asrc

Working Paper

The *Report on Carcinogens*: What Went Wrong; What Can Be Done to Fix It

Richard B. Belzer
Belzer@RegulatoryCheckbook.org

Regulatory Checkbook
PO Box 319
Mount Vernon, VA 22121
(703) 780-1850

Revised: November 3, 2011
Comments welcome

REGULATORY CHECKBOOK

I. Executive Summary

There is nothing in principle wrong 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 fatal than they were when the "war on cancer" was announced by President Richard Nixon in 1971 (Kohler et al. 2011).

This paper documents how the *Report on Carcinogens* (RoC) has failed to live up to what Congress intended. It was supposed to provide an objective reference that could be used widely to inform rational decision-making by individuals, families, and governments alike. It has become a tendentious, unresolvable argument between two competing schools of thought.

The first school says that the purpose of cancer risk assessment is to estimate cancer risk, thereby providing useful insights about cancer etiology, the conditions under which exposure to various substances initiate or promote cancer, and scientifically-based estimates of relative potency and human risk, all of which could be used to enable families and government officials alike to make rational decisions balancing risks and benefits. To accomplish this purpose, RoC listings would be based on the most objectively possible characterization of scientific knowledge relevant to actual human exposure.

The second school says that the purpose of cancer risk assessment it to manage potential certain risks in a highly precautionary manner. To accomplish this purpose, RoC listings would be based on near worst-case circumstances so that potential threats to human health are rarely, if ever, permitted to occur, irrespective of their relative magnitude or whether these risks are accompanied by human benefit. New scientific knowledge suggesting that a substance isn't an important human carcinogen at environmentally relevant doses would be subject to very demanding tests of proof.

In the RoC, the second school has clearly prevailed, in a rout. The criteria the NTP uses to decide whether to list substances in the RoC are designed and implemented to ensure that substances are listed unless the absence of carcinogenic effects can be assured, perhaps beyond any reasonable doubt. Listing in the RoC sets in

motion numerous precautionary default risk management rules and procedures at the international, federal and state level.

This has several adverse effects that are not usually accounted for. For example, markets respond quickly and often ruthlessly when place substances tagged by the RoC as carcinogens. To avoid regulatory burdens and potential tort liability, often they are removed as inputs in products and industrial processes. Sometimes this may lead to an unexpected gain in both efficiency and reduced public health risk, but neither outcome is either guaranteed or known to be empirically common. Input substitutions and process changes very often lead to higher consumer prices, and often result in lower product quality or fewer desirable product attributes. Input substitution can perversely cause cancer risk to increase if, for example, the carcinogenic effects of the replacement substance are unknown. Meanwhile, because the RoC is only a labeling exercise, it virtually no value for estimating cancer risk or informing risk-based decision-making. It is an empirical question, to date unanswered, whether the benefits to society of the RoC justify its costs.

This paper concludes with specific recommendations for statutory and administrative reforms of the RoC process that would improve both its scientific quality and its practical utility for rational risk-benefit decision-making. Some reforms only Congress can make; others are within the discretion of the NTP Director. Given the NTP's recent history, it is not clear that the NTP is culturally capable of reform without congressional action.

II. Introduction

The NTP was established in November 1978 by the Secretary of the Department of Health and Human Services for the purpose of conducting laboratory tests of chemicals primarily for carcinogenicity (Office of Technology Assessment 1987, 147). These studies were not designed for or intended to inform regulatory decision-making. Separately, the NTP was assigned the responsibility of producing an annual (subsequently biennial) RoC. NTP has now issued 12 such reports, the l11th in 2004 and the 12th in 2011. The NTP's inability to adhere to the statutory schedule is partly the result of technical complexity, to which the law is insensitive, and the extent to which each RoC has become highly controversial. The fact that seven years passed between the last two biennial reports strongly suggests that

there is a wide gap between what Congress intended the RoC to be and what the NTP produces. This paper argues Congress (which legislated ambiguously) and the NTP (which implemented this ambiguous statutory language in politicized ways that drained it of scientific legitimacy) share responsibility for failure.

Nominations for listing or delisting are provided by federal agencies, though now the public may contribute suggestions. Decisions concerning which nominations to advance rest solely with the government (National Toxicology Program 2011h). The law provides the NTP with two statutory categories: "known human carcinogen" and "reasonably anticipated to be a human carcinogen." Though the law called for annual (and then biennial) reporting, implying that listings would change as warranted by new scientific information, in practice listing decisions tend to be permanent. The hurdle for delisting appears to be exceedingly high.[1]

If the purpose of the NTP's laboratory studies was exploratory, the purpose of the *RoC* has always been to provide a plausibly scientific justification for regulation (de la Cruz 2009, 171; Office of Technology Assessment 1987). Yet each substance in the *RoC* is assigned to a category defined with probabilistic terms that lacking probabilistic definition, classified based on non-transparent criteria, and often founded on a controversially selective database. For these reasons, and many others, the *RoC* will continue to be a highly controversial document produced through a process that is

[1] 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 7th 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 (2002) revised its classification downward to Group 3 ("inadequate evidence in humans"; "limited evidence in experimental animals"). 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, was more than 20 years' long. Moreover, the NTP did not delist the substance so much as change its definition to exclude fiberglass.

incompatible with the era of modern risk analysis that began roughly in 1983 (National Research Council 1983) and the information quality era that began in 2002 (Office of Management and Budget 2002)

The chief opponents of the RoC process and its outcomes are those businesses and business trade organizations involved in the production or use of target chemicals. This is an inevitable result of how the NTP implemented its statutory charge, which was to err strongly in favor of over-classification (and hence to support over-regulation). Had the NTP implemented the law to err strongly in favor of under-classification (and hence to support under-regulation), the chief opponents would be environmental and public health activist groups. Science—the pursuit of knowledge about what is—is the clear casualty of either bias, as bias reduces science to a weapon in the pursuit of policy—the pursuit of what ought to be. An interesting question to consider is where the various actors would stand if the RoC process were revised to be strictly scientific and scrupulously neutral with respect to risk management preferences. Could competing interests be persuaded to allow science to answer only those questions for which it is best suited, and to transfer policy disputes to more appropriate forums?

This paper consists of five major sections. Section III provides background on the NTP's program of laboratory testing. It shows that the practice of making strong policy inferences on the basis of limited scientific evidence predates the NTP by decades. Section IV describes the statutory directive the NTP is supposed to implement to produce the RoC. It includes important provisions that the NTP ignores or misinterprets. Section V describes how the NTP implemented the law in a manner that ensures many false positives in order to avoid false negatives.[2] Section VI provides cases studies of naphthalene, styrene, and formaldehyde. Being recent, they should illustrate the best performance that the public can expect from the NTP, but the evidence is convincing that public expectations should be modest. The naphthalene case shows that the NTP does not keep abreast of scientific advancements that affect the validity of its prior decisions. The styrene and formaldehyde cases show that the NTP is highly resistant to scientific information that contradicts its predetermined

[2] In this context, *false positives* consist of deeming a substance a human carcinogen when it isn't; *false negatives* consist of failing to deem a substance a human carcinogen when it is.

conclusions. The formaldehyde case provides supporting evidence for the hypothesis that listing decisions are actually made by the government at the nomination stage, not after the scientific evidence has been collected, analyzed, and weighed.

Section VII concludes the paper with an array of recommendations for reform premised on the assumption that the RoC ought to adhere to high standards of scientific quality and have practical utility for rational risk management decision-making. Several of these recommendations are directed to Congress because they require statutory changes. Others are offered to the NTP, however, whose leadership surely understands the problems described in this report and may be interested in engaging in serious reform, if for no other reason to discourage Congress from asking whether the benefits of producing the RoC justify the costs.

III. The Statutory Antecedents of the Report on Carcinogens

The Department of Health and Human Services (DHHS) 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, typically rodents, conducted under the auspices of DHHS' National Cancer Institute (NCI) since 1961.[3] These functions moved to the new National Toxicology Program (NTP) in 1978 (Office of Technology Assessment 1987, 16), where it was designed to be a profit center within the federal government.[4]

A. How the NTP Interprets the Results of Its Laboratory Experiments

[3] 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."

[4] 42 U.S.C. § 241(b)(1): "In carrying out this paragraph, the Secretary shall consult with entities of the Federal Government, outside of the Department of Health and Human Services, engaged in comparable activities. The Secretary, upon request of such an entity and under appropriate arrangements for the payment of expenses, may conduct for such entity studies and testing of substances for carcinogenicity, teratogenicity, mutagenicity, and other harmful biological effects." A similar profit center also was established in support of human nutrition research. See 42 U.S.C. § 241(b)(3).

Figure 1: NTP's Definition of Carcinogenicity Results

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 because of major qualitative or quantitative limitations cannot be interpreted as valid for showing either the presence or absence of carcinogenic activity.

Source: National Toxicology Program (2005b).

The NTP classifies the results of its carcinogenicity experiments in a format that implies, but does not actually include, scientific content of a probabilistic nature, as shown in Figure 1. Each category contains scientific information (e.g., "studies … showing a dose-related increase of malignant neoplasms"), but the NTP couches each fragment of scientific information with ambiguous, non-scientific caveats (e.g., "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 institutional interests of the NTP.

The descriptors for these categories also have ordinal or metric qualities that imply probabilistic content. Mathematically, "clear evidence" is superior to "some evidence," which is superior to "equivocal evidence," which is superior to "no evidence." But the lines

dividing these categories are murky at best, and wholly subject to taste, politics, and agency bureaucratic interests. In short, it is a myth that carcinogen classifications are based on science. Rather, they are based on policy judgments only roughly constrained by science.

At the outset, these judgments may well have varied substantially across toxicologists. Over time, however, as a record was built of how studies had been interpreted, this record became the functional equivalent of legal precedent. If the output of an experiment on Substance H (e.g., a chronic 2-year rodent bioassay[5]), Substance H looked more like Substances A, D, F, and G than like Substances B, C, and E, it was virtually certain to be assigned the same classification as the former and not the latter. In this way, assignments developed a pattern of consistency that, over time, has become confused with scientific accuracy.[6]

Finally, it should be noted dose—the key determinant of human health risk[7]—has a severely circumscribed role in how test results are described. A laboratory experiment that shows carcinogenic "activity" in a rodent at doses thousands of times greater than human

[5] The chronic 2-year bioassay is a 104-week controlled laboratory experiment in which different groups mice or rats are exposed to two or three different doses of a test agent, with another group unexposed to serve as a control. At the end of the 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 (e.g., potent genotoxicity) and mechanisms that are not (e.g., cancer subsequent to organ toxicity resulting from frank poisoning).

[6] It should be noted that the variable being scaled is actually quite broad. It is not cancer that is being measured 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".

[7] 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."

experience earns the descriptor "clear evidence of carcinogenicity" but may have no contribution to human cancer incidence.[8]

B. The Chronic 2-Year Bioassay 40 Years Later

The chronic 2-year bioassay has been called the "gold standard" for toxicology, but it appears that this endorsement has more to do with tradition than the value of the scientific information they produce. The highest dose in a bioassay is called the Maximum Tolerated Dose (MTD).[9] As far back as 1979, at least one senior government official expressed the hope that better tests would begin to replace it by 1985 (Office of Technology Assessment 1987, 147). In 1993, a committee of the National Research Council raised profound doubts about the scientific value of studies based on the MTD because they yield produce so many false positives (National Research Council 1993).[10] More recently, there has been an earnest appeal for a fundamental change in direction (National Research Council 2007).

Changing direction will be difficult for several reasons. Chronic 2-year bioassays have become a cottage industry for the NTP and numerous private contractors, with several guaranteed markets including the NTP's own RoC.[11] Rentseeking aside, there is

[8] To be clear, 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 (2006a, 18)

[9] The "maximum tolerated dose" is defined "[t]he highest dose of a drug or treatment that does not cause unacceptable side effects" (National Cancer Institute 2011). The definition of "unacceptable" varies (Eaton and Klaassen 2001, 29).

[10] In the "Bible" of modern toxicology, the design and conduct of the chronic 2-year bioassay is described with at least as much attention devoted to its scientific limitations. The conclusion is one of undisguised resignation: "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." See Pitot III and Dragan (2001, 294).

[11] The inability of government laboratories to shoulder the original load is recounted by the Office of Technology Assessment (1987, 148-149).

considerable support for maintaining the existing protocol because hundreds of substances have been subjected to it, and if the protocol were substantially changed, comparisons with this historical record would be problematic. Established practice also may be defended to avoid having to confront the question whether past expenditures were worthwhile. Meanwhile, the NTP testing laboratory operates as both a monopoly producer and the dominant consumer of its own products. Almost any positive result in a chronic 2-year bioassay means that a substance satisfies the RoC's criteria for being deemed "reasonably anticipated to be" a human carcinogen. No other data are needed to complete this showing, and it appears that there are no data, including human data, that could be produced to rebut it.[12]

For this reason, the chronic 2-year bioassay appears to be in no danger of being replaced by higher quality science that predicts human cancer risk with greater selectivity. In 1979, when the end of the chronic 2-year bioassay was "just around the corner," the NTP had conducted about 190 such studies. As of September 15, 2011, it has conducted another 400 more (National Toxicology Program 2011d). The business of conducting these studies has become thoroughly entrenched—financially, bureaucratically, and intellectually—so the number of chronic 2-year bioassays performed is limited only by federal appropriations and regulatory requirements.

IV. The Statutory Design of the Report on Carcinogens

Many things could be responsible for cancer, but the statutory text requires the NTP to adopt a highly selective view of cancer etiology. Only "substances" (predominantly man-made chemicals) matter. To the extent that cancer is primarily a byproduct of DNA mutations due primarily to aging, this is not relevant.[13] The RoC is a tool for establishing a scientific-looking predicate for regulating

[12] In Section V, it is shown that the RoC listing criteria provide no way for the absence of cancer in humans to overcome a determination based on animal data that a substance is "reasonably anticipated to be" a human carcinogen.

[13] 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 de la Cruz (2009)

chemicals; it is not about learning what actually causes cancer in humans.

That the chronic 2-year bioassay evolved to become a quasi-regulatory instrument is clear from contemporaneous reports about the NTP's founding and original mission. These laboratory tests were devised to serve an exploratory scientific purpose, but the establishment of the NTP and the requirement to produce the RoC meant that they would be used for the much more demanding purpose of guiding (if not controlling) regulatory decision-making:

> In the 1960s, government agencies, especially the National Cancer Institute (NCI), used animal tests to predict carcinogenicity, though at first to learn more about the relation between chemical structure and carcinogenicity and not for regulatory purposes (Office of Technology Assessment 1987, 147).

Today, though study design is largely unchanged, the NTP does not even discuss the limited scientific utility of its laboratory studies. Rather, the agency promotes how other agencies have used NTP laboratory studies to justify dozens of regulatory standards (National Toxicology Program 2011b).

This is but one example in a regular pattern of information quality abuse. Low-resolution experiments are funded strictly for exploratory or methodological purposes. Positive results were then stretched just beyond what the original data quality could support, such as for the ostensibly limited purpose of providing a "virtually safe dose"[14]—an amount below which risk is not zero but is widely agreed

[14] Faustmann and Omenn (2001, 95) define the virtually safe dose (VSD) as "the lower 95 percent confidence limit on a dose that gives an 'acceptable level' of risk (e.g., upper confidence limit for 10^{-6} excess risk)". They go on to comment that VSDs "are believed to represent conservative, protective estimates." That is, the VSD includes, in a non-transparent way, the policy judgments of its designers concerning (1) what is an "acceptable risk" (i.e., one excess cancer case per million lifetimes) and (2) the amount of precaution that individuals, households, and public and private risk managers ought to take with respect to avoiding or preventing such risks (i.e., the difference between the 95 percent confidence interval and the unbiased best estimate, which is never reported).

to be "low enough to be safe."[15] Once familiarity with VSDs had been established, however, their intentionally embedded biases were forgotten, allowing low-resolution experiments to be used to characterize human cancer risk and become the basis for regulatory standard-setting and regulatory benefit-cost analysis.[16]

This is how the RoC operates. The chronic 2-year bioassay—a low-resolution laboratory experiment—yields evidence of "carcinogenic activity" in animals. This evidence is assumed to be relevant to humans at the same very high doses, then it is further assumed that effects are proportional to dose absent proof that it is not. Under the NTP's definition, the substance automatically becomes "reasonably anticipated to be" a human carcinogen. The RoC is controversial precisely because this model disregards scientific knowledge about how risk varies by dose ands how humans differ from animals. The resulting product has little or no utility for objectively estimating human cancer risk at environmentally or occupationally relevant doses, a fact the NTP acknowledges in the preamble to the RoC but not in the substance profiles.[17]

[15] Like "sufficient" evidence, the "virtually safe dose" is another attempt to make policy preferences look scientific.

[16] To be valid, benefit-cost analysis requires unbiased estimates of all input parameters. When benefits consist of avoided cancer risks, benefit estimates are automatically biased upward by an amount equal to the unknown difference between the expected value and 95 percent upper-bound unit risk estimates, *even if every other assumption, datum, and model specification is unbiased.*

[17] The preamble states: "The RoC does not present quantitative assessments of the risks of cancer associated with these substances. Thus, the 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." As for what purpose the RoC actually serves, the preamble describes it as "an informational scientific and public health document that identifies and discusses [substances] that may pose a hazard to human health by virtue of their carcinogenicity." See National Toxicology Program (2011e, 3, emphasis added). The combination "scientific and public health" means the NTP knows that the RoC consists of policy judgments wrapped in science.

Figure 2: The RoC Definition of Human Carcinogen Must Include Both Hazard and Exposure

> (4) 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…

A. The Statutory Threshold for Classification as a Human Carcinogen

The relevant statutory text consists of three clauses. Clause (A) sets forth a two-part threshold for classifying a substance as a human carcinogen in the RoC; this is reproduced in Figure 2. To be so classified, a substance must be (1) either known or reasonably anticipated to be a carcinogen and (2) a significant number of persons residing in the United States must be exposed to it. These statutory tests are difficult to implement scientifically because they contain crucial non-scientific language. However, they are easy to implement in a way that looks scientific to non-scientists.

1. Clause (A), Part (i): What does it mean to be a "known" human carcinogen?

To be a "known" human carcinogen implies a strong, specific, and selective statistical association in humans supported by clear biological theory and evidence relevant to humans. Such an association would have all of the attributes one would expect from an implication of virtual scientific certainty.[18] Except in the extraordinary

[18] The most famous criteria for determining the causality of an association were published by Bradford Hill (1965): strength, consistency of results across independent studies, specificity to the target population, temporality, biological gradient (i.e., dose-response), biological plausibility,

circumstance of high-quality human data from a controlled experiment,[19] multiple high-quality human studies showing robust results would be required before most independent scientists would consider drawing the inference that an observed relationship is causal.

It also implies nothing that would counsel doubt or skepticism, such as the presence of contradictory evidence or important limitations in study design or data quality. That is, there cannot be any studies of similar or greater quality showing that the observed cancer incidence has other identified or scientifically plausible origins, nor can there be high-quality studies that show the absence of a carcinogenic effect from the substance of interest. In short, to be scientifically deemed a "known" carcinogen, the scientific database must contain strong evidence refuting the null hypothesis of no-effect; this evidence must meet very high internal quality standards; and it must be compellingly consistent.

While there is no scientific way to define this threshold, nothing prevented the NTP from defining it in an arbitrary (but sensible) way that would make classification decisions transparent and reproducible. Moreover, the NTP could have set this threshold high enough that material scientific controversies were exceedingly rare. To ensure that classification decisions were based on science, the NTP could have defined "known" human carcinogens such any determinations was always subject to refutation if confronted with contrary new evidence.[20] That is, the NTP could have ensued that there was always

---

coherence with known facts, consistency with experimental evidence, consistency of evidence by analogy. It is surprisingly common to see Hill's criteria cited affirmatively in cases where they support the inference of causality only weakly, or not much at all. It is likely that Bradford Hill's criteria are cited more often than his famous paper is read. It is certain that Bradford Hill's criteria are difficult to reproduce because they cannot be applied without the exercise of judgment.

[19] 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.

[20] If there were no way that evidence could move a substance out of the "known" category, then the assignment would be a matter of faith rather than science.

a hypothesis that, if refuted, would cause the "known" classification to fall away and for the NTP to happily abandon it.[21]

2. Clause (A), Part (i): What does it mean to be "reasonably anticipated to be" a human carcinogen?

Wherever the threshold for a "known" human carcinogen is set, we can be sure that the threshold for being "reasonably anticipated to be" a human carcinogen is lower. Indeed, the legislative history of the statute provides useful insights concerning what Congress intended. First, the term "suspected" carcinogen has been used in an earlier bill, but in the final bill this was replaced with "reasonably anticipated to be" a carcinogen "in order to make it absolutely clear in the statute that there must be reasonable ground for designating a substance as a putative carcinogen" (de la Cruz 2009, 2). The definition of "reasonableness," of course, is eternally elusive and Congress did not give additional; guidance concerning what it considered to be "reasonable."[22]

As in the case of "known" human carcinogens, it is essential that evidence from human data show a strong, robust and consistent association. Where the threshold is less stringent is in the toleration of equivocal and contrary evidence. Some equivocal evidence is tolerable, provided it is from studies of lower quality, but there should be little or no contrary evidence.

A harder question is what to do when environmentally or occupationally relevant human data are not available or are inadequate. Science provides no answers. The default practice has been to simply assume that substances which cause cancer in animals at high doses also cause cancer in humans at low doses (International Agency for Research on Cancer 2006b, 18). While there are some instances in which this extrapolation across dose and species is supported by scientific evidence, there are many instances in which there is little such evidence, or even evidence indicating that the

[21] As Section V.B shows, the NTP has never elucidated any way for a listing decision to be scientifically refuted.

[22] Congress considered three reporting thresholds ("known" only, "known" and "reasonably anticipated," and "known" and "suspected"). It rejected both extremes. A fair description of the way the NTP has implemented the statute is it has interpreted "reasonably anticipated to be" as the same thing as "suspected."

assumption is likely to be false. Assumptions are not scientific data; they are used in lieu of scientific data, so assumptions alone cannot legitimately result in a scientific determination that a substance that causes cancer in animals is "reasonably anticipated to be" a human carcinogen.

Indeed, the phrase "reasonably anticipated to be" has limited scientific content. It cannot be operationalized without a weight-of-evidence (WoE) scheme. But all WoE schemes require the exercise of judgment, so if classification is to be scientific then the scheme must be strictly scientific. That is, it must be designed and implemented so as to exclude judgments of a nonscientific nature, such as scientists' risk management preferences, an agency's bureaucratic interests, and the like. It also must be transparent (i.e., all material aspects must be fully disclosed) and reproducible (i.e., it must yield similar outcomes for different situations with similar facts).[23] A WoE scheme that is not transparent or reproducible has no scientific credibility.[24]

3. Clause (A), Part (ii): What does it mean for a significant number of persons residing in the United States to be exposed?

Risk requires both hazard and exposure. Having tackled the hazard question, and classified substances that are "known" to be human carcinogens or "reasonably anticipated to be" human carcinogens, the NTP now must investigate the extent of human exposure in the United States because the list must include only those substances to which a significant number of persons residing in the United States are exposed. Because exposure means the ratio of mass or volume per unit of time (or the cumulative amount over a time period), neither mass nor volume by itself qualifies as an exposure unit. The NTP also cannot rely on historical data (e.g., "persons who were exposed"), or figures of a hypothetical nature (e.g., "persons

[23] Each of these problems is exacerbated by the addition within the statutory directive of an infinitely elastic qualifier. A substance qualifies if it may be "reasonably anticipated" to be a carcinogen. Taken literally, this qualifier could be abused to expand the domain so that hardly anything is not "reasonably anticipated."

[24] As Section V shows, the NTP has never elucidated a WoE scheme for making listing decisions.

who may be exposed"), or on figures obtained elsewhere (e.g., "persons who are exposed in China").

This requires describing the exposure distribution, which can be done scientifically, and defining the term "significant," which cannot. A faithful implementation of the law involves performing the first task and setting clearly defined and intuitively reasonable policy thresholds for "significant." In the RoC, the NTP does not describe the distribution of exposure in the United States, nor has it ever stated policy thresholds for the meaning of "significant number of persons". The NTP generally relies on mass, volume, and historical or hypothetical figures.in lieu of what the law requires.

B. Other Information that Must Be Included in the Report

For each substance listed, the statute directs the NTP to report an estimate of the number of persons exposed and the cancer-reduction benefits achieved by federal regulation. The statute also directs the NTP to flag any substance listed for which there may be a gap in federal regulatory standards.

1. Estimates of the number of persons exposed

For each substance that makes the list, Congress directed the NTP to estimate the number of persons exposed and provide information about the "nature" of their exposure, as shown in Figure 3. The requirement to quantify the number of persons exposed serves a critical purpose, which is to ensure that the NTP focuses on big problems and isn't distracted by minutiae. The NTP has the discretion to decide how many persons qualify as a "significant" number, but it must does not have the discretion to keep this figure a secret.

Figure 3: Substances Listed in the RoC Must Be Accompanied by Exposure Data

> (4) The Secretary shall publish a biennial report which contains—
>
> …
>
> (B) information concerning the nature of such exposure and the estimated number of persons exposed to such substances…

In any event, the legislative history supports a broad interpretation of this task. Congress intended for the NTP to quantify the numbers of persons exposed even if it did not direct it to estimate individual exposure. Although the statutory text is silent, it is logical to infer that the estimate that the NTP is directed to produce should be unbiased. It serves little purpose to require the NTP to disclose an estimate and simultaneously allow any estimate to be sufficient.

Somewhat greater uncertainty surrounds what Congress intended when it sought information about the "nature" of exposure. Given that the language was written in the late 1970s, however, the two most important "natures" would have been the environmental and occupational domains. This is consistent with Congress' clear focus on human cancer risk. No cancer risk exists in the absence of exposure. Hazards have limited relevance to human cancer risk if they require doses that do not occur in environmental or occupational settings.

2. Estimation of the cancer-reduction benefits from federal regulation

For each substance listed that has one or more federal regulatory standards, the statute directs the NTP to report the extent to which regulation has "decreased the risk to public health." As shown in Figure 4, the NTP is permitted to include health effects other than cancer.

The importance of this provision is clear from the legislative history. Congress specifically wanted the RoC to include "where possible, estimates [of] the magnitude of the risk each [substance] poses" (de la Cruz 2009, 9).

The text clearly includes the quantification of health effects cases, but it also can be read to permit the monetization of the value of avoiding cancer. 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 quantification of health effects. 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.[25]

[25] 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 Viscusi (1979) Moreover, the discipline of

Figure 4: Estimation of Regulatory Benefits

> (4) The Secretary shall publish a biennial report which contains—
>
> …
>
> (C) a statement identifying—
>
> …
>
> (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…

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 sister agencies such as the Environmental Protection Agency (EPA) and the Occupational Safety and Health Administration (OSHA). Nonetheless, the statutory text clearly provided an 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 given the invitation.

But the NTP did not exercise its statutory authority to establish, build, and sustain these new missions. One possible explanation is that the NTP was (and remains somewhat) a virtual agency, funded and staffed mostly by the National Institute of Environmental Health Sciences (NIEHS) but with significant voluntary contributions from other agencies. This structural design could have prevented the NTP from taking expanding its mission into areas already occupied by other agencies.[26]

---

benefit-cost analysis properly applied was widely popular among environmentalists; see Berkman and Viscusi (1973).

[26] The potential loss of interagency funding seems likely to be a more likely explanation than mere interagency opposition. Also, to exercise this authority, the NTP would have had to secure approval at several higher levels

3. Identification of potential gaps in federal regulation

The statute also directs the NTP to flag substances on the list for which there are no federal regulatory standards, as shown in Figure 5. The NTP 's role is to identify potential gaps in regulatory coverage; it is not to opine on the merits or desirability of such standards.

The NTP does not devote any significant effort to this task. Substance profiles contain sections listing applicable regulations and guidelines. They do not identify gaps in federal regulation.

V. How the NTP Implemented Its Statutory Authority

Statutes rarely implement themselves, and agencies typically follow certain established practices when implementing new statutory authorities.

A. The NTP Grafted its Existing Testing Framework into the RoC

One predictable pattern is to incorporate new authority within an existing mission or regulatory framework. The NTP grafted into the new RoC authority its established scheme for classifying the results of laboratory tests, shown above in Figure 1. This had obvious bureaucratic benefits, such as dramatically increasing the scope of the

Figure 5: Identification of Regulatory Gaps

> (4) The Secretary shall publish a biennial report which contains—
>
> …
>
> (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…

within DHHS. It might not have been perceived to be in DHHS' 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 additives regulations promulgated by the Food and Drug Administration.

market for its laboratory tests. It also allowed the NTP to ensure its own laboratory studies secured favorable treatment, if not unrebuttable priority, in its (undisclosed) WoE scheme.[27]

A. Where in the RoC Does Science End and Policy Begin?

A substance might be considered a "known" carcinogen if the vast majority of scientists believe that it is, but any such agreement is a statement about shared subjective probabilities or policy preferences rather than scientific fact. To "reasonably anticipate" a future effect is to make a statement with an undefined but lesser probability. Does a probability of 51% qualify as "reasonably anticipated," or must the probability exceed 90%? 95%? More? Do the circumstances matter? For example, would it be appropriate to designate a substance "reasonably anticipated" to be a human carcinogen if the probability was negligible at environmentally relevant doses but exceeded 95% at a dose one million times greater?

Because "known" and "reasonably be anticipated" are not scientific terms, they cannot be determined through the application of science. Congress may have intended the NTP to obtain scientific determinations, but in fact the text invites the NTP to subordinate science to policy. Of course, the NTP could have structured the assignment process in ways that, if not scientific, would at least have been transparent in the way they used science and exercised policy discretion. Instead, as shown below, the NTP adopted what Wagner (1995) calls a "science charade." Outcomes are made to appear as if they are founded on science but in fact are grounded on undisclosed policy judgments and ratified by scientific experts who are required to conform their scientific determinations to these policy judgments.

[27] Though beyond the scope of this paper, this hypothesis is testable. Results from NTP bioassays are often difficult to interpret or even highly controversial. However, the RoC could be examined systematically to determine if there are any instances in which the NTP allowed one of its own positive laboratory tests to be rebutted with other evidence, particularly strong evidence of no effects from human studies. If it has not, then it can be inferred that the NTP has an inherent conflict of interest in performing the RoC function at the same time that it performs laboratory tests.

B. The NTP's Definition of a "Known" human Carcinogen

The NTP interpreted this statutory directive in ways that appear to be highly inconsistent with the text of the law. Faced with the task of having to define criteria for classification that incorporate the usual attributes of causality (such as set forth by Bradford Hill (1965)) it instead devised criteria, shown in Figure 6, that are merely tautological.

Whereas the statutory text implies the establishment of a casual relationship between a listed substance and human cancer, the NTP's definition skirts the question of causality entirely. 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 presumed to "indicate[s] a causal relationship."

Figure 6: NTP's Definition of a "Known" Human Carcinogen

> 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.
>
> *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.

This becomes clear when the sentence construction is examined carefully. 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 (U.S. Government Printing Office Style Board 2008, 201, Rule 8.40 on comma usage; "to set off parenthetic words, phrases, or clauses"). Thus, the full meaning of the "known" human carcinogen criterion is 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.

Substances that the NTP defines as "known" human carcinogens must per se have sufficient evidence; otherwise, it would not be "known".[28]

This tautology is rendered even more obscure by the absence of a transparent and reproducible definition of "sufficient." The term has no scientific meaning, though scientists and physicians have made numerous attempts to construct them. Meanwhile, "sufficient" does have meaning in other areas—conveniently in law, for example—where "sufficient" defines the threshold burden of proof for criminal or civil litigation purposes. Because those thresholds vary depending on context, what constitutes sufficiency also depends on context. Thus, evidence may be "sufficient" in civil litigation if meets a "preponderance of evidence" standard (typically "more likely than not," or greater than 50% likelihood). In criminal litigation, however, evidence is sufficient only if it meets the "beyond reasonable doubt" standard. For our purposes, we can use as an approximation for this 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%.

Without knowledge of the NTP's evidentiary standard, however, we cannot know what it means for evidence to be "sufficient." We do not know if the NTP requires evidence to be "beyond a reasonable doubt" (≥95%), or a "preponderance of the evidence" (> 50%). Indeed, the NTP's evidentiary standard could be well below 50%. For all we know, the NTP might be applying a "beyond reasonable doubt" standard in which the null hypothesis is the substance is presumed to

[28] Some may mistakenly view this analysis of grammar as an exercise in pedantry. However, writing regulatory language is an art form that follows prescribed rules. Agency rule-writers know these rules and must be presumed to follow them absent persuasive evidence to the contrary. In this case, the appearance of the comma is crucial. Without the comma, "sufficient" evidence would have to meet a second test—the indication of a causal relationship. With the comma, however, the second test is implied by the existence of "sufficient" evidence, rendering the criterion tautological. A case could be made hat the NTP interprets the text as if the comma were missing, but such a case would have to show that the NTP makes an explicit showing of causality before it deems performs a substance a "known" carcinogen. In the listing decision for formaldehyde, discussed in Section VI.C below, the NTP did not make a showing of causality. It acted as if the comma rendered the subsequent clause a parenthetical element.

be a carcinogen, and thus it is the duty of negative evidence to show that there is less than 5% chance that the substance is not a carcinogen. Or maybe even a 1% chance.

In addition to the utter opacity of the NTP's definition, a second glaring defect is the absence of a transparent WoE scheme. How many positive human studies are needed? How strong must they be? How are negative or equivocal data taken into account? The NTP criterion discloses answers to none of these questions. The NTP reserves to itself the discretion to consider any data it wants, to exclude any data it wants, and to evaluate that data in accordance with ad hoc criteria that may include overriding policy judgments.[29]

C. The NTP's Definition of a "Reasonably Anticipated to be" Human Carcinogen

To be listed as "reasonably anticipated" to be a human carcinogen, a substance must satisfy at least one of a series of paths shown in Figure 7. Superficially, it appears that there are only three such paths, but a careful review of the text shows that there are myriad paths by which the NTP would define a substance as "reasonably anticipated to be" a human carcinogen.

1. Path (A): "Limited" evidence from human data

Option (A) is the only path involving human data, and it begins with all the defects of the "known" carcinogen criterion, except that it requires only "limited" rather than "sufficient" evidence. We do not know the evidentiary standard the NTP uses for "sufficient" evidence, so we also do not know the evidentiary standard for "limited" evidence. We can only presume that it is lower.[30]

---

[29] Instead of a WoE scheme (which requires all data to be accounted for), the NTP appears to use a "strength-of-evidence" scheme (which counts only positive data). IARC's scheme is stated as "strength of evidence," but IARC states that this terminology is used for consistency rather than to describe its actual practice (International Agency for Research on Cancer 2006b, 2; "[IARC] continues the previous usage of the phrase 'strength of evidence' as a matter of historical continuity, although it should be understood that Monographs evaluations consider studies that support a finding of a cancer hazard as well as studies that do not")

[30] We actually do not know for sure that "limited" evidence truly is less than "sufficient," because the NTP has never published the WoE scheme is uses to make these determinations.

Note that the parenthetical element following the comma is different, and much more expansive, than in the case of the "known" human carcinogen definition. But just as causality is grammatically implied by "sufficient" evidence in the definition of a "known" human carcinogen, causality is grammatically implied here as well. What is different is the description of the nature of causality implied. For "known" carcinogens, conflicting evidence or doubts about the strength of positive evidence appear not to exist. For "reasonably

Figure 7: NTP's Definition of a "Reasonably Anticipated" Human Carcinogen

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; or
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
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.

*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.

anticipated to be" carcinogens, questions about conflicting evidence or doubts about the strength of positive evidence are acknowledged to exist, but they are assumed to be unpersuasive.[31]

2. Path (B): "Sufficient" evidence in animals

A second set of paths involves the existence of "sufficient" evidence of carcinogenicity in animals. As before, the sentence structure utilizes a comma before the relative clause, thereby grammatically rendering it a parenthetical element. This parenthetical element contains three different examples that are conventionally interpreted as three alternative conditions, any one of which is enough to trigger the "reasonably anticipated to be" classification. A careful read shows that there are actually eight versions of these three paths:

- 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
- A combination of malignant and benign tumors in one species to an unusual degree with regard to age at onset

A substance with animal data fitting any one of thee eight conditions is deemed to be "reasonably anticipated to be" a human carcinogen. The ambiguous words that are integral parts of some of these paths (i.e., "increased" incidence," "unusual degree") are not defined.

[31] "Alternative explanations … could not be adequately excluded." Circling back to the "known" human carcinogen definition, this implies that alternative explanations were adequately excluded in that definition. The NTP does not define a method for doing this, nor does it explain how it does this in practice.

3. Path (C): "Less than sufficient" evidence in humans or animals

This path is a catchall for situations in which the NTP believes that a substance ought to be designated as "reasonably anticipated to be" a human carcinogen, but the data are too weak, too controversial, or too burdened by negative or equivocal data. Any one of 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"

Both paths appear to be forms of guilt-by-association that obviate the usual need to present and defend evidence. Nothing needs to be known about the actual carcinogenicity of the substance, and scientific knowledge about the substance cannot trump information-free assumptions made about it. Path (C) does not allow for other information to create a presumption against human carcinogenicity, and there is nothing to prevent the exercise of policy judgment to masquerade as science.

When the eight options in Path (B) and the two options in Path (C) are taken together, it becomes clear that the NTP defines the "reasonably anticipated to be" category rather capaciously. Indeed, it may be surprising that any substance escapes being deemed "reasonably anticipated to be" a human carcinogen. In short, the NTP listing criteria imply that Paracelsus was wrong: The molecule makes the poison, not the dose. A substance escapes listing if no amount of molecules is enough to secure a confession.

D. Negative and Equivocal Evidence

Except for Path (C) in which no empirical evidence of carcinogenicity is required, only evidence supporting the inference of carcinogenicity truly matters. Negative and equivocal evidence apparently play a minor and highly subordinate role. The NTP's listing criteria provide no transparent way through which a presumption of human carcinogenicity—either "known" or "reasonably anticipated to be"—might be rebutted.

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.

E. Peer Review and Public Participation

The NTP's procedures appear to put a high value on peer review and public comment, but this appearance is misleading.

1. Procedures used for the 11th RoC

Public participation and peer review were conducted in three disjointed steps. First, the NTP and other federal agencies submitted nominations for RoC review. No public participation was permitted at this crucial step. A panel of senior NIEHS/NTP scientists (the NIEHS/NTP Review Group, or "RG1") would be established and charged with "determin[ing] if the information provided [in the nomination] indicated that the nomination warranted further consideration by the NTP." If it did, a draft background document would be prepared and the nomination would be proposed in the Federal Register for public comment (National Toxicology Program 2005a). Because a presumptive listing decision could be inferred with little difficulty, it is clear from public comments that, once a substance was nominated for listing, listing was essentially certain. Comments were prepared on the science and submitted primarily to preserve the ability to object to a listing decision on scientific grounds.[32]

A second review would be performed by the NTP Executive Committee's Interagency Working Group ("RG2"), but like the RG1, it was not required to respond to public comments or defend contested scientific claims. Once the RG2 accepted the background document, it became "the final document of record" and the basis for its recommendation whether to proceed with listing (National Toxicology Program 2005a).

[32] The opportunity for policy agendas to drive the review process is made clear: "[B]ackground documents were prepared with the assistance of a consultant or a panel of consultants who have expertise and/or knowledge for the specific nomination…" See (National Toxicology Program 2005a).

These reviews were closed to the public and neither independent nor external. A review of their work products shows that they were not especially rigorous. These defects, combined with the lack of procedural transparency, generally confirmed the public impression that the NTP's peer review process lacked scientific integrity and accountability.[33]

Only after all of these internal steps had been completed would the NTP invite external peer review by a subcommittee of its Board of Scientific Counselors (BSC). Comments from the public were accepted and limited opportunity was provided for public presentations during BSC meetings. The NTP's process was ambiguous concerning whether the BSC had any responsibility for reviewing public comments and addressing the issues they raise, or even an obligation to read them. Reviewing the minutes of BSC meetings shows that they are conducted as colloquies between the NTP and the BSC. Technical errors and misrepresentations of the science made by the NTP presenter could not be challenged except by BSC members, who often were the least informed people in the room and were by temperament or selection bias likely to be very trusting of the NTP's expertise.

A typical BSC meeting would occur over two days and address multiple substances, so the intensity of review must be regarded as superficial. For example, the BSC panel that reviewed the proposed listing of naphthalene reviewed seven other proposed listing determinations Also, the charge to the BSC consisted of a mix of science and policy (National Toxicology Program 2005a). While the committee had the latitude (if not the time) to delve into the science, its primary function was to decide whether to ratify the NTP's proposed listing decision, which as noted previously in this Section, consists of a policy decision hiding behind a scientific façade.[34]

---

[33] Given commenters obvious need to refrain from antagonizing government employees holding such enormous power, it is unsurprising that many comments would overflow with flattery. What is more surprising is how many comments display unusual candor. By far, it is the university scientists hired by industry to perform independent reviews of draft background documents and proposed substance profiles who penned the most strongly worded scientific criticisms.

[34] There is no evidence from the meeting minutes for the naphthalene review that the panel raised any questions about the scientific merit of the

2. Procedural reforms instituted for the 12th RoC

The NTP had described its process as "undergo[ing] a multi-step, scientific review … that includes several opportunities for public comment" (National Toxicology Program 2005a). It experienced withering criticism for lacking independence, rigor, timeliness, and transparency, and protecting scientific deficiencies (National Toxicology Program 2004a).[35] The NTP Director responded by engaging the public in a search for acceptable procedural reforms (National Toxicology Program 2006). When they were proposed for comment, the reforms elicited superficial praise combined with concerted opposition to a number of specific provisions such as the NTP's intention to conduct the BSC's review of the draft substance profile in secret (National Toxicology Program 2006). This particular aspect of the reform proposal likely doomed any chance the NTP had of restoring public confidence in its scientific and political integrity.[36]

The NTP's revised process includes four main changes (National Toxicology Program 2011c). First, the public is now invited to submit nominations for listing.[37] Second, the RG1 "review" is now identified more accurately, if ambiguously, as an internal "evaluation." Third, peer review begins when an external panel reviews the draft background document[38] with a two-part charge:

---

criteria it was charged with applying. See National Toxicology Program (2002b, 5-8).

[35] NTP listing decisions and its process were the subject of six information quality challenges between 2002 and 2004, leading the Office of Management and Budget to make a number of recommendations for process reform in furtherance of its information quality program (Graham 2004).

[36] Because the BSC is a federally chartered advisory group, its deliberations are required by law to be conducted in public. It is anyone's guess how the NTP Director and senior staff thought they could evade this legal requirement.

[37] Revealingly, no provision was made in the revised process for nominations for delisting. Subsequently, the NTP subsequently said that delisting was implied (National Toxicology Program 2007), but the text of the revised process makes clear that it is intended only to support listings by containing no content relevant to delisting.

[38] The draft background document is limited to "publicly available, peer-reviewed sources" except for sections on human exposure and "other relevant data". Many public commenters objected to these exclusions. The

> The expert panel is first charged to peer review the background document. Once the peer review is complete, the NTP asks the expert panel to (1) to apply the RoC listing criteria to the relevant scientific evidence and make a recommendation regarding the listing status for the candidate substance and (2) to provide the scientific justification for that recommendation (National Toxicology Program 2011h).

Thus, the charge to peer reviewers is unchanged but now highly transparent with respect to its policy elements. Peer reviewers, who are presumably selected for their scientific credentials, are tasked with ascertaining whether the agency's policy determination can be plausibly ratified by an appeal to science. The peer review panel's job isn't to conduct an independent and objective review of the science and leave policy matters to policy officials, as federal guidelines on per review require.[39]

As before, the RoC subcommittee of the BSC conducts a second ostensibly scientific peer review that begins only after the NTP's decision has been all but made. This late timing means the BSC bears an implicit burden of proof to show why the proposed decision should be reversed.

3. Problematic aspects of the NTP's peer review process

Although the BSC is appointed by the Secretary of Health and Human Services to advise NTP on science (Sebelius 2010), in practice, the role of its RoC subcommittee is to provide a science-based ratification of the NTP's (and the Secretary's) policy decisions. The NTP says that it "makes available to the BSC all relevant information" including public comments (National Toxicology Program 2011g), but this assurance may not be meaningful. Minutes of the June 2010 RoC

NTP defended the exclusion of exposure data on the ground that they usually were not published in peer-reviewed journals; it did not defend the exclusion of "other relevant data". See National Toxicology Program (2007).

[39] Federal guidelines require scientific peer review panels to limit their reviews to science and address whether the scientific information they are asked to review is objective. See Office of Management and Budget (2005, 2675; "Peer reviewers shall be charged with reviewing scientific and technical matters, leaving policy determinations for the agency. Reviewers shall be informed of applicable access, objectivity, reproducibility and other quality standards under the Federal laws governing information access and quality.")

subcommittee meeting during which three draft substance profiles were reviewed indicate that it is standard practice not to disclose data that do not support listing.[40]

Evidence that the BSC peer review is an intentionally muddled process, in which panel members are charged with simultaneously reviewing the scientific record and ratifying the agency's policy decisions, is clear from the charge:

> The BSC is charged to determine whether the scientific information cited in the draft substance profile for a candidate substance is technically correct, clearly stated and supports the NTP 's policy decision regarding its listing in the RoC (National Toxicology Program 2011f, emphasis added)

Contrary to NTP claims, this violates federal peer review guidance.[41] The BCS certainly can evaluate whether a background document is "technically correct" and it can determine whether the information in a document is "clearly stated." But it cannot verify that a background document "supports the NTP's policy decision." Policy decisions are just that—policy decisions—and while policy decisions can be informed by science to a greater or lesser extent, asking external scientists to opine on whether science "supports" policy is little different than asking the BSC if its members share the same policy views.

[40] From the BSC's review of the draft substance profile for glass wool: "[BSC Member] Dr. [Mitzi] Nagarkatti asked whether animal studies had been conducted in species other than rats and hamsters, and if so, why they were not included. NTP staff member Dr. [Gloria] Jahnke replied that there had been studies in guinea pigs, as well as inhalation studies in monkeys, that had been negative. She explained that she did not include negative results in her presentation, as it is the practice to only report studies that support the listing recommendation". See National Toxicology Program (2010, 16).

[41] The NTP revised its procedures in part to comply with government-wide peer review guidance issued by the Office of Management and Budget. See National Toxicology Program (2011b). However, OMB's guidelines prohibit the use of peer review to validate agencies' policy choices. See footnote 39. The NTP dutifully forbids peer reviewers from disagreeing with its policy choices, but it expressly seeks their endorsement.

The NTP goes so far as to prohibit the BSC from disagreeing with the NTP's policy views, which the NTP helpfully makes clear in case there might have been confusion on this point:

> The BSC is not asked to review the NTP 's decision regarding listing status (National Toxicology Program 2011f).

4. The process reforms installed for the 12th RoC did not succeed

The NTP has been sharply criticized for withholding information from the public, committing material scientific error, ignoring public comments pointing these errors, and permitting gross conflicts of interest in its external peer review process. In 2004, the Office of Management and Budget formally requested that the NTP adopt significant procedural changes in hopes that they would "further instill public confidence." The available evidence suggests that the procedural reforms the NTP put in place for the 12th RoC were too limited in scope and generally unsuccessful for achieving even modest objectives.

The NTP's peer review process has two structural defects that ensure failure: peer review does not occur until a point in the process when it is too late to correct error and the charge to reviewers is to support the NTP's decisions, not to correct error. The inevitable result of these structural defects is that peer review at the NTP does not (and probably cannot) improve scientific quality. As a bureaucratic strategy, this process makes sense if and only if the NTP's interest in science is conditional on whether it supports policy. Such a strategy may be intentional, for de facto listing determinations may be made as early as the nomination stage. After all, the principal reason why some substances are nominated is that a federal agency wants it listed. A serious peer review program would subject nominations to external, independent, scientific examination, as a prerequisite for advancing to the review stage.

F. Discrepancies Between the Statute and the NTP's Implementation

There are several ways in which the NTP implemented its statutory charge in ways that are inconsistent (if not in conflict) with the statute.

1. The NTP ignores the statutory requirement to limit listing to substances to which a significant number of persons are exposed

Contrary to its statutory directive, the NTP does not limit the list to substances to which a "significant" number of persons are exposed (see Figure 2). There are two prongs to the statutory text—"significant" number of persons and actual exposure. For each substance, a decision is reached at the nominations stage—i.e., well before peer review—that the number of persons exposed is "significant." The term is never defined. A significant number of persons is exposed if the NTP says so; nothing more, and nothing less.

2. The NTP ignores the statutory requirement to estimate the number of persons residing in the United States who are exposed

The NTP does not provide useful information about the number of persons residing in the United States who are exposed. It discusses different "natures" of exposure (i.e., environmental and occupational), contrary to its statutory charge (see Figure 3). In lieu of addressing exposure as the law requires, the NTP relies on various proxies such as the mass or volume of production, usage, emissions, or disposal.[42] For example, the NTP substance profile for naphthalene relies on recent mass data from EPA's Toxic Release Inventory (TRI), surveys of occupational exposure from past decades, and estimates from worst-case conditions (National Toxicology Program 2011e, 277-278). Like naphthalene, the substance profiles for styrene and formaldehyde are do not attempt to make any showing that a "significant" number of persons are exposed (National Toxicology Program 2011e, 387-388, 199-201).

When snippets of exposure data are disclosed, the RoC does not place these levels in perspective, such as, for example, comparing them with the exposures in the animal and/or epidemiologic studies on which the decision to list ostensibly depends. Substance profiles also do not include estimates of the number of persons residing in the United States exposed at any particular level. The NTP does not interpret probabilistic terms in probabilistic ways.

[42] Mass and are inherently deficient volume proxies for exposure. What makes them especially interesting is that they imply that regulations promulgated to reduce cancer risks are completely ineffective.

The NTP provides no quantitative or semi-quantitative estimate of the probability that a "known" or "reasonably anticipated" carcinogen substance actually causes cancer at environmentally or occupationally relevant doses. The listing criteria also do not quantitatively or semi-quantitatively define any of the crucial terms within the definitions (e.g., "sufficient," "increased incidence," "an unusual degree," a "well-defined, structurally related class"). Although the NTP uses the probabilistic term "likely," this term is not defined in probabilistic ways (e.g., 51% chance; 75% chance; 95% chance). In short, a substance is "reasonably anticipated" to cause cancer if the NTP says it is; nothing more, nothing less.

3. The NTP implicitly acknowledges that substances determined to be "known" and "reasonably anticipated to be" human carcinogens are not necessarily carcinogens in the real world

The preamble to the RoC contains some very interesting disclaimers. For example, one disclaimer very nearly admits that its determinations are merely theoretical or hypothetical. Rather, the designations "known" and "reasonably anticipated to be" human carcinogens only indicate the potential for human cancer risk under exposure conditions the NTP declines to identify:

> [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 (National Toxicology Program 2011e, 3).

The NTP then attempts to make a virtue of this fundamental disconnect between theory and reality by denying any responsibility to have made realistic decisions. The NTP shifts this responsibility to other agencies:

> [F]ormal risk assessments are the responsibility of the appropriate Federal, state, and local health regulatory and research agencies (National Toxicology Program 2011e, 3).

G. IARC's Alternative Classification Scheme

Figure 8: IARC Cancer Classifications

Group 1	*Carcinogenic to humans*
Group 2A	*Probably carcinogenic to humans*
Group 2B	*Possibly carcinogenic to humans*
Group 3	*Not classifiable as to its carcinogenicity to humans*
Group 4	*Probably not carcinogenic to humans*

Standing outside the NTP, but not at a very far intellectual distance, is the International Agency for Research on Cancer (IARC), a United Nations organization that performs much the same function. It was established years before the NTP, and it therefore already had a classification scheme in place when the NTP opened for business. This scheme has substantial similarities and a few differences.

Group 1 parallels the NTP's "known" category, but it is more expansive because it permits assignment based on "sufficient" evidence from animal data plus "strong" evidence that the substance acts through a "relevant mechanism of carcinogenicity."[43] Group 2 maps to the NTP's "reasonably anticipated" classification, though IARC is more transparent than the NTP about its breadth.[44] IARC classifications include many of the same descriptors used by the NTP

[43] "This category is used when there is *sufficient evidence of carcinogenicity* in humans. Exceptionally, an agent may be placed in this category when evidence of carcinogenicity in humans is less than *sufficient* but there is *sufficient evidence of carcinogenicity* in experimental animals and strong evidence in exposed humans that the agent acts through a relevant mechanism of carcinogenicity." See International Agency for Research on Cancer (2006b, 22, emphasis in original).

[44] "This category includes agents for which, at one extreme, the degree of evidence of carcinogenicity in humans is almost *sufficient*, as well as those for which, at the other extreme, there are no human data but for which there is evidence of carcinogenicity in experimental animals." See International Agency for Research on Cancer (2006b, 22).

as well as some that are not (e.g., "inadequate"). None of these descriptors has a transparent definition.

The most potentially interesting difference in the IARC scheme is its use of expressly probabilistic descriptors ("probable," "possible") instead of legalese ("reasonably anticipated to be"). But IARC's probabilistic descriptors do not, like those used by the NTP, in fact transparent or have any reproducible probabilistic content.

The NTP and IARC are sometimes competitors, sometimes collaborators, but in many respects they appear to be redundant. They might provide a governmental example of Hotelling's Law, which holds that competitors differentiate their goods and services as little as possible in order to maximize demand (Hotelling 1929).This can be observed by reviewing the NTP's substance profiles, and especially the reports of its peer reviews. If IARC has already taken a position, for example, an NTP review panel will often find it persuasive as a substitute for an independent review of the underlying scientific evidence.[45] If an IARC listing is that influential, the value added provided by the NTP is negligible.

VI. Case Studies

Case studies can illustrate phenomena of interest and suggest lines of inquiry for future research, but they must be used with care for making inferences. The three case studies discussed briefly in this Section illustrate the problems described in previous Sections. They show how the NTP continues to experience serious difficulties producing a RoC with minimally acceptable scientific quality. Two of the substances (formaldehyde and styrene) were newly reviewed for the 12th edition. They show how the NTP responds when highly qualified third parties contest its scientific claims. The third (naphthalene) was listed for the first time in the 11th edition (2004), but new peer reviewed reports raised serous doubts about the scientific merit of NTP's initial classification. It shows how effective the NTP is on its own in identifying and processing new scientific information.

[45] See, e.g., the minutes of the June 2010 BSC meeting (National Toxicology Program 2010), where prior decisions by IARC figured prominently in the review of each of the proposed listings.

These cases should approximate the best that the NTP can be reasonably anticipated to accomplish. For formaldehyde and styrene, the quality improvements resulting from the procedural reforms installed in advance of the 12th edition should be evident. For naphthalene, the NTP's capacity to keep up with the scientific literature, thus giving practical meaning to its commitment to due diligence, should be apparent.

Unfortunately, the case studies reveal that recent NTP process reforms have failed to improve scientific quality. The NTP confirmed its critics' impression that it is incapable of (or unwilling to) give any perceptible weight to negative and equivocal data no matter how technically accurate. The NTP continues to be hindered by bureaucratic or ideological blinders resulting from how the agency has implemented its statutory charge.

A. Case Study #1: Naphthalene

In the 11th RoC, the NTP deemed naphthalene "reasonably anticipated to be" a human carcinogen (National Toxicology Program 2004b). This determination was based on the results of chronic 2-year bioassays in mice (National Toxicology Program 1992) and rats (National Toxicology Program 2000).

In the mouse bioassay, a statistically significant increase in benign lung tumors was observed in females but not in males.[46] In the rat bioassay, a statistically significant increase in benign lung tumors was observed in males but not in females.[47] These results alone satisfy a strict reading of the second option in Path (B) ("a combination of malignant and benign tumors in multiple species"), but only because of benign tumors and the extraordinarily high doses to which rodents the were subjected.

[46] At 0, 10, and 30 ppm, alveolar/bronchiolar adenoma incidence in female mice was 5/69 (7%), 2/65 (3%), and 28/134 (21%). The incidence of bronchiolar carcinoma at these doses was 0/69 (0%), 0/65 (0%), and 1/134 (1%).

[47] At 0, 10, 30, and 60 ppm, alveolar/bronchiolar adenoma incidence in male rats was 0/49 (0%), 6/49 (12%), 8/48 (17%), and 15/48 (31%). No bronchiolar carcinomas were reported.

The rat bioassay also observed a statistically significant increase in olfactory neuroblastomas,[48] a rare cancer in both animals and humans.[49] This alone was sufficient to trigger listing as a "reasonably anticipated" human carcinogen under the seventh option in Path (B) ("a combination of malignant and benign tumors in one species to an unusual degree with regard to type of tumor").

Thus, the NTP had two independent grounds under its listing criteria for deeming naphthalene "reasonably anticipated to be" a human carcinogen.[50] Given the way the NTP interprets data, it should not have been a surprise to anyone when it proposed to do this.[51] NTP scientists would not have raised questions about whether the laboratory experiments were appropriately designed, particularly the propriety of the 30 ppm and 60 ppm doses, because other NTP scientists designed and implemented the experiments.[52] NTP scientists would not have expressed concern about the absence of lung carcinomas because the listing criteria count benign tumors as if they are malignant. NTP scientists would not have been concerned about whether the mode of action at such high doses is relevant to humans because evidence about mode of action can only be used to support a

[48] At 0, 10, 30, and 60 ppm, olfactory neuroblastoma incidence in male rats was 0/49 (0%), 0/49 (0%), 4/48 (8%), and 3/48 (6%); in females the incidence was 0/49 (0%), 2/49 (4%), 3/49 (6%), and 12/49 (25%).

[49] In a 2004 external review draft cancer risk assessment, EPA stated "[a]pproximately 300 human cases of olfactory epithelial neuroblastomas were reported in the world research literature between 1924 and 1985." See U.S. Environmental Protection Agency Office of Research and Development (2004, 23, fn 1).

[50] For comparison, the highest occupational exposures appear to be a thousand-fold or more lower, with all other human exposures much lower (Griego et al. 2008; Price and Jayjock 2008).

[51] Industry alleged that a decision to list "would be the first substance to be listed based on 'clear evidence' in one species of experimental animal [i.e., rats], 'some evidence' in one sex of a second species [i.e., mice], and that is not genotoxic." See (Price 2003, 3). Verifying this claim is beyond the scope of this paper.

[52] The consultant the NTP brought in to provide subject matter technical support was head of the NTP's bioassay technical support group. See National Toxicology Program (2002a, 55-56), and thus conflicted with respect to what turned out to be the most salient scientific question—whether the NTP's experimental design was defective

listing, never to reject one. Finally, NTP scientists would not have thought it relevant that the virtual absence of olfactory neuroblastoma in people residing in the United States in any way compromised its conclusion that naphthalene was "reasonably anticipated" to be a human carcinogen in the real world, because the NTP acknowledges—indeed, proudly claims—that real world is irrelevant to listing decisions.

1. Public comment and peer review

Industry commented on the proposed nomination (Price 2001), the draft background report (Price 2002a), prior to the meeting of the NTP Executive Committee (Price 2002b), and prior to the agency's nominal date for final decision (Price 2003). Industry raised numerous scientific arguments, the strongest of which was that cytotoxicity resulting from massive inflammation was the almost certain mechanism through which cancer occurred (Price 2001), an opinion shared by a subsequently convened independent review panel. Industry also argued that anatomical, physiological, and metabolic differences between rodents and humans argued against extrapolating from rodents to humans with respect to the inhalation path, and that if the WoE schemes of other authorities were applied, naphthalene would not be listed.

These comments had no perceptible effect on the NTP because they appear to have been beside the point. As noted earlier, under the NTP's listing criteria, human causality is automatically assumed once cancer in animals is observed; the circumstances and conditions under which carcinogenic effects occur in test animals does not matter; and there is no published way to rebut the inference of causality. Nor does it matter how these data would fare in a WoE scheme, for the NTP does not have published WoE guidelines. The NTP's implicit WoE scheme is analogous to a checkbox: if animal data for a relevant category are present, the criterion for inclusion is met and additional data are superfluous. If any single box in Path (B) is checked, for example, classification is predetermined and all other data are irrelevant.

2. NTP process failures

Perhaps the most striking feature of the NTP's naphthalene review was the breakdown of normal procedure, which industry commented on with surprising candor:

> Unfortunately, the RG-1 review occurred before publication of the Draft Background Document, the RG-2 review occurred after publication of the background document but before the date tor receipt of public comments, and the BSC RoC Subcommittee based its decision apparently in large measure on information newly introduced at the Subcommittee meeting and not made available as of this date by NTP either on its website or in a revised background document or even in minutes of the meeting. This simple recitation of the calendar of the nomination, review, and now proposed listing of naphthalene makes evident that there has been no sincere effort to engage public stakeholders in the process, and no effort to ensure that "[NTP's] three scientific review committees are basing their decisions on the same basic material augmented by the additional public comments obtained during the review process." Indeed, members of the public who were not physically present at the Subcommittee meeting are not even aware that a substantial part of the apparent basis for the Subcommittee's recommendation is not part of the public record, was not shared prior to the Subcommittee meeting with either Subcommittee members or the public, and has not been made publicly available to those who may be interested in submitting comments in response to the January 22, 2003 Federal Register notice regarding nomination of naphthalene as *Proposed for Listing in the Report on Carcinogens, Eleventh Edition*.[53]

A review of the record shows other substantive errors, such as the incorrect presentation of study results to peer reviewers by senior NTP personnel,[54] and material process errors, such as the decision of the NTP to allow the RoC subcommittee chairman to strep down from that role in order to make new scientific claims that were not part of

[53] Price (2003, 1-2, quoting from a March 11, 2003 letter from NTP Director Kenneth Olden in response to previous process complaints; reference omitted). Reviews cannot be "augmented" by comments received after the reviews are completed.

[54] The RoC program director verbally delivered the NTP's report to the BSC's RoC subcommittee and stated that adenomas of the rat respiratory epithelium were so rare as to never have been seen in historical controls. See National Toxicology Program (2002a, 63). In the mid 1990s, the rate of spontaneous lung tumors in rats was reported to be 2.4% in males and 1.4% in females (Pitot III and Dragan 2001, Table 8-30).

the NTP review, had not been peer reviewed, were not available to the public, were at odds with the raft substance profile, but nevertheless were persuasive to the peer review panel (National Toxicology Program 2002a, 99-129). At several points along the way, responsible agency management committed to effective independent and external scientific peer review would have terminated the review and restarted the process with different personnel and much closer supervision. This did not happen.

3. Genuinely independent and external scientific peer review

In 2006, four independent panels of scientists unaffiliated with the NTP or industry conducted a targeted review of specific scientific aspects of the primary scientific database for naphthalene (Belzer et al. 2008).[55] One panel reviewed the NTP bioassays (North et al. 2008). This panel concluded that the results observed in both experiments were artifacts of defective study design. In particular, the panel said that the higher exposures in these studies exceeded the maximum tolerated dose (MTD). They noted that in both studies the incidence of inflammation at the highest dose was at or near 100%, and the inferred that this "clearly compromised" the animals' well being. Further, in the rat study, what they saw indicated that at very high doses, naphthalene was killing cells directly. They inferred that tumors, to the limited extent they were observed, had occurred secondary to frank toxicity and almost certainly would not have occurred otherwise (North et al. 2008, S9). Other panels that reviewed the scientific evidence for alternative modes of action reinforced this conclusion (Bogen et al. 2008; Brusick 2008; Brusick et al. 2008).

4. The predicates for effective use of scientific information

A critical, quality-oriented review of the data might well have led the NTP to a different decision. However, the NTP does not normally review positive data critically, and there is no public evidence that the NTP did so in this case. Yet there are obvious doubts about each of the cancers that the NTP said were "reasonably anticipated" to be caused

[55] The Naphthalene State-of-the-Science Symposium did not review secondary materials, such as government risk assessments or policy determinations such as NTP and IARC listing decisions. It also avoided all policy matters such as the derivation of cancer potency estimates and the like.

by naphthalene. Despite extraordinarily high exposures, few lung carcinomas were observed. Perhaps benign tumors would have become malignant, but the fact that they didn't seems important. The rare incidence of olfactory neuroblastoma does not exonerate naphthalene, but it's hardly strong evidence that exposure to levels a thousandth or less is "reasonably anticipated" to be a likely cause of the very small number of such cancers that are diagnosed each year.

A quality-oriented review of the data is not part of the NTP's scientific review because making a case for causality is not required to make a listing decision. As noted above, the RoC criteria assume causality in humans from the existence of positive animal studies. Listing decisions are not complicated questions because the complexities of science are not part of the NTP's inquiry.

Concerns about study design could have been flagged by the NTP during its preparation of the background document but were not, if for no other reason than the study design was the NTP's own invention. Even though the defects of this design were clear after the completion of the mouse bioassay (published 1992), the NTP repeated the same design in the rat bioassay (published 2000), except with the addition of a highest dose twice as large as the highest dose administered to mice[56]. It appears that the testing side of the NTP saw its purpose as to do whatever was required to obtain a positive effect. Meanwhile, the RoC side of the NTP thought that critically reviewing these studies—especially since they were performed by the NTP itself—was not part of its charge. These discrepancies highlight the inherent conflict of interest arising when one wing of the NTP has at least the nominal responsibility to review the work of another. As long as the NTP owns both functions, it should be assumed that its own studies would be treated with reverence.

B. Case Study #2: Styrene

The NTP added styrene to the RoC in the 12th edition, concluding that it was "reasonably anticipated" to be a human

[56] Public discussion during the Symposium revealed that the decision to double the highest dose in the rat study appears was the result of technological improvements in the ability to sustain higher levels in experimental settings. It was not based on any consideration of rat biology or physiology, nor was it based on a short-term experiment to ascertain the likely MTD.

carcinogen. It based this determination on "limited" evidence from occupational studies in humans (see Path (A) on page 23) and "sufficient" evidence of carcinogenicity in animals. Most weight was given 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 (National Toxicology Program 2011e, 383-384).[57]

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:

> 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 (National Toxicology Program 2011e, 383).

The standard model of carcinogenesis assumes that cumulative exposure is the correct exposure metric. In the study by Kogevinas et al. (1994), which the NTP found dispositive, a statistically significant association was observed with average but not cumulative styrene exposure. This erects a serious hurdle to inferring causality, because it violates the presumptive mode of action.[58] The NTP resolved the matter by implicitly assuming the existence of an unknown mode of action sufficient to avoid disregarding the association with average exposure as spurious:

---

[57] 1,3-butadiene was first listed by the NTP as a "known" carcinogen in the 5th RoC (1989).

[58] The conventional model of carcinogenesis posits that risk is a function of cumulative exposure. If this model is correct, then the association with average exposure is spurious. If average and not cumulative exposure is indeed the cause, then a different model for carcinogenesis is needed.

> 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 (National Toxicology Program 2011e, 383).

1. Public comment

The Styrene Information and Research Center (SIRC)[59] commented on the proposed nomination (Snyder 2004), the draft background document (Snyder 2008a), the NTP expert panel recommendation (Snyder 2008b), prior to BSC review (Snyder 2009a), and in a pair of last ditch efforts to stop the final decision based on shared professional scientific affiliation (Bus and Cruzan 2009) and deference to a recently published report of a National Research Council committee (Banton 2011).

In these comments industry raised numerous scientific arguments, the strongest of which concerned the way the NTP interpreted the epidemiologic studies and evidence showing that results in rodents would not be replicated in humans because of known physiologic and metabolic differences (Cruzan et al. 2009; European Union Chemicals Agency 2008).

With respect to epidemiology, the principal author of the crucial styrene-butadiene rubber worker study objected to NTP's interpretation of her work as showing a causal relationship (Delzell 2008, 2009). The NTP also inferred a causal exposure-response relationship in the main reinforced-plastics industry study. But the authors themselves interpreted their work much more modestly, saying only that their results "leave the question open of whether an excess risk of neoplasms of the lymphatic and hematopoietic tissues occurs among workers exposed to styrene" (Kogevinas et al. 1994, 260). In short, the NTP interpreted weak epidemiological evidence as "limited," but NTP's listing criteria define this as sufficient for listing a substance as "reasonably anticipated to be" a human carcinogen because "alternative explanations, such as chance, bias, or confounding factors, could not adequately be excluded."[60]

---

[59] Numerous others commented as well; the Styrene Information and Research Council is representative and its comments are comprehensive.

[60] A dispute arose within the BSC's RoC subcommittee over the meaning of "credible": "It was pointed out by Dr. Friedman-Jimenez that the

To escape this designation as "reasonably anticipated to be" a human carcinogen via Path (A) (see page 23), industry tried to prove that the assumption of causality was, literally, incredible. Industry relied on mechanistic data obtained from an experiment using a "knock-out" mouse (Banton, Cruzan, and Bus 2010). It had been hypothesized that this gene, which is minimally expressed in humans, explains the positive results observed in mouse experiments. 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 mice were specific to mice and not relevant to humans, the knock-out mice would not display sensitivity to styrene. And that is what occurred: knock-out mice experienced no dose-related increase in lung toxicity, even from high doses of styrene.

These results had not been peer reviewed in time for the 12th RoC, but their specific relevance to the NTP's determination could not have been greater. Indeed, the experiment was as close as science ever gets to proving the absence of a human cancer hazard. Had the NTP been open to new science, it would have at least postponed a decision until the 13th edition. It did not. The NTP ignored this pathbreaking research and proceeded to the listing as if the experiment had never been performed.[61]

2. NTP process failures

Industry also identified numerous violations of the RoC process and generally accepted scientific practices (Snyder 2008b, 2008c, 2009a, 2009b; Banton 2011). This 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 to support preferred inferences; the use of non-peer reviewed information first introduced by members of a peer

---

dictionary indicates that credible means 'reliable, trustworthy, believable,' and that a causal association between styrene exposure and increased cancer in humans was not 'reliable, trustworthy or believable'." See Snyder (2009a, 2). Clearly, Friedman-Jimenez interpreted "credible" to require a higher standard of proof than the NTP, but did so based on an English dictionary rather than any scientific principle or authority.

[61] Ironically, the particular knock-out mouse used in this study was developed by a researcher funded by NIEHS, the NTP's parent organization.

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.

3. Process confusion

These comments led to no perceptible effort by the NTP to examine the procedural complaints and cure confirmed irregularities and defects. As for the scientific evidence, the dispute can be reduced to two 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 to be" human carcinogen). Commenters persistently misconstrued the RoC listing criteria as creating a separate requirement tor the NTP to make a showing of causality. That may be standard scientific practice, but it is not a grammatically correct reading of the text of the listing criteria and it is not what the NTP actually did.

Second, industry and academic commenters persistently assumed that it was possible to produce scientific evidence that would rebut the NTP's inference (actually, its assumption) of causality. But the assumption of causality is not scientifically rebuttable under the listing criteria. The criteria contain no evidentiary standard for rebuttal, so no amount or quality of evidence is sufficient to rebut.

Third, commenters may have understood that NTP determinations were strictly policy decisions, but they acted as if these decisions were strictly scientific. Such confusion would be justified by the NTP's persistent assertion that science ruled the roost, and thus there was no need for a transparent and reproducible policy rule for determining when the assumption of causality is rebutted. Having such a rule would have converted causality from a superfluous parenthetical element in each listing criterion into a substantive one. That, in turn, would reduce the NTP's policy discretion to decide whether to list. For the NTP, the authority to conserve its discretion to make policy decisions based on science when science is supportive, and without the interference of science when it isn't, could have been at least as important as the substance of the actions it takes. Ironically, by contesting only the science industry unwittingly supported the fiction that the NTP's determinations were based on science. That is, the NTP may have benefitted politically from the immense effort industry

devoted to rebutting the NTP's scientific case. Public officials in neither the Executive branch nor the Congress want to ever be seen as interfering with science, so as long as this was a "scientific" issue their hands are tied.

In sum, the proximate cause of industry's confusion is the NTP's refusal to be candid about the controlling role of nonscientific policy judgment in its listing decisions. Although it empanels scientists to conduct certain reviews, the task set before them is to provide policy advice from behind a scientific façade, opining whether some body of evidence is "sufficient" or "limited," terms that have no scientific meaning. When offered the opportunity to "make policy," few scientists can resist the temptation.[62] Armed with this "scientific" advice, the NTP can plausibly claim that its decisions are grounded in science.

What the styrene case shows is the NTP permits weak evidence of carcinogenic effect to be counted in favor of listing but considers strong evidence of confounding an inadequate basis for exclusion. This is consistent with a policy model in which a high rate of false positives is acceptable in order to avoid even a low rate of false negatives. That is, the NTP strongly prefers to list many substances as human carcinogens that are not in order to ensure that virtually no substances that are human carcinogens are missed.

C. Case Study #3: Formaldehyde

The NTP determined that formaldehyde is a "known" human carcinogen based on "sufficient" evidence from studies in humans and cited "supporting data on mechanisms of carcinogenesis" (National

[62] See the example in the naphthalene case study in which the chairman of the BSC implored his colleagues to list on the ground that doing so would advance the cause of public health. In the styrene case, it was alleged by industry scientists that BSC members allowed their policy preferences to override their low regard for the quality of the draft substance profile: "Our overall observation of the meeting was that the BSC *did not* offer a ringing scientific endorsement of the styrene Draft Substance Profile, but did not feel they had a meaningful way to disagree with the document… but were concerned that recommending against listing would mean either that no further research would be conducted on styrene, or that it would be seen as giving styrene a clean bill of health." See Bus and Cruzan (2009, 1-2).

Toxicology Program 2011e, 195-205). The listing claims causality using the language of the parenthetical element of the listing criterion:

> Causality is indicated by consistent findings of increased risks of nasopharyngeal cancer, sinonasal cancer, and lymphohematopoietic cancer specifically myeloid leukemia among individuals with higher measures of exposure to formaldehyde (exposure level or duration), which cannot be explained by chance, bias, or confounding (Ibid., 195).

The two long-term NCI epidemiologic studies are cited as "most informative" because they "are the only studies that evaluated quantitative exposure-response relationships." Also cited for special attention are studies of rare sinonasal and nasopharyngeal cancer. Numerous other positive studies are cited but appear to have been given little or no weight. No negative epidemiologic studies are said to exist.

4. Public comment

The Formaldehyde Council[63] commented on the proposed nomination (Natz 2005), the draft background document (Natz 2009), the NTP expert panel recommendation (Natz 2010b), and prior to BSC review (Natz 2010a). These comments and comments from its consultants raised numerous scientific issues, including complaints that the NTP had fundamentally misinterpreted formaldehyde toxicokinetics (Andersen 2010), mode of action data (Golden 2009; Andersen 2009), and the epidemiologic studies, some of which were incorrectly reported as well (Collins 2009; Mundt 2009; Marsh 2009). These comments had no apparent effect on the NTPs characterization of the science.

The NTP's claim that there were studies purporting to identify mechanisms by which formaldehyde would cause leukemia drew especially strident criticism (Golden 2009, 3-4), as did the NTP's conclusion that formaldehyde is a "known" cause of leukemia despite no evidence that it is transported at unusual concentrations via the bloodstream to reach distal tissues, as required by basic human biology (Andersen 2010, 3).[64] Whereas the NTP claimed that mode of

[63] Numerous others commented as well; the Formaldehyde Council is representative and its comments are comprehensive.

[64] "Neither the acetal nor the thioacetal represent ways in which significant amounts of formaldehyde could enter the circulation and reach

action data supported its conclusion that formaldehyde was a "known" cause of leukemia, commenters said these data all but proved that leukemia is biologically infeasible (Golden 2009; Andersen 2009). The boundaries of this scientific dispute could not be greater, and obviously they extended to the policy determination at hand: the NTP said mode of action data were "sufficient" to conclude that formaldehyde is "known" to cause leukemia and comments implied they were "sufficient" to conclude this was impossible.

1. NTP process failures

The Formaldehyde Council also alleged numerous violations of the RoC process and the NTP's general unwillingness to address and resolve scientific controversies. When controversies arose because of missing data, disputes over statistical methods, and conflicting results from the main epidemiological study (Marsh 2009; Mundt et al. 2010; Mundt 2009), the NTP does not appear to have attempted to resolve them. The NTP, its internal review panel, or both gave weight to those analyses of the data purporting to show positive effects (e.g., Hauptmann et al. 2003, 2004) and dismissed or excluded analyses of the same data that did not show statistically significant associations or identified uncontrolled confounders (Marsh 2010, citing multiple peer reviewed papers). Thus, even with respect to the same epidemiologic data, the NTP gave weight only to those analyses with positive results.

The NTP also declined to resolve the crucial scientific dispute concerning the mechanism, if any, by which formaldehyde could cause lymphohematopoietic cancers such as leukemia. The NTP's position appears to have been that while such a mechanism is unknown (National Toxicology Program 2011e, 199), it should be assumed to exist because of associations were reported in epidemiological studies (197)(197)(197). This short-cut was made possible by the superfluous role causality plays in the NTP's listing criteria, and it was "validated" by the absence of any disciplined procedure within the NTP listing

---

distant tissues. The panel needs to justify this statement since it is contrary to our extensive understanding of formaldehyde chemistry and biochemistry. In addition the comment in the last sentence says that high endogenous levels represent a challenge for extrapolation. They certainly do. The challenge for the panel should have been to provide any reasonable argument that inhaled formaldehyde can in any way cause biologically appreciable increases in tissue concentration at sites remote from the epithelial cells lining the respiratory tract." See Andersen (2010, 3).

process for demonstrating causality, such as by applying a generally accepted rubric (e.g., Bradford Hill 1965).

2. The NTP listing was transparently policy-based but not transparent

The NTP's practices were inherently anti-scientific. They allowed the NTP to ignore as irrelevant an extensive body of established scientific evidence provided by commenters strongly suggesting that there is no biological mechanism by which formaldehyde could cause lymphohematopoietic cancers. The NTP was abetted by its peer reviewers, which upheld the agency at every turn and, like the NTP, declined to substantively address the scientific controversies.[65] The NTP also attempted to dismiss (though not refute) a National Academy of Sciences review of formaldehyde that interpreted the scientific evidence the same way (National Toxicology Program 2011a).

The NTP managed the public comment process for formaldehyde as a necessary duty to be endured, not an opportunity to improve the quality of the science. It managed the peer review process to create the appearance of science to make its policy decisions opaque. This is unmistakable from the cursory and dismissive nature of its 11-page response public comments (National Toxicology Program n.d., 9-20 plus references) and its 6-page dismissal of the relevant parts of the National Academy review (National Toxicology Program 2011a).

Unresponsiveness is a frequently observed characteristic of agency response-to-comments documents generally. Agencies respond only to only a subset of issues; their characterizations of the issues in dispute are often tendentious or incorrect; and their replies are often unresponsive. This may be a particularly acute problem for the NTP because it is governed by a form of matrix management that could deny it the ability to make its own decisions. NIEHS has nominal control because it provides the majority of the funds. However, several other agencies belong to the NTP collaborative and supply funding. It is plausible that the NTP cannot correct some errors because of

[65] See, e.g., Marsh (2010, "The focus of my comments is the blatant and unsubstantiated omission in the Expert Panel Report of several of my recent peer-reviewed publications dealing with our reanalyses of the National Cancer Institute (NCI) cohort study of formaldehyde-exposed workers"), and Andersen (2010), quoted in endnote 64.

resistance from these other agencies, which play an undisclosed role in the review process.[66]

VII. Conclusions and Recommendations

The reason why the RoC arouses such controversy should be obvious. Although Congress appears to have intended the RoC to be a scientific compendium, the NTP has implemented its statutory charge in a manner that at best intermingles, but more realistically fully subordinates, science to policy. Despite the trappings, the RoC cannot be fairly construed as a scientific document. It utilizes scientific information, but at its core it is wholly an exercise of Executive policy judgment and, because of its lack of transparency, implicit regulatory decision-making without accountability. The text of the law invited this controversy by using language that is not, and can never be made to be, scientific. But much responsibility rests with the NTP, which has cultivated controversy by implementing a science charade. When it decided to reform its procedures to respond to critics, the NTP made only superficial changes and left the fatally defective structure in place. It streamlined a failed system rather than fix it.

D. Key Problems to be Solved

From the discussions of the law and its antecedents in Sections III and IV, the NTP's approach to implementation in Section V, and the three case studies in Section VI, several conclusions are justified. Five are about the circumscribed role of role of science in the process; the sixth concerns whether science actually matters at all.

[66] A good example may be the NTP's determination to hold onto the claim that mechanistic studies support the inference that formaldehyde could circulate in the bloodstream and damage stem cells in bone marrow. The NTP did not need this claim to list; it needed it only to include leukemia within the domain of "known" cancers. This is an immaterial issue to the NTP, but probably not to others. There is evidence that EPA, a key member of the NTP consortium, cares deeply about maintaining this mechanism even though its case has been independently judged to be speculative (National Research Council 2011, 26-27, 35). EPA may have prevented the NTP from making corrections in order to preserve its ability to cite this elsewhere, such as in rulemaking.

1. Five fatal scientific defects

First, RoC listings are at best hazard-based; they are not risk-based, even though the language of listing implies a statement about cancer risk. The dose at which a carcinogenic effect is observed in humans or animals is critically important to a listing's practical utility for risk assessment and risk-based decision-making, but dose has essentially no bearing on the NTP's decisions. The NTP's disclaimer that risk assessment is not its job might be defensible bureaucratically, but it is incompatible with the production of a product that has practical utility for risk-based decision-making sufficient to justify the costs of producing it.

Second, the listing criteria make only the mildest distinction between carcinogenic effects observed at or below human experience and doses many thousands of times greater in laboratory animals. Uncertainty about the relevance of animal tests will always be with us, and science could be brought to bear in ways that reduce it, but the RoC listing process has no room for advances in scientific knowledge on this crucial margin.

Third, the RoC listing criteria simply abdicates any responsibility for evaluating causation. The RoC accomplishes this by simply assuming that epidemiological association is causation. In this regime, ignorance trumps knowledge because it is sufficient to accomplish the agency's purposes.

Fourth, the RoC listing process relies on an undisclosed weight-of-evidence (WoE) scheme that puts no apparent weight on equivocal and negative evidence. This is a fatal defect on both technical and procedural grounds. Scientifically, no responsible authority would counsel ignoring equivocal and negative evidence. Procedurally, the absence of transparent and reproducible WoE guidelines renders the RoC listing process politically illegitimate.

Fifth, the RoC listing process includes no procedure whereby NTP determinations believed to be false positives can be reversed by scientific challenge. The NTP has no criteria that could be used to predict how it would process the information contained in any such challenge or the criteria it would use to evaluate it. A listing procedure that is immune to advances in scientific knowledge is not merely inadequate; it is anti-scientific.

2. One fatal policy defect

It is clear that the NTP's listing process contains too much scientific controversy for the RoC to be considered scientific. Rather, what the NTP has created is a method for making regulatory policy decisions Congress never authorized in a way that appears scientific in order to evade accountability. The façade of science is useful to the NTP because the authority of science is publicly respected; witness the propensity of Congress to delegate hard policy decisions to the National Academy of Sciences. But it is also politically feared: When ostensibly scientific reports such as the RoC are challenged, it is easy to accuse the challenger with the grave political sin of "interfering with science."

Nonetheless, the NTP's effort to secure political legitimacy by hiding regulatory policy decisions behind science is wearing thin. The styrene and formaldehyde cases studies, in particular, strip away the façade of science from the RoC process and expose the politics that control the NTP's listing process. Faced with evidence that its strong and ostensibly scientific inferences had no scientific foundation, the NTP declined even to attempt a defense. Instead, the NTP battened down its bureaucratic hatches to ride out the storm, knowing that eventually calmer weather would return. The NTP proved incapable of or unwilling to allow its policy decisions to be informed by the best available science.

E. What Can Congress Do to Fix This?

To make the RoC the science compendium Congress appears to have intended, Congress would need to legislate significant reforms. None of these reforms involves radical surgery. Here are six suggestions.

1. Direct the NTP to make its binary determinations conditional on dose

This reform would revive the portion of the statutory change that the NTP has largely ignored: exposure. The law directs the NTP to limit the RoC list to substances for which a "significant" number of persons residing in the United States are exposed, The NTP has interpreted this to mean exposure or any number of people at any level. The law directs the NTP to estimate the number of persons who are exposed at various levels, a task it does not perform at all.

Congress could make RoC listings somewhat exposure-based by adding a second dimension to the NTP's existing reporting scheme, illustrated in Table I with hypothetical scientific conclusions. This would require the NTP to make its listing decisions conditional on the level of exposure.

Table I: Exposure-Based Classifications (with illustrative hypothetical determinations)

	Exposure Domain		
	Ambient Environmental	US Occupational	High-Dose Laboratory
"Known" Human Carcinogen	✗	✗	✓
"Reasonably Anticipated to Be" a Human Carcinogen	✓	✓	✗

The advantage of this approach is obvious. For the public, it would yield much more useful information than the current classification scheme. The NTP also would benefit by getting it out of the business of making decisions that cannot accommodate even minimal scientific complexity. Substance listings would finally take account of the fundamental principle of toxicology—that the dose, not the molecule, makes the poison.

2. Direct the NTP to include potency in its listing decisions

When the RoC process was set in motion more than 30 years ago, little was known about the mechanisms of carcinogenesis. It may have made sense in the face of that ignorance to describe cancer in binary terms. That has not been true for many years, however.

Extensive mechanistic research has enabled scientists to offer useful insights concerning the relative potency of substances that, under certain conditions, may cause cancer. For the same reason that the practical utility of the RoC depends on the level of exposure to

which it applies, potency also matters. It is misleading to report substances with the same carcinogenicity label when their potencies vary by orders of magnitude.

A problem that Congress would have to address is that the conventional practice in cancer potency estimation is to derive values that are purposely biased to avoid understating potential risk.[67] This practice confuses science with policy and it makes valid comparisons across substances impossible. A scientific practice would be to derive unbiased estimates of cancer risks and leave to risk managers the discretion to decide what to do about them.

The solution is for Congress to explicitly direct the NTP to make its potency estimates unbiased. Where uncertainties exist that cannot be resolved without policy judgment, multiple potency estimates should be derived and presented without policy judgments embedded. Tools exist for eliciting the subjective, but nevertheless scientific views of experts. The NTP could learn to use them but it must ensure that nonscientific policy judgments are excluded.

3. Replace problematic risk descriptors or provide guidance concerning how to interpret statutory them

If Congress wants the RoC to be a scientific compendium, it must abandon its reliance on nonscientific descriptors like "known" and "reasonably anticipated to be" human carcinogens. A better approach is explicitly state alternative levels of concern in units scientists understand, such as probabilities. A useful example is provided by the Intergovernmental Panel on Climate Change, which adopted the scheme set forth in Table II. While all of the descriptors on this list may not be necessary, it is clear that more than two is needed (IARC uses four; EPA uses five) and that mapping descriptors to probabilities makes the process more sensible and the output easier for everyone to understand. Explicit minimum probabilities are essential to faithfully interpret the statutory text in a transparent and reproducible manner.

[67] These practices are often misleadingly described as "conservative," a term that lacks a clear meaning. The closest dictionary definition of "conservative" in this context is "cautious and on the low side." In the cancer risk context, "conservative means "cautious and on the "high side." It is no wonder that the public is confused.

Even if Congress were disinclined to replace the existing problematic descriptors "known" and "reasonably anticipated to," it could significantly improve the quality of the RoC by providing the NTP with guidance concerning how to interpret them. The mapping of probabilities and descriptors in Table II has the advantage of closely approximating public intuition, an essential attribute. In daily use, one does not "reasonably anticipate" that it will rain if the Weather Service says the likelihood of precipitation is, say, 30%. The NTP appears to deem substances "reasonably anticipated" human carcinogens when the likelihood is well south of that probability.

4. Direct the NTP to establish a strictly scientific weight-of-evidence scheme for making listing decisions

Every classification system requires a scheme for making weight-of-evidence (WoE) determinations. To be scientifically and politically legitimate, WoE schemes must be fully transparent and reproducible. The NTP has a WoE scheme, but it has never disclosed it.

Congress could induce a great deal of quality improvement in the RoC if it required the NTP to make its WoE scheme public. That instruction alone would be insufficient, however, because the WoE scheme that the NTP would disclose would be neither transparent nor reproducible.[68] Thus, Congress would have to accompany a disclosure requirement with parallel transparency and reproducibility requirements. Even better, 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.

5. Sunsetting to Encourage Revision

The NTP listing process operates somewhat like a high stakes poker game that is played only once in most cases. This creates perverse incentives both in the research community and within the NTP. Once a substance is deemed a human carcinogen, the incentive to conduct further research is seriously attenuated. The NTP's delisting process is cumbersome, expensive, extraordinarily time-consuming, and fraught with uncertainty about how the agency will process scientific information that conflicts with its prior decisions.

[68] If the NTP had a transparent and reproducible WoE scheme for listing decisions, it would have disclosed it long ago.

Within the NTP, the incentive to stay abreast of new science becomes severely weakened as action on the next round of nominations occupies the available resources. Listing decisions become ossified, with all bureaucratic energy devoted to expanding the list rather than ensuring its accuracy.

These problems could be reduced if the NTP had an affirmative obligation to review its previous listings on a set schedule. Under the current process, however, reviews of listed substances must pass a significant bureaucratic gauntlet. They must be nominated for review, and be accepted by the NTP and its federal partners. This seems very unlikely to occur except in extraordinary circumstances.

6. Direct the NTP to faithfully comply with applicable Information Quality Guidelines

The NTP's various reports are covered by applicable Information Quality Guidelines (IQG) (Office of Management and Budget 2002; National Institutes of Health 2002). The IQG requires, among other things, that scientific information disseminated by federal agencies be substantively and presentationally objective. Adhering to these guidelines would dramatically improve the scientific quality of NTP work products and reduce the propensity of the agency to make

Table II: Mapping Descriptors to Probabilities

Probabilistic Term	Likelihood of the occurrence/outcome
Virtually certain	> 99% probability
Very likely	> 90% probability
Likely	> 66% probability
About as likely as not	33 to 66% probability
Unlikely	< 33% probability
Very unlikely	< 10% probability
Exceptionally unlikely	< 1% probability

sweeping policy decisions based on incorrect information. Unfortunately, the NTP's history of responding to information quality error correction requests has not been satisfactory (Graham 2004).

Currently there is no practical way to compel the NTP to adhere to these guidelines. Agencies are given a weak presumption that can be rebutted by persuasive evidence, but under existing rules it is the agencies themselves who must be persuaded. The IQG's administrative procedures for error correction thus have limited effectiveness because they are constrained by a built-in conflict of interest. External enforcement is handicapped because federal courts to date have been unwilling to permit legal challenges to agency noncompliance.

Congress could achieve a highly significant reform simply by making adherence to the IQG a statutory imperative.

7. Fundamental Change

Each of the previous options involves nothing more than tinkering at the legislative margins. But Congress should reconsider whether the RoC, especially as it is currently structured, still serves a genuine public need for which there are no adequate substitutes. A case can be made for having multiple authorities create lists such as the RoC provided that they are competitively engaged in providing a useful, high-quality product. The RoC and its counterparts within the federal government and elsewhere do not appear to be competing, however. For naphthalene and formaldehyde, it appears that the NTP may have been primarily motivated by classification decisions made by IARC, with which there was a real or perceived need to conform. For styrene, the NTP went well beyond what IARC had done (Group 2B, "possibly" carcinogenic to humans) but it seems to have heavily relied on (or perhaps over interpreted) some of IARC's inferences about certain studies. Congress also should consider whether the production of lists such as the RoC constitutes an essential governmental function in an era of many competing goals and constrained resources.

F. What Can the NTP Do on its Own?

There also are many ways the NTP could improve the quality of the RoC without appealing to Congress for a change in its mandate.

1. Institute genuine, not superficial, process reforms

The case studies are not dispositive but they indicate that NTP can be cavalier about procedure, particularly with regard to public participation. In each of the three cases illustrated, there were substantial complaints about the lack of due process and agency unresponsiveness that are easily disentangled from disputes about science and its interpretation. In the case of naphthalene, which was first reviewed for the 11th RoC, the NTP issued its draft background document for public comment after the expert panels had completed their reviews. Significant new peer reviewed science was published before the 12th RoC, but the NTP staff either did not notice it or were uninterested in updating the substance profile to make a change in the listing possible.

Process problems acknowledged to have occurred in the production of the 11th RoC were the reason changes were made for the 12th edition. The styrene and formaldehyde cases show that these changes were, at best, ineffectively implemented. From the outside, it appears that the problem lies with an internal culture that is insular, defensive of its policy prerogatives, and envious of the superior technical competence that often can bring to bear by those who disagree. For process reforms to be effective, they must actively seek to achieve cultural change within the NTP. Each of the technical reforms suggested below would make this easier.

2. Adopt a streamlined probabilistic mapping of the statutory language

The NTP could adopt a version of Table II based on the existing statutory language, such as provided in Table III. The choice of minimum percentages is a policy decision, but it should be guided by how the terms are understood intuitively.

3. Adopt a probabilistic mapping of evidentiary descriptors

Table III: Suggested Mapping Existing Carcinogen Classifications to Probabilities

Terminology	Likelihood of the occurrence/outcome
"Known"	≥ 95% probability
"Anticipated to be"	≥ 75% probability

"Sufficient," "limited, " "less than sufficient," and similar descriptors have no recognizable scientific content. They obfuscate more than they inform. The NTP could replace them with a subset of the descriptors in Table IV below. No longer would the NTP encounter endless argument about what its descriptors actually mean.

This is not to say that there wouldn't be disputes about whether a particular strand of evidence deserves, say, "high confidence" (~80% likelihood) or "medium confidence" (~50% likelihood). Such disputes are inevitable. But it is immeasurably more useful to have a debate over likelihoods expressed in probabilistic terms than over adjectives that can only be interpreted subjectively. To the extent that some scientists consider "sufficient" to mean ~90% likelihood and

Table IV: Mapping Strength-of-Evidence Descriptors

Terminology	Degree of confidence in being correct
Very High confidence	At least 9 out of 10 chance
High confidence	About 8 out of 10 chance
Medium confidence	About 5 out of 10 chance
Low confidence	About 2 out of 10 chance
Very low confidence	Less than 1 out of 10 chance

others believe that ~50% likelihood will do, the use of probabilistic descriptors would eliminate that confusion.

4. Establish a genuinely *scientific* peer review process

The NTP's peer review process has fundamental flaws that the reforms put in place for the 12th RoC did not even begin to address. Reviews consist of a confused mix of limited scientific review structured to produce a consensus judgment that the NTP's proposed policy decision has "enough" scientific support in order to proceed. In practice, the burden of proof rests with a majority of reviewers to show that it does not. The scientific component of the review is structured around the NTP's synthesis of positive evidence favoring listing; it is not a review of the full scientific record or even a balanced subset of it, and the review of original scientific data and analyses is generally not part of the task. The NTP staff acts as a filter substantially controlling the information made part of the review.

Once a consensus opinion has been reached that the NTP synthesis is not inadequate, reviewers are tasked with providing policy advice all the while knowing exactly what advice their client wants to receive. This advice consists of making a pair of binary choices—first, whether to list; and second, whether to list as "known" or "reasonably anticipated."[69] These decisions are constrained by the RoC listing criteria, with which review panels are not permitted to dissent, even on scientific grounds.[70]

Reviewers may couch their policy advice in scientific language, and indeed the more scientific the language the more attractive it is to the NTP, but there is no way to escape the fact that peer reviewer's—scientists—are providing policy advice when they recommend a listing decision. By virtue of being scientists, typically

[69] If the existing listing is "reasonably anticipated," then the first binary decision has already been made.

[70] An unwillingness to be constrained by the RoC listing criteria on the ground that it is nonscientific presumably would result in non-selection in the first instance. Members may disagree about whether evidence is "sufficient" or "limited," but they may not disagree with the assumption that "carcinogenic activity" in two high-dose animal tests automatically means a substance is "reasonably anticipated to be" a human carcinogen.

tenured employees of a major university, they are granted social standing and implicit political authority far beyond their expertise.[71]

The scientific quality of the BSC review process is far too superficial to be effective. Members cannot genuinely review all the scientific literature presented in NTP draft substance profiles, and they are supposed to also have reviewed the often extensive public comments beforehand. A review of meeting minutes suggests that BSC reviews are mostly exercises in ratifying completed agency work products. Sometimes members put up a bit of a fuss, but in the end they dutifully capitulate.

Even if these defects are ignored, the NTP's peer review process has other significant internal defects. Many NTP peer review panels include members with institutional or intellectual conflicts of interest. A representative from the California Environmental Protection Agency was allowed to serve as an expert panel member for formaldehyde despite the fact that her agency was known to have already taken a position (Denton 2010). A senior Cal-EPA scientist served on the styrene expert panel despite the fact her agency was known to be preparing to list it under Prop 65. A senior styrene epidemiologist also was permitted to serve on this panel, thus being placed in the position of indirectly reviewing her own work. As noted above in the naphthalene case study, the BSC review lost scientific credibility when its chairman used it to transparently advance his preferred public policy. Most disturbingly, none of his colleagues complained; apparently, none of them even recognized the clear impropriety of using an ostensibly scientific peer review as a forum for policy advocacy.

The NTP should seriously consider replacing its existing peer review model with a fundamentally different one, one that scrupulously excludes all policy issues from its domain. In short, it should stop asking scientists for policy advice, and convince scientists and stakeholders that it has done so. As long as the NTP's peer reviews seek policy advice, it will be impossible to overcome the perception (if not the reality) that the ratification of agency policy judgments is the real reason the NTP sponsors peer review.

[71] This phenomenon is not unique to the NTP's peer review process. Congress frequently resolves its inability to reach difficult political decisions by delegating the matter to the National Academy of Sciences.

Operationally, peer review needs to expressly permit, if not require, scientists to interrogate the data and methods directly. Where there are competing teams of epidemiologists, for example, a better way to ascertain which one has the better case is to let them debate it in public session in a setting where all stakeholders can participate, there is ample time, and the agency does not exert control. Ideally, this should be done before the NTP staff produces its draft background document, not afterwards, for once an agency has set forth its views in print it becomes exceedingly difficult for it to change course.[72]

Another approach to peer review is the science audit.[73] Instead of reviewing a draft document to ascertain whether it is "technically correct, clearly stated and supports the NTP 's policy decision,"[74] a practice that requires reviewers to mistakenly search for consensus above all other things, a science audit would focus on transparency, reproducibility, and accountability. Transparency requires that every step be fully disclosed and explained. Reproducibility requires that properly trained individuals can utilize the same data and methods and achieve substantially the same result. In combination, they mean that every item in a scientific conclusion can be traced back to its component parts, with every material uncertainty in that process clearly identified and its implications described as quantitatively as possible.

Neither the current listing process, nor the current peer review process, nor the new "response to comments" step that the NTP added for the 12th RoC, have any of these desirable properties. Adopting the reforms suggested here would not be effective without some form of scientific auditing, but to be fair, the NTP's existing listing and peer review processes would never survive one.

VIII. References

[72] For an example of this alternative peer review process successfully implemented, see Belzer et al. (2008).

[73] The proposals here are heavily influenced by work of Judith Curry to fix the problems that beset peer review of global climate change research.

[74] (National Toxicology Program 2011h) As noted above, the NTP should abandon the practice of seeking policy advice—asking if the document "supports the NTP's policy decision."

Andersen, Melvin E. 2011. *Mode of Action Studies Do Not Support Distant Site Carcinogenic Targets for Formaldehyde*. The Hamner Institutes for Health Sciences, October 19, 2009 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Andersen20091016.pdf.

———. 2011. *Comments on the Recommendation from the Expert Panel Report (Part B) on Formaldehyde, 74 Fed. Reg. 67,883 (December 21, 2009)*. The Hamner Institutes for Health Sciences, February 2, 2010 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Anderson20100202.pdf.

Banton, Marcy. 2011. *NAS Formaldehyde Report Supports SIRC Concerns with Deficiencies of NTP Report on Carcinogens Process*. Styrene Information and Research Center, May 16, 2011 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2010/PublicComms/SIRC20110516.pdf.

Banton, Marcy, George Cruzan, and James Bus. 2011. *Letter to Linda S. Birnbaum Re: Knock-Out Mice Research Supports Conclusion that Styrene is NOT a Carcinogen; Request to Remove Styrene from 12th Report on Carcinogens Based on Lack of Scientific Validation*. Styrene Information and Research Center, October 5, 2010 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/Styrene/SIRC20101005.pdf.

Belzer, Richard B., James S. Bus, Ercole L. Cavalier, Steven C. Lewis, D. Warner North, and Richard C. Pleus. 2008. The naphthalene state of the science symposium: Objectives, organization, structure, and charge. *Regulatory Toxicology and Pharmacology* 51 (2(1)):1-5.

Berkman, Richard L., and W. Kip Viscusi. 1973. *Damming the West*. New York: Grossman Publishers.

Bogen, Kenneth T., Janet M. Benson, Garold S. Yost, John B. Morris, Alan R. Dahl, Harvey J. Clewell III, Kannan Krishnan, and Curtis

J. Omiecinski. 2008. Naphthalene metabolism in relation to target tissue anatomy, physiology, cytotoxicity and tumorigenic mechanism of action. *Regulatory Toxicology and Pharmacology* 51 (2(1)):27-36.

Bradford Hill, Austin. 1965. The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine* 58:295-300.

Brusick, David. 2008. Critical assessment of the genetic toxicity of naphthalene. *Regulatory Toxicology and Pharmacology* 51 (2(1)):37-42.

Brusick, David, Mitchell S. Small, Ercole L. Cavalieri, Dhrubajyoti Chakravarti, Xinxin Ding, David G. Longfellow, Jun Nakamura, Eleanor C. Rogan, and James A. Swenberg. 2008. Possible genotoxic modes of action for naphthalene. *Regulatory Toxicology and Pharmacology* 51 (2(1)):43-50.

Bus, James S., and George Cruzan. 2011. *Letter to Linda S. Birnbaum RE: Finalization of NTP Report on Carcinogens Review of Styrene*, April 29, 2009 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/june/SIRC20090429.pdf.

Collins, James J. 2011. *Comments on the Draft Background Document on Formaldehyde Exposure and Risk of Cancer*. Dow Chemical Company,, October 16, 2009 [cited September 3, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Collins20091016_E.pdf.

Cruzan, George, James Bus, Marcy Banton, Ralph Gingell, and Gary Carlson. 2011. *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 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/SIRC_RTP-S-09-000231.pdf.

de la Cruz, Peter. 2011. *Report on Carcinogens--Legislative History*. Keller and Heckman LLP April 23, 2009 [cited September 6, 2011]. Available from

http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/Styrene/SIRC20090501.pdf.

Delzell, Elizabeth. 2011. *Comments on the RoC Background Document for Styrene and Related Documents*, October 10 2008 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/SIRC20081023A.pdf.

———. 2011. *Letter to Barbara Shane, Executive Secretary, National Toxicology Program Board of Scientific Counselors*, February 5, 2009 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/20090205Delzell.pdf.

Denton, Joan. 2011. *Letter to Dr. Ruth Lunn, Director, Report on Carcinogens Center*. California Offic of Environmental Health Hazard Assessment, January 14, 2010 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Denton20100114.pdf.

Eaton, David L., and Curtis D. Klaassen. 2001. Principles of Toxicology. In *Casarett & Doull's Toxicology: The Basic Science of Poisons*, edited by C. D. Klaassen. New York: McGraw-Hill Medical Publishing Division.

European Union Chemicals Agency. 2011. *European Union Risk Assessment Report: Styrene*. European Union Chemical Agency, June 2008 [cited October 15 2011]. Available from http://echa.europa.eu/doc/trd_substances/styrene/rar/trd_rar_uk_styrene.pdf.

Faustman, Elaine M., and Gilbert S. Omenn. 2001. Risk Assessment. In *Casarett & Doull's Toxicology: The Basic Science of Poisons*, edited by C. D. Klaassen. New York, N.Y.: McGraw-Hill Medical Publishing Division.

Golden, Robert. 2011. *Comments on the Report on Carcinogens Draft Background Document for Formaldehyde, September 3, 2009, U.S. Department of Health and Human Services, Public Health Service, National Toxicology Program: Biological Mechanisms for Formaldehyde-Induced Leukemia*. ToxLogic LLC, October 16, 2009 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Golden20091016_C.pdf.

Graham, John D. 2011. *Letter to Dr. Elias A. Zerhouni, Director, National Institutes of Health*. Office of Management and Budget, November 16, 2004 [cited September 20, 2011]. Available from http://www.reginfo.gov/public/prompt/nih_ntp111604.pdf.

Griego, Fumie Y., Kenneth T. Bogen, Paul S. Price, and Douglas L. Weed. 2008. Exposure, epidemiology and human cancer incidence of naphthalene. *Regulatory Toxicology and Pharmacology* 51 (2(1)):22-26.

Hauptmann, Michael, Jay H. Lubin, Patricia A. Stewart, Richard B. Hayes, and Aaron Blair. 2003. Mortality From Lymphohematopoietic Malignancies Among Workers in Formaldehyde Industries. *Journal of the National Cancer Institute* 95 (21):1615-1623.

———. 2004. Mortality from Solid Cancers among Workers in Formaldehyde Industries. *American Journal of Epidemiology* 159 (12):1117-1130.

Hotelling, Harold. 1929. Stability in Competition. *Economic Journal* 39 (153):41-57.

International Agency for Research on Cancer. 2002. *Man-made Vitreous Fibres*, *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon, FR: International Agency for Research on Cancer.

———. 2006a. *Formaldehyde, 2-Butoxyethanol and 1-tert-Butoxypropan-2-ol*, *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Lyon, FR: International Agency for Research on Cancer.

———. 2006b. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Lyon, FR: International Agency for Research on Cancer.

Kogevinas, Manolis, Gilles Ferro, Aage Andersen, Tom Bellander, Marco Biocca, David Coggon, Valerio Gennaro, Sally Hutchings, Henrik Kolstad, 8lngvar Lundberg, Elsebeth Lynge, Timo Partanen, and Rodolfo Saracci. 1994. Cancer Mortality in a Historical Cohort Study of Workers Exposed to Styrene. *Scandinavian Journal of Work, Environment & Health* 20 (4):251-261.

Kohler, Betsy A., Elizabeth Ward, Bridget J. McCarthy, Maria J. Schymura, Lynn A. G. Ries, Christie Eheman, Ahmedin Jemal, Robert N. Anderson, Umed A. Ajani, and Brenda K. Edwards. 2011. Annual Report to the Nation on the Status of Cancer, 1975–2007, Featuring Tumors of the Brain and Other Nervous System. *Journal of the National Cancer Institute* 103 (9):1-23.

Marsh, Gary M. 2011. *Comments on National Toxicology Program Draft Background Document for Formaldehyde For Presentation at the NTP Report on Carcinogens, Expert Panel Meeting Research Triangle Park, NC, November 2, 2009*, November 2, 2009 [cited September 3, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Marsh20091016_A.pdf.

———. 2011. *Comments on the Recommendation from the Expert Panel Report (Part B) on Formaldehyde, 74 Fed. Reg. 67,883 (December 21, 2009)*. University of Pittsburgh, February 8, 2010 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Marsh20100208.pdf.

Mundt, Kenneth A. 2011. *A Meta-Analysis of Formaldehyde Exposure and Risk of Leukemia and Nasopharyngeal Cancer*. ENVIRON International Corporation, October 16, 2009 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Mundt20091016_D.pdf.

Mundt, Kenneth A., Philip Cole, Richard Irons, Gary M. Marsh, and Jack S. Mandel. 2011. *Letter to Linda S. Birnbaum Re: NTP Evaluation and Classification of Formaldehyde for the 12th Report on Carcinogens*, November 12, 2010 [cited September 3, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2010/PublicComms/Mundt_etal_2010Nov12.pdf.

National Cancer Institute. 2011. *Dictionary of Cancer Terms; "Maximum Tolerated Dose"*. Narional Cancer Institute 2011 [cited October 10, 2011]. Available from http://www.cancer.gov/dictionary/?CdrID=546597.

National Institutes of Health. 2007. *HHS Guidelines for Ensuring the Quality of Information Disseminated to the Public; National*

Institutes of Health. U.S. Department of Health and Human Services, December 13, 2006 2002 [cited November 12, 2007]. Available from http://aspe.hhs.gov/infoquality/Guidelines/NIHinfo2.shtml.

National Research Council. 1983. *Risk Assessment in the Federal Government: Managing the Process*. Washington, D.C.: National Academies Press.

———. 1993. *Issues in Risk Assessment*. Washington, D.C.: National Academies Press.

———. 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, D.C.: National Academies Press.

———. 2011. *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde*. Washington, D.C.: National Academies Press.

National Toxicology Program. 1992. Toxicology and Carcinogenesis Studies of Naphthalene (CAS NO. 91-20-3) in B6C3Fl Mice (inhalation Studies). Research Triangle Park, N.C.: National Toxicology Program Technical Report Series.

———. 2000. Toxicology and Carcinogenesis Studies of Naphthalene (CAS no. 91-20-3) in F344/N Rats (inhalation studies). Research Triangle Park, N.C.: National Toxicology Program Technical Report Service.

———. 2002a. Condensed Transcript; Board of Scientific Counselors Report On Carcinogens (ROC) Subcommittee Meeting, November 19, 2002. Stephenson, Va.: County Court Reporters, Inc.

———. 2011. *National Toxicology Program Board of Scientific Counselors Report on Carcinogens Subcommittee Meeting*. National Toxicology Program, November 19-20, 2002b [cited September 29, 2011]. Available from http://ntp.niehs.nih.gov/ntp/htdocs/Liaison/111902.pdf.

———. 2011. *NTP Public Meeting Report On Carcinogens (RoC) Review Process*. National Toxicology Program, January 27, 2004a [cited September 30 2011]. Available from http://web.archive.org/web/20100605093823/http://ntp.niehs.nih.gov/ntp/Meetings/2004/12229-3niehsconference.pdf.

———. 2004b. *Report on Carcinogens*. 11th ed, *Report on Carcinogens*. Research Triangle Park, N.C.: National Toxicology Program.

———. 2011. *Nomination Review Process for the 11th Report on Carcinogens (RoC)*. National Toxicology Program, March 30, 2005a [cited October 10, 2011]. Available from http://web.archive.org/web/20051031042143/http://ntp.niehs.nih.gov/ntpweb/index.cfm?objectid=B2580696-F1F6-975E-726E312CCF078AAE.

———. 2011. *NTP Study Reports; Definition of Carcinogenicity Results*. National Toxicology Program, May 12, 2005b [cited October 10, 2011]. Available from http://ntp.niehs.nih.gov/?objectid=07027D0E-E5CB-050E-027371D9CC0AAACF.

———. 2011. *Proposed Review Process for the 12th Report on Carcinogens*. National Toxicology Program, August 17, 2006 [cited September 29, 2011]. Available from http://web.archive.org/web/20060930221330/http://ntp.niehs.nih.gov/index.cfm?objectid=720162B0-BDB7-CEBA-FE2B27BBA2785BA5.

———. 2011. *NTP Response to Comments Received on the Draft Report on Carcinogens Review Process* 2007 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/?objectid=FA9343EC-F1F6-975E-7143BACBFE5CE2C2.

———. 2011. *Summary Minutes June 21-22, 2010 NTP Board of Scientific Counselors*. National Toxicology Program, Jujne 21-22, 2010 2010 [cited September 26 2011]. Available from http://ntp.niehs.nih.gov/ntp/About_NTP/BSC/2010/June/Minutes20100622.pdf.

———. 2011. *Addendum to the 12th Report on Carcinogens* 2011a [cited October 3, 2011]. Available from http://ntp.niehs.nih.gov/ntp/roc/twelfth/Addendum.pdf.

———. 2011. *How NTP Studies are Used to Protect Human Health*, June 14, 2011 2011b [cited September 26 2011]. Available from http://ntp.niehs.nih.gov/index.cfm?objectid=03612A12-9F5F-C336-79B4709B8013F338.

———. 2011. *NTP Report on Carcinogens Review Process*. National Toxicology Program, October 3, 2011c [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/?objectid=FA925F34-F1F6-975E-775C81773747D452.

———. 2011. *NTP Study Reports; Long-Term Study Reports & Abstracts*. National Toxicology Program, September 15, 2011 2011d [cited September 23 2011]. Available from http://ntp.niehs.nih.gov/index.cfm?objectid=084801F0-F43F-7B74-0BE549908B5E5C1C.

———. 2011e. *Report on Carcinogens (12th Edition)*. Research Triangle Park, N.C.

———. 2011. *Report on Carcinogens; NTP Report on Carcinogens Review Process*. National Toxicology Program, June 15, 2011 2011f [cited September 23 2011]. Available from http://ntp.niehs.nih.gov/?objectid=FA925F34-F1F6-975E-775C81773747D452.

———. 2011. *Report on Carcinogens; NTP Report on Carcinogens Review Process*. National Toxicology Program 2011g [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/images/12thProcess-large.jpg.

———. 2011. *Review Process for the 12th Report on Carcinogens*. National Toxicology Program, June 15, 2011h [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/?objectid=FA925F34-F1F6-975E-775C81773747D452.

———. 2011. *NTP Response to Issues Raised in the Public Comments for Candidate Substances for the 12th Report on Carcinogens* n.d. [cited October 1 2011]. Available from http://ntp.niehs.nih.gov/ntp/roc/twelfth/2011/ResponsePublicComments2011.pdf.

Natz, Betsy. 2011. *Formaldehyde - Proposed Nomination for Review in the 12th Report on Carcinogens*. Formaldehyde Council, November 17, 2005 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/B_NatzFCI.pdf.

———. 2011. *Comments on the National Toxicology Program's Draft Background Document for Formaldehyde*. Formaldehyde Council,

October 16, 2009 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Natz20091016.pdf.

———. 2011. *Letter to Lori White Re: Comments for Consideration by the NTP Board of Scientific Counselors on the Draft Substance Profile for Formaldehyde, 75 Fed. Reg. 21,003 (April 22, 2010)*. Formaldehyde Council, June 7, 2010a [cited September 6, 2011].

———. 2011. *Letter to Ruth Lunn RE: Comments on the Recommendation from the Expert Panel Report (Part B) on Formaldehyde, 74 Fed. Reg. 67,883 (December 21, 2009)*. Formaldehyde Council, February 11, 2010b [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/NTP/RoC/twelfth/2009/November/Public_Comments/Natz20100211.pdf.

North, D. Warner, Kamal M. Abdo, Janet M. Benson, Alan R. Dahl, John B. Morris, Roger Renni, and Hanspeter Witschi. 2008. A review of whole animal bioassays of the carcinogenic potential of naphthalene. *Regulatory Toxicology and Pharmacology* 51 (2(1)):6-14.

Office of Management and Budget. 2002. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication. *Federal Register* 67 (36):8452-8460.

———. 2005. Final Information Quality Bulletin for Peer Review. *Federal Register* 70 (10):2664-2667.

Office of Technology Assessment. 1987. *Identifying and Regulating Carcinogens*. Washington, D.C.: U.S. Government Printing Office.

Pitot III, Henry C., and Yvonne P. Dragan. 2001. Chemical Carcinogenesis. In *Casarett & Doull's Toxicology: The Basic Science of Poisons*, edited by C. D. Klaassen. New York: McGraw-Hill Medical Publishing Division.

Price, Courtney M. 2011. *Comments of the Naphthalene Panel of the American Chemistry Council in Response to NTP's Request for Comments on the Nomination of Naphthalene for Possible Listing in the Report on Carcinogens*. American Chemistry Council,

September 24, 2001 [cited October 4, 2011]. Available from http://web.archive.org/web/20090826033525/http://ntp.niehs.nih.gov/ntp/roc/pbcarchive/11th/naphth/price-nap-09-24-01.pdf.

———. 2011. *Comments on National Toxicology Program Draft Report on Carcinogens Background Document for Napthalene, 26 August 2002*. American Chemistry Council, October 2, 2002a [cited September 23, 2011]. Available from http://web.archive.org/web/20090826031951/http://ntp.niehs.nih.gov/ntp/roc/pbcarchive/11th/naphth/price-10-03-02.pdf.

———. 2011. *Letter to C.W. Jameson Re: National Toxicology Program (NTP) Executive Committee Working Group for the Report on Carcinogens (RG2) Naphthalene Review (October 2, 2002)*. American Chemistry Council, November 4, 2002b [cited September 23, 2011]. Available from http://web.archive.org/web/20090826033930/http://ntp.niehs.nih.gov/ntp/roc/pbcarchive/11th/naphth/price-11-04-02.pdf.

———. 2011. *Letter to C.W. Jameson Re: "National Toxicology Program (NTP), Call for Public Comments on 10 Nominations; Proposed for Listing in the Report on Carcinogens, eleventh Edition; Feder register, January 22, 2003 (Vol. 68. No. 14)*. American Chemistry Council, November 4, 2003 [cited September 23, 2011]. Available from http://web.archive.org/web/20090826033409/http://ntp.niehs.nih.gov/ntp/roc/pbcarchive/11th/naphth/price-nap-03-24-03.pdf.

Price, Paul S., and Michael A. Jayjock. 2008. Available data on naphthalene exposures: Strengths and limitations. *Regulatory Toxicology and Pharmacology* 51 (2(1)):15-21.

Sebelius, Kathleen. 2011. *Charter: National Toxicology Program Board of Scientific Counselors* October 28, 2010 [cited October 13, 2011]. Available from http://ntp.niehs.nih.gov/ntp/About_NTP/BSC/2011/Charter_BSC_2011.pdf.

Snyder, Jack. 2011. *Re: Nomination of Styrene for Review for 12th Report on Carcinogens*. Styrene Information and Research Center, July 16 2004 [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/ntp/NewHomeRoc/RoC12/snyderj-07-16-04.pdf.

———. 2011. *Letter to Ruth M. Lynn RE: Comments on NTP Report on Carcinogens, Draft Background Document for Styrene*. Styrene Information and Research Center, July 7, 2008a [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/Snyder_SIRCJuly7.pdf.

———. 2011. *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, 2008b [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/SIRC20081023.pdf.

———. 2011. *Letter to Samuel H. Wilson*. Styrene Information and Research Center,, June 13, 2008c [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/SIRC_NTP_06-13-08_508sj2.pdf.

———. 2011. *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, 2009a [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/files/20090206SIRC.pdf.

———. 2011. *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, 2009b [cited September 6, 2011]. Available from http://ntp.niehs.nih.gov/ntp/roc/twelfth/2009/Styrene/SIRC20091022.pdf.

U.S. Environmental Protection Agency Office of Research and Development National Center for Environmental Assessment. 2004. Toxicological Review of Naphthalene (CAS No. 91-20-3) In Support of Summary Information on the Integrated Risk Information System (IRIS); External Review Draft. Washington, D.C.: U.S. Environmental Protection Agency.

U.S. Government Printing Office Style Board. 2008. *Style Manual: An Official Guide to the Form and Style of Federal Government Printing* Washington, D.C.: U.S. Government Printing Office.

Viscusi, W. Kip. 1979. *Employment Hazards*. Cambridge, Mass.: Harvard University Press.

Wagner, Wendy E. 1995. The Science Charade in Toxic Risk Regulation. *Columbia Law Review* 95 (7):1613-1723.