Regulation under Uncertainty
Use of the Linear No-Threshold Model in Chemical and Radiation Exposure

Dima Yazji Shamoun, Edward Calabrese, Richard A. Williams, and James Broughel

September 2016
Abstract

This paper examines the use of the linear no-threshold (LNT) model in chemical and radiation exposure. The LNT model assumes that exposure to any level of a chemical or radiation is harmful, down to even the last molecule. Used primarily to be “public health protective,” the model has been the backbone of chemical and radiation risk regulation for many decades. Given the current state of science and risk management tools, we challenge the notion that using the LNT as the default model is public health protective. First, more and more research has uncovered dose-response relationships that reveal either a threshold or, more importantly, a hormetic response, where exposure to low doses of a hazard actually yields health benefits. Second, given these more realistic alternative dose-response models, risk management tools including risk-risk analysis and health-health analysis show that regulating down to extremely low levels can have negative health consequences when ancillary risks are considered. Risk-risk analysis focuses on how reductions in target risks can lead to increases in risk from substitute chemicals or activities. Health-health analysis explores how costs of compliance are borne in part by consumers who are forced to reduce their own private risk-mitigating activities. Overestimating risk, a common feature of the LNT model, upsets the careful balancing of risks required of risk managers.

JEL codes: D62, D81, I18

Keywords: linear no-threshold, dose-response, health and safety regulations, benefit-cost analysis, risk assessment

Author Affiliation and Contact Information

Dima Yazji Shamoun
Associate Director of Research
Center for Politics and Governance
Lecturer, Economics Department
University of Texas at Austin
dshamoun@utexas.edu

Edward Calabrese
Professor of Toxicology
University of Massachusetts at Amherst

Richard A. Williams
Director, Regulatory Studies Program
Mercatus Center at George Mason University

James Broughel
Research Fellow, State and Local Policy Project
Mercatus Center at George Mason University

All studies in the Mercatus Working Paper series have followed a rigorous process of academic evaluation, including (except where otherwise noted) at least one double-blind peer review. Working Papers present an author’s provisional findings, which, upon further consideration and revision, are likely to be republished in an academic journal. The opinions expressed in Mercatus Working Papers are the authors’ and do not represent official positions of the Mercatus Center or George Mason University.
1. Introduction

The linear no-threshold (LNT) model has been the standard risk assessment model used for both chemical and radiation exposure for decades, particularly for low-dose exposure. While many assume that using the model provides for a public health–protective risk management decision, this remains to be proven. This paper challenges the notion that the LNT model is protective of the public health under conditions of uncertainty, in particular with modeling low-dose exposure. The meaning of “public health–protective” becomes less clear when there are offsetting increases in risk either to the target population or to an entirely different population.

We argue that there are three related assumptions, which are central to many risk assessments, that may lead to poor public health decisions: the LNT assumption, which might be thought of as a zero threshold assumption; the zero substitution effect assumption; and the zero income effect assumption. The LNT model, a widely applicable dose-response model in risk assessment—especially in cancer risk assessment—hypothesizes that exposure to even a single molecule of a hazard is sufficient to induce harm. By contrast, a threshold model assumes that exposure up to a certain dose is harmless, and a hormetic model hypothesizes that exposure to low doses of stressors is protective (i.e., beneficial) and only becomes harmful at higher doses.

The zero substitution effect assumption is that there are no risk-risk tradeoffs and, thus, a reduction in a target risk yields no unintended increases (or decreases) in other risks. Finally, the zero income effect assumption is that there are no health-health tradeoffs, meaning that
regulatory efforts to mitigate a target risk yield no offsetting increases in personal risks when private income is reduced by regulatory spending on health and safety.

If any one of these assumptions (or a combination of them) is found to be false, then public health may be compromised. The use of the LNT model, especially with its emphasis on conservatism, may lead to choices that increase the expected cost of risk-risk and health-health tradeoffs. Its widespread use could, for example, contribute to a culture among regulators whereby focus is aimed narrowly at target risks, but to the exclusion of countervailing risks, without consideration of diminishing marginal returns to public risk-reduction attempts, and in ignorance of private risk-reduction efforts.

We begin this paper with a background discussion of the history and origin of the LNT model. We then present a brief review of the recent scientific literature on hormesis, DNA repair, preconditioning, and adaptive responses in biology, challenging the foundational validity of linearity. Finally, we conclude with a discussion of tradeoff analysis, namely risk-risk analysis (RRA) and health-health analysis (HHA), which sheds light on the role of unintended consequences and opportunity costs in magnifying the potential health consequences of using the LNT model. Despite its widespread use, the LNT model is due for a reevaluation. In addition, because much of the health effect we are discussing occurs in the very low dose range, dose-response uncertainty, risk-risk tradeoffs, and health-health tradeoffs should be analyzed as part of risk management to improve public policy decisions and outcomes.

2. Background of the LNT Model

When estimating the risk from exposure to chemical hazards, neither epidemiological nor animal studies generally provide dose-response data in the relevant region for the average
human level of exposure, that is, the low-dose region. Due to the limitations of existing study protocols, extrapolations to possible responses in the relevant low-dose region are usually made from the level of response observed in the high-dose region. The LNT model assumption, which roughly connects the lowest dose-response point observed in animal studies to the origin, is the most common model used for extrapolation. For cancer risk assessments, in particular, it is the regulatory default,¹ and, in effect, it implies that there is no safe threshold for exposure to a carcinogen; exposure to even a single molecule of a carcinogen could cause harm proportional to the dose.

The adoption of the LNT model for cancer risk assessment stands at odds with the founding principle of toxicology that “the dose makes the poison.” To quote Paracelsus in full: “All substances are poisons; there is none which is not a poison. The right dose differentiates a poison from a remedy” (Kirsch-Volders, Aardema, and Elhajouji 2000). Yet the LNT model assumption eliminates consideration of a threshold and focuses only on the level of “presumed” poisonous effects. Recent work, however, continues to affirm the presence of repair, as the body has “demonstrated response to mitigate or eliminate [the] damage” from low dose radiation. (Sacks, Meyerson, and Sigel 2016).

Abandonment of the previously held threshold assumption constituted a significant paradigm shift in toxicology. Although “extraordinary claims require extraordinary proof,”² the LNT model was accepted as the default model for cancer risk assessment by US regulatory agencies without an extraordinary justification. In the following two subsections we provide a

¹ Some noncancer risk assessments make use of the LNT model as well, though they are more aberrational than customary. For example, the Environmental Protection Agency applies the LNT model to estimate the risk of exposure to low doses of some air pollutants, such as fine particulate matter (PM$_{2.5}$) and more recently to ozone (O$_3$).

² The original quote may have been from Marcello Truzzi (1987): “The more extraordinary a claim, the heavier is the burden of proof demanded.”
brief history of the LNT model and its application by regulatory agencies to ionizing radiation and then to chemical carcinogens in general (a historical summary is provided in the table in the appendix).

2.1. Adoption of the LNT model in the Assessment of Risk of Ionizing Radiation

Ever since the publication of Darwin’s 1859 work, *On the Origin of Species by Means of Natural Selection*, the question as to the cause of genetic change by which natural selection takes place has occupied the biology community (Calabrese 2013b). Evolution was seen to be driven by random mutations to individual genes, which would then be passed on to future generations (Muller 1922). But what was inducing the mutations?

Geneticists raced to discover this mechanism of evolution. They applied stressors ranging from temperature to ionizing and nonionizing radiation. In 1927, Nobel Prize–winning geneticist Hermann J. Muller initiated a discussion on the possibility that X-rays could lead to heritable mutations. Though the doses he used in his study were extremely high (200,000 times the background dose), he found a significant mutation rate, which led Olson and Lewis (1928) to speculate that naturally occurring ionizing radiation may be the process behind evolution (Calabrese 2013b).

Even given linearity in the low-dose region, however, the inducible mutation theory was ambitious, as it hypothesized that small doses of natural radiation could explain the full extent of evolution driven by genetic mutation. Despite the need for extraordinary proof and the emergence of several studies rejecting the LNT model interpretation (Patterson 1928), genetic mutation in response to ionizing radiation came to be the common assumption, requiring a new framework to accompany it.
This novel framework would emerge from a collaboration of geneticists and physicists. Before Muller’s theory of inducible mutation, medical physicists had envisioned that each cell has a sensitive area, or a heart, and that when the heart dies, the cell dies. According to this theory, known as target theory, cells are dosed with radiation that may result in “hits” that can then kill the cell. A cell can potentially survive a number of radiation doses, as not every dose will hit the heart (of a cell); also the heart may withstand several hits before it dies. Thus, target theories are modeled as \( x \)-hit target theory, where “\( x \)” denotes the number of hits it takes to kill the heart of the cell. For example, a single hit theory implies that the heart will be killed on the first hit (Nomiya 2013)—that is, response occurs in proportion to the dose.

Applying target theory to radiation-induced mutation advanced both the state of target theory and the LNT model for ionizing radiation. The formal justification of the linear dose response within the target theory framework appeared in an influential paper by radiation geneticist Timoféeff-Ressovsky and physicists Delbruck and Zimmer (1935). The paper hypothesized a binary reaction mechanism where an observable response (i.e., mutation) takes place when units of energy are absorbed (or ionized) by the target region (the particularly sensitive region, or the heart) of a gene. Once an X-ray treatment excites an electron in the target region of the gene, a permanent effect takes place in the form of a mutation (Calabrese 2013b). The units of energy are generally referred to as hits and thus the target theory of ionizing radiation is often referred to as the one-hit target theory.

The one-hit target theory of mutation stood at odds with the general physiological understanding of the time that the elimination of one molecule out of a very large number of molecules does not generate an observable effect. Even after the 1953 discovery of the structure of DNA (Watson and Crick 1953) came to replace most of what had been assumed about gene
structure (e.g., its molecular stability), the one-hit target theory continued to be applied. And in 1956 the theory even made its way to the Biological Effects of Ionizing Radiation I (BEIR I) committee formed by the National Academy of Sciences. There the geneticists comprising the BEIR I panel made the seminal recommendation to switch from a threshold model to a linear model to estimate the risk of mutation from ionizing radiation (Calabrese 2013b).

Still, as understanding of low-dose-induced DNA repair and recovery was making its way through the scientific community, some challenged the BEIR I decision. These challenges, however, did not succeed in reversing the BEIR I decision, and regulatory agencies in America and around the world followed BEIR I’s lead, adopting linearity in cancer risk assessment (Calabrese 2013b).

2.2. Adoption of the LNT Model in the Assessment of Risk of Other Stressors

In 1961 Nathan Mantel and W. Ray Bryan used a probit model to estimate the risk of developing cancerous tumors when exposed to carcinogens (Calabrese 2013b). They recommended a “safe dose” of 1 in 100 million. The common regulatory “tolerated” level of risk from exposure to carcinogens traces its origin to this publication. This safe-dose recommendation was adopted by the Food and Drug Administration (FDA) in its publication of the 1973 risk guidelines, but it was modified to 1 in 1 million in 1977.³ In 1979, the FDA revised its cancer risk assessment policy, replacing the probit model used by Mantel and Bryan with the LNT model.⁴

---
³ The 1 in 1 million level was the threshold below which no regulatory action was necessary.
⁴ In fact, there have been many mathematically based models used to extrapolate from high to low dose for carcinogenesis. One significant method used by the EPA early on came from K. S. Crump (1984). There were two-stage models (Armitage and Doll 1957), three-stage models (Neyman and Scott 1967) and the one-hit model from Moolgavkar and Venzon (1979) and Moolgavkar and Knudson (1981). These are discussed in Thorsland, Brown, and Charnley (1987). A more general discussion can be found in Anderson and the Carcinogen Assessment Group of the EPA (1983). For a thorough analysis of the history and evolution of dose-response modeling, see Calabrese (2013b).
The EPA took several measures in the 1970s to limit exposure to carcinogens. In its 1976 proposed guidelines on carcinogenic risk, the EPA recommended the use of quantitative risk assessments to estimate the risk of exposure to carcinogens. Based on limited epidemiological evidence on ionizing radiation and the link between smoking and lung cancer, the EPA also endorsed the use of the one-hit model (and thus a linear dose response) (Calabrese 2013b). According to EPA Administrator Douglas Costle, the one-hit model was chosen due to its conservative nature, that is, its perceived bias toward overestimation of risk in the presence of uncertainty (EPA 1976). Overestimation of risk was (and still is) considered consistent with the agency’s mission to protect public health from environmental chemical exposures. A later publication suggested that wide application of the LNT model in regulatory risk assessment was due in part to its attractiveness to regulators, namely, “It is easy to apply and . . . it will generate an upper bound on the unknown, underlying cancer risk in most instances.” (Office of Science and Technology Policy 1986). And the timing for the regulation of chemical carcinogens was simply right, following as it did on the heels of ionizing radiation, a mutagen with a readily available and widely used framework of analysis. So while the one-hit model was initially proposed for the mutational effects of ionizing radiation, it eventually became the default model for all chemical carcinogens.

In 1977 the Safe Drinking Water Committee (SDWC) of the National Academy of Sciences (NAS) recommended to the EPA the adoption of the LNT model in cancer risk assessment (Calabrese 2013b). The EPA followed this recommendation in 1979 in its assessment of the risk of chloroform in drinking water (Environmental Protection Agency 1979). The

---

5 The EPA’s website has a Quantitative Risk Assessment for Exposure to Vinyl Chloride (Kuzmack and McGaughy 1975) and Interim Procedures and Guidelines for Health Risk and Economic Impact Assessments of Suspected Carcinogens (Train 1976).
SDWC expressed skepticism on the grounds that the LNT model did not incorporate biological characteristics of the animal studies nor did it anticipate “newer developmental methodologies” (Calabrese 2013b). As a result, the SDWC briefly withheld its endorsement of the LNT model only to endorse it again in 1983, since the model was still in use by the EPA. From there, the LNT model became the default methodology for the assessment of risk of chemical carcinogens. These endorsements and the application of the LNT model, first by the FDA (1979) and then by the EPA (1979), were foundational steps in the history of regulatory risk assessments.

3. Recent Developments in Dose Response

Regardless of the reasons why regulatory agencies initially decided to use the LNT model, the debate should now be on whether there is sufficient evidence to justify maintaining its use. As we argue, there is mounting evidence in biology and toxicology (as well as risk management theories) to support reevaluation of the choice of dose-response model to optimize public health. The LNT model is difficult, if not impossible, to validate and, therefore, integrating other default models may allow for conducting validation exercises. Evidence of alternative dose-response models (e.g., hormesis) and biological mechanisms (e.g., DNA repair, preconditioning, and adaptive response) suggest that adherence to the LNT model may be imprudent, as it prevents public policy from achieving its full potential in protecting public health.

In fact, due to these issues of validation and plausibility, the Nuclear Regulatory Commission has recently started examining the validity of the LNT model as compared to the hormetic model for ionizing radiation (Nuclear Regulatory Commission 2016). In the next section we briefly outline three challenges from toxicology and biology to the LNT model,
namely, validation issues, hormesis as an alternative model, and, finally, research on DNA repair, preconditioning, and adaptive responses in biology.

3.1. Validation Issues

As noted above, it is extremely difficult, if not impossible in some instances, to validate the dose-response function at low doses, since thousands of subjects are required to uncover either a small response or a relatively infrequent event. This is particularly true when the adverse effect, such as cancer, occurs in both the test and the control group (Scala 1991). This task is made even harder when one potential response in the test group is a decrease in the incidence of the adverse event, that is, a hormetic response. To uncover such an effect would require a study design that would allow for such a response. Another difficulty for a dose-response researcher, and the more familiar one, is extrapolation. Extrapolation problems exist for both animal and human (epidemiological) studies. Even the most sophisticated epidemiological and animal studies are incapable of detecting low levels of risk, for example, below 1 percent, and so these risks must be imputed based on data at higher doses.

The validation issue is further magnified with the LNT model, as it predicts proportional risk to ever smaller and smaller doses. Much of the current justification for using LNT as the default dose-response model for exposure to ionizing radiation and chemical carcinogens is rooted in epidemiological studies. However, epidemiological studies are difficult to reproduce, hard to map to the general population due to the presence of confounders, and are often focused on cases where the population in question is exposed to high dose levels (Taubes 1995). Examples of such cases are studies of the effect of ionizing radiation that rely on evidence from radiation exposure following Hiroshima, Nagasaki,
Chernobyl, and Fukushima; occupational radiation studies; and medical studies on highly exposed individuals (Calabrese and O’Connor 2014).

Such high-dose exposure events and studies are therefore unsuited for extrapolation to the relevant day-to-day low-dose events like the use of X-rays and CT scans for medical purposes (Berrington de González et al. 2009). Even some of the more recent articles in the medical literature that predict high rates of disease and cancer-related deaths due to medical imaging in the United States rely on extrapolation from high-dose exposure to radiation (Berrington de González et al. 2009; Abbott 2015).

Moreover, there is a sizable stock of scientific research (epidemiological and medical) suggesting the possibility of a threshold model for radiation exposure for doses below 100 mSv (Ropeik 2013), while other studies have detected a beneficial response to low-dose exposure. For example, four epidemiological studies of subjects who are naturally exposed to background radiation did not detect any increase in cancer risk, with one study detecting a positive response to low-dose radiation (Tao et al. 1999). Another study on the effect of radon exposure revealed beneficial effects to low-dose exposure (Cohen 1995). These results were affirmed in another more recent study on radon exposure, which detected the possibility of positive effects from low doses of radiation on lung cancer (Thompson et al. 2008). A multiple-country analysis of occupational exposure to X-rays and gamma-rays in nuclear power plants also did not detect negative health effects from exposure in workers; instead it showed a rate of all cancer mortality lower in the exposed workers relative to the general population (Cardis et al. 2007). A quick search on Google Scholar for hormesis alone generates 23,800 articles.

---

6 The lack of statistical significance in these studies is nonetheless important, as it means that the effect of exposure to low-dose radiation on cancer risk is not different from zero. This finding of non-significance may imply a possible threshold and not an LNT model.
Much of the aforementioned research, which was unable to validate a linear response, also relies on epidemiological, occupational, and ecological investigations, which naturally suffer from the same shortcomings as the studies supporting linearity. Yet, regulatory risk assessment has lacked a systematic review of the evidence in support of each model. Such a review could shed light on the weight of evidence in support of each model while accounting for study design and quality. For example, lack of a systematic review is illustrated by the seventh committee on the Biologic Effect of Ionizing Radiation (BEIR VII) that attributed the beneficial response in the multiple-country study of occupational exposure to X-rays and gamma-rays to a “healthy worker effect and unknown differences between nuclear industry workers and the general public” (Calabrese and O’Connor 2014). These kinds of assertions are not helpful when equally plausible alternative explanations exist, but are ruled out without any review of the existing evidence.

The difficulty of validating models at very low doses drove the Health Physics Society and the American Association of Physicists in Medicine (AAPM) to conclude in December 2011 that the effects of radiation at very low doses (50–100 mSv) are either too minuscule to detect or virtually nonexistent. As a result, the two organizations issued statements recommending against quantitative estimation of health risks for doses of radiation below 50 mSv annually or below 100 mSv above that of background radiation in a lifetime. In the words of the AAPM (2011),

Risks of medical imaging at patient doses below 50 mSv for single procedures or 100 mSv for multiple procedures over short time periods are too low to be detectable and may be nonexistent. Predications of hypothetical cancer incidence and deaths in patient populations exposed to such low doses are highly speculative and should be discouraged. These predictions are harmful because they lead to sensationalistic articles in the public media that cause some patients and parents to refuse medical imaging procedures, placing them at substantial risk by not receiving the clinical benefits of the prescribed procedures.
As outlined, the low-dose region is one saturated with uncertainties and the choice of one model to estimate health risks gives "a false sense of precision" (Office of Management and Budget 2007) where none currently exist. Given that the use of the LNT model may lead to poor public health decisions, then integrating it with other plausible dose-response models moves us closer to optimizing public health protection (Calabrese et al. 2015).

The issue of model uncertainty and model validation in the low-dose region has been a challenge for decades. Ever since the publication of Risk Assessment in the Federal Government: Managing the Process in 1983 by the National Research Council (NRC), the choice of the low-dose model must be given to the one with the most biological plausibility (National Research Council 1983).

In the next subsection, we present recent developments in biology that support another low-dose model, namely, hormesis, or the biphasic dose-response model. In addition to having more biological support than the LNT model, hormesis, if correct, casts doubt on the supposed conservative nature of LNT.

3.2. Hormesis

In contrast with the LNT model, the hormetic dose-response model is a biphasic model where direction of response is not constant across doses. While response to exposure to a high dose of some substance may indeed be proportional to dose (i.e., harmful), response to exposure to a low dose of the same substance may be inversely related to dose (i.e., protective). In other words, exposure to a low dose of a carcinogen may—up to a certain threshold—lower the risk of developing a particular cancer. These characteristics are sometimes described as low-dose stimulation and high-dose inhibition.
As previously mentioned, all dose-response models encounter validation problems in the low-dose region; hormesis faces this issue as well. In hormesis, the hormetic effect is generally modest, that is, 30–60 percent greater than control values (Calabrese and Baldwin 2003). Given the small ratio of signal to noise and the modest effect, it is difficult to replicate hormesis and to distinguish between a threshold and a hormetic model in the low-dose region respectively (Calabrese and Mattson 2011). Considering, however, the significance of health implications of correctly identifying the type of dose-response model, efforts to design better studies have continued. As described in one paper, “The use of different default models has important implications in many areas, including the establishment of limits for chemical exposures” (Calabrese 2008).

Recent advances in clinical studies have begun to allow researchers to overcome some of the aforementioned obstacles. For example, shifting focus from whole-animal to cell-level investigation has allowed for more doses to be tested and results to be replicated, in addition to both allowing results more relevant to humans and to relying less on extrapolation (FDA 1993). These and other recent advances suggest that the dynamics of the low-dose region may be more nuanced than is predicted by the default LNT model.

Hormesis has been found to make more accurate predictions than both the LNT and threshold models using large independent data sets (Calabrese and Baldwin 2003). Some research has provided an explanation for the mechanism of action of hundreds of hormetic dose responses, suggesting that hormesis may be more of a rule than an exception. This claim was extended to both cancer and noncancer end points and is said to be independent of the biological model and the stressors tested (Calabrese and O’Connor 2014).
Studies in toxicology have revealed hormetic dose responses for both ionizing radiation and chemical carcinogens. One estimate for chemicals found a hormetic response in 37–50 percent of chemicals tested and also found that the hormetic responses exceeded those of the threshold by 2.5 to 1 (Calabrese and Baldwin 2003). In fact, a hormetic response is detected in nearly 2,000 chemical agents from a broad range of chemical classes (Calabrese et al. 2008, Calabrese 2013a). Some of the studies showing a beneficial health effect of ionizing radiation at low levels of exposure (discussed in the previous subsection) may also be an example of a hormetic dose response.

A major FDA-funded study (the mega-mouse ED01 study), which included 24,000 animals exposed to a known carcinogen (2-acetylamino-fluorene, a derivative of fluorene), found evidence supporting a hormetic, or biphasic, dose response (Bruce et al. 1981). Additionally, a reassessment of the effect of DDT in an animal study—on which regulatory agencies had based their risk assessment—revealed a hormetic dose response (Sukata et al. 2002). Hormesis has also been detected in exposure to low doses of air pollutants, namely particulate matter (Cox 2012).

The LNT model is often argued for and justified on the basis that it is a conservative approach (EPA 2005). However, hormesis alone casts doubt that adherence to linearity is necessarily conservative as we intervene to maintain lower doses. As recent research on model

---

Some claim that the LNT model is not conservative. For example, Bailar et al. (1988) argue that a supralinear dose-response relationship is possible for some chemicals. Others have argued that the human population is heterogeneous in its susceptibility to cancer risks (Finkel 2014), such that some individuals will experience higher than average cancer responses. Bailar et al. (1988), however, did not consider the possibility of a J-shaped dose response in his study due to its lack of support at the time. Now, however, ample support for a J-shaped dose response is available, as mentioned above. Regarding variation in human susceptibility, at least for the purposes of calculating benefits in a benefit-cost analysis, it is the mean response in the population that should be considered. Some individuals will no doubt experience higher than average cancer responses, just as others will be lower than average. As will be discussed in more detail below, taking an upper bound of risk that accounts for humans having higher than average susceptibility or having a higher exposure is not conservative because there is a balance to be struck between target risks and the risks associated with risk-risk tradeoffs and health-health tradeoffs. Such balancing is impossible when upper bounds are used in place of mean population responses. Further, research on the integration of hormesis and the LNT model shows that setting a protection standard based on the response of the most sensitive populations can lead to a net negative health outcome (see Calabrese et al. 2016).
uncertainty suggests (Calabrese et al. 2015), the optimal hormetric response occurs at the nadir of the hormetric curve, which is illustrated in figure 1. As argued, the dose corresponding to a $10^{-4}$ response according to the LNT model is roughly aligned with the dose yielding the optimal hormetric response. Therefore, seen in light of model uncertainty and if the hormetric model is correct, then pushing exposure to a dose smaller than the dose corresponding to a $10^{-4}$ response as predicted by the LNT model will yield net health harm. Taking bladder cancer as an example, the health gains achieved by pushing exposure to a dose corresponding to a $10^{-6}$ LNT response (i.e., 100 bladder cancers less than a dose corresponding to a $10^{-4}$ LNT response), will be dwarfed by the health harm induced by eliminating the potential for protective hormetric effects (i.e., 3,150 more incidences of bladder cancer) (Calabrese et al. 2015).

**Figure 1. Model Uncertainty and Health Protection when Accounting for Hormesis**

3.3. DNA Repair, Preconditioning, and Adaptive Responses in Biology

When the LNT one-hit model was first proposed, it was assumed that a single change of DNA could initiate the carcinogenesis process and damage could not be reversed. In other words, DNA repair was ruled out. Scientific understanding has come a long way since then. In addition to recent developments indicating that displacing a large number of molecules is required to affect a mutational event (Weiss 1944), several types of cells are now found to successfully repair mutated DNA (Hanawalt 1994). And even if a carcinogen can initiate a carcinogenesis process in a linear fashion, the development of tumors may not necessarily follow. For instance, in one study Driver, White, and Butler (1987) demonstrated that a single administration of the mutagen/carcinogen dimethylnitrosamine (DMN) induced a linear dose response for renal mesenchymal DNA adducts (early cancer process stage), as well as for mesenchymal foci (later cancer process stage), observations consistent with the LNT model. However, the linear transition to the occurrence of tumor formation was not observed, as the foci at the lower doses failed to proceed to the tumor stage, yielding a threshold, rather than a linear dose-response relationship.

A similar point was made in a 1990 paper by Ames and Gold. The authors argued that cell division plays an important role in the carcinogenesis process, as cell division increases the vulnerability of DNA to mutation. Since animal testing is very expensive, rodents are generally subjected to chronic doses of hazards in order to better detect a carcinogenic effect. However, when high doses of a carcinogen are being administered in an acute manner—causing the destruction of some cells—then cell division is the natural bodily reaction to replace these dead cells, making DNA mutation more likely. As Ames and Gold have observed,

By causing chronic cell division, a high percentage of all chemicals might be expected to be carcinogenic at chronic, near-toxic doses. . . . About half of all chemicals tested
chronically at the MTD [maximum tolerated dose] are carcinogens. The fact that about 40% of rodent carcinogens are not mutagens is consistent with our understanding of the important role of cell division in carcinogenesis. Although toxicity at or near the MTD often induces cell division, below a certain dose no such effect is observed. (Ames and Gold 1990)

Research on DNA repair offers a significant challenge to the LNT paradigm: The notion of self-repair is inherently inimical to a linear theory. But while it could be argued that DNA repair does not on its own resolve the debate over the dose-response model, recent biological research on preconditioning and adaptive response seems to make a convincing case for hormesis.

Preconditioning and adaptive response research explores whether a low dose of a stressor induces a protective reaction in the body against higher doses of the same stressor and, in some cases, higher doses of other stressors. In other words, low doses of a stressor can increase resilience and promote survivability in the environment. Stressors can vary from environmental pollutants to chemical carcinogens to exercise to intermittent fasting. The ability of organisms to react adaptively to low doses of stressors has recently been argued to play a fundamental role in evolution (Mattson and Calabrese 2010). In fact, preconditioning and adaptive response is challenging two fundamental implications of the LNT model, namely that dose is cumulative and damage is irreversible (Calabrese 2015, Calabrese 2016).

Research on preconditioning and adaptive response is now proposing less invasive methods, both to treat present diseases and to prevent susceptibility to future ones. Recent studies argue that low doses of X-rays can induce a protective effect to treat pneumonia by promoting an anti-inflammatory response (Calabrese, Dhawan, and Kapoor 2014). Moreover, low-dose radiotherapy is argued to be highly effective on patients with shoulder tendinitis or bursitis (Calabrese, Dhawan, and Kapoor 2014). Low-dose X-rays have also been asserted not only to initiate an adaptive response to higher doses of radiation but also to nonradiation stress, such as oxidative damage,
which constitutes a major cause of diabetic complications. Low-dose radiation has been found to induce a maximal protective effect against kidney damage in diabetic patients (Shao et al. 2014).

Other research examines low-dose light therapy administered to the front lobe of the brain to stimulate brain and muscle activity and to sharpen memory (Hayworth et al. 2010). Additionally, low-level light therapy (LLLT) has been shown to be an effective treatment against subsequent heart attacks, and when administered to patients before surgery, it can promote healing of surgical wounds. In addition, LLLT administered on normal muscles may increase the amount of physical work that can be performed by extending the time that the muscle can function comfortably before fatigue starts (Agrawal 2014).

A comprehensive review of all the recent research on preconditioning and adaptive response and the biological basis of hormesis is beyond the scope of this paper, but one such study is Calabrese (2008). It is clear, however, that hormesis and preconditioning play substantial roles in public health. While massive uncertainties may fog up the low-dose region and make model selection a challenging endeavor, biological plausibility—as advocated by regulatory agencies and the NRC for many decades—must be the tiebreaker.

4. A Methodology to Alleviate the Uncertainty of Regulation in the Low Dose

Guidelines from the National Academy of Sciences can assist when reevaluating critical assumptions such as the LNT model. In addition, the NRC has dedicated numerous publications to risk assessment over the past three decades. For example, in 2009 the Council released Science and Decisions: Advancing Risk Assessment in which an entire chapter was dedicated to the “Selection and Use of Defaults.” Choosing scientific defaults has been defined as “trans science,” that is, “questions which can be asked of science and yet which cannot be answered by science” (Wagner
By their nature, then, many of the default assumptions on which regulatory agencies generally rely in their risk assessments have been subject to controversy over the years (National Research Council 2009). This problem has been recognized in NRC publications dating back to the 1983 *Risk Assessment in the Federal Government: Managing the Process*—the famous Red Book—and the 1994 *Science and Judgment in Risk Assessment*. In the chapter on defaults in the 2009 publication, the NRC makes the case for selecting sound default assumptions as summarized in the following four recommendations (National Resource Council 2009):

1. Have a clear choice of defaults to prevent inconsistency resulting from an ad hoc interpretation of the data across the agency’s analysis. Further, a default assumption may be well chosen in general, but it is necessary to maintain flexibility in the application of defaults, as substance-specific data may justify a departure from defaults.

2. Invoke defaults for the steps of the risk assessment where it is necessary to make “inferences beyond those that can be clearly drawn from the available data or to otherwise fill common data gaps.” “Inferences are needed when underlying biologic knowledge is uncertain or absent.”

3. Maintain criteria “available for judging whether, in specific cases, data are adequate for direct use or to support an inference in place of a default.”

4. Report and compare alternative risk estimates in the presence of a “comparably plausible” alternative assumption; abandon a default assumption in favor of an alternative assumption when the latter is determined to be “clearly superior” to the former, that is, “its plausibility clearly exceeds the plausibility of the default.”

The NRC makes the analogy between the “clearly superior” standard for alternatives to the legal concept of “evidence beyond reasonable doubt.” A similar analogy can be drawn for
this point where “comparably plausible” can be interpreted as the legal parlance “preponderance of evidence,” or the 50 percent range of plausibility. The two points can be reasonably summarized as follows: when an alternative is comparatively plausible, quantitative model uncertainty should be characterized and presented in the risk assessment; on the other hand, when an alternative is clearly superior, it should, then, replace the default. The NRC further clarifies the clearly superior standard by saying, “The term clearly superior should not be interpreted quantitatively, but the committee notes that statistical P values can also be used as an analogy. For example, rejecting the null in favor of the alternative only when P < 0.05 could be viewed as insisting that the alternative hypothesis is ‘clearly superior’ to the ‘default null.’”

In a manner consistent with the recommendations from the NRC outlined above, regulatory agencies can make a well-justified fresh assessment of their LNT default assumption. Though choosing a default may be necessary in cases where data is lacking, the NRC encourages abandoning a default for an alternative when evidence accumulates and identifies the alternative as a more appropriate assumption. To follow an objective process for determining the appropriate default, regulatory agencies should consider both bodies of evidence validating the LNT, threshold, and hormetic models. Specifically, regulatory bodies can base their decision on a systematic review of evidence methodology\(^8\) to determine whether hormesis is a “comparatively plausible” or “clearly superior” alternative model to LNT.

If neither the LNT nor the hormetic model are deemed “clearly superior,” and the systematic review instead reveals them to be “comparatively plausible,” then regulatory agencies

\(^8\) Systematic review of evidence, instead of weight of evidence, is the latest recommendation from the NRC (National Research Council 2011).
can develop a quantitative model uncertainty analysis in their risk assessment and update their protection standards accordingly.\textsuperscript{9}

The LNT model has long been the model of choice for cancer (and since 2009, for PM\textsubscript{2.5}) risk assessment. Choosing and adhering to a particular dose-response model may have been necessary for many reasons: to ensure consistency in analysis and avoid ad hoc interpretation of the data; to prevent halting valuable scientific inquiry in the face of scientific uncertainty or lack of technical ability; or to ensure protection of public health and safety when knowledge and consensus are lacking. As argued in this paper, however, since certain assumptions may drive much of the results of a risk assessment, periodic reflection on the choice of assumptions is necessary to ensure that the resulting risk management decision is optimal, given the existing information.

5. Implications of Tradeoff Analysis

The analysis of tradeoffs is foundational to economics and sound decision-making. Tradeoff analysis looks at the consequences of making a choice or taking an action. Every choice taken eliminates another choice that could have been taken instead, and every choice taken has both intended and unintended consequences. Tradeoff analysis, therefore, attempts to calculate how the weight of the intended consequences of an action taken compares to the weight of the unintended consequences of that action as well as the weight of consequences of forgone alternative actions.

Below we will discuss two types of tradeoffs, namely, risk-risk and health-health tradeoffs, which are essential for consideration in any risk analysis based on an LNT hypothesis.

\textsuperscript{9} One proposal on how LNT and hormetic models can be harmonized to maximize public health protection is suggested in Calabrese, Shamoun, and Hanekamp (2015).
5.1. Risk-Risk Tradeoffs

The doctrine of better safe than sorry is commonly invoked to justify the use of the LNT model because the “conservative” LNT is more likely to overestimate average risk than a threshold or a hormetic model, but it isn’t so simple. Any regulation of risky behavior can push consumers into other, sometimes riskier, behavior. Thus, it is important not to develop tunnel vision, focusing only on the risk at hand. Risk policies must always take risk tradeoffs into account and, at a minimum, ensure that there are no negative public health consequences.10

A risk-risk tradeoff happens when risk-reducing actions increase (or decrease) a non-target risk at the same time that a target risk is decreased. These changes in non-target risks—so-called countervailing and coincident risks—are usually unintended but are also often discoverable. Any risk management action will cause people to make different choices, whether because of a change in relative prices or because of a need to employ a different technology (Williams and Thompson 2004).

Risk-risk analysis (RRA) is a formal analytical framework that compares reductions in target risks with unintended increases or decreases in other risks resulting from the mitigation efforts. Countervailing risks are the negative side effects of risk mitigation efforts, while coincident risks are those risks that are likely to fall in tandem with the target risk. A popular example of a risk-risk tradeoff is the increase in the risk of a stomachache as a consequence of taking aspirin to reduce the risk of a headache continuing (Graham and Wiener 1995).

RRA frequently involves both risk assessment and economic analysis, so it must involve a combined effort of risk assessors and economists (Williams and Thompson 2004). Risk-risk

---

10 There may be an overall positive public health change resulting from a risk decision that may still fail a benefit-cost test because of non-health-related costs.
tradeoffs add to the uncertainty surrounding the choice of a dose-response model because they complicate the effort to identify a public health-protective policy. Just as there are many low-dose response functions that could be derived from high-dose animal studies, so too there could be many behavioral responses induced by a new regulation. Exposure to new risks as one takes actions to avoid proscribed risks can turn a regulatory action into a public health hazard.11

Thus, the issue of whether changes in exposure to risks are producing public health negative or positive outcomes is complicated. As we move to reduce exposure to one hazard, other risks will increase; the crucial risk management question is whether countervailing risks will increase by more than the targeted and coincident risk reductions. We suspect, as do Graham and Wiener, “that risk tradeoffs are quietly hindering the effectiveness of the national campaign to reduce risk” (Graham and Wiener 1995). If we ignore these countervailing risks, we increase the chances of moving in the wrong public health direction. The uncertainty with LNT models acknowledged by considering risk-risk tradeoffs is illustrated in figure 2.

Often, a countervailing risk will result from people using a substitute compound for the one being regulated. Looking at figure 2, if we presume that the target and the substitute both have LNT dose-response curves, then our concern is how the reduction in exposure to the target hazard (from A to B)—which results in a change in risk (here a decrease in response from 1 to 2)—compares to the risk posed by the use of the substitute compound (here an increase in dose from C to D with an increase in response from 3 to 4). The issue becomes even more complicated when there is the possibility that the target or the substitute compound or both might possibly have a hormetic dose-response function. This possibility is illustrated in figure 3.

11 For example, an FDA warning label requirement for raw unpasteurized juice resulted in juice being pasteurized or ceasing to be produced rather than in the addition of the warning labels (Food and Drug Administration 1998).
Figure 2. Uncertainty Created by Risk-Risk Tradeoffs, Assuming an LNT Model

Figure 3. Uncertainty Created by Risk-Risk Tradeoffs, Assuming Hormesis
The stars at the bottom of the two curves represent the apex of the hormetic effect—that is, the optimal health response point. In this example, a decrease in exposure to the target compound from A to B not only decreases risk, it also increases the likelihood of a positive hormetic response, although past the optimal level. For the substitute compound, moving from C to D loses both the protective hormetic response and increases risk. Predicting the net effect on risk requires a great deal of information. Without such information, uncertainty may be so great as to make it unclear whether reductions in exposure to the target risk are producing public health positive or negative outcomes. Furthermore, the uncertainty involved in the shape and position of these functions on the two-dimensional dose-response plane makes decisions to improve public health considerably more uncertain.  

5.2. Health-Health Tradeoffs

As a general rule, the lower the level at which the mandated exposure to a risk is set, the higher the marginal cost of that mandate is likely to be, due to the economic phenomenon known as diminishing marginal returns. Since a large percentage of regulatory costs are translated into higher prices for goods and services, consumers will have lower real incomes and thus be less able to afford reducing the risks most relevant to them. The lower the levels of exposures chosen, the more it costs to comply (per unit of risk reduced) and the resulting higher prices reduce expenditures on private risk mitigation.

A subset of risk-risk analysis, known as health-health analysis (HHA), focuses on those countervailing risks that occur when regulatory costs reduce private expenditures that address

---

12 Of course, consideration of the combined effects of multiple stressors, i.e., additive, antagonistic, and synergistic effects on either risk or hormetic effects, only further complicates the uncertainty.
personal health risks (Lutter and Morrall 1994). This effect alone can render a policy harmful to public health and, as discussed below, can also be regressive.

Vanderbilt Professor Kip Viscusi estimates that for each additional dollar of income earned (or lost), people tend to increase (or reduce) health-related expenditures by 10 cents; that is to say, an individual’s marginal propensity to spend on health is roughly equal to 10 percent (Viscusi 1994). As people are obligated to incrementally spend more and more resources complying with regulations addressing public risks, they will respond by reducing expenditures on mitigating risks that they face in their private lives. At some point, if one takes enough income away from people and these losses are spread out across a large enough group, countervailing risks will increase by an amount sufficient to result in expected fatalities. One estimate of the magnitude of burden sufficient to induce one expected fatality is $92 million in 2016 dollars (Viscusi 1994).13

Such fatalities are not likely to be distributed evenly across society. Ralph Keeney has shown that such cost-induced fatalities fall disproportionately on those with lower incomes, including some minority groups (Keeney 1994). Conversely, if health and safety goods have diminishing marginal effectiveness, then spending the first dollars yields the largest return (e.g., spending on doctor’s visits before spending on a car with a rear-view camera), which, in turn, means that the dollars spent by the lower-income population are the most effective at reducing health and mortality risks. As such, it is important to consider distributional effects in terms of who is bearing the costs and who is enjoying the benefits of risk mitigation. This is a more compelling

---

13 This is known as a “statistical fatality” and refers to the adding up of small probabilities of death to one. That is, if 1,000 people stop making expenditures that will prevent a 1 in 1,000 risk of death, then there is the expectation that one “statistical death” will occur, although the identity of the deceased is unknown. The $92 million estimate is adjusted for inflation from $50 million in 1990 dollars, using the Consumer Price Index.
reason why inter-individual variability may be important. Indeed, presidential executive orders currently in effect also require agencies to consider such distributional impacts of regulations.\textsuperscript{14}

Since many of the benefits of reducing target risks will accrue to concentrated groups of the exposed population, while dispersed populations will realize increases in countervailing risks in addition to the costs of regulatory action,\textsuperscript{15} policies with these kinds of differential impacts may be more likely to yield negative public health outcomes in the aggregate.

5.3. Risk and Health Tradeoffs in Practice

An example of how risk-risk and health-health tradeoffs can inform a decision to manage pathogenic risks comes from the consumption of raw oysters. Raw oyster consumption, especially from the warm waters of the Gulf of Mexico, results in approximately 30 deaths each year and more than twice that number of illnesses (Kuchler et al. 1999). One option to reduce this risk would be to restrict consumption of raw oysters during certain months of the year (e.g. March through November) when the pathogen is present at high doses. With perfect enforcement, this would essentially eliminate the target risk of vibrio Vulnificus, the pathogen in question. But two tradeoffs arise.

The first is a risk-risk tradeoff from switching to substitutes, that is, what people eat instead of raw oysters. All foods contain some risk from exposure to microbial, chemical, nutritional, and physical hazards, and there may be other kinds of raw seafood, such as sushi, with which people would replace oysters. One must account for the risks posed by these substitutes.

\textsuperscript{14} See, for example, President Clinton’s Exec. Order No. 12866, 3 C.F.R. 76 (1993); President Obama’s Exec. Order No. 13563, 3 C.F.R. 58 (2011).

\textsuperscript{15} For example, US ethanol rules increased corn prices, which reduced purchasing power for lower-income households around the world (Abdukadirov 2015). The general phenomenon of concentrated benefits and dispersed costs is discussed in Olson (1965).
The second is a health-health tradeoff from reduction in income. Because the typical oyster harvester’s job skills are not readily transferable, these individuals would suffer an income loss—perhaps for prolonged periods—if oyster consumption were restricted (Kuchler et al. 1999). Research by Ralph Keeney and others has shown how income loss can cause health problems due to increased alcoholism, depression, and even suicide. Such income effects can lead to a reduction in expenditures meant to reduce personal risks, such as buying safer cars, living in safer neighborhoods, purchasing smoke detectors and baby gates, paying for preventive medical visits, and other risk-reducing products (Keeney 1994).

Pesticide standards are another nuanced example. If banning certain pesticides forces a switch to more expensive pesticides, the price of fruits and vegetables will increase (Gray and Graham 1997). Higher-priced fruits and vegetables may induce marginal consumers to switch to a cheaper but less healthful substitute. The inframarginal consumers, on the other hand—those who elect to keep eating fruits and vegetables despite the higher price—are now made poorer and less able to address their personal risks. Farmers’ incomes may suffer as well, due to the higher production costs or a net decrease in demand.

6. Conclusion

Risk assessments were originally meant to give risk managers information that would allow them to choose policies that would unambiguously reduce risks and thereby protect public health. Risk assessments for both radiation and chemical exposure that employ conservative defaults, most particularly the LNT model, seemed to provide a ready-made safe level of exposure to a target risk to achieve this goal. The so-called “safety factors” were also meant to be conservative divisors to accomplish the same effect.
But as regulation has expanded and regulatory exposure limits have reached lower and lower levels, it is no longer possible to ignore the evidence of the biological implausibility of the one-hit model as well as the increasing evidence in favor of hormesis. A default model that inaccurately characterizes risk is a problem not just because the model could be wrong, but also because it could lead to adverse consequences to public health. This follows from the fact that risk management choices must take into account the health consequences of countervailing risks and health-health tradeoffs. These tradeoffs, in some cases, can be sufficient to offset the positive effects of target risk reductions, a consequence that becomes more likely when already-low target risks are overestimated.
### Appendix: Major Historical Points Leading to the Adoption of the LNT Model

<table>
<thead>
<tr>
<th>Year</th>
<th>Author/institution</th>
<th>Event</th>
</tr>
</thead>
</table>
| 1859 | Charles Darwin    | • Publishes *On the Origin of Species*.  
• Initiates interest in the biological community to determine the cause of genetic change that drives natural selection. |
| 1927 | Hermann J. Muller | • X-rays induce mutation in fruit flies. |
| 1928 | Olson and Lewis   | • LNT model proposed to account for evolutionary changes.  
• Follows Muller's discovery that X-rays can induce mutations in fruit fly germ cells. |
| 1930 | Hermann J. Muller | • Develops proportionality rule (i.e., linear dose response) for ionizing radiation-induced mutagenicity. |
| 1935 | Timoféeff-Ressovsky et al. | • Application of radiation target theory for mutagens.  
• Use target theory to propose a one-hit theory for ionizing radiation-induced mutation. The hit mechanism is used to explain the LNT dose response. |
| 1956 | Biological Effects of Ionizing Radiation Committee (BEIR I), Genetics Panel | • Proposes the use of the linear dose-response model for germ cell mutation, using the “doubling rule.” |
| 1961 | Mantel and Bryan | • Develop carcinogen risk assessment model based on the probit model.  
• This is undertaken to advise US government agencies on chemical risk assessment. |
| 1973 | FDA               | • Proposes a probit-based quantitative risk assessment method for cancer risk based on the 1961 Mantel and Bryan paper. |
• Recommends a linear dose-response model. |
| 1977 | FDA               | • Retains the Mantel-Bryan model with some modifications.  
• Acceptable risk value is changed to $10^{-6}$. |
• This recommendation is significant, given the widespread multimedia regulatory functions of EPA. Within two years of the recommendation, EPA applies LNT model to the regulations of trihalomethanes (e.g., chloroform) in drinking water. |

*continued on next page*
<table>
<thead>
<tr>
<th>Year</th>
<th>Author/institution</th>
<th>Event</th>
</tr>
</thead>
</table>
| 1979 | FDA               | • Replaced the modified Mantel-Bryan model with the LNT model for carcinogen risk assessment, based on the following reasons:  
  o Linear procedure is least likely to underestimate risk.  
  o Linear extrapolation does not require complicated mathematical procedures.  
  o No arbitrary slope is needed to carry out linear extrapolation.  
  o Several significant limitations had been found with the application of the Mantel-Bryan model. |
| 1979 | EPA               | • Establishes a national drinking water standard for trihalomethanes (including chloroform).  
  • This is based on an LNT methodology as recommended by the US NAS Safe Drinking Water Committee (1977). |

Note: Table is constructed from discussion in Edward J. Calabrese, 2013. “Origin of the Linearity No Threshold (LNT) Dose-Response Concept.” Archives of Toxicology 87 (9): 1621–33.
Bibliography


