are not scientific terms, so substances cannot be assigned to them through scientific means. A substance might be considered a *known* or *reasonably anticipated* carcinogen if the vast majority of scientists consider it so, but any such agreement is a statement about shared subjective probabilities or policy preferences, not scientific fact. Of course, the NTP could have devised listing criteria that approximated a plausibly scientific scheme, such as by clearly specifying in probabilistic terms what the two categories meant. For example, it could have determined what probabilities are implied by the terms *known* and *reasonably anticipated*, and asked the independent scientists who provide advice whether they believed the scientific evidence was strong enough to exceed these thresholds. Instead, the NTP has practiced what a staunch advocate for stringent standards has called a “science charade,” exaggerating the contributions made by science to avoid accountability for underlying policy decisions.⁴¹

The law limits listings to substances where there is significant human exposure for listing, but the NTP does not implement this provision. The second part of Clause (A) in the statutory charge limits the RoC to only those substances to which a “significant number of persons residing in the United States are exposed.” To fulfill this element of its statutory directive, the NTP must investigate and estimate the extent of human exposure in the United States. This cannot be done merely by counting mass or volume. Exposure is the *ratio* of mass or volume per unit of time, and dose—a more biologically relevant concept than exposure—is the ratio of mass or volume for a defined biological unit (for example, body weight) per unit of time. The statutory charge also cannot be met by relying on historical data (persons who were exposed sometime in the past), figures of a hypothetical nature (persons who may be exposed), or figures obtained elsewhere (persons who are exposed in China). The law is clear: It must be actual exposure occurring now in the United States.⁴²

There are three tasks the NTP needed to complete to fulfill this prong of Clause (A):

- Define “a significant number of persons residing in the United States”;
- Define a *de minimis* cancer risk level; and
- Estimate for each candidate substance the number of persons in the United States exposed above the *de minimis* cancer risk level.

If and only if this number was *significant* would the substance be eligible for listing.

The first two of these tasks are exercises of policy discretion delegated by Congress, while the third is strictly scientific. Congress gave the NTP the policy discretion to define *significant* to be a large or small number of people, and perhaps to make it general or context-specific, and a scientific mandate to estimate the number of people who qualify.⁴³

The NTP has performed none of these tasks. Instead of estimating the number of persons in the United States actually exposed, the NTP reports mass or volume and sometimes historical or hypothetical numbers of persons potentially exposed. If mass or volume are great enough, the NTP simply assumes that a *significant* number of persons are exposed. Moreover, the NTP does not estimate the cancer risk to which they are allegedly exposed or define a *de minimis* cancer risk level.

Estimates of the actual number of persons exposed. Clause (B) directs the NTP to actually estimate this number and provide information about the *nature* of their exposure. The legislative history supports a broad interpretation of this task.⁴⁴ The requirement to quantify the number of persons exposed serves a critical purpose, which is to ensure that the NTP focuses on high-priority substances and is not distracted by minutiae.⁴⁵ As for the *nature* of exposure, it is reasonable to infer that Congress intended the NTP to focus on environmental and occupational cancer risks because it was these circumstances on which Congress was focused at the time it enacted the law.

The NTP does not estimate the actual number of persons exposed. As noted above, it relies on mass and volume indicators in lieu of exposure indices. For example, the NTP substance profile for styrene includes estimates of production volume as if production volume implied human exposure.⁴⁶

On the requirement to estimate risk reduction achieved by regulation, the NTP has been utterly silent.

Identify regulatory gaps and estimate cancer-reduction benefits. Clause (C) directs the NTP to identify regulatory gaps and report the extent to which regulation “decreases the [cancer] risk to public health.” The importance of this provision is clear from the legislative history. Congress wanted the RoC to include “where possible, estimates [of] the magnitude of the risk each [substance] poses.”⁴⁷ This text can be read to permit the NTP to also estimate the value of regulations intended to reduce cancer risk.

Methods for valuing health effects were certainly primitive in the late 1970s, when Congress enacted the law authorizing the RoC, but the same is true with respect to the estimation of cancer risk. Presumably, Congress expected the quality of this information to improve over time, as the NTP gained experience and quantitative methods in public health and economics improved.⁴⁸ To be sure, the NTP did not have the institutional capacity in 1978 to quantify the number of cancer cases (or other effects) prevented by regulation, much less estimate the value of risk reduction achieved. And a strong case can be made that it would have encountered strenuous resistance from other agencies, such as the Environmental Protection Agency and the Occupational Safety and Health Administration. Nonetheless, the statutory text clearly provided at least a wide open invitation to the NTP to expand into these areas with congressional blessing, and public choice theory suggests that it would have tried to do so.

But the NTP did not take advantage of this opportunity to establish, build, and sustain these new missions. One possible explanation is that the NTP’s RoC program was—and remains somewhat—a virtual agency, staffed and funded mostly by the National Institute of Environmental Health Sciences but with significant contributions from other agencies. This structural design could have prevented the NTP from expanding its mission into areas already occupied by other agencies, by giving it fewer of the incentives for institutional expansion that agencies usually face.⁴⁹

In any event, the NTP has ignored both of these statutory requirements. In lieu of the requirement to identify potential regulatory gaps, each substance profile includes a list of applicable regulations and guidelines. On the requirement to estimate risk reduction achieved by regulation, the NTP has been utterly silent.

Recently proposed revisions to the review process. The NTP recently proposed to modify its review procedures for the 13th edition, “to enhance transparency and efficiency and to enable the NTP to publish the RoC in a timelier manner.”⁵⁰

The most significant of these process changes are: 1) elimination of one round of public comment; 2) use of ad hoc peer review panels of unknown character, composition, independence, transparency, and charge; and 3) elimination of any duty to respond to public comment. Each of these changes is likely to reduce transparency and efficiency, and markedly increase controversy.

In addition, the proposed process revisions include a material change in substance. The NTP proposes to “tailor” its review procedures to allow it to rely on a variety of nonstandardized inputs, without any safeguards in place to ensure that they satisfy applicable information quality standards (including transparency, reproducibility, and objectivity), adequate provision for public access, or meaningful public participation. As it is described, “tailoring” would permit the NTP to accept and even rely upon nonscientific input, including nonscientific input of a frankly political nature.

What Can Congress Do?

To make the *Report on Carcinogens* a genuine science compendium, Congress needs to legislate significant reforms. Here are six complementary suggestions.

- ***Direct the NTP to make its determinations conditional on exposure or dose.*** As noted above, the NTP completely ignores exposure or dose in making its determinations. That severely undermines the practical utility of the RoC and arguably renders its determination useless or even misleading. This reform would provide listing decisions that have value from a public health standpoint and more closely approximate original congressional intent.
- ***Direct the NTP to include potency in its listing decisions.*** The NTP makes no distinction between strong and weak carcinogens. This ignores more than 30 years of scientific knowledge—much of it taxpayer-funded—that has provided valuable insights concerning the relative potency of different carcinogens. Relative potency matters for the same reason that the practical utility of the RoC depends on dose or level of exposure. It is misleading to report substances with the same carcinogenicity label when their capacity to cause cancer varies by orders of magnitude.

*“Tailoring”
would permit the
NTP to accept
and even rely
upon nonscientific
input, including
nonscientific
input of a frankly
political nature.*

If Congress wants the RoC to be a scientific compendium, it has to abandon its reliance on nonscientific descriptors such as known and reasonably anticipated.

- ***Replace problematic risk descriptors or provide guidance concerning how to interpret them.*** If Congress wants the RoC to be a scientific compendium, it has to abandon its reliance on nonscientific descriptors such as *known* and *reasonably anticipated*. A better approach is to explicitly state alternative levels of concern in units scientists understand, such as probabilities, and provide enough alternative categories to allow for distinctions commensurate with the evidence. One approach already in use elsewhere has seven descriptors, each mapped to a specific probability: virtually certain (> 99 percent), very likely (> 90 percent), likely (> 66 percent), about as likely as not (33-66 percent), unlikely (< 33 percent), very unlikely (< 10 percent), and exceptionally unlikely (< 1 percent).⁵¹ Such a scheme would enable advances in scientific knowledge to be reflected in classification decisions. There would be spirited debate about which category is best for individual substances at different exposures or doses, but it would be better to have a debate about adjacent probabilistic descriptors than about how to scientifically interpret legal terms. Moreover, differences of opinion among scientists would be highly informative concerning the range of uncertainty. For example, if each independent expert were highly confident of the probability category to which a substance should be classified, but the experts differ concerning which category that is, then the public would know that the assignment depends less on scientific knowledge than on which scientists are asked to provide their opinions.
- ***Direct the NTP to establish a strictly scientific weight of evidence scheme.*** Congress could direct the NTP to devise a new WoE scheme that is transparent, reproducible, and strictly science-based. This last requirement is essential to restore science as the foundation for listing decisions. A WoE scheme that allows NTP's own chronic two-year bioassays to trump other science, or allows nonscientific factors to intrude or even dominate, is really a weight-of-politics scheme.

- ***Sunset listings to encourage revision.*** The current process is anti-scientific because it encourages the NTP to review each substance once, then bolt the door to prevent the intrusion of inconvenient, new scientific knowledge. Congress can overcome this by giving the NTP a nondiscretionary duty to review its previous listings on a set schedule or allow them to expire. Under the current process, reviews of listed substances must pass a significant bureaucratic gauntlet. They must be nominated for review *and* be accepted by the NTP *and* by its federal partners. This seems very unlikely to occur except in extraordinary circumstances.
- ***Direct the NTP to affirmatively comply with applicable Information Quality Guidelines.*** The NTP's various reports are covered by Information Quality Guidelines (IQG) issued by both the Office of Management and Budget and the Department of Health and Human Services.⁵² These guidelines require, among other things, that scientific information be transparent, reproducible, and substantively and presentationally objective. In practice, however, agency commitments to adhere to these guidelines are not enforceable and are thus unmet. Problems with NTP compliance became clear almost immediately after the IQGs became effective in 2002. Numerous requests for correction were filed within the first two years, several of which were subsequently appealed without success.⁵³ An enforceable statutory mandate to adhere to information quality principles would dramatically improve the scientific quality of the NTP's background documents and substance profiles. It also would indirectly compel the NTP to disclose its existing weight of evidence scheme, subject it for the first time to critical scientific review, and publicly reveal the extent to which NTP decisions are determined by nonscientific considerations. Congress could achieve a highly significant reform simply by making adherence to the IQG a statutory imperative.

Each of these proposals could accomplish a lot with only modest legislative changes. Of course, Congress also could reconsider whether there is

An enforceable statutory mandate to adhere to information quality principles would dramatically improve the scientific quality of the NTP's background documents and substance profiles.

a genuine public need for listings such as the RoC, and if so, whether the RoC is needed to satisfy it. The production of a science compendium is not an inherently governmental function, and there are nongovernmental organizations that also provide such references. In an era when the federal government faces many competing goals and increasingly constrained resources, it is possible that the RoC served a useful purpose 30 years ago but is no longer needed.

Appendix: Styrene Case Study

The National Toxicology Program (NTP) added styrene to the *Report on Carcinogens* in the 12th edition, concluding that it was a *reasonably anticipated* human carcinogen. It based this determination on *limited* evidence from occupational studies in humans (see discussion of Path (A) beginning at page 8) and *sufficient* evidence of carcinogenicity in animals. It gave the most weight to a pair of epidemiological studies of workers in the European reinforced-plastics industry, described by the NTP as showing “significantly higher risks (or elevated risks approaching statistical significance),” and a multi-plant cohort study of styrene-butadiene rubber workers, described by the NTP as suggesting an exposure-response relationship between styrene and non-Hodgkin’s lymphoma (NHL) and NHL-chronic lymphocytic leukemia (NHL-CLL) that could not be explained by butadiene exposure.⁵⁴

Causality assumed. Consistent with the grammatical construction of the listing criteria, the NTP did not make a showing of causality with respect to any of these studies. It treated causality as a presumptive default incapable of being rebutted:

Causality is not established, as the possibility that the results were due to chance or to confounding by exposure to other carcinogenic chemicals cannot be completely ruled out. However, a causal relationship between styrene exposure and cancer in humans is credible and is supported by the finding of DNA adducts and chromosomal aberrations in lymphocytes from styrene-exposed workers.⁵⁵

The standard model of carcinogenesis assumes that risk is a function of cumulative lifetime exposure.⁵⁶ Therefore, cumulative exposure is the correct exposure metric. However, in the main study on which the NTP relied, a statistically significant association was observed with average but not cumulative exposure.⁵⁷ This posed a serious barrier to inferring causality. If the conventional model is correct, then the observed association with average exposure is either spurious or a different theoretical model of carcinogenesis is needed to explain it. The NTP resolved the matter by retaining the conventional model of carcinogenicity but assuming the existence of an unknown mode of action:

Without *a priori* knowledge, it is difficult to know which exposure metric is most appropriate for evaluating causality, so a positive relationship observed with any exposure metric is a concern.⁵⁸

Public comment. The Styrene Information and Research Center and several consulting academics⁵⁹ commented on the proposed nomination,⁶⁰ the draft background document,⁶¹ the NTP expert panel recommendation,⁶² and prior to Board of Scientific Counselor (BSC) review.⁶³ Industry also made a pair of last ditch efforts to stop the final decision. One letter appealed to a shared high-level professional scientific affiliation; sender and recipient were both past presidents of the Society of Toxicology.⁶⁴ A second letter asked the NTP to defer to a recently published report of a National Research Council committee that, although directed at an EPA draft risk assessment, was highly critical of the scientific case the NTP was relying upon to infer human cancer risk.⁶⁵

In these comments, the industry raised numerous scientific issues. These included the allegation of frank error in the way the NTP interpreted the epidemiologic studies and a willful refusal to consider and rebut evidence showing that results in rodents would not be replicated in humans because of known physiologic and metabolic differences.⁶⁶ The principal author of the crucial styrene-butadiene rubber worker study objected to NTP's interpretation of her work as showing a causal relationship.⁶⁷ Where the NTP inferred a causal exposure-response relationship in the main reinforced-plastics industry study, the researchers themselves interpreted their results much more modestly.⁶⁸ In short, the NTP simultaneously applied an exceptionally low threshold for *limited* evidence of carcinogenesis and an extraordinarily high evidentiary threshold for rebuttal.

To escape the *reasonably anticipated* designation, the industry tried to prove that human cancer risk was biologically infeasible using a “knock-out” mouse experiment.⁶⁹ A knock-out mouse is one that has been genetically altered to remove the specific gene of concern, in this case one that produces the enzyme CYP2F2. If the positive results observed in laboratory experiments in mice were specific to mice and not relevant to humans, the knock-out mice would not display sensitivity to styrene. That is precisely what occurred. Knock-out mice experienced no dose-related increase in lung toxicity, even from high doses of styrene, effectively proving that the results observed in mice were specific to mice and not applicable to humans.

These results had not been peer reviewed in time for the 12th edition of the RoC, but their specific relevance to the NTP's determination are obvious. Had the NTP wanted to prevent issuing a false positive determination, it could have decided to review the knock-out mouse study with care and postpone deciding whether to list until the 13th edition. It did not. The listing went forward as if this research had never been performed.⁷⁰ This is consistent with the inference that positive animal data are enough for the NTP to define a substance as a *reasonably anticipated* human carcinogen unless human carcinogenicity can be proved impossible beyond a reasonable doubt.⁷¹

NTP process failures. Industry representatives also identified numerous violations of the RoC process and generally accepted scientific practices.⁷² They included:

- The limited time available for stakeholders to present information in public meetings;
- The alleged delegation of the task of writing portions of the background document to the author of one of the studies on which the NTP intended to rely;
- Cherry-picking of data, models and analyses to support preferred inferences;
- Use of non-peer reviewed information first introduced by members of a peer review panel;
- The NTP review panel's decision to base its recommendations on their own non-peer reviewed re-analyses; and
- The NTP's nondisclosure of written comments from BSC members.

Process confusion. Industry and academic comments led to no material change in the NTP's characterization of the science, or any perceptible effort by the NTP to examine the procedural complaints and address confirmed irregularities and defects. As for the scientific evidence, the dispute can be reduced to three related issues.

First, as noted above, under the RoC listing criteria, causality is automatically assumed once a determination that evidence from human studies is either *sufficient* (for a *known* human carcinogen) or *limited* (for a *reasonably anticipated* human carcinogen). Industry and academic commenters persistently misconstrued the RoC listing criteria as creating a separate requirement for the NTP to make a showing of causality. That may be standard scientific practice, but it is not a correct reading of the text of the listing criteria and it is not what the NTP actually did.

Second, industry and academic commenters erroneously assumed that it was possible to rebut the NTP's assumption of causality through the presentation of scientific evidence. But neither the listing criteria nor the review process provide any way to do this. The evidentiary standard for rebuttal is not known; it is not even known whether the NTP has such a standard.

Third, commenters may have understood that NTP determinations were strictly policy decisions, but hoped that these policy decisions could be swayed by science. This hope would have been fostered by the NTP's persistent assertion that science ruled the roost. But it appears to have been an incorrect reading of the true role of science in NTP determinations. Rather than being informed by science, and thus potentially influenced by scientific evidence, the NTP appears to have made a policy decision based on nonscientific considerations that styrene ought to be listed. By telegraphing that science would rule decision making, however, the NTP successfully led industry to devote all of its resources to the wrong battlefield.

Why statutory reform is needed. This diversionary tactic has obvious bureaucratic and political advantages to the agency. First, it reinforces the NTP's narrative that it is the repository of relevant scientific expertise. All other scientists, including many with extraordinary credentials in their fields, are thus reduced to mere supplicants.

Second, it conserves the NTP's discretion to choose its preferred policy outcome in a way that inhibits political accountability. Public officials, whether in the Executive branch or Congress, never want to be seen as "interfering with science." The NTP ties these officials' hands by professing that its policy decisions are actually scientific. Ironically, by contesting the NTP's actions as if they were the products of scientific analysis and review, the industry may have unwittingly increased the NTP's capacity to make policy decisions under the cover of science and strengthened its capacity to resist policy level scrutiny.

That said, it is not clear what else could have been done to motivate the NTP to change course. A scientific case against listing was always necessary, so extraordinary expenditures on scientific review and research should be viewed as having been necessary. That expenditures on science were not enough, and perhaps never could have been enough, means that a different legislative charge to the NTP would have been needed to achieve a science-based outcome.

The needed statutory charge is one that explicitly maximizes the role of science in decision making and provided effective means of ensuring accountability. Each of the recommendations made in the final section of the main paper, had they been in place when the NTP reviewed styrene, would have reduced the NTP's capacity to perform its science charade.

Notes

- 1 Office of Technology Assessment, *Identifying and Regulating Carcinogens*, Washington, D.C.: U.S. Government Printing Office, 1987, p. 147.
- 2 Pub. L. 95-622, 92 Stat. 3435.
- 3 Pub. L. 103-43, § 2009.
- 4 42 U.S.C. § 241(b)(4)(D). For the 12th edition, the NTP permitted the public to make suggestions.
- 5 A recent example of delisting is glass wool (i.e., fiberglass), which the NTP listed as *reasonably expected to be* a human carcinogen in the seventh edition of the RoC (1994), apparently in response to a previous decision by the International Agency for Research on Cancer (IARC) to list it as Group 2B (*possibly carcinogenic to humans*). After an extended industry research effort, IARC revised its classification downward to Group 3 (*inadequate evidence in humans; limited evidence in experimental animals*). International Agency for Research on Cancer, *Man-made Vitreous Fibres, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* (Lyon, FR: International Agency for Research on Cancer, 2002). In 2004, industry nominated glass wool for delisting from the RoC, and in 2011 NTP modified its substance profile in the 12th RoC to exclude varieties of glass wool that are not biopersistent in the lung. Cf. “Glass Wool (Respirable Size)” in the 11th RoC and “Certain Glass Wool Fibers (Inhalable)” from the 12th RoC. The path—from not being labeled, to being labeled as a *possible* human carcinogen, then *reasonably anticipated to be* a human carcinogen, to once again not being labeled—extended over 20 years. Moreover, the NTP did not delist the substance so much as change its definition to exclude fiberglass.
- 6 The longer version of this paper is available at <http://www.rbbelzer.com/working-papers.html>.
- 7 42 U.S.C. § 241(b)(1): “The Secretary shall conduct and may support through grants and contracts studies and testing of substances for carcinogenicity, teratogenicity, mutagenicity, and other harmful biological effects.”
- 8 Office of Technology Assessment, *Identifying and Regulating Carcinogens*, p. 16.
- 9 National Toxicology Program, *NTP Study Reports; Definition of Carcinogenicity Results*, National Toxicology Program, May 12, 2005, <http://ntp.niehs.nih.gov/?objectid=07027D0E-E5CB-050E-027371D9CC0AAACF>.
- 10 It should be noted that the variable being defined is quite broad. It is not “cancer” that is being described but “carcinogenic activity.” This distinction is usually lost on non-toxicologists, who readily but incorrectly infer that carcinogenic “activity” is the same thing as “cancer.”
- 11 The fundamental principle of toxicology was enunciated by Philippus Aureolus Theophrastus Bombastus von Hohenheim (“Paracelsus”), 1493–1541: “All things are poison, and nothing is without poison; only the dose permits something not to be poisonous.”
- 12 The relevance of high-dose animal experiments to humans exposed to low doses is simply assumed. IARC, which predates the NTP and whose classification model the NTP largely follows, though with somewhat different language, makes this clear: “[I]n the absence of adequate data on humans, it is biologically plausible and prudent to regard agents and mixtures for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans” [boldface in original, internal citations omitted]. See International Agency for Research on Cancer. 2006. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon, FR: International Agency for Research on Cancer, p.18.
- 13 At the end of the two-year bioassay experiment, and often at specific intermediate points, animals are sacrificed and carefully examined pathologically for evidence of malignant or benign tumors, precancerous lesions, and other effects believed to be possible precursors of cancer. Doses are purposefully chosen to be very high so as to maximize the experiment’s “sensitivity,” i.e., the test’s ability to detect an effect if an effect is biologically plausible. High sensitivity has its costs, most notably low “selectivity,” i.e., the test’s inability to discriminate between mechanisms that may be relevant to humans at low doses (potent genotoxicity) and mechanisms that are not (cancer subsequent to organ toxicity resulting from frank poisoning).
- 14 Bruce N. Ames, Renae Magaw, and Lois Swirsky Gold, “Ranking Possible Carcinogenic Hazards,” *Science* Vol. 236, 1987, pp. 271-280. “Extrapolation from the results of rodent cancer tests done at high doses is routinely attempted by regulatory agencies when formulating policies attempting to prevent future cancer. There is little sound scientific basis for this type of extrapolation, in part due to our lack of knowledge about mechanisms of cancer induction, and it is viewed with great unease by many epidemiologists and toxicologists.”
- 15 Henry C. Pitot III and Yvonne P. Dragan, “Chemical Carcinogenesis,” in *Casarett & Doull’s Toxicology: The Basic Science of Poisons*, C. D. Klaassen, ed., New York: McGraw-Hill Medical Publishing Division, 2001, p. 29. Despite these criticisms and problems, the chronic 2-year bioassay continues to be the major basis for regulatory action in this country and in many countries throughout the world.”
- 16 The “maximum tolerated dose” is defined as “[t]he highest dose of a drug or treatment that does not cause unacceptable side effects”; see National Cancer Institute, *Dictionary of Cancer Terms*; “Maximum Tolerated Dose,” 2001, <http://www.cancer.gov/dictionary/?CdrID=54659>. The definition of “unacceptable” varies. See David L. Eaton and Curtis D. Klaassen, “Principles of Toxicology,” in *Casarett & Doull’s Toxicology*, op. cit. note 15, pp. 29-30.
- 17 National Research Council, *Issues in Risk Assessment*, Washington, D.C.: National Academies Press, 1993. A false positive occurs when the null hypothesis of no effect is rejected even though it is true.
- 18 National Research Council, *Toxicity Testing in the 21st Century: A Vision and a Strategy*, Washington, D.C.: National Academies Press, 2007.

- 19 Office of Technology Assessment, op. cit. note 1, p. 147.
- 20 For a list of all chronic two-year bioassays performed, see National Toxicology Program, *NTP Study Reports; Long-Term Study Reports & Abstracts*. National Toxicology Program, September 15, 2011, <http://ntp.niehs.nih.gov/index.cfm?objectid=084801F0-F43F-7B74-0BE549908B5E5C1C>.
- 21 The inability of government laboratories to shoulder the original load is recounted by the Office of Technology Assessment, op. cit. note 1. Today, this cottage industry has an overwhelming rent seeking interest in opposing reform.
- 22 When Congress enacted this statute in 1978, it appears to have believed that most cancer has an environmental origin and is preventable “if we identify causative agents, and avoid them, eliminate them from the environment, or modify the individual’s response to them, or reverse or arrest the biological effects that may result in cancer.” See Peter de la Cruz, “Report on Carcinogens—Legislative History,” Keller and Heckman LLP, April 23, 2009, <http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/Styrene/SIRC20090501.pdf>. At that time, the prevailing view among scientists was that chemicals in the environment and the workplace combined were responsible for about 6 percent of cancer fatalities with tobacco and diet responsible for 30 percent (range: 25-40 percent) and 35 percent (range: 10-70 percent), respectively. See Richard Doll and Richard Peto, *The Causes of Cancer*, Oxford University Press, 1981.
- 23 42 U.S.C. § 241(b)(4).
- 24 Though the statutory text does not say so explicitly, this text is understood to mean that listing is limited to *human* carcinogens.
- 25 National Toxicology Program, *Report on Carcinogens Listing Criteria*, June 15, 2011, <http://ntp.niehs.nih.gov/?objectid=47B37760-F1F6-975E-7C15022B9C93B5A6>. The asterisk following the word “human” refers to a footnote that reads: “This evidence can include traditional cancer epidemiology studies, data from clinical studies, and/or data derived from the study of tissues or cells from humans exposed to the substance in question, which can be useful for evaluating whether a relevant cancer mechanism is operating in humans.”
- 26 National Toxicology Program, op. cit. note 25.
- 27 We do not know the evidentiary standard the NTP uses for “sufficient” evidence, and we also do not know the evidentiary standard for *limited* evidence. We can only presume that it is lower. Ambiguities about the meaning of *sufficient*, *limited*, and similar nonscientific descriptors are discussed beginning below.
- 28 IARC is explicit about this assumption. See International Agency for Research on Cancer, op. cit. note 12, p. 18: “[I]n the absence of adequate data on humans, it is biologically plausible and prudent to regard agents and mixtures for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans.” Assumptions such as this are often called *science policy defaults*, an oxymoron linked to a non sequitur. *Science* and *policy* are sets with no intersecting territory. A *default* is an assumption used when data are not available, but in conventional practice precautionary defaults trump all but the most extraordinary data.
- 29 National Toxicology Program, op. cit. note 25.
- 30 A classic example of guilt-by-association arose in the infamous Salem Witch Trials (1692-93), where accusations of witchcraft against one person in a family quickly extended to other family members, including in one case a four-year old girl. Other aspects of the Trials have eerie parallels. For example, the evidence presented, which qualified as scientific in its day, was distant by “orders of magnitude” from verifiable human experience. Like high-dose animal tests, some confessions were secured through torture. Causality was assumed upon the presentment of “sufficient” evidence in this case, so-called “spectral” evidence from dreams and visions, “witch cakes,” “touch test” experiments, and ostensibly eyewitness testimony. Some accusations appear to have been motivated by politics or preexisting personal animus. Compared to the pace of RoC listings, the trials were models of efficiency, with the period between accusation and execution being as little as three months.
- 31 For our purposes, we can use as an approximation for this latter standard the conventional rule in classical statistics, which requires that the likelihood of rejecting the null hypothesis when it is in fact true be less than 5 percent.
- 32 Each of these problems is exacerbated by the addition within the statutory directive of a highly elastic qualifier. A substance qualifies if it *may be* “reasonably anticipated” to be a carcinogen. When the modal verb *may* precedes the auxiliary verb *be*, it reduces likelihood from certainty to possibility. It is a weaker statement to say that a substance *may be* reasonably anticipated to be a human carcinogen than to say that it *is* reasonably anticipated to be one.
- 33 Normally an experiment in humans would be highly unethical. However, quasi-experiments are routinely performed in humans with chemotherapeutic agents, some of which may cause cancer.
- 34 For the most famous criteria for determining the causality of an association. See Austin Bradford Hill, “The Environment and Disease: Association or Causation?” *Proceedings of the Royal Society of Medicine* Vol. 58, 1965, pp. 295-300. Bradford Hill’s criteria include: strength, consistency of results across independent studies, specificity to the target population, temporality, biological gradient (i.e., dose-response), biological plausibility, coherence with known facts, consistency with experimental evidence, consistency of evidence by analogy. It is surprisingly common to see his criteria cited affirmatively in cases where they support the inference of causality only weakly, or not much at all. It is likely that they are cited more often than his famous paper is read. It is certain that they are difficult to reproduce because they cannot be applied without the exercise of judgment.
- 35 de la Cruz, op. cit. note 22, p. 2 (quoting Rep. Paul G. Rogers [D-Fla.]). Congress considered three reporting thresholds ([1] *known* only, [2] *known* and *reasonably anticipated*, and [3] *known* and *suspected*). It rejected both extremes.
- 36 U.S. Government Printing Office Style Board, 2008, p. 201 (Rule 8.40 on comma usage; “to set off parenthetical words, phrases, or clauses”).

- 37 See Henry Campbell Black, *Black's Law Dictionary*, abridged 5th Edition, St. Paul, Minn.: West Publishing, 1983, p. 822. **“Weight of evidence.** The balance of preponderance of evidence; the inclination of the greater amount of credible evidence, offered in a trial, to support one side of the issue rather than the other. It indicates clearly to the jury that the party having the burden of proof will be entitled to their verdict, if, on weighing the evidence in their minds, they shall find the greater amount of credible evidence sustains the issue which is to be established before them. Weight is not a question of mathematics, but depends on its effect in inducing belief.”
- 38 National Toxicology Program, *Report on Carcinogens Listing Criteria*, op. cit. note 25.
- 39 National Toxicology Program, *Report on Carcinogens*, 12th Edition, Research Triangle Park, N.C.: National Toxicology Program, 2011, p. 3.
- 40 Ibid. The NTP listing criteria imply that Paracelsus was wrong: The molecule makes the poison, not the dose. This is a major source of scientific controversy in the RoC process. Science requires hypotheses that are capable of being refuted. Hypotheses that cannot be refuted are equivalent to dogma—their truth or falsity may be readily apparent to adherents and critics, respectively, but they are immune to challenge using the tools of the scientific method.
- 41 Wendy E. Wagner, “The Science Charade in Toxic Risk Regulation,” *Columbia Law Review*, Vol. 95 No. 7, 1965, pp. 1613-1723. Wagner defines a “science charade” as a case “agencies exaggerate the contributions made by science in setting toxic standards in order to avoid accountability for the underlying policy decisions,” a practice she says “impedes the ability of agencies to promulgate science-based toxic risk standards,” p. 1614.
- 42 Production mass and volume are inherently deficient proxies for exposure. What makes them especially interesting in this context is they imply that regulations promulgated to reduce emissions are completely ineffective at reducing cancer risk.
- 43 Whatever policy choices the NTP made only had to be permissible constructions of the statute. This standard is the second prong of the Supreme Court’s *Chevron* test. See *Chevron U.S.A, Inc. v. Natural Resources Defense Council*, 467 U.S. 837 (1984).
- 44 de la Cruz, op. cit. note 22 (quoting Rep. McGuire and H.R. Rep. No. 95-1192, a precursor bill).
- 45 Although the statutory text is silent, it is logical to infer that the NTP’s estimate of actual exposure was supposed to be unbiased. It would serve little purpose to require the NTP to disclose an estimate for the purpose of setting priorities and simultaneously allow biased estimates to be sufficient. Unless all biased estimates are biased by an identical amount, a rank ordering of biased estimates yields a biased rank ordering.
- 46 National Toxicology Program, *Report on Carcinogens*, 12th Edition, op. cit. note 38, pp. 387-388. The substance profile also reports ambient exposure is below 1 ppb. For comparison, in the principal occupational epidemiological study that the NTP controversially interpreted as showing “substantial” evidence of cancer caused by styrene, “high” exposures were 200,000 times higher (~200 ppm).
- 47 de La Cruz, op. cit. note 22, p. 9.
- 48 In the late 1970s, controversies about the valuation of health effects were far off in the future. The seminal work in the valuation field would be published the next year. See W. Kip Viscusi, *Employment Hazards*. Cambridge, Mass.: Harvard University Press, 1979. Moreover, the discipline of benefit-cost analysis properly applied was widely popular among environmentalists. See Richard L. Berkman and W. Kip Viscusi, *Damming the West*, New York: Grossman Publishers, 1973.
- 49 The potential loss of interagency funding seems likely to be a more likely explanation than mere interagency opposition, which federal agencies routinely learn how to overcome. Another possible explanation is *intra*-agency opposition. To exercise this authority, the NTP would have had to secure approval at several higher levels within the Department of Health and Human Services, no doubt including the Secretary. It might not have been perceived to be in the Department’s overall interest to encourage a small agency buried deep within the National Institutes of Health to be in the business of regularly estimating the health benefits of, say, food additive regulations promulgated by the Food and Drug Administration.
- 50 National Toxicology Program, 2011. Proposed National Toxicology Program (NTP) Review Process for the Report on Carcinogens: Request for Public Comment and Listening Session, 76 FR 67200.
- 51 This scheme was adopted by the Intergovernmental Panel on Climate Change, though there is little evidence that it actually implemented it in a transparent and reproducible manner.
- 52 Office of Management and Budget, “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication,” 67 FR 8452-8460. Department of Health and Human Services, “Guidelines for Ensuring the Quality of Information Disseminated to the Public,” *National Institutes of Health, U.S. Department of Health and Human Services*, last revised December 13, 2006, <http://aspe.hhs.gov/infoquality/Guidelines/NIHinfo2.shtml>.
- 53 See John D. Graham, “Letter to Dr. Elias A. Zerhouni, Director, National Institutes of Health,” Office of Management and Budget, November 16, 2004, http://www.reginfo.gov/public/prompt/nih_ntp111604.pdf. Seven requests for correction (were related to the RoC prior to the Graham letter: [1] anthraquinone (November 17, 2002), [2] nickel (April 11, 2003), [3] styrene (June 12, 2003), [4] naphthalene (April 1, 2004), [5] the RoC’s peer review procedures (June 28, 2004), [6] vinyl chloride (July 1, 2004), and [7] atrazine (July 16, 2004). Subsequent to the Graham letter, a second request for correction was filed with respect to styrene (October 26, 2009). See Department of Health and Human Services, “Information Requests for Corrections and HHS’ Responses,” last revised October 14, 2011, <http://aspe.hhs.gov/infoquality/requests.shtml>.

- 54 National Toxicology Program, *Report on Carcinogens*, 12th Edition, op. cit. note 38, pp. 383-384; 1,3-butadiene was first listed by the NTP as a “known” carcinogen in the 5th RoC (1989).
- 55 *Ibid.*, p. 383.
- 56 Pitot and Dragan, op. cit. note 15.
- 57 Manolis Kogevinas, Gilles Ferro, et al, “Cancer Mortality in a Historical Cohort Study of Workers Exposed to Styrene,” *Scandinavian Journal of Work, Environment & Health* Vol. 20 No. 4, 1994, pp. 251-261.
- 58 National Toxicology Program, *Report on Carcinogens*, 12th Edition, op. cit. note 38, p. 383.
- 59 Numerous others commented as well; the Styrene Information and Research Council is representative and its comments are comprehensive.
- 60 Jack Snyder, “Re: Nomination of Styrene for Review for 12th Report on Carcinogens,” Styrene Information and Research Center, July 16, 2004, <http://ntp.niehs.nih.gov/ntp/NewHomeRoc/RoC12/snyderj-07-16-04.pdf>.
- 61 Jack Snyder, Letter to Ruth M. Lynn. RE: Comments on NTP Report on Carcinogens, Draft Background Document for Styrene, Styrene Information and Research Center, July 7, 2008, http://ntp.niehs.nih.gov/files/Snyder_SIRCJuly7.pdf.
- 62 Jack Snyder, Letter to Ruth M. Lynn RE: Comments on NTP Report on Carcinogens, Styrene Expert Panel’s Listing Status for Styrene & Scientific Justification, Styrene Information and Research Center, October 23, 2008, <http://ntp.niehs.nih.gov/files/SIRC20081023.pdf>.
- 63 Jack Snyder, Comments for the Consideration of the NTP Board of Scientific Counselors on (1) Draft Substance Profile for Styrene and (2) Deficiencies in NTP’s Review of Styrene, Styrene Information and Research Center, February 6, 2009, <http://ntp.niehs.nih.gov/files/20090206SIRC.pdf>.
- 64 James S. Bus and George Cruzan, “Letter to NTP Director Linda S. Birnbaum, RE: Finalization of NTP Report on Carcinogens Review of Styrene,” Styrene Information and Research Center, April 29, 2009, <http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/june/SIRC20090429.pdf>. Bus was president of the Society of Toxicology in 1996-97; Birnbaum was president in 2004-05.
- 65 Marcy Banton, “NAS Formaldehyde Report Supports SIRC Concerns with Deficiencies of NTP Report on Carcinogens Process,” Styrene Information and Research Center, May 16, 2011, <http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2010/PublicComms/SIRC20110516.pdf>.
- 66 Cruzan, Bus, et al, “Mouse Specific Lung Tumors from cyp2f2-Mediated Cytotoxic Metabolism: An Endpoint/Toxic Response Where Data from Multiple Chemicals Converge to Support a Mode of Action,” *ToxWorks*, January 27, 2009, http://ntp.niehs.nih.gov/files/SIRC_RTP-S-09-000231.pdf. European Union Chemicals Agency, *European Union Risk Assessment Report: Styrene*, European Union Chemical Agency, June 2008, http://echa.europa.eu/doc/trd_substances/styrene/rar/trd_rar_uk_styrene.pdf.
- 67 Elizabeth Delzell, “Comments on the RoC Background Document for Styrene and Related Documents,” October 10, 2008, <http://ntp.niehs.nih.gov/files/SIRC20081023A.pdf>. and Delzell, “Letter to Barbara Shane, Executive Secretary, National Toxicology Program Board of Scientific Counselors,” February 5, 2009, <http://ntp.niehs.nih.gov/files/20090205Delzell.pdf>.
- 68 Kogevinas, et al., op. cit. note 55, p. 260: “[T]hese findings leave open the question of whether an excess risk of neoplasms of the lymphatic and hematopoietic tissues among workers exposed to styrene.”
- 69 Banton, op. cit. note 63; Cruzan, et al., op. cit. note 64; Bus and Cruzan, op. cit. note 62.
- 70 Ironically, the particular knock-out mouse used in this study was developed by a researcher funded by the National Institute of Environmental Health Sciences, the NTP’s immediate parent organization. The NTP ignored, with equal vigor, industry- and agency-funded science that contradicted its predetermined conclusions.
- 71 Proof beyond reasonable doubt that positive data from animals are irrelevant to humans may not be enough. During BSC review, one member is reported to have acknowledged that the NTP’s definition of a *reasonably anticipated* human carcinogen is “self-contradictory, ambiguous, and difficult to implement in a transparent and objective way”; that “the definition relies too heavily on the word ‘credible’ as it introduces a large, subjective element into the decision process”; and that “much of the controversy reflects the diversity in understanding the definition of limited evidence.” Contrary to the BSC’s charge, which forbids the panel from questioning the scientific merit of the NTP’s definitions, this panel member is reported to have “thought it might be worthwhile to revisit the definitions to address both the scientific challenge and the public health values of the RoC.” See National Toxicology Program Board of Scientific Counselors, *Summary Minutes for February 24, 2009 Meeting*, p. 27, http://ntp.niehs.nih.gov/ntp/About_NTP/BSC/2009/February/minutes022409.pdf. Other BSC members accepted the NTP’s definitions without complaint and appear to have interpreted *reasonably anticipated* to mean that carcinogenicity in humans could not be ruled out. *Ibid.*, pp. 24-32. Contemporaneous references to the meeting by others cite remarks with similar themes that are not reported in the meeting minutes (Bus and Cruzan, op. cit. note 64, p. 2.), so questions about the accuracy and completeness of the minutes when compared to a verbatim transcript must be taken into account when evaluating the published report of BSC reviews.
- 72 Banton, op. cit. note 63; Snyder, op. cit. notes 60 and 61; Jack Snyder, Letter to Samuel H. Wilson, Styrene Information and Research Center, June 13, 2008, http://ntp.niehs.nih.gov/files/SIRC_NTP_06-13-08_508sj2.pdf. Jack Snyder, Letter to Linda S. Birnbaum RE: NTP Report on Carcinogen Process Failures on the Use of Publicly Available, Peer-Reviewed Evidence, Styrene Information and Research Center, October 22, 2011, <http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/Styrene/SIRC20091022.pdf>.

About the Author

Dr. Richard B. Belzer is President of Regulatory Checkbook, a Virginia-based nonprofit corporation whose twofold mission is to improve the quality of scientific and economic information used in regulatory decision making and monitor agency compliance with laws and executive directives intended to accomplish this goal. Regulatory Checkbook does not take positions on substantive policy matters.

From 1988 to 1998, Dr. Belzer served as an economist in the Office of Management and Budget's Office of Information and Regulatory Affairs, where he earned multiple performance awards. He was twice elected Treasurer of the Society for Risk Analysis (1998, 2000) and Secretary/Treasurer of the Society for Benefit Cost Analysis (2008, 2010). He is also a Fellow of the Cecil and Ida Green Center for the Study of Science and Society. He earned his doctorate from Harvard University in 1989.

The Competitive Enterprise Institute is a non-profit public policy organization dedicated to the principles of free enterprise and limited government. We believe that consumers are best helped not by government regulation but by being allowed to make their own choices in a free marketplace. Since its founding in 1984, CEI has grown into an influential Washington institution.

We are nationally recognized as a leading voice on a broad range of regulatory issues ranging from environmental laws to antitrust policy to regulatory risk. CEI is not a traditional “think tank.” We frequently produce groundbreaking research on regulatory issues, but our work does not stop there. It is not enough to simply identify and articulate solutions to public policy problems; it is also necessary to defend and promote those solutions. For that reason, we are actively engaged in many phases of the public policy debate.

We reach out to the public and the media to ensure that our ideas are heard, work with policymakers to ensure that they are implemented and, when necessary, take our arguments to court to ensure the law is upheld. This “full service approach” to public policy makes us an effective and powerful force for economic freedom.



Competitive Enterprise Institute

1899 L Street, NW
12th Floor
Washington, DC 20036
202-331-1010
Fax 202-331-0640
www.cei.org

Issue Analysis is a series of policy studies published by the Competitive Enterprise Institute. Nothing in *Issue Analysis* should be construed as necessarily reflecting the views of CEI or as an attempt to aid or hinder the passage of any bill before Congress. Contact CEI for reprint permission. Additional copies of *Issue Analysis* may be purchased through CEI's publications department (pubs@cei.org or 202-331-1010).