BELLE NEWSLETTER

Biological Effects of Low Level Exposures

A Publication of the Northeast Regional Environmental Public Health Center, University of Massachusetts, School of Public Health, Amherst, MA 01003
Vol. 12, No. 1, March 2004, ISSN 1092-4736

ECONOMICS AND HORMESIS

INTRODUCTION

In many discussions on the assessment of toxicological foundations of hormesis, questions have been raised about the possible societal impact of hormesis especially in the area of environmental regulations. Such discussions have quickly focused on whether and how hormesis could influence the costs of various regulations. Consequently, it seemed appropriate and timely to have an independent and scholarly exploration of the topic of "The Economic Implications of Hormesis". To achieve this goal Professor James K. Hammitt, Harvard University, was invited to offer a "white paper" on this topic. His manuscript was then sent to a number of recognized experts in relevant areas to offer expert commentary. Dr. Hammitt was offered the opportunity to respond to the expert commentaries. These collective articles now follow in this issue of the BELLE Newsletter.

As is the case in past publications, the BELLE Editorial Office invites comments on these papers by the readership and will consider the publication of representative comments in a forthcoming issue of the Newsletter.

Finally, we encourage the readership to offer any suggestions for future topics to be addressed by the Newsletter.

Edward J. Calabrese, Ph.D.
Professor of Toxicology Director, NREPHC
Environmental Health Sciences
University of Massachusetts
Amherst, MA 01003
Phone:413-545-3164
Fax: 413-545-4692
Email:edwardc@schoolph.umass.edu

TABLE OF CONTENTS

ECONOMIC IMPLICATIONS OF HORMESIS

James K. Hammitt
Center for Risk Analysis,
Department of Health Policy and
Management, and Department of Environmental Health
Harvard School of Public Health
718 Huntington Ave.
Boston, MA 02115
Phone: 617-432-4030
Fax: 617-432-0190
E-mail: jkh@harvard.edu

ABSTRACT

The implications of hormesis for decision making about control of environmental exposures are examined. From an economic perspective, environmental exposures should be controlled to a level that optimizes health effects and minimizes control costs. The possibility that substances are, or may be, hormetic poses no fundamental challenge for economic analysis. In contrast with the linear no-threshold model, optimal control may be either less or more stringent under the hormetic model, depending on the incremental control cost. When exposure levels or exposure-response functions differ across individuals or are uncertain, the optimal population-level control of exposure must balance possible benefits and harms to individuals against control costs. Economic-incentive-based regulatory instruments, such as tradable permits, are likely to offer less improvement relative to command-and-control regulations under a hormetic model than under a linear no-threshold model.

1. Introduction

Most environmental-health regulations are premised on the notion that a substance is consistently harmful, beneficial, or benign. Substances that are harmful at some exposure are usually treated either as being harmful at all exposures or as harmful at exposures above some threshold and benign at exposures below the threshold. For substances which cause cancer and some other toxicants (e.g., airborne fine particulate matter), the linear no-threshold model is standard. This model assumes the probability of an adverse effect is proportional to exposure and that there is no positive threshold below which the probability of harm vanishes.

An alternative hypothesis, hormesis, describes the situation in which a substance is harmful at high exposure but beneficial at low exposure. There are many well-known examples of substances with hormetic properties, such as minerals which are essential nutrients at low exposure and toxic at high exposure. Proponents of the hormetic hypothesis have compiled evidence to suggest that the hormetic model provides a more realistic description of experimental exposure-response data for many substances than either the linear no-threshold or threshold models provide (Calabrese and Baldwin 2001a; 2001b; 2002).

This essay examines the implications of the hormetic hypothesis for environmental regulation from an economic perspective. The economic perspective takes as its goal the maximization of human welfare, including human health. In identifying the implications of the hormetic hypothesis for regulation, it is useful to distinguish several contexts depending on: (1) whether the decision is for an individual or a population, and (2) whether the level of exposure and the exposure-response function are known or uncertain. Environmental regulation occurs in the context of decision making for a population where the exposure-response function is uncertain, but it is helpful in understanding this situation to begin by considering simpler cases.

The following section describes the economic approach to decision making in the context of an individual and a single substance when both the exposure-response function and the individual's exposure are known. Section 3 relaxes the assumption that the exposure-response function and exposure are known and describes the method of decision making under uncertainty for an individual. In Section 4, the context is broadened to examine the case of decision making for a population where it is assumed that both exposure and the exposure-response function may differ between individuals, and may also be known imperfectly. Section 5 describes the implications of hormesis for the use of economic-incentive-based regulatory mechanisms, such as tradable emission permits. Conclusions are in Section 6. For convenience, the toxicant is described as a substance but the discussion applies equally to radiation and other agents that may exhibit linear no-threshold or hormetic exposure-response functions. The distinction between exposure and dose is not central to the analysis and the discussion focuses on exposure. The results can be readily transferred to cases in which the dose is the concept of interest.

2. Individual with Known Exposure-Response Function and Exposure

The fundamental objective in economic decision making for an individual is to maximize the individual's well being by increasing benefits and reducing harms. In choosing her level of exposure to a substance that is harmful at some exposure levels, an individual would seek to maximize the net benefits of alternative exposure levels, defined as the difference between beneficial and

harmful effects on her well being. These effects include both health consequences and other effects, such as the cost of controlling exposure (which consumes resources that could be put to other beneficial uses).

Comparing health effects with the costs of altering exposure is straightforward if both health effects and control costs can be measured in a common unit, such as a monetary value. The monetary value of a health improvement to an individual can be defined by her "willingness to pay" (WTP) for the improvement, i.e., the largest amount of money she would be willing to give up in exchange for the health improvement. If the cost of the health improvement is smaller than her WTP for it, then she is made better off by paying the cost and obtaining the improvement. Otherwise, she is better off foregoing the health improvement and using the money it would cost to control exposure in ways she considers more valuable. Since WTP describes the amount of money an individual would exchange for a health improvement, it depends on her wealth and the other demands on her resources. There is a well-developed theory and set of empirical methods for estimating WTP to reduce the probability and/or the severity of adverse health effects (Hammitt 2000a; Freeman 2003). Empirical estimates have been obtained for the value of many types of health improvements, with most attention directed toward the value of reducing mortality risk (Viscusi and Aldy 2003).

The benefits and costs of limiting exposure to a substance are illustrated in Figure 1. The zero point of the ordinate describes the individual's well being if she is exposed to none of the substance and incurs no cost to control her exposure to it. In the absence of control, her exposure is equal to the uncontrolled exposure level e_U and she incurs no control costs. As she reduces her exposure (moving from right to left in the figure), control costs increase. Control costs are defined as the opportunity cost of the resources consumed (or behavioral changes required) to reduce exposure in the least costly manner. Controlling exposure requires that activities be altered in some way, and the opportunity cost is the increase in well being that the resources could provide if devoted to their best alternative use. Countervailing risks caused by actions to reduce exposure can also be considered as part of the cost (Graham and Wiener 1995). Control costs are assumed to increase at an increasing rate. This assumption implies that the marginal cost of control (i.e., the absolute value of the slope of the control-cost function) increases as exposure is progressively reduced, as illustrated in Figure 2. The assumption that additional reductions in exposure become increasingly costly is reasonable since one would initially adopt the most cost-effective methods for reducing exposure (i.e., those that cost the least per unit of exposure reduction), resorting to less cost-effective methods only as needed to attain the desired exposure reduction.

Under the linear no-threshold assumption, harm to health is proportional to exposure. At any positive exposure level, reducing exposure reduces harm. The individual would seek to minimize total harm, equal to the sum of the monetary value of the harm from exposure[1] and the control costs. Under the linear no-threshold model, the exposure that minimizes total costs is e_L*. This optimal exposure level can be identified either as the lowest point on the curve defined by the sum of health harms plus control costs in Figure 1, or as the point where the marginal benefit of additional control (i.e., the incremental reduction in harm, which is equal to the slope of the exposure-response function) equals the marginal cost of additional control, illustrated in Figure 2. Under the linear no-threshold model, the marginal benefit of control (i.e., the slope of the exposure-response function) is constant. For exposure levels larger than e_L*, the marginal benefit of further control is larger than the marginal cost, and so reducing exposure reduces harm more than it increases cost. In contrast, for exposure levels smaller than e_L*, the marginal benefit of additional control is smaller than the marginal cost, and so total harm can be reduced by relaxing controls and allowing exposure to increase.

Under the hormetic assumption, the exposure-response curve is j-shaped, producing beneficial effects at low exposures and adverse effects at high exposures. Note that whether the beneficial effects of low exposure and the adverse effects at high exposure are changes in the probability or severity of the same health effect or of completely different effects (e.g., fluoride reduces dental caries at low dose but is lethal at high dose) is irrelevant to the analysis, so long as the relevant health effects are measured on a common scale, such as willingness to pay. Similarly, whether the low exposure benefits are directly induced consequences of exposure or result from stimulating a compensatory or defense mechanism is also irrelevant to the analysis.

The hormetic exposure-response function in Figure 1 is drawn to be consistent with the linear no-threshold function at the uncontrolled exposure level e_U. This is appropriate because, in most cases, the most reliable estimates and perhaps the only response data are for relatively high exposure levels. Both the linear no-threshold and hormetic exposure-response functions must be consistent with these observations and can differ only in the low-exposure region where response cannot be (reliably) measured. Hence the magnitude of the harm at any exposure level less than the uncontrolled exposure is smaller (and perhaps less than zero) under the hormetic model than under the linear model, as illustrated in Figure 1.

While the harm is smaller under the hormetic than the linear model, the optimal level of exposure, which minimizes the sum of harm and control costs, may be larger or smaller under the hormetic model than under the linear no-threshold model (Cross 2001). For small exposure levels, reductions in exposure will reduce harm under the linear model but may have little effect or may even reduce beneficial effects under the hormetic model. For these exposure levels, the optimal control is

less stringent under the hormetic model than under the linear model. In contrast, for high exposure levels the optimal control may be more stringent under the hormetic model. This follows because the constraint that both hormetic and linear exposure-response functions are consistent with the observed, high-exposure response implies that the hormetic exposure-response function must be steeper, on average, than the linear exposure-response function for exposures larger than the exposure at which the substance produces no harm or benefit, the zero-equivalent point e_0. At least for relatively high exposure levels, the marginal benefit from reducing exposure is larger under the hormetic model than under the linear model, and so more stringent controls may be warranted.

If the hormetic function is convex (curving upward) for all exposure levels from zero through e_U, then there is a unique exposure level e_E at which the hormetic function is parallel to the linear function. This level may be larger or smaller than e_0, but must be larger than the exposure level at which the harmful effects of the substance are minimized (i.e., beneficial effects are maximized), e_M. As illustrated in Figure 2, the marginal benefit of reducing exposure under the hormetic model is equal to the marginal benefit of reducing exposure under the linear no-threshold model when exposure equals e_E. At larger exposure levels, the hormetic exposure-response function is steeper, and so the marginal benefit of reducing exposure is larger than under the linear model. At exposure levels smaller than e_E, the hormetic exposure-response function is flatter than the linear function and so the marginal benefit of control is smaller. At the exposure level e_M, which maximizes the beneficial effect of exposure to the substance, the slope of the exposure-response function (the marginal benefit of control) is zero. If the cost of reducing exposure to e_M were positive, it would not be optimal to reduce exposure all the way to this health-maximizing level.[2] For exposures between e_M and zero, the marginal benefit of reducing exposure is less than zero and so controls that reduce exposure to this level would be excessive.

Whether the optimal exposure is larger or smaller under the hormetic model than under the linear no-threshold model depends on whether the marginal cost of controlling exposure is relatively high or low. Specifically, if the marginal cost of control is smaller than the marginal benefit at exposure level e_E, as illustrated by the curve labeled "marginal cost" in Figure 2, then the optimal exposure under the hormetic model e_H* is larger than the optimal exposure under the linear model, e_L*, and so the optimal control is less stringent under the hormetic than under the linear model. In contrast, if the marginal control cost at e_E is larger than the marginal benefit, as illustrated by the curve labeled "higher marginal cost," then the optimal exposure level under the hormetic model e_{H1}* is smaller than the optimal exposure under the linear model, e_{L1}*, and so the optimal control is more stringent under the hormetic model.

Note that the exposure level e_0 at which, under the hormetic model, exposure produces neither harm nor benefit as compared with zero exposure, is irrelevant to determining the optimal exposure level. As noted above, the exposure level at which the linear and hormetic exposure-response functions are parallel, e_E, may be larger or smaller than the zero-equivalent level e_0. While it may be informative to compare the benefits or harms of exposure to the case of zero exposure, the economic framework focuses on the incremental benefits and costs of altering exposure. Whether the effects are measured relative to zero exposure or to some other comparator does not alter the differences in costs and benefits between exposure levels and so it does not influence the solution. The choice of comparator is somewhat like the choice among inertial frames of reference in physical mechanics—the results are equivalent across frames.

3. Individual with Uncertain Exposure-Response Function or Exposure

In contrast to the idealized situation examined in Section 2, in practice neither the exposure-response function nor the exposure level for an individual are likely to be known with certainty. Hence the health consequences of a decision to control exposure are uncertain.[3]

The standard economic approach to decision making under uncertainty, expected utility (Raiffa 1968; Clemen 1996), assumes that uncertainty about the consequences of a decision can be represented as a probability distribution over the possible consequences. That is, while the precise health consequence resulting from a particular exposure level is unknown, the decision maker can identify the range of possible health outcomes and assign probabilities representing how likely each is to occur. In choosing an exposure level, an individual is not choosing between health consequences, but rather choosing between alternative lotteries (i.e., probability distributions on health consequences).[4]

Expected-utility theory implies that preferences between lotteries can be summarized by comparing numerical summaries—the expected utilities—of the lotteries. The expected utility of a lottery is just the expected value of the utility of the consequences, i.e., the sum over all possible outcomes of the utility of the outcome multiplied by the probability of the outcome's occurrence. The utility of an outcome is a number which represents how desirable the outcome is and is scaled in a way that depends on the individual's preferences for risk. When the harms from different exposure levels are simply the probabilities of a common health effect, these probabilities can serve as utility values. When the harms are health effects that differ in severity, WTP to avoid the harm can serve as a utility value (perhaps after transforming it to account for risk preference), or health-utility scales such as Quality Adjusted Life Years (QALYs) can be used (Pliskin *et al.* 1980; Bleichrodt *et al.* 1997; Hammitt 2002).

The expected utility framework permits the analysis in Section 2 to be extended to the case in which the exposure or exposure-response function are uncertain. For simplicity, assume that the harm measured in Figure 1 can be interpreted as the utility of harm, as would be the case if harm is the probability of a specified health effect. Assume as well that the control costs are measured on the same utility scale, or that both the utility of the health effects and the costs are measured in monetary units so that they can be added together.

First, consider the case in which the exposure-response function is known, but the exposure level associated with any particular level of control is uncertain. For example, assume that the individual can choose a control level, c, but her uncertainty about the efficacy of control is such that she believes that actual exposure will be uniformly distributed between c – E and c + E (i.e., all exposure levels between these bounds are equally likely to result). The costs of control are given by the cost function in Figure 1, and are assumed to be known exactly and to depend only on the chosen level c. The optimal control level minimizes the sum of the control costs and the expected harm from exposure. The expected harm can be calculated as the harm at each possible exposure level (between c– E and c + E) multiplied by its probability of occurrence, which is simply the average harm for exposure levels between c – E and c + E.

If the exposure-response function is the linear no-threshold function, the expected harm is equal to the harm at exposure level c, and so the optimal control level would be exactly equal to the optimal exposure level in the case where exposure is certain. If the marginal control cost is given by the curve labeled "marginal cost" in Figure 2, then the optimal control level $c^* = e_L^*$. Since the marginal benefit of control is constant and does not depend on the exposure level, setting the control level so that the marginal cost equals this constant marginal benefit ensures that the marginal cost and marginal benefit are equal, regardless of the realized exposure level.[5] In this case, uncertainty about the efficacy of control does not influence the decision about how much control is appropriate.

If the exposure-response function is the hormetic function illustrated in Figure 1, the average harm for exposure levels between c – E and c + E is not equal to the harm at exposure level c. It is greater than the harm at exposure level c, but equal to the harm for some exposure level in the interval between c – E and c + E. The optimal control level can be identified by equating the expected marginal benefit of control with the marginal cost. If the marginal harm is a linear function of exposure, as illustrated for the hormetic model in Figure 2, then the expected marginal benefit of control is equal to the marginal benefit of reducing exposure at c, and so the optimal level of control is again equal to the optimal level in the case where there is no uncertainty about the efficacy of control. In contrast, if the marginal harm from exposure is a convex function of exposure (i.e., the marginal harm curves upward), then the expected marginal benefit of control is larger than the marginal benefit of control at exposure level c. This implies the optimal level of control will be more stringent, i.e., the optimal control level c* is smaller than the level e_H^* that would be optimal if the efficacy of control were certain. Alternatively, if the marginal harm is a concave function of exposure (i.e., the marginal harm curves downward), then the expected marginal harm is smaller than the marginal harm at exposure level c, and so the optimal control c* is larger (less stringent) than e_H^*.[6]

Second, consider the case where the exposure level is known but the exposure-response function is uncertain. In general, there may be a wide range of exposure-response functions that are consistent with theory and empirical evidence about the substance. For simplicity, consider the case where the only uncertainty is about whether the hormetic or the linear no-threshold model is correct (i.e., uncertainty about the slope of the linear exposure-response function and the exact shape of the hormetic exposure-response function is assumed to be negligible). This uncertainty can be summarized by the probability p that the linear no-threshold model is correct and the complementary probability 1 – p that the hormetic model is correct.

As before, the optimal control level can be determined by equating the expected marginal benefit of control with the marginal cost of control. The expected marginal benefit is a weighted average of the marginal benefit under the linear and hormetic models with weights equal to p and 1 – p, respectively. Hence the expected marginal benefit is between the expected benefit under the two models and the optimal exposure level is between the levels that would be optimal under each model. Assuming the marginal cost is given by the curve labeled "marginal cost" in Figure 2, the optimal exposure depends on the probability that each model is correct. If the hormetic model is much more plausible than the linear model, then p is close to 0, the expected marginal benefit is close to the marginal benefit under the hormetic model, and the optimal exposure level is close to e_H^*. Alternatively, if the linear model is much more plausible, p is close to 1, the expected marginal benefit is close to the marginal benefit under the linear model, and the optimal exposure level is close to e_L^*.

The optimal exposure level depends on both the relative plausibility of each model and the relative slopes of the corresponding exposure-response functions. The relative slopes of the exposure-response functions differ by exposure level. In cases where the exposure level e_U (for which response data are available) is large compared with the exposure level e_0 (at which there is no adverse effect), the slopes of the hormetic and linear exposure-response functions will generally be similar for exposures between e_U and e_0. In this region, uncertainty about which model is more accurate will tend to have little effect on the expected marginal benefit of reducing exposure.

For many substances of regulatory concern, exposures are likely to be in this range where the marginal benefits of control are similar under the hormetic and linear models. The no-observed-adverse-effect level (NOAEL) is a toxicological parameter that is estimated for many substances and is intended to identify an exposure level close to the zero-equivalent point e_0 at which adverse effects, if any, are minimal. The average slope of the hormetic exposure-response function between e_0 and e_U is equal to the slope of the linear no-threshold exposure-response function divided by $[1 - e_0 / e_U]$. Hence if e_0 is small compared with e_U, the slopes of the hormetic and linear exposure-response functions between e_0 and e_U will be nearly equal, e.g., if e_0 is 10-fold smaller than e_U, the average slope of the hormetic exposure-response function is about 10 percent larger than the slope of the linear exposure-response function. In many cases, the estimated NOAEL is at least 10-fold smaller than the exposure level at which the response is relatively well estimated. For these substances, uncertainty about whether the hormetic or linear model is more accurate may not substantially affect the estimated marginal benefit of controlling exposure.

Even if the expected marginal benefit is not sensitive to uncertainty about which model is more accurate, the optimal exposure level may be. The optimal exposure level is identified as the level at which the expected marginal benefit of control equals the marginal cost, and so the effect of uncertainty about the expected marginal benefit depends on the slope of the marginal-control-cost function. If the marginal cost of reducing exposure rises sharply with more stringent control, then uncertainty about the expected marginal benefit will have little effect on the optimal exposure level. In contrast, if the marginal cost of control is approximately constant over a wide range of exposure levels, a small change in the estimated marginal benefit of reducing exposure can yield a relatively large change in the optimal level of control. In Figure 2, if the marginal benefit of reducing exposure were increased from its level under the linear no-threshold model, the optimal exposure level would fall below e_L* by an amount that is inversely proportional to the slope of the marginal-cost function.

In contrast, for exposure levels comparable to and smaller than the zero-equivalent point e_0, uncertainty about which model is correct can be more influential. In this region, the relative slopes of the exposure-response functions depend on the magnitude of the beneficial effect associated with the most beneficial exposure level e_M under the hormetic model, relative to the magnitude of the harm associated with that exposure level under the linear model. If the maximum beneficial effect of exposure is negligible (i.e., if the hormetic exposure-response function is similar to a threshold model having no effect at exposures less than e_0), then the expected marginal benefit of reducing exposure is equal to the marginal benefit under the linear model multiplied by the probability p that the linear model is correct. In this case, uncertainty about which model is correct reduces the expected marginal benefit of reducing exposure in proportion to the probability assigned to the hormetic model. Unless the probability that the linear model is correct is quite small (e.g., less than 10 percent), the expected marginal benefit is of the same order of magnitude as the marginal benefit under the linear model. In contrast, if the beneficial effect under the hormetic model is non-neglible, the marginal harm from reducing exposure below the most beneficial level e_M may be comparable to or larger than the marginal benefit from reducing exposure under the linear model. In this case, uncertainty about which model is correct could be extremely important to choosing an optimal exposure level.

Typical estimates of the exposure level e_M at which beneficial effects are maximized are reported to be between one-third and one-sixth as large as the no-observed-adverse-effect level (NOAEL) (Calabrese and Baldwin 2001b), with beneficial effects observed for exposures between 10 and 20 fold below the NOAEL (Calabrese and Baldwin 2001a). The magnitude of the beneficial effect is often 30 to 60 percent of the effect level at zero exposure (Calabrese and Baldwin 2001b) which suggests that the beneficial effects of exposure may be sensitive to the background rate of the health effect. These observations are not sufficient to determine the relative slopes of the hormetic and linear exposure-response functions for exposure levels comparable to e_M and below, and so it is difficult to determine how important uncertainty about the model could be to evaluating the optimal exposure levels below a NOAEL. For carcinogens, which are assumed not to have a NOAEL, uncertainty about whether the substance is hormetic could have a very large effect on the optimal exposure level if the exposure level that is optimal under the linear model is comparable to or smaller than the zero-equivalent level e_0 under the hormetic model.

If both the exposure level and the exposure-response function are uncertain, the same approach can be used to identify the optimal control level. Uncertainty about the slope of the linear no-threshold model and about the exact shape of the hormetic model can be characterized by considering a set of possible exposure-response functions and assigning probabilities to each. The expected marginal benefit of control is equal to the probability that each model is accurate multiplied by the expected marginal benefit, conditional on that model, given uncertainty about the exposure level. The optimal control is the level at which the marginal cost of more stringent control is equal to the expected marginal benefit of control.

4. Population with Variable or Uncertain Exposure-Response Functions or Exposures

Environmental regulation operates at the population level, typically by limiting concentrations of designated substances in environmental media or releases of substances to the environment. If either exposure levels or

exposure-response functions differ among members of the population, it is not possible to set regulations in a manner to optimize every individual's exposure.

Reductions in environmental concentrations of toxic substances are public goods. Unlike private goods, where the owner can control who directly benefits from the good, public goods benefit everyone in the population and it is not feasible to exclude anyone from benefiting. The economic approach to determining the optimal level of a public good is to compare the total benefits to everyone in the population with the total cost, and to choose the level of provision that maximizes the difference between total benefits and costs.

Conventional economics begins with the presumption that individuals can determine their preferences over different combinations of health, income, and other factors that influence their well being, but that there is no objective or defensible method for comparing changes in well being among individuals. That is, while everyone may prefer a reduction in the risk of developing cancer this year, there is no principled basis for saying whether one person or another benefits more from this reduction. Given this limitation, the concept of Pareto improvement becomes central. A Pareto improvement is defined as any change that improves the well being of some people without worsening the well being of anyone. Since it depends only on whether individuals benefit or are harmed, determining whether a change is Pareto improving does not require interpersonal comparison of changes in well being. Changes that are Pareto improving appear relatively unobjectionable, although not completely so. Because the concept foregoes comparison across individuals, it ignores issues such as the degree of inequality of well being in a population. Changes that benefit only people who are already very well off are Pareto improving but increase inequality, and may be viewed as not improving social welfare (Rawls 1971).

The concept of Pareto improvement provides only a partial ordering over possible changes in social conditions, since it requires either that the change harms no one, or that it benefits no one. If some individuals benefit and others are harmed, the change cannot be classified by this criterion. To compare situations in which some people benefit and others are harmed, applied welfare economics relies on the concept of a potential Pareto improvement and the Kaldor-Hicks compensation criterion (Stokey and Zeckhauser 1978). Under this approach, a change which benefits some people and harms others is defined to be a social improvement if those who benefit could compensate those who are harmed (by paying money to them) in such a way that no one would be harmed by the combination of the change and payment of compensation. A potential Pareto improvement is a change that can be transformed into a Pareto improvement by coupling it with a set of compensation payments. Equivalently, a change is defined as a social improvement by this criterion if the harms from the change can be distributed in a way such that no individual is harmed on net.

Note that the payment of compensation is purely hypothetical. There is no requirement that the compensation actually be paid, and thus changes which are potential Pareto improvements do in general harm some people. A primary justification for adopting this concept as a definition of social improvement is that, if the approach is followed over a broad set of social decisions, the people who benefit and those who are harmed in each case will not consistently be the same, and so over many decisions everyone may be made better off if decisions are based on this approach than on some alternative approach. Moreover, if some people are consistently harmed, their well being may be improved at less cost to others by direct compensation payments rather than by making social decisions that are not potential Pareto improvements.

In order to determine whether a change is a potential Pareto improvement, it is sufficient to sum the monetary values of benefits and harms across the population. If the total value of the benefits exceeds the total value of the harms, then it is possible to identify a pattern of payments that compensate for all harm to individuals without causing others to be worse off than they would be without the change. For evaluating a reduction in environmental exposure to toxic substances, the economic approach is to add the monetary values of the benefits from reducing exposure across all members of the population, and to maximize the difference between these total benefits and the total costs of controlling exposure.

Under the linear no-threshold model, adding health benefits across a population is straightforward since the marginal benefit of reducing exposure to an individual is constant and does not depend on the individual's current exposure level.[7] Even if the slope of the exposure-response function differs among individuals, total benefits can be calculated using the average slope of the exposure-response function in the population.[8]

Under the hormetic model, adding health benefits is more complicated because the health benefit to an individual depends on her initial exposure level and her exposure-response function. If the exposure-response function is identical across individuals and individual exposure levels are all greater than the most beneficial level e_M, then reducing exposure will improve health for everyone in the population and the magnitude of the gain may be estimated using only the average slope of the exposure-response function for the relevant range of exposures (accounting for any correlation between the magnitude of exposure reduction and slope of the exposure-response function across individuals). If some individuals are exposed at levels below e_M and others are exposed at levels above e_M, then reducing ambient levels of the substance will benefit some individuals and harm others. In this case, estimating the total health effect requires distinguishing individuals or subpopulations by the slope of the exposure-response function at their initial exposure levels and determining how the slope

changes as their exposures are diminished. The total health effect can still be evaluated as the sum of the beneficial and adverse effects across individuals.

If exposure-response functions differ across individuals, the same principles apply. In this case, whether exposures are equal or different across individuals, the slope of the individuals' exposure-response functions at their initial exposure levels will typically differ in magnitude and sign. Estimating the net health effect of reducing exposure in the population will require distinguishing subpopulations that are benefited or harmed by different amounts, and cumulating these individual effects into a population total health effect.

5. Economic-Incentive Instruments for Regulation

Recent years have seen growing use of regulatory mechanisms that rely on economic incentives. These are an alternative to more traditional command-and-control regulations which typically specify either a technology to be used to reduce emissions or a performance standard, such as a limit on total emissions. Under certain conditions, economic-incentive mechanisms can provide the same environmental benefit as command-and-control regulations at lower total cost (e.g., Hanley *et al.* 1997; Lesser *et al.* 1997).

The primary forms of economic-incentive instruments are tradable permits and taxes. Tradable permits have been more frequently used in practice, notably for limiting SO_2 emissions from power plants in the United States. Taxes and tradable permits can reduce the cost of controlling emissions relative to a command-and-control alternative by encouraging firms that can reduce emissions at relatively low cost to make larger reductions than firms facing higher emission-control costs. Imposing a tax on emissions creates an incentive for firms to adopt all measures to reduce emissions that cost less per unit of emission reduction than the tax. Similarly, it encourages firms to pay the tax rather than adopt emission-control measures that cost more per unit emission reduction than the tax. In this way, an emission tax provides an incentive to reduce emissions using relatively low-cost measures, and to forego relatively high-cost measures, thereby minimizing the total cost of emission reduction across firms.

Tradable permits provide a similar result. Under this approach, permits to emit a certain quantity of a substance are distributed to firms, by direct allocation, government auction, or other mechanism. These permits are legally tradable, and so firms that can reduce emissions at relatively low cost have an incentive to reduce emissions below their permitted level and to sell their excess permits to other firms for which it is more costly to reduce emissions. The total quantity of emissions is determined by the quantity of permits allocated and the market price of the permits is determined by the marginal cost to different firms of reducing emissions. From a firm's perspective, the market price of tradable permits is equivalent to a tax on emissions, as the firm profits by adopting measures to reduce emissions that cost less than the permit price (allowing it to purchase fewer permits or sell excess permits) and has no incentive to adopt control measures that cost more than the price of additional permits.

In principle, command-and-control regulations can be designed to control emissions at the same cost as an economic-incentive mechanism if the regulatory agency has access to information on the marginal cost of reducing emissions for all firms. One benefit of the economic-incentive mechanisms is that this least-cost solution can be achieved even when the regulatory agency lacks this detailed information. Moreover, economic-incentive mechanisms can provide stronger incentives for development of low-cost methods of reducing emissions (including changes in products and manufacturing processes). They do so by providing a monetary incentive for firms to find ways to further reduce emissions, even if emissions are already low. Under a command-and-control system, a firm that is in compliance with an emission limit or a requirement to have specified control equipment may have no incentive to develop methods for further reducing emissions (Milliman and Prince 1989).

Economic-incentive mechanisms reduce the total cost of controlling emissions by allocating the emission reductions across firms in a least-cost manner. In contrast, command-and-control regulations can specify the emission level firm-by-firm. The environmental consequences of allowing firms to reallocate emission reductions depend on the substances involved. For pollutants that are well mixed in the environment, such as CFCs and other stratospheric-ozone-depleting substances, and CO_2 and other greenhouse gases, the environmental consequences are completely independent of the location of the release, and so any system that produces the same total quantity of emissions in a time period will have the same environmental consequences. In contrast, economic incentive mechanisms to limit the total release of a substance with more local effects, e.g., fertilizer runoff into ponds, might result in very large concentrations of the contaminant in some ponds and very little in others. In this case, the environmental consequences of a total quantity of contaminant released to the environment would depend on the geographic distribution of release.

Under the linear no-threshold model, the total health effects within a population depend solely on the change in total exposure. The marginal damage associated with a unit of exposure is identical, and so a system of tradable exposure permits or an exposure tax could be anticipated to reduce total control costs while providing the same total health benefit as a command-and-control system. An optimal system would set the exposure tax at a level equal to the marginal benefit of reduced exposure, or would set the total quantity of exposure permits at a level such that the market-clearing price of permits was equal to the marginal benefit of reduced exposure.

Under the hormetic model, the total health effects of a reduction in population exposure depend on the distribution of changes in exposure levels among the population. If exposure reductions are concentrated among highly-exposed individuals, total health benefits will be relatively large. If reductions are concentrated among individuals having low exposure, the health effects will be smaller and may be adverse. The marginal harm associated with a unit of exposure is not constant but depends on whose exposure is altered. Under the hormetic model, a simple economic-incentive mechanism having a single tax or single type of tradable permit would not be anticipated to work as well as under the linear model. A more complicated system, in which the tax or quantity of permits required per unit of exposure varies across subpopulations in proportion to the marginal benefits of reducing exposure could provide superior outcomes but would be substantially more complicated to develop.

In practice, it is usually easier to regulate emissions of a substance to the environment rather than human exposures to the substance. Regardless of whether the exposure-response function is linear or hormetic, the relationship between emissions and exposures may be nonlinear if it depends on environmental fate and transport mechanisms and other factors. To the extent that the marginal health damage per unit emission differs across emission sources, economic-incentive mechanisms will perform less well relative to command-and-control alternatives, whether the exposure-response function is linear or hormetic.

Differences in the temporal scale of emission control and of the exposure that is relevant to health effects may also complicate the assessment of regulatory mechanisms. If the health effects of exposure to a given quantity of a substance depend on how the exposure is distributed over time (e.g., as a large bolus or as a continuous low-level exposure), the exposure-response function will depend on the temporal pattern of exposure. If the shape of the exposure-response function depends on the temporal pattern of exposure, responses to the same substance might conceivably be consistent with the linear no-threshold model for some exposure patterns and with the hormetic model for other exposure patterns. Similarly, the cost of controlling exposure may depend on the time pattern of exposure, e.g., it may be less costly to control average annual exposure than to control peak hourly exposure, because the peak may be influenced by accidents, random failure of control equipment, anomalous weather conditions, and other factors, while the average annual exposure can be modified by adopting lower-than-planned emissions later in the year if needed to offset higher-than-planned emissions that occur earlier. The relationships between exposure levels, health effects, and control costs represented in Figures 1 and 2 are implicitly conditioned on a relevant time pattern of exposure.

6. Conclusion

The possibility that many substances of regulatory concern affect health adversely at high exposure but beneficially at low exposure does not pose any fundamental challenges to economic approaches for decision making. Compared with the case in which a linear no-threshold exposure-response function can be assumed, however, the possibility that a substance does or may produce hormetic effects substantially complicates evaluation of the benefits of reducing exposure and estimation of the optimal level of control. The complication arises because the benefit of reducing an individual's exposure to the substance depends on her current exposure level. Decisions that influence exposure levels in a population are likely to differentially affect individuals whose exposure levels and possibly exposure-response functions differ because of environmental, behavioral, genetic, or other factors. Population-level decisions must balance benefits of reducing exposure to individuals whose exposure is above the level at which they would receive the greatest health benefit from the substance against harms to individuals whose exposure is less than the level that would best promote their health. Because the beneficial effect of reducing exposure depends on individual exposure levels, regulatory instruments that rely on economic incentives, such as tradable permits, offer less benefit relative to command-and-control regulations than when a linear no-threshold exposure-response function obtains.

The economically-optimal regulation of a substance may be either more or less stringent if the exposure-response function is hormetic rather than linear no-threshold. Under the hormetic model, if the marginal cost of reducing exposure is relatively low it is optimal to regulate the substance stringently so that the marginal benefit of further exposure reduction is small. This implies the exposure level will be set near the level at which the beneficial effect of exposure is maximized. At this low exposure level, the marginal benefit of further control is larger under the linear no-threshold model than under the hormetic model, and so it is desirable to reduce exposure even further under the linear model, increasing the marginal cost of control until it reaches the (constant) marginal benefit of exposure reduction under the linear model. Alternatively, if the marginal cost of reducing exposure is relatively large, then it is optimal to regulate only modestly under the hormetic model, since the marginal cost of further reducing exposure would exceed the marginal benefit. At this relatively high exposure level, the marginal benefit of regulation under the hormetic model exceeds the marginal benefit under the linear model, and so even this modest degree of control is excessive under the linear model. When marginal control costs are relatively large, the economically optimal regulation is less stringent under the linear than under the hormetic exposure-response function.

FOOTNOTES

Figure 1. Health effects and costs as a function of exposure

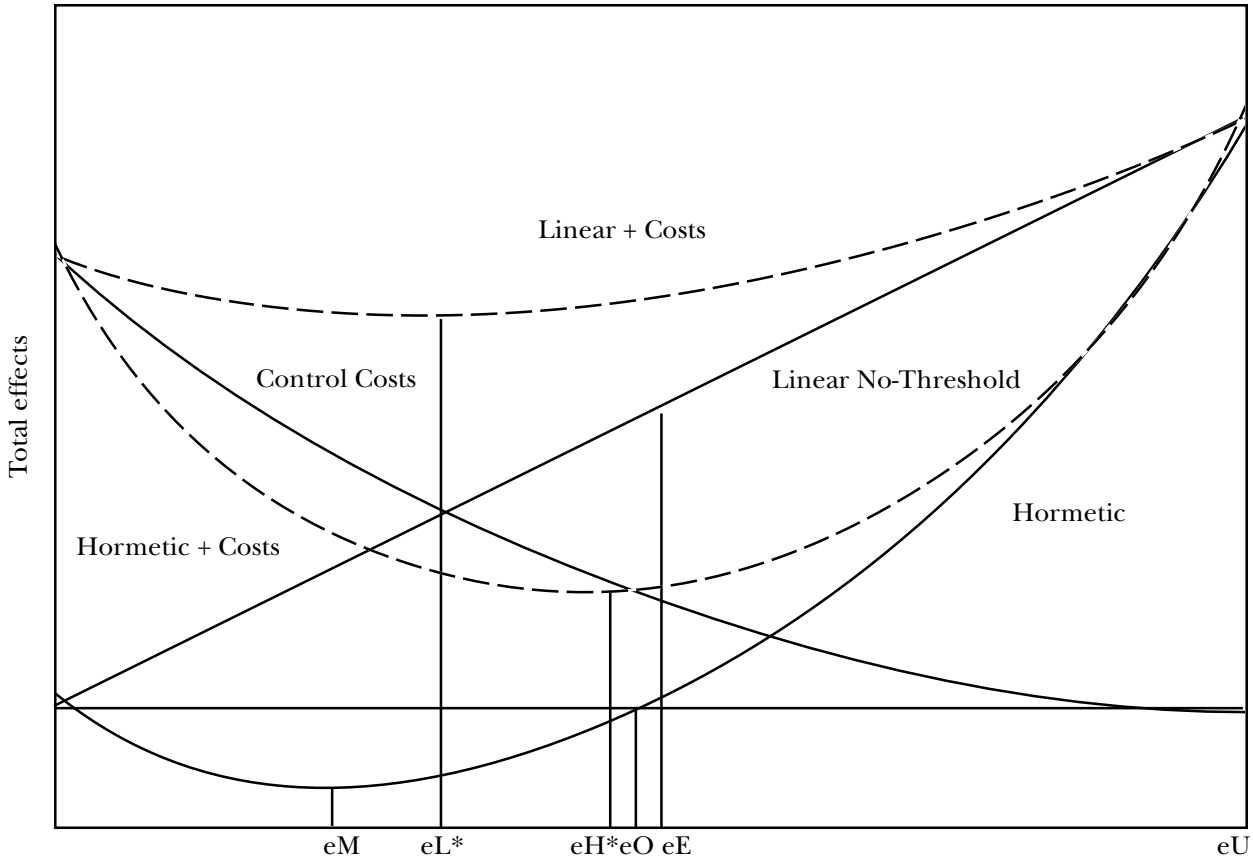
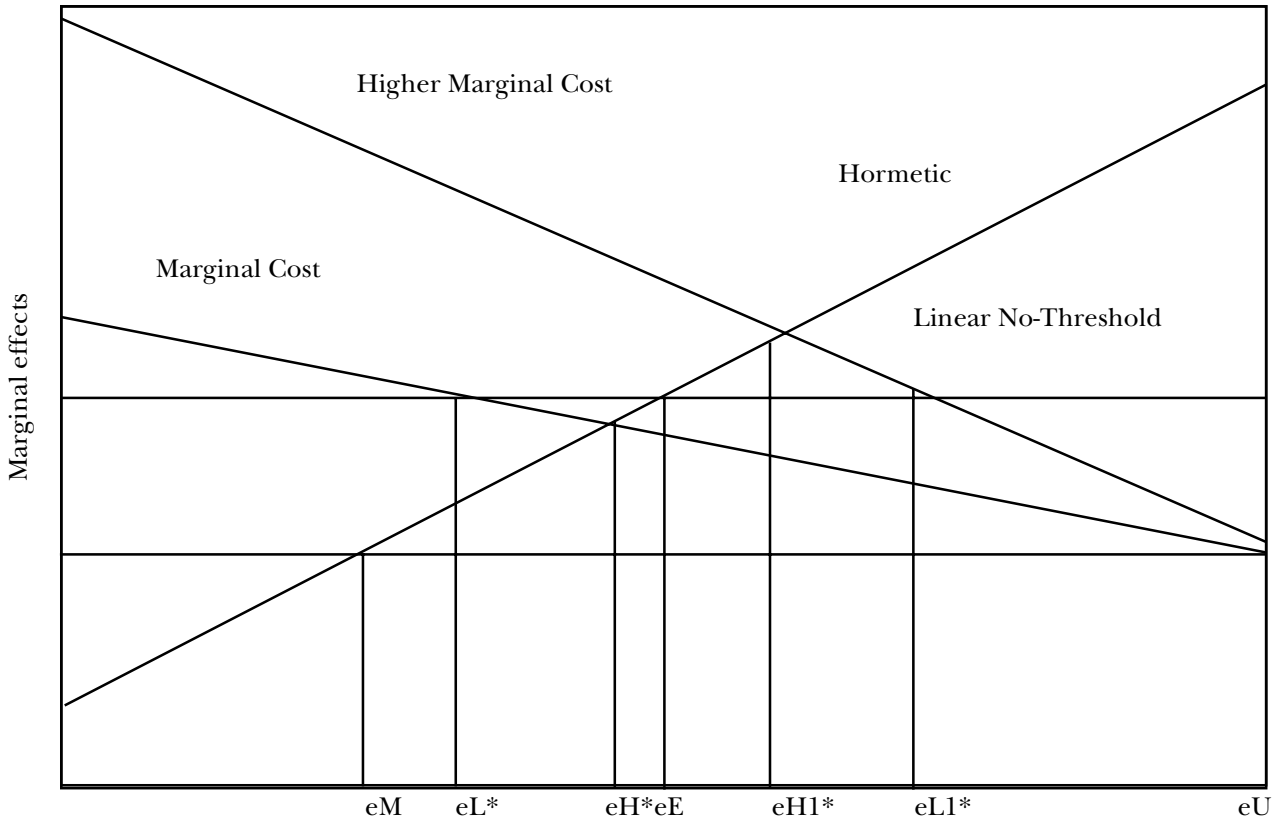


Figure 2. Marginal health effects and marginal costs as a function of exposure



[1] For simplicity, assume the monetary value of harm is proportional to the harm. If harm is the probability of an adverse health effect, the monetary value of the harm will be approximately proportional to the probability, at least for the small probabilities that are characteristic of many exposures to environmental toxins (Hammitt 2000b; Corso et al. 2001). If harm is the severity of the adverse effect, the assumption that the monetary value is proportional to the magnitude imposes a restriction on how severity is measured.

[2] Similarly, if it were costly to increase exposure from some lower background level to eM, it would not be optimal to increase exposure all the way to eM.

[3] The control costs may also be uncertain, but incorporating that possibility adds little to the analysis, and so for simplicity control costs are assumed to be known.

[4] The concept of probability employed in this formalization is Bayesian or subjective probability. Subjective probability characterizes an individual's degree of belief about the event in question (i.e., that a specific health effect will result). To the extent that different people may have different beliefs about whether an event is likely to occur, they may assign different subjective probabilities to the same event. The Bayesian concept of probability differs from the frequentist concept, which characterizes the frequency with which a particular outcome results in a series of repeated trials (e.g., the fraction of laboratory animals that will show a specific health effect in a defined exposure condition). The Bayesian concept can more easily incorporate uncertainty about factors that influence high-to-low dose and interspecies extrapolation.

[5] This result is similar to a seminal result concerning the choice between controlling the quantity of exposure (e.g., by limiting quantities emitted to the environment) and the price of exposure (e.g., by imposing a tax on emissions) when control costs are uncertain. Weitzman (1974) found that controlling the price is preferred if marginal control costs are more sensitive than marginal benefits to exposure level, and that controlling exposure is preferred in the alternative case. In the present example, marginal control costs depend on exposure but marginal benefits do not, so a tax on exposure would yield the optimal exposure.

[6] The dependence of the optimal control level on the curvature of the marginal benefit function is an example of the concept of prudence (Kimball 1990). These results do not require that uncertainty about exposure given a control level is uniform, but only that the expected value of the exposure is the control level c.

[7] To the extent that the monetary value of health improvement differs among individuals, because of differences in income or other factors, the marginal benefit of reducing exposure will differ across individuals. In practice, differences in preference and WTP for health improvement are ignored, in part because it is often viewed as inappropriate to count health benefits differently depending on who benefits. Moreover, if health benefits are distributed in the population in a manner that is not correlated with WTP, then total benefit is more easily calculated using the population total health improvement and mean WTP.

[8] If the slope of the exposure-response function and the change in exposure are correlated across individuals, the total health benefit is not equal to the product of the average slope and the average reduction in exposure, but it can be calculated taking account of the correlation.

REFERENCES

Bleichrodt H, Wakker P, and Johannesson M. 1997. Characterizing QALYs by risk neutrality. *Journal of Risk and Uncertainty* 15:107-114.

Calabrese EJ and Baldwin LA. 2001a. Hormesis: u-shaped dose responses and their centrality in toxicology. *TRENDS in Pharmacological Sciences* 22:285-291.

Calabrese EJ and Baldwin LA. 2001b. U-shaped dose-responses in biology, toxicology, and public health. *Annu. Rev. Public Health.* 22:15-33.

Calabrese EJ and Baldwin LA. 2002. Applications of hormesis in toxicology, risk assessment and chemotherapeutics. *TRENDS in Pharmacological Sciences* 23:331-337.

Clemen RT. 1996. *Making hard decisions* (2d ed.). Duxbury Press, Belmont, CA.

Corso PS, Hammitt JK, and Graham JD. 2001. Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. *Journal of Risk and Uncertainty* 23:165-184.

Cross FB. 2001. Legal implications of hormesis. *BELLE Newsl.* 9(2) and *Human and Experimental Toxicology* 20:122-128.

Freeman AM III. 2003. *The Measurement of Environmental and Resource Values:Theory and Methods* (2d Ed.). Resources for the Future, Washington, DC.

Graham JD and Wiener JB. 1995. *Risk v. Risk.* Harvard University Press, Cambridge, 1995.

Hammitt JK. 2000a. Valuing mortality risk: theory and practice. *Environmental Science and Technology* 34:1396-1400.

Hammitt JK. 2000b. Evaluating contingent valuation of environmental health risks: the proportionality test. *Association of Environmental and Resource Economists Newsletter* 20(1):14-19.

Hammitt JK. 2002. QALYs versus WTP. *Risk Analysis* 22:985-1001.

Hanley N, Shogren JF, and White B. 1997. *Environmental Economics in Theory and Practice.* Oxford University Press, Oxford.

Kimball MS. 1990. Precautionary saving in the small and in the large. *Econometrica* 58:53-73.

Lesser JA, Dodds DE, and Zerbe RO Jr. 1997. *Environmental Economics and Policy.* Addison-Wesley, Reading, MA.

Milliman SR and Prince R. 1989. Firm incentives to promote technological change in pollution control. *Journal of Environmental Economics and Management* 17:247-265.

Pliskin JS, Shepard DS and Weinstein MC. 1980. Utility functions for life years and health status. *Operations Research* 28:206-224.

Raiffa H. 1968. *Decision Analysis.* Addison-Wesley, Reading, MA.

Rawls J. *A Theory of Justice.* 1971. Harvard University Press, Cambridge.

Stokey E and Zeckhauser RJ. 1978. *A Primer for Policy Analysis.* Norton, New York.

Viscusi WK and Aldy JE. 2003. The value of a statistical life: a critical review of market estimates throughout the world. *Journal of Risk and Uncertainty,* 27:5-76.

Weitzman ML. 1974. Prices vs. quantities. *Review of Economic Studies* 41:477-491.

ECONOMIC IMPLICATIONS OF HORMESIS IN POLICY MAKING

Charles Griffiths
U.S. EPA
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Phone: 202-566-2288
Fax: 202-566-2338
E-mail: griffiths.charles@epamail.epa.gov

Hammitt (2003) offers and excellent presentation of the economic perspective of how one can calculate the optimal exposure of a hormetic substance (that is, one that provides beneficial effects at low doses but harmful effects at high doses). The second section of his paper succinctly describes how an individual's optimal exposure to a substance with a known hormetic exposure-response function (that is, one that is U-shaped) differs from the optimal level if the exposure-response function were linear with no threshold. Presuming that the government would choose to prescribe this level of exposure, this is the optimal regulatory level.

The optimal regulatory level is obtained by minimizing the sum of the monetized harmful effects of a given exposure and the costs of controlling exposure to that point. Mathematically, regulation is set where the marginal cost of control is equal to the slope of the exposure-response function. The difference in optimal level between the hormetic function and the linear no-threshold model depends upon the slope of the respective exposure-response curves. The optimal level is lower (that is, regulation is more strict) under the hormetic model at higher doses, where the hormetic curve is steeper than the linear curve. Regulation is less strict under a hormetic model at lower levels, where the slope of the hormetic curve is less than the linear function. Higher marginal costs make it more likely that the hormetic model will imply stricter regulation than the linear model.

The critical question is if hormesis is the correct model for a given substance. Ultimately, it must fall to toxicologists, epidemiologists, and other non-economists to make this determination. One of the reasons that hormesis is so controversial is because it deals with an area of the exposure-response curve where these experts feel that they are less capable of measuring a response. However, if hormesis is determined to be the correct model, then, as Hammitt (2003) points out, there is no fundamental problem for the economist to incorporate this information. The estimation of benefits and optimal control is more complicated, but there is no reason why, from an economic welfare-maximizing perspective, we can not include the potential benefits of low exposure in our calculations.

On the surface, the results of determining that a substance displays hormesis seem relatively uncontroversial. If the hormetic exposure-response curve is steeper than the linear curve, then the marginal benefits of reducing exposure are greater than under the linear model, and the optimal regulatory level is more strict. If the hormetic curve is flatter, then the detrimental effect of a substance is substantially less than that implied by the linear curve. In other words, hormesis appears to imply stricter regulation or less harm. Most government economists, though, know that regulatory decisions are (and should be) made including factors other than the economically optimal level. One of these factors is public concern, and there seems to be some public reluctance to assuming hormesis.

There are a number of possible reasons for this public concern. One possibility is that determining a substance to be hormetic will always imply a lower level of risk for any given exposure. One might argue that precaution dictates that we default to a model that produces the highest level of risk. Assuming a linear model when hormesis is valid, however, raises Portney's (1992) "happyville" problem, where the government must decide whether to regulate a chemical that is of public concern but, in fact (according to risk assessors), poses no real risk. The benefits of regulation in such a situation are unclear.

Another possible concern is that assuming hormesis will weaken regulatory standards. As pointed out above, this is not necessarily true. The optimal level could be more strict under hormesis if the slope of the hormetic curve is steeper than the linear curve. Alternatively, Heinzerling and Lechleider (2000) have argued that the beneficial portion of exposure to a hormetic substance is only a "very tiny sliver" of the exposure-response curve. If so, then the average slope of the linear and hormetic curves prior to the beneficial region will be very similar, and hormesis may have little effect on the optimal regulatory level. The real concern is where the optimal regulatory level under hormesis is less strict than the liner no-threshold model, the region where the hormesis curve is relatively flat. This, however, is a very difficult region for economists to calculate benefits, particularly if it deals with regulatory levels below the RfD. By definition, the RfD it is a region where (with uncertainty spanning perhaps an order of magnitude or greater) the dose "is likely to be without appreciable risk of deleterious health effects during a lifetime." (Subcommittee on Chronic Reference Doses 2000) It is difficult, from an economic perspective, to measure a change in benefits if a change in exposure leads to no appreciable increase in deleterious effects.

A third possible reason for public concern may be a fundamental opposition to a non-zero optimal level of

exposure. The linear no-threshold model generally implies a positive optimal exposure. However, if we accept Hammitt's (2003) construction of benefits proportional to changes in exposure, then it is technically possible to justify zero emissions assuming very low marginal costs. Under hormesis, on the other hand, there is always a positive optimal level of exposure. In fact, unless marginal costs fall to zero, the optimal point is above the exposure where beneficial effects alone are maximized. This is because, from an economic standpoint, if we care about the harm (or benefit) caused by exposure to a substance, then we should care about the cost imposed by controlling the exposure. This is one of the fundamental aspects of economics, explicitly trading off the consequences, costs and benefits, of a given action.

If all we care about are the negative effects measured by a single health endpoint, then the RfD may be appropriate. The RfD defines the level of exposure below which there is no appreciable risk of this endpoint. Hormesis maintains the single endpoint, but accepts beneficial as well as deleterious effects. The economist wonders why we should not consider all health endpoints and all costs and benefits. This is why Hammitt (2003) correctly states that costs can include behavioral changes, and counterveiling risk and benefits can be "changes in the probability or severity of the same health effect or of completely different effect." This moves us beyond the definition of hormesis (Calabrese and Baldwin 2002), but explains why certain chemicals do not appear in lists of hormetic substances (Calabrese, Baldwin, and Holland 1999) but tend to enter the debate. For example, fluorine (soluble fluoride) is widely touted as providing positive benefits of reducing dental caries at low levels (ADA 1999) but has harmful effects of dental fluorosis at higher levels and has an EPA-calculated RfD (US EPA 1989). Technically, these are not the same health endpoint, but often enter the debate about hormesis. From an economic standpoint, if we are going to consider hormetic benefits, then we should include the all of the benefits (and costs) of the substance. This includes what the substance is being used for. When all of these factors are included, and what Hammitt (2003) is describing, is benefit-cost analysis with a hormetic exposure-response curve.

Conducting this benefit-cost analysis for the individual is one task. The extension of this task, and one faced by policy makers, is to ask if the consideration of all benefits and costs can be extended to society. If we accept the Kaldor-Hicks compensation criterion (sometimes referred to the "potential Pareto criterion" (US EPA 2000)) then hormetic effects can be included in the benefit-cost calculation of regulation. But, as Hammitt (2003) points out, the individual's level of exposure becomes important. Zinc provides an example. At high doses, zinc can potentially lead to copper deficiency anemia and has a RfD 0.3 mg/kg-day (US EPA 1992). This protects the majority of individuals from the negative effects, but the RfD is below the Recommended Daily Allowance for pregnant and lactating women (Poirier and Dourson 1999). Any policy change affecting zinc would need to account for the negative effect on the majority, but also incorporate the hormetic effects for the sensitive subpopulation.

Unfortunately, while Hammitt (2003) lays out the correct economic optimum, the exercise may be extraneous for standard policy making. First, the law often limits policy for potentially hormetic substances to setting appropriate RfDs. As mentioned above, the RfD is not an optimal exposure level, it is a protective threshold. Under the Food Quality Protection Act (Public Law 104-170 1996), tolerances for dietary, water, and residential exposures must be set to assure a reasonable certainty of no harm, without consideration of costs. Even when explicit risk/benefit tradeoffs can be made, such the certain occupational and ecological exposures under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), the policy effort tends to be more of a search for the least costly way to achieve the lowest margin of exposure for a particular endpoint, not an optimization problem. A second reason why the optimum described above may be moot for conventional policy is because regulatory efforts such as tolerance setting and labeling have not traditionally been considered to be subject to Executive Order 12866 (President 1993). This is the executive order that requires a benefit-cost analysis for "major" regulation or regulation that raises novel policy issues. Absent a paradigm shift for more benefit-cost analysis, the economically optimal level of regulation may be of little use for standard command-and-control policy.

This is not to say that the economically optimal level is inconsequential. In fact, it may take on an increasingly important role if the EPA chooses to adopt more voluntary policy measures. Again, zinc provides and example. The subpopulation who could potentially benefit from zinc's hormetic effects above the RfD were advised to voluntarily seek their optimal level (Poirier and Dourson 1999). Deviating from direct hormetic effects, a similar discussion surrounds encouraging certain at risk groups to consume moderate amounts of alcohol (Klatsky 2003). Hammitt's (2003) article offers the economist an excellent starting point for benefits evaluation under these circumstances.

FOOTNOTE

[1] Charles Griffiths is an economist at the National Center for Environmental Economics. The views expressed here are entirely those of the author and do not necessarily represent the views of the U.S. Environmental Protection Agency.

REFERENCES

ADA (American Dental Association). 1999. *Fluoridation Facts*. Accessed on July 10, 2003. Downloaded from http://www.ada.org/public/topics/fluoride/facts-toc.html.

Calabrese, Edward J., Linda A. Baldwin, and Charles D.

Holland. 1999. "Hormesis: A Highly Generalizable and Reproducible Phenomenon with Important Implications for Risk Assessment" *Risk Analysis*. 19(2): 261-81.

Calabrese, Edward J. and Linda A. Baldwin. 2002. "Defining Hormesis" *Biological Effects of Low Level Exposures*. 10(2) (February): 25-31.

FIFRA (Federal Insecticide, Fungicide and Rodenticide Act), 21 U.S.C. 301 et seq.

Hammitt, James K. 2003. "Economic Implications of Hormesis" *Biological Effects of Low Level Exposures*. 12(1) (February). Forthcoming.

Heinzerling, Lisa and Robert J. Lechleider. 2000. "Hormesis and the Law" *Biological Effects of Low Level Exposures* Newsletter. 9(2) (January).

Klatsky, Arthur L. 2003. "Drink to Your Health? *Scientific American*. 288(2) (February): 75-81.

Poirier, Kenneth A. and Michael L. Dourson. 1999. "Is the Current Risk Assessment Paradigm Used by U.S. EPA and Others Compatible with the Concept of Hormesis?" *Biological Effects of Low Level Exposures*. 8(1) (July).

Portney, Paul R. 1992. "Trouble in Happyville" *Journal of Policy Analysis and Management*. 11(1): 131-2.

President. 1993. "Executive Order 12866 of September 30, 1993: Regulatory Planning and Review." *Federal Register*. 58(190) (October 4): 51735.

Public Law 104-170 (Food Quality Protection Act). 1996. 104th Congress, August 3.

Subcommitte on Chronic Reference Doses. 2000. "Derivation of Reference Doses" *Journal of Toxicology and Environmental Health, Part A*. 59(5/6) (March): 297-301.

US EPA (U.S. Environmental Protection Agency). 1989. *Fluorine (soluble fluoride) (CASRN 7782-41-4)*. Accessed on July 10, 2003. Downloaded from http://www.epa.gov/iris/subst/0053.htm.

US EPA (U.S. Environmental Protection Agency). 1992. *Zinc and Compounds (CASRN 7440-66-6)*. Accessed on July 22, 2003. Downloaded from http://www.epa.gov/iris/subst/0426.htm.

US EPA (U.S. Environmental Protection Agency). 2000. *Guidelines for Preparing Economic Analyses*. Washington, D.C.: Office of the Administrator. EPA 240-R-00-003.

COMMENT ON "ECONOMIC IMPLICATIONS OF HORMESIS" BY JAMES K. HAMMITT

Richard D. Morgenstern
Resources for the Future
Quality of the Environment Division
1616 P Street, NW
Washington, DC 20036-1400
E-mail: morgenst@rff.org

James Hammitt considers the economic implications of trying to reduce human exposures to so-called hormetic materials, i.e., substances known to be harmful at high levels but beneficial at lower levels. This differs from the typical regulatory situation where substances that are harmful at some exposure levels are treated as harmful at all exposures or as harmful at exposures above some threshold and benign at lower levels. Hammitt considers two different situations: 1) those involving individuals with known exposures and exposure-response functions; 2) those involving populations with variable or uncertain exposures or exposure-response functions. He also examines the potential benefits of trying to use tradable permits or other economic incentive measures to control hormetic substances.

At the outset, it is important to emphasize the relatively limited state of scientific knowledge about the hormesis hypothesis. As Hammitt notes, there are many well known examples of substances with hormetic properties, such as minerals which are essential nutrients at low levels and toxic at higher exposures. Proponents have compiled evidence to suggest that for a number of substances the hormetic model provides a more realistic description of experimental exposure-response data than either the linear no-threshold or the threshold models. At the same time, critics note the growing evidence that some developing organisms appear to be extremely sensitive to low levels of toxicants that have no known effect on adult animals. They also point to the relatively homogeneous animal populations from which the hormetic-relevant data are drawn and, even within such populations, the limited information available about the shape of the dose-response functions. If, for example, further research finds the potential benefits at low-level exposures are relatively modest, and possibly only relevant to a small subpopulation over a narrow exposure range, then the policy significance of the hormesis hypothesis may be quite limited. Similarly, if the harms associated with regulating such substances are minor and can be readily and inexpensively offset by other means, e.g., dietary supplements, then the regulatory significance may be further limited. Notwithstanding these important and still unresolved scientific issues, it is worthwhile to consider the economic implications for regulating substances that have (hypothetical) hormetic properties.

For the case involving individuals with known exposure-response functions and exposures, Hammitt convincingly demonstrates that optimal control under a hormetic model may be either more or less stringent than under the linear no-threshold model, depending on the incremental control costs. With the regulatory goal set to maximize net benefits, then in the case of known exposures and exposure-response functions the hormetic model calls for relatively more stringent controls compared to the no-threshold model when marginal abatement costs are low, and relatively less stringent controls when the reverse holds.

The second situation, involving populations with variable or uncertain exposures or exposure-response functions, is considerably more complex. As Hammitt notes, decisions that influence exposure levels in a population are likely to differentially affect individuals whose exposure levels and possibly exposure-response functions may vary because of environmental, behavioral, genetic, or other factors. If some individuals are exposed at beneficial levels and others at harmful levels then reducing ambient exposures will clearly aid some and harm others. At the same time, if certain population subgroups, e.g., fetuses or infants, have particularly low thresholds for harm, then reducing overall ambient exposures will likely result in gains to this group for all but the most stringent control levels. In essence, the hormesis hypothesis potentially transforms the classic risk-cost tradeoff involved in most environmental decisionmaking into one of a risk-risk trade-off.

The notion of a risk-risk trade-off is certainly not a new one for environmental regulators. In the case of drinking water chlorination, for example, both the reduction in infectious disease and the increased cancer rates associated with adding chlorine to the water supply were expressly considered by the EPA in setting the standards. Other examples that readily come to mind include the addition of iodine to salt or of fluoride to drinking water. As Hammitt notes, conceptually the benefit-cost methodology applicable to this situation is straightforward. Of course, the data intensity and analytic complexity of the computations are likely to be greater for hormetic substances than for those conforming to the linear, no threshold model. As always, the total (population) impacts would be evaluated as the sum of the estimated beneficial and adverse effects across individuals. A strict benefit-cost calculus would involve *balancing* possible benefits and harms to individuals against control costs. While such balancing does

involve the well-known issues associated with interpersonal utility comparisons, regulators are quite familiar this problem. This is not to deny the potential for controversy – particularly if the individuals perceived to suffer harm are part of a group deemed to be especially worthy of assistance.

Undoubtedly, what distinguishes drinking water chlorination and similar cases from the situation that the proponents of the hormesis hypothesis have is mind is the relative magnitudes of the health benefits and harms involved, along with the associated uncertainties. For example, what if there was a potential for large numbers of cancers to result from the addition of chlorine to drinking water? And what if an already disadvantaged group was particularly susceptible? From the perspective of benefit-cost analysis two outcomes are possible: either the net benefits are shown to justify regulation or not. However, if the hormesis hypothesis holds and large numbers of people are adversely affected by either emission controls or the absence of such controls, then those adversely affected may have a strong basis for opposing the regulatory decision, whether or not it is consistent with the benefit-cost analysis. In other words, in the presence of hormesis, the job of the regulator potentially becomes more difficult than in the case of a no-threshold pollutant. Heretofore, the regulator was balancing the health benefits of regulation against the economic costs to industry and/or society – itself a difficult task. With a hormetic substance, she must now balance the health benefits of regulation to some possibly ill-defined sub population against the health damages of regulation to some other possibly ill-defined sub population. In addition, she must consider the economic costs.

Hammitt's third issue involves the implications of the hormesis hypothesis for the use of emissions trading or other economic regulatory instruments. He correctly argues that because of concern about 'hot spots' or other highly differentiated spatial effects, economic instruments are likely to yield relatively fewer efficiency gains for the regulation of hormetic substances compared to the traditional command-and-control approach. At the same time, several caveats are in order. In fact, economic instruments are rarely used to regulate toxic substances, the ones most likely to exhibit hormetic properties. The most obvious exception is the use of information programs to limit exposures to substances listed in the Toxics Release Inventory, although even that depends on whether one defines such a program as an economic instrument. More commonly, economic instruments are applied to pollutants which travel long distances and/or mix a great deal in the environment. Since toxics do not generally meet these criteria, concerns about using economic instruments in this application are largely academic. Having said that, it is important to note that hybrid instruments, which combine both command-and-control and economic approaches, are growing in use. In most cases, the presence of hormetic substances are likely to present only slightly greater complications when there are also highly differentiated spatial impacts compared to those substances which do not exhibit hormesis. Thus, noting the limitations for regulating hormetic substances via economic instruments, one should be careful not to close the door all the way.

Overall, it is clear that hormesis hypothesis presents certain new challenges for environmental regulation. While the fundamental economic methods of analyzing hormetic substances are well known, and the regulatory agencies already have some experience in addressing substances which involve risk-risk tradeoffs, in previous cases the harmful effects of regulation have been judged to be relatively modest. If more extensive harm from reducing low-level exposures to known toxicants is documented for different population subgroups, then the overall complexities of regulating such substances, including the data and analytic requirements, will surely increase. The complexities will be even greater if simple dietary or other antidotes are not readily available. Hammitt has laid out the economic implications quite clearly. We now await more robust scientific information.

ECONOMIC IMPLICATIONS OF HORMESIS: SOME ADDITIONAL THOUGHTS

V. Kerry Smith
Director, CENREP
University Distinuguished Professor
and Resources for the Future Fellow
North Carolina State University
4223 Nelson Hall
Campus Box 8109
Raleigh, NC 27695-8109
Phone: 919-513-3761
Fax: 919-515-1824
E-mail: kerry_smith@ncsu.edu

Mary F. Evans
Assistant Professor
Department of Economics
University of Tennessee
Knoxville, TN
Phone: 865-974-1700
E-mail: mevans12@utk.edu

Introduction

Hammitt (2003) has prepared an excellent summary of the implications of hormesis for the application of economic principles to environmental decision-making. His primary focus is on how uncertainty about the exposure-response function(s) affects the efficiency properties of policy instruments. These functions can describe direct human health effects or they might involve impacts on plants or animals that people care about. Because we generally agree with his analysis and conclusions, our comments will focus on additional issues that either follow from his paper or were not completely developed.

Our discussion begins by suggesting additional policy insights might be added to Hammitt's description of the problems posed by hormesis by considering cases where similar phenomenon arise elsewhere in environmental economics. It also comments on the difficulties hormesis poses for benefit measurement and closes with an alternative rationale for applying the efficiency principles Hammitt discusses.

Policy Design Lessons from "Hormesis-Like" Responses with Environmental Amenities and Disamenities

Hormesis relates to dose-response (equivalently exposure-response or concentration-response) functions that are characterized by either a U-shaped or an inverted U-shaped pattern (Calabrese and Baldwin 2001). Such phenomena are not unique to health effects. They pose important questions for the design and selection of policy instruments and the measurement of the benefits from intervention. Three examples help to illustrate the potential similarities between hormesis and more familiar environmental policy issues.

There is growing recognition in evaluations of land use policy in the U.S. and Europe that agricultural landscape can provide amenities to residential homeowners if they do not live "too close." Odor, dust, pesticide drift, and a variety of negative externalities can be experienced by *immediate* neighbors, but if the residence is located a *short* distance away then agricultural fields are perceived as creating "open space" amenities. This link between distance and open space amenities resembles a hormetic relationship. Of course, in this case, distance serves as a proxy intended to capture how the landscape in a residential area changes with greater proximity to farmland.

Another example arises in economic models of congestion that describe the implications of alternative use patterns for recreation sites. There are a number of possible situations where congestion affects users' enjoyment of recreational activities. Over some levels of total use and for some types of recreation, each person's experience can be enhanced by having other users in the same area. A beach trip will yield greater satisfaction (for some users) due to the presence of others. The beneficial effect may arise from the enjoyment associated with "people-watching". Nonetheless, there may be a threshold. At some point more users in the same time period will detract from the experience. Too many people limit beach and ocean use, infringe on "space", and create other external negative effects on proximate users.

Our last example involves high-risk recreation such as rock climbing, whitewater rafting, advanced skiing, etc. In these cases, the amenity value of the resource depends on the skill of the user. Beginners in each of these activities tend to prefer safer climbs. Advanced climbers prefer more challenging terrain, to which the beginner would arguably assign a negative value. Similarly, easy climbs are boring to the advanced user (see Jakus and Shaw 1996). The same type of description could be applied to recreationists with varying skills who participate in whitewater rafting and skiing. In each case, skill and experience alter an individual's demands for the attributes of a resource. Combining the evaluations of a low skill and a high skill recreationist reveals hormetic-like behavior.

These examples are not intended to divert importance from hormetic exposure-response functions

associated with the health impacts of pollutants. Rather, our objective is to call attention to completely different situations with "similar" diversities in response to a beneficial (or harmful) stimulus. They may help in gauging whether the policy responses used in these quite different situations provide lessons for policies involving pollution and health.

Several common factors stand out in our examples - the role of heterogeneity among people in understanding these other hormetic-like responses, the definition and measurement of the source for the hormetic response, and the response itself. As a rule, the policy interventions for these cases have selected instruments that allow the affected individuals to undertake diverse responses. For example, information about the difficulty of sites for skiing, rock climbing, and whitewater rafting allows people with different skill and experience to sort among them. It creates the market equivalent of differentiated products.

In the case of rural amenities, information is also a key issue, alerting those living near operating farms to the implications of close proximity so that their choices reflect full information[1]. The case of congestion at recreation facilities can also be handled by defining capacity limits, specifying access conditions (e.g. first come, first served up to limit, differential pricing, etc.), and disseminating information about the attributes of different sites.

How might these relate to the case of the hormetic relations involving health and pollution? We believe they raise questions about uniform or "one size fits all" policies. Such uniform policies seem to be the primary focus of Hammitt's comments. He suggests more complex policies regulating the source of the problem with a tax or quantity of permits per unit of exposure that takes account of varying impacts across sub-populations. He regards these policies as having potential, but also emphasizes the complications associated with their implementation.

Why not consider the role of information to the exposed populations so they can adjust, sorting in response to unobserved (to the policymaker) heterogeneity in their gains or losses from exposures? Once we acknowledge the prospect and potential for such averting behavior, a strategy facilitating diverse responses seems more desirable. The logic for it can be seen to parallel the approaches used in our examples from outside the domain of health / pollution impacts.

Measuring Incremental Benefits from Controlling Exposure to Harmful Stimuli

One of the challenges posed by hormesis not developed in Hammitt's paper concerns the measurement of the incremental benefit of reducing exposure. Hammitt's framework for benefit measurement implicitly accepts the health effect / valuation strategy that EPA (1999) has adopted for most benefit-cost analyses. That is, a concentration-response relationship usually derived from epidemiological research is used to estimate the cases of illness or premature death avoided as a result of reductions in ambient concentrations of pollutants. In most cases, the analysis focuses on a sensitive group and then assumes the effects of each pollutant being studied will be absent for other groups. Beneficial effects of pollutants are rarely considered. Cases avoided are then converted to benefits using unit values, such as measures of the value of fatality risk reductions from labor market studies (e.g. economic measures labeled as the value of statistical lives, "VSL").

Once we acknowledge that the exposed population can adapt, then a revealed preference model is needed to recover the net outcome. In this case, the effect realized depends on actual exposure, after accounting for each individual's adjustment to situations creating a potential for being exposed to the pollutant(s). Hormesis greatly complicates this process because revealed preference models that allow identification of individuals' marginal benefit functions tend to separate the detection of amenity and disamenity effects into separate behavioral responses. The first exploits complementarity relationships with private goods and the second substitution relationships.

For example, improved air pollution increases time outdoors (complementary activities) and reduces air conditioning, materials maintenance, etc. (substitution activities). If we suppose the relationship switches from amenity to disamenity at different points for different people, then we face significant difficulties in measuring behavioral responses. Our models rely on maintained assumptions that impose either complementary or substitution relationships between private goods and the external effect outside their direct control. As a rule, we assume these patterns do not switch. Our methods for analysis do not look for the switching point or consider strategies to detect behavior consistent with it[2]. The challenges in responding to this implication of hormesis are exceptionally great.

Potential Pareto Criteria and Hormesis

Hammitt describes the conventional argument for judging the efficiency properties of policy by comparing expected aggregate benefits to aggregate costs. This strategy implicitly accepts that situations with positive aggregate net benefits would allow gainers to compensate losers. We agree. However, it may also be helpful to consider an alternative conceptual rationale. Suppose we consider how people would choose to have the public sector allocate expenditures on risk reduction if they did not know their personal circumstances in terms of wealth, sensitivity to the pollutant, etc. Pratt and Zeckhauser's (1996) analysis of the implications of this rhetorical question, under an expected utility hypothesis, calls for a focus on incremental willingness to pay for risk reductions. Ideally, people would continue to adapt, if given opportunities, until their personal incremental costs of averting risk equaled their incremental willingness to pay for the risk change. Using the public criteria that equates incremental public expenditures for risk reductions to marginal willingness to pay assures an

efficient level of public and private averting behavior (i.e. individual marginal cost and the marginal cost of public intervention are equalized because both equal marginal willingness to pay). While this is a conceptual ideal, it does serve to motivate policies that allow heterogeneous agents to adapt personally in lieu of imposing a uniform policy on all.

FOOTNOTES

[1.] Implicit in this statement is a "grandfathering" of rights to use land in particular ways. Depending on these rights and what are accepted as permissible external effects, then information and recognition of the impacts of land use changes become the foci of policy.

[2.] Smith and Evans (2003) discuss this in the context of relationship between weak complementarity and weak substitution, extending earlier work on weak complementarity by Smith and Banzhaf (2004).

REFERENCES

Calabrese Edward J. and Baldwin Linda A. 2001. Hormesis: u-shaped dose responses and their centrality in toxicology. *TRENDS in Pharmacological Sciences* 22(June):285-291.

Hammitt James K. 2003. Economic implications of hormesis. Working paper. Center for Risk Analysis, Harvard University, May.

Jakus Paul M. and Shaw W. Douglass. 1996. An empirical analysis of rock climbers' response to hazard warnings, *Risk Analysis* 16(4):581-586.

Pratt John W. and Zeckhauser Richard J. 1996. Willingness to pay and the distribution of risk and wealth. *Journal of Political Economy* 104(August):747-763.

Smith V. Kerry and Banzhaf H. Spencer. 2004. A diagrammatic exposition of weak complementarity and the Willig condition. *American Journal of Agricultural Economics* (May).

Smith V. Kerry and Evans Mary F. 2003. Reconciling health and amenity consequences of pollution: insights from intersecting indifference curves. Working paper. Center for Environmental and Resource Economic Policy, North Carolina State University, July.

U.S. Environmental Protection Agency. 1999. *The Benefits and Costs of the Clean Air Act Amendments of 1990-2010. Report to the U.S. Congress* (November).

HORMESIS, HOTSPOTS, AND EMISSIONS TRADING

Jonathan B. Wiener*
Law School, Nicholas School, Sanford Institute,
and Center for Environmental Solutions,
Duke University
Box 90360
Durham NC 27708-0360
Tel: 919-613-7054
Fax: 919-613-7231
Email: wiener@law.duke.edu

Instrument choice has been a major topic of environmental law and environmental economics. But to date we have not addressed whether instrument choice should be sensitive to the particular toxicology of each pollution problem. How does the dose-response relationship for a pollutant affect the optimal choice of policy instrument to regulate that pollutant?[1] James Hammitt's insightful article[2] analyzes many of the policy implications of a hormetic or J-shaped dose-response function (in which a substance is harmful at high doses but beneficial at low doses). Examples of hormesis may include vitamins, essential minerals, sunshine, red wine, oxygen, salt, and perhaps many other substances, possibly even radiation and some pollutants. Most of Hammitt's article deals with the implications of hormesis for standard setting: how much protection against exposure the government should require. In the last section of the article, Hammitt briefly addresses the implications of hormesis for instrument choice. Given a standard to be achieved, should the government use command-and-control instruments such as technology mandates, or fixed limits on the quantity of emissions at each source, or market-based incentive instruments such as emissions taxes or tradable emissions allowances? And, how is that choice affected by the shape of the dose-response function?

Economic theory and experience show that if the cost of emissions abatement varies across sources, emissions taxes and tradable emissions allowances will generally achieve a given standard at less cost (or achieve greater protection at a given cost) than will technology mandates or fixed emissions limits.[3] Hammitt argues, however, that tradable emissions allowances are likely to have less of an advantage over alternative policy instruments if the dose-response function is hormetic than if it is linear. He says: "Under the hormetic model, a simple economic incentive mechanism having a single tax or single type of tradable permit would not be anticipated to work as well as under the linear model."[4]

Here I analyze more thoroughly Hammitt's conjecture about the effect of hormesis on emissions trading. First, I distinguish the effects of hormesis from the effects of local emissions. Hormesis is an attribute of the dose-response or exposure-response relationship. Hotspots are an attribute of the emissions-exposure relationship. Some pollutants may be hormetic and cause local emissions-exposure effects; others may be hormetic without causing local emissions-exposure effects. It is only the local exposure effects of emissions that poses a problem for emissions trading. Second, I show that the conditions under which emissions trading would perform less well or even perversely under hormesis depend on how stringent a level of protection is set. Only when the regulatory standard is set at the nadir of the hormetic curve would emissions trading be seriously perverse (assuming other restrictive conditions as well), and such a standard is unlikely. Moreover, the benefits of the overall program may justify the risk of small perverse effects around this nadir. Third, I argue that hotspots can be of concern for two distinct reasons, harmfulness and fairness. Last, I argue that the solution to these problems may not be to abandon market-based incentive instruments and their cost-effectiveness gains but to improve them further by moving from emissions trading and emissions taxes to risk trading and risk taxes. In short, I argue that hormesis does not pose a major obstacle to emissions trading or emissions taxes, but to the extent that hormesis does pose such a problem, a shift toward risk trading or risk taxes is the superior route.

A. Policy Instruments and Dose-Response

Market-based incentive instruments such as emissions taxes and tradable emissions allowances reduce the cost of achieving a given level of protection (or achieve greater protection at a given cost) by allowing the degree of abatement to vary across sources. This variation is often termed intersource flexibility, "where flexibility," or locational flexibility. In contrast, fixed emissions limits allow each source flexibility in its internal abatement or compliance method (intrasource or "how" flexibility) (as do tradable emissions allowances), but no flexibility to vary emissions across sources. (Fixed emissions limits can allow "where" flexibility within the source; this is typically called a plantwide "bubble" approach to aggregate emissions from the source. Tradable emissions allowances expand the bubble to encompass more than one plant or firm across a region, thereby expanding the scope for intersource "where" flexibility.) Technology mandates dictate specific abatement methods at each source and thus allow neither "where" flexibility nor "how" flexibility. If the cost of emissions abatement varies across sources, then where flexibility enables more abatement to be undertaken at sources where it is less costly, and less abatement to be undertaken at sources where it is more costly. The result is the same aggregate

level of abatement, but at lower total cost, and with some heterogeneity in the ultimate level of emissions across sources.

Consider a hypothetical lake with the pollutant smox emitted by four sources evenly spaced around its shore, A, B, C, and D. As shown in Table 1, assume that without any required controls, each source emits 20 units of smox, for a total of 80. Now the government imposes limits to achieve a 50% aggregate reduction to a new total of 40 units of smox. The cost and location of this abatement depends on the policy instrument the government chooses. If the government imposes a technology mandate (smox scrubbers) predicted to cut emissions by 50% (to 10 units at each source for a new total of 40), the cost per source will be $100, for a total cost of $400. If however the government imposes fixed emissions limits of 10 units per source, each will again emit 10 for a total of 40; source A will install the smox scrubber at a cost of $100, but sources B and C will change their materials inputs to reduce emissions at a cost of $50 each, and source D will change its process method to reduce emissions at a cost of only $20, so the total cost will be $220. "How" flexibility saves $180. If alternatively the government imposes tradable emissions allowances, issuing 10 allowances to each source, then again the total emitted will be 40. But now the amount emitted at each source may vary from 10. Source A will seek to avoid the costly smox scrubber and instead purchase 10 allowances from the other sources at some cost less than the scrubber's $100. Source D will seek to sell additional abatement at a price above its low cost of abatement (say, $25 for the next 10 units). If A buys 10 allowances from D, then A emits 20, D emits zero, and B and C emit 10 each, for a total of 40. A will pay somewhere between $26 and $99 to D to purchase D's 10 allowances; assume A pays $30 to D, saving A $70 while earning D $5. B and C will each spend $50 as before. The total abatement cost will be $145 ($50 each at B and C, plus $20 at D for the first 10 abated and $25 for the second 10 abated; the extra $5 paid by A to D is a transfer). "Where" flexibility saves $75 compared to fixed emission limits and $255 compared to technology mandates. A similar result would obtain if the government imposed an emissions tax; source A would pay the tax while source D abates to avoid the tax and sources B and C abate partially and pay partially.

The example above omits the administrative costs of emissions monitoring and enforcement, which would be necessary under all three policy instruments. It also omits risk-risk tradeoffs with emissions of other pollutants or into other media. It also shows A purchasing allowances to return to 20 units of emissions and not higher, on the intuition that if A's unregulated emissions were 20, A would not now find it more profitable to emit more than 20 (at the cost of additional allowances or tax payments); otherwise A would have emitted more than 20 in the unregulated case (for free). This observation suggests an upper-bound to the bunching that may occur under emissions trading: it is highly unlikely that sources will purchase allowances to raise their emissions above their unregulated levels. Given that purchasing allowances (or paying the tax) is costly, it seems almost inevitable that bunching cannot exceed the unregulated (free) level of emissions, all other factors held constant. Of course, a source's unregulated business-as-usual emissions might rise over time as its product output grows, so a source that initially emitted 20 might later want to emit 25 or more. Adopting a fixed emissions limit would prevent that growth, but technology mandates could permit such growth to occur at all sources, and tradable allowances could permit such growth to occur at some sources while reductions (to meet the

Table 1. Illustrative Emissions and Costs under Alternative Policy Instruments.

Source	No regulation		Technology		Fixed Emissions		Tradable Emissions	
	Smox emitted	Abatement Cost	Smox emitted	Abatement Cost	Smox emitted	Abatement Cost	Smox emitted	Abatement Cost
A	20	0	10	100	10	100	20	0
B	20	0	10	100	10	50	10	50
C	20	0	10	100	10	50	10	50
D	20	0	10	100	10	20	0	45
Total	80	0	40	400	40	220	40	145

aggregate cap) occur at other sources. Also, if emissions trading were introduced as a replacement for previously imposed fixed emissions limits, then emissions at source A might grow to exceed the previous (regulated) level, but emissions at other sources would have to fall more sharply as A purchased allowances. But usually emissions trading is adopted to meet a more stringent cap that reduces aggregate emissions considerably below previous requirements, so increases at any source above the previous level are unlikely.

So far the analysis has ignored the dose-response function; it has implicitly assumed that every unit of exposure is equally harmful (constant marginal harm from exposure), and that every unit of emissions yields one unit of exposure (constant marginal exposure from emissions). In that case, aggregate harm depends only on aggregate emissions.

The claim is often made, however, that under the locational flexibility allowed by emissions trading, "bunching" of higher emissions in one or some locations, like source A in the example above, could create "hotspots" (near source A) that increase total harm (or decrease total harm less than intended). Of course, as emissions allowances are bunched at A, they must be "drained" from other locations where emissions are overcontrolled to sell allowances, like source D in the example above, and perhaps the net effect of draining and bunching is a wash. (The same pattern of bunching and draining could occur under emissions taxes.)

But the situation is more complex and depends on the dose-response function, illustrated in Figure 1. Consider four possibilities.

(1) Under a linear dose-response function (each unit of exposure causes one unit of harm), with constant marginal harms of exposure, and constant marginal exposure from emissions, bunching and draining (heterogeneity in emissions) would not increase total harm compared to a uniform distribution. They would offset each other. This is the standard initial case analyzed above.

(2) Under a supra-linear dose-response function, which is monotonically increasing (not hormetic) but has increasing marginal harms (each added unit of exposure causes more harm than the prior unit), and with constant marginal exposure from emissions, bunching would yield rising harms, but draining would yield declining harms, compared to a uniform distribution of emissions. Thus in the example above, under emissions trading as compared to the fixed emissions limits of 10 at each source, the increase at source A from 10 to 20 would raise harms more than if A and D each emitted 10; but the decline at source D from 10 to zero would also reduce harms more than if A and D each emitted 10. The net effect is uncertain, and depends on whether the buyers (bunching) and sellers (draining) are above or below the average level of emissions. (a) If buyers increase emissions above the average and sellers reduce emissions below the average, and the dose-response function has increasing marginal harms, then the harms from bunching above the average will likely exceed (grow more steeply than) the benefits from draining below the average. Divergence of buyers and sellers away from the average of the supralinear dose-response function will increase harms on net. (b) If however buyers increase emissions from below the average toward the average, and sellers reduce emissions from above the average toward the average, then the result will be convergence toward the average and the benefits of draining may exceed (harms fall more steeply than) the harms from bunching. Convergence of buyers and sellers toward the average of the supralinear dose-response function will decrease harms on net. (In the smox example above we began with uniform emissions of 20 at each source, but in reality the buyers and sellers could start with different emissions, and then either diverge or converge through emissions trading depending on their costs of abatement.)

The supra-linear dose-response case might describe the SO2 emissions trading system adopted in 1990 to reduce acid rain. Some sources of SO2 in the Midwest emit plumes that yield acid deposition in relatively sensitive ecosystems such as the Adirondack Mountains, while other sources in other parts of the US emit plumes that land in relatively insensitive areas such as the Atlantic Ocean. If the emissions trading program had resulted in undercontrol (bunching) by allowance buyers whose emissions are deposited into the sensitive Adirondacks (above average, steeper marginal harms), and overcontrol (draining) by allowance sellers whose emissions would thereby be removed from eventual deposition into the insensitive Atlantic (below average, flatter marginal harms), the net result could have been an increase in harm compared to fixed emissions limits. Congress was aware of this concern but nonetheless adopted a national SO2 emissions trading market for several reasons, including the prediction that the pattern of actual control costs at SO2 sources would yield the reverse: more allowance selling (draining) at sources that deposit into the Adirondacks and more allowance buying (bunching) at sources that deposit into the Atlantic or other places – a happy coincidence of control costs and wind patterns yielding a net benefit from emissions trading. That is, the SO2 market resembled case 2(b) above, with draining occurring on a steeper part of the dose-response function, and bunching occurring on a flatter part of the dose-response function, hence yielding net benefits despite any bunching. Another consideration favoring the adoption of the SO2 trading program was the prediction that the aggregate 50% reduction in emissions from the 1987 level would help the Adirondacks far more than any bunching might detract from that benefit.

The supra-linear dose-response case could also describe an emissions trading program in which some emissions affect sensitive subpopulations who are increasingly damaged by higher emissions. Some have raised this point as an environmental justice concern about sensitive subpopulations — including low-income and

minority populations — who are residents located near sources of toxic materials, such as emissions trading involving mercury or industrial chemicals.[5] The same two factors that made trading a net benefit in the SO2 case could also be applicable in the environmental justice cases, but that is an empirical question, depending on the pattern of control costs and thus of buying (bunching) and selling (draining) and the corresponding marginal harms.

The basic point here is that the concern about the harmfulness of bunching and hotspots must arise, if at all, from a supra-linear dose-response function. If the dose-response function is linear, as noted above under (A), then bunching and draining due to emissions trading would have no net effect on harmfulness. (Fairness is another matter, to which I return below.)

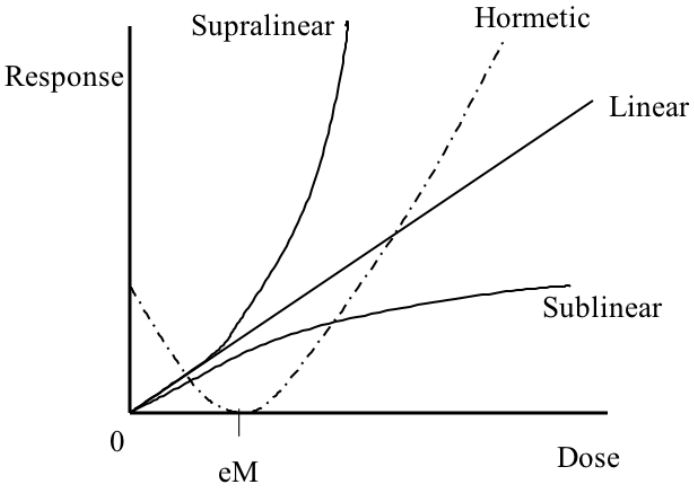
(3) If the dose-response function is sublinear, it flattens out and has declining marginal harms with increasing exposure. In this case bunching (from buying allowances) and draining (from selling allowances) would have the converse orientation to the supralinear case. In the sublinear case, trading would be beneficial on net if buying (bunching) and selling (draining) diverge from the average, so that bunching occurs where the dose-response curve is rising but flattening out, and draining occurs where the dose-response curve is falling more steeply. But trading would be harmful on net if buyers and sellers converge toward the average of the sublinear dose-response curve, i.e. if bunching occurs where the dose-response function is rising more steeply and draining occurs where the dose-response curve is falling slowly.

(4) Under a hormetic or J-shaped dose-response function, harm increases with increasing exposure above a certain positive exposure level (the nadir of the function, labelled eM in Hammitt's graphs for the harm-minimizing exposure[6]), but harm also increases with decreasing exposure below that point (eM). If there are constant marginal exposures from emissions, and if the regulatory standard (the number of allowances) is set precisely at eM, and if all sources begin at eM, then under hormesis, bunching (undercontrol to emit above eM via buying allowances) would yield increasing harms *and* draining (overcontrol to emit below eM and selling allowances) would *also* yield increasing harms, compared to fixed uniform emissions by all sources at eM. Under these restrictive conditions, bunching is harmful because it ascends the positively sloped portion of the J-shaped curve to the right of eM; this is the same problem as under a supra-linear but monotonically increasing dose-response function (when bunching moves sources above the average, as discussed above under case 2(a)). Under the above restrictive conditions, the twist is that draining (overcontrol) is *also* harmful because it ascends the negatively sloped portion of the J-shaped curve to the left of eM.

If however emissions trading starts with some sources emitting above eM and some emitting below eM, and if trading results in buyers bunching from below eM upward toward eM, and sellers draining from above eM downward toward eM (which depends on their starting points and their costs of abatement), then emissions trading would converge to eM and to the level of harm under fixed emissions limits set at eM.

In our smox example above, if the harm-minimizing level of exposure (eM) occurred with emissions of 10 at each source (note that this assumes local effects of emissions), then the unregulated emissions of 20 at each source would be harmful, and the move to fixed emissions limits of 10 at each source would be optimal. Under emissions trading, the increase at A from 10 to 20 would be harmful, and the decline at D from 10 to zero would *also* be harmful because under hormesis the harms also rise as exposures fall below eM. If 10 is the optimal harm-minimizing level of emissions (eM) (and emissions translate directly into exposures), and if each deviation of 10 emissions units in either direction increases harm by, say, 10 units of harm (a V-shaped hormetic curve which is linear at 45 degree angles on each side of eM), then before regulation the aggregate harm from emissions of 20 at each of four sources was 4x10 = 40. With fixed emissions levels of 10 at each source, the aggregate harm is zero. With emissions trading as illustrated in the smox example above, the harm is 10 from A emitting 20, zero from each of B and C emitting 10, and 10 from D emitting zero, for a total harm of 20. If the hormetic curve is flatter than this (rises more slowly) on each side of eM, the aggregate harm under emissions trading would be smaller than 20.

Figure 1. Dose-Response Curves



B. Hormesis and the Local Effects of Emissions are Distinct Attributes

The first implication of this analysis is that the effects of hormesis must be distinguished from the local effects of emissions. Some pollutants may be hormetic and have local emissions effects, while others may be hormetic without their emissions causing local effects. Hormesis is an attribute of the dose-response or exposure-response

relationship. Hotspots are an attribute of the emissions-exposure relationship. It is only the local impact of emissions that poses a problem for emissions trading. In the example above under case (4), emissions trading yielded increasing harms under hormesis at both source A (increasing emissions) and source D (decreasing emissions) only because we assumed that the optimal emissions at each source (eM) were 10 and that deviations from that point would yield harms. Some pollutants pose no such local effects from their emissions. If smox affected the lake only because of its aggregate concentration in the entire lake and not its local concentration near each source, then the reduction from 80 to 40 in the lake would still be far above the eM of 10 and on the positively sloped portion of the dose-response curve, and the emissions trade between A and D would not affect harms at all. Whether emissions were at 20 or 10 or 0 at each source would not matter to harms by itself; only the total smox in the lake from all sources combined would matter.

Hammitt cites CFCs and CO2 as substances that mix globally in the atmosphere and therefore the emissions of which have no local effects: the harms of these emissions are utterly independent of where they are emitted and depend only on the total quantity in the earth's atmosphere.[7] The ultimate effects of ozone depletion and global warming would be geographically uneven and would have larger effects in some locales than others (e.g. the ozone "hole" over Antarctica, more skin cancer in some populations than others, greater changes in temperature and precipitation in some places, and greater harms from sea level rise in some places), but these impacts would arise from global processes after the CFCs and CO2 are emitted and mixed into the global atmosphere, and would not be affected by the geographic pattern of the initial emissions. (A caveat: extremely large local releases of CO2 could have local effects, such as asphyxiation of ocean life due to sudden sea floor outgassing, but such releases are extremely rare and are not the kind associated with fossil fuel combustion or emissions trading.) Because their harms of emissions (and benefits of abatement) are independent of the location of emissions, both CFCs and CO2 are therefore excellent candidates for emissions trading, including international emissions trading. The U.S. used a tradable allowance system to help phase out CFCs. The costs of CO2 abatement vary considerably, so the cost savings from emissions trading are quite large.[8]

But although CO2 emissions have no local effects to speak of, CO2 does have a hormetic dose-response function (that is, where response means harm; CO2 seems to have a monotonically increasing effect on temperature, but a hormetic effect on human and ecological welfare). Whatever one thinks of predictions that rising levels of CO2 in the atmosphere may cause global warming, it is fairly clear that declining levels of CO2 in the atmosphere below some level would also be harmful. The preindustrial level of CO2 in the atmosphere, circa 1800, was about 275 ppm; the current level in 2003 is about 375 ppm. Exposure rising above 400 ppm is predicted to raise global average temperature. But exposure below, say, 250 or 200 or 150 ppm could lower global average temperature dangerously and also suppress plant photosynthesis. Just as the historical record suggests that rising CO2 is associated with warming, it also suggests that falling CO2 is associated with cooling and ice ages. Cause and effect are not fully established — perhaps warming and cooling cause changes in CO2 concentrations rather than the other way around – but if CO2 influences the climate, then it clearly exhibits hormesis, and the optimal exposure (eM) is probably somewhere around 250 to 350 or 450 ppm (although this point is controversial and difficult to test), with rising harms on either side.

Thus, hormesis alone does not pose any problems for emissions trading. A necessary condition for hormesis to detract from the success of emissions trading is that emissions themselves have local effects. The "where" flexibility offered by emissions trading is of concern only when the location of emissions matters to harm. Not all cases of hormesis are also cases of emissions-dependent local effects.

C. Dependence on the Level of Protection

The above analysis shows how emissions trading can under certain restrictive conditions perform less well or even perversely if the standard is set at the nadir of the hormetic curve (eM). In the example above, case (4), emissions trading yielded increasing harms (compared to fixed emissions limits at each source) because the harms were dependent on local emissions, and because both the buyer (A) and seller (D) deviated from the optimal emissions point (eM) at 10. Emissions-dependent local harm, however, is only one of the necessary conditions for this perverse result to obtain. In addition, it is also necessary that the standard have been set at or near the harm-minimizing point (eM) on the hormetic curve.

If the standard had been set at a higher level than eM, such as at the point Hammitt labels eL* or eH* (the optimal points for the linear and hormetic functions, respectively, when setting marginal benefits equal to marginal costs of control), then modest deviations from the standard by emissions allowance buyers (bunching) and sellers (draining) would still be on the positively sloped portion of the dose-response curve, and the analysis would be similar to that for the supra-linear curve as described above in case (2). Bunching would be harmful but draining would be beneficial, and the net effect would depend on where the bunching and draining occur on the curve. If by contrast the standard were set below eM (such as a very stringent level of control for, say, red wine or vitamins, set below the optimal level of that substance for long-term health), then draining would be harmful but bunching would be beneficial, up to eM.

Thus, in addition to depending on emissions having local effects, the influence of emissions trading on net

benefits under hormesis depends fundamentally on what level of protection is set and how that level compares to the nadir of the curve (eM). The perverse results are worst when the standard is set precisely at eM, and the perversity diminishes as the standard moves away from eM.

Yet it is unlikely that the standard would be set at eM. As Hammitt shows, if there are positive marginal costs of abatement, then under hormesis the optimal level of control is still above (to the right of) eM, at a point like eH*. On that analysis, the point eM would only be the optimal point to set the standard if the costs of abatement were zero, which seems implausible. At eH*, the hormetic curve is supralinear, and thus the impacts of draining are still beneficial (so long as they do not go so far as to slip below eM). In the example above, if eM were at 3 units of smox emissions and the government set the standard at eH* = 10, and if A decided to buy only 6 units from D rather than all 10, then the resulting emissions would be A=16, B=10, C=10, D=4. A's bunching would increase harm above the alternative of fixed emissions limits for each form at 10, but D's draining would reduce the harm compared to 10. Only if D sells more than 7 would D slip below eM = 3 and incur the perverse harms of too few emissions. It would be useful to review the emissions trading programs adopted to date, such as for leaded gasoline, SO2 (acid rain), NOx, and now CO2, to see if any of these programs set the standard so low near eM (and also had effects dependent on local emissions) such that locational flexibility risked perverse increases in harm from both bunching and draining; and if the standards were set above eM, then the effects of bunching and draining should still be netted out along the linear or supralinear dose-response function.

Some policies may of course set the level of protection at or below eM, even if considering cost would suggest a less stringent standard. If the standard for a hormetic substance or agent is set below eM, thereby depriving society of the health benefits of increased exposure up to eM, it is unlikely that emissions trading would be used to implement the policy because the level would be so close to zero that there would be little scope for overcontrolling emissions and selling allowances. Moreover, even if emissions trading were employed for a standard set at or below eM, it could turn out that the sellers were the larger emitters (above eM), and buyers were the lower emitters (below eM), so that emissions trading yields convergence to the optimal eM rather than divergent bunching and draining away from eM.

Thus the only situation in which hormesis could render emissions trading ineffective or perverse is when emissions have local effects, when the standard is set at or near eM, when eM is sufficiently far from zero to enable selling of allowances, and when allowance sellers are emitting below eM and buyers are emitting above eM.

Further, the relevant comparison is not just to the alternative of fixed emissions limits at each firm, but also to the level of unregulated emissions (eU). The aggregate reduction in harm from unregulated eU to the regulated level would often dominate any increases in harm that might ensue from the heterogeneity in local emissions allowed by emissions trading. That is, the difference between eU and the standard set at eH* (or eL*) may be so great that any bunching and draining around the standard would be negligible in comparison. This was one of the reasons for establishing a national SO2 trading system in 1990, as noted above. If the benefits of the aggregate reduction from eU to the standard are large enough, they can outweigh any harms from locational flexibility at every locale, thus yielding net benefits at every locale. Of course, fixed emissions limits could try to ensure greater net benefits at the sensitive locales subject to bunching. But fixed emissions limits would forego the benefits of draining at other locations (unless the draining slips below eM), and they would forego the cost savings from emissions trading. Because the cost savings from emissions trading may be the key to enabling government to set more stringent standards than under fixed emissions limits (as they arguably were in the SO2 case), the increase in protection of sensitive locations from employing fixed emissions limits may be outweighed by the increase in aggregate benefits to all locations (including the sensitive locations) from employing emissions trading. To illustrate using the SO2 case: if the cost savings obtained by shifting from fixed emissions limits to emissions trading enabled the program to achieve, say, a 10 million ton cut in national SO2 emissions rather than a 6 million ton cut in emissions at the same or lower overall cost, but emissions trading risked a 0.5 million ton increase due to bunching at sources that would deposit in the Adirondacks (although actually the pattern of abatement costs suggested draining at such sources, as noted above), then the extra 4 million tons of aggregate emission reduction from emissions trading as compared to fixed emissions limits could yield a net benefit to the Adirondacks despite the bunching. To illustrate using the smox example above: if the cost savings from emissions trading ($75) were partially or fully used to obtain more stringent emissions reductions (i.e., by spending up to $220 under emissions trading to equal what would have been spent under fixed emissions limits), thus reducing aggregate emissions not just to 40 but, say, to 24 (6 allowances allocated to each source instead of 10), and if smox is not hormetic or its eM is low, then the extra harms to the area near source A from any bunching due to A's purchase of, say, the 6 allowances from D could be outweighed by the benefits to all — and even to A alone — from the extra aggregate reduction achieved (16 units).

D. Fairness

Bunching and hotspots may be of concern for two possible reasons, harmfulness and fairness. So far we have addressed only harmfulness. Fairness concerns may remain even if harmfulness is not affected. In our initial

smox example, fixed emissions limits left each source with 10 units of emissions, whereas emissions trading left A with 20, B and C with 10 each, and D with zero.

Assume that the dose response function is linear (with one unit of smox generating one unit of harm), and that local emissions have local effects. Thus under no regulation, the harms are 20 at each of the four source locations and 80 in total. Under fixed emissions limits, the harms are 10 at each and 40 total. Under emissions trading, the harms are 20 at A, 10 each at B and C, and zero at D; the total harm remains 40. This is case (1) analyzed above in which emissions trading makes no difference to aggregate harm; the dose-response function is not supralinear or sublinear or hormetic, so bunching and draining have no disproportionate effect on harms. Harms are 40 under both policies, and no source has increased its total emissions above the unregulated level of 20. But now the local harms differ, and the community near source A may complain of unfairness because it suffers harms of 20 while the communities near B and C suffer 10 each and the community near D suffers zero. (The community could be the human population or, say, the aquatic population of organisms living in the lake near the source's discharge pipe. Whether a non-human ecosystem can raise fairness concerns is another story; for now I assume it can.)

For example, perhaps smox causes impaired childhood brain development (as does lead (Pb)) and attendant lost IQ, so that before all four communities were equally (and seriously) impaired, but now A's children will gradually fall behind B, C and D over time. The community at A might feel that this is unfair even if A does not increase emissions above its pre-regulation level of 20 (which as argued above it would not), so that the community near A is at least experiencing no greater emissions than it did before the policy. Although the other communities now experience lower emissions than they did before the policy, they also incur the costs of abatement. And the aggregate benefit is significant. Still, equality of future opportunity (freedom from unequal smox impairment) may be worth something to each community, in addition to the aggregate reduction in smox. If so, although the community near A is not experiencing increased emissions, it is harder to say that it is no worse off than before, given that it has not enjoyed any reduction in emissions (it is still at 20) and now it will lag the other three communities in IQ over time (and its source is paying to purchase some allowances — but that might save jobs in A that are being cut in D). This fairness concern might be restated as a claim that the policy is not Pareto-improving for each community. On the other hand, it is hard to argue that every community should be exposed to equal risk. No one argues that police or fire protection should aim to ensure equal mortality rates across all communities. If the cost savings from emissions trading enabled a tighter overall standard, then foregoing trading to maintain uniform fixed emissions levels would be less unfair to community A at the cost of greater injury to all (more impairment of other children). Perhaps community A could be compensated through side payments or through the income tax system. People near A could use the exit option to move to B, C or D. Or community A could use its voice to induce source A to reduce emissions further, or the government to tighten the aggregate level even further.

If the dose-response function is supralinear, such as because of more sensitive subpopulations at some locations, and if those subpopulations are subject to higher emissions due to bunching, then those subpopulations may raise a fairness concern that is compounded by increasing harmfulness. This would be the case if the community near A were also sensitive to greater injury from smox than other groups. This fairness concern is more compelling than the fairness concern associated with linear damages, because now the fairness concern is combined with a greater vulnerability to harm. This is the type of fairness concern often posed in environmental justice cases, where a subpopulation with lower incomes, perhaps lower nutrition and education, and perhaps victims of historical discrimination on the basis of race or ethnicity, argues that it would be subject to a higher exposure and that it is more vulnerable to damages from each unit of exposure. Of course, such concerns can be raised wholly apart from emissions trading systems, such as the siting of new industrial facilities.

On the other hand, if the sensitive subpopulation is located where the source subject to emissions trading is not bunching but draining, such as at source D, then the combined fairness and harmfulness concern would be neutralized or reversed. This was predicted to occur for the Adirondacks under the acid rain trading program, as lower-abatement-cost SO2 sources in the Midwest would overcontrol (drain) and sell allowances to higher-abatement-cost sources elsewhere which would undercontrol (bunch) and emit into other less sensitive deposition sites.

If the dose-response function is hormetic, and if the standard is set at or near the harm-minimizing level (eM), then the fairness concerns are complicated by the possibility of harm from both higher and lower exposures than eM under emissions trading. As discussed above, however, it seems unlikely that the standard would be set at eM.

Now assume that the harms depend only on the aggregate level of smox in the lake, not on the local emissions at each source. Then the community near source A should have no fairness concern. A does emit more than the others, but the local community is not affected any more or less by these emissions than everyone else. Indeed, the community near A may be accused (rightly or not) of unfairness by others around the lake who incur the harms of the aggregate 40 and see A emitting fully half that. (Of course A did so quite legally pursuant to the government's policy design, and the aggregate reduction to 40 — which is all that matters in

this case — has been achieved at cost savings to A and profits to D compared to fixed emissions limits.)

E. From Emissions to Risk

To the extent that hormesis poses a problem for market-based incentives such as emissions trading and emissions taxes, which as I have argued above is limited, the solution to these problems may not be to abandon market-based incentive instruments and their cost-effectiveness gains, but to improve them further by moving from emissions trading and emission taxes to risk trading and risk taxes.

One way to avoid the potentially perverse effects of emissions trading or taxes under hormesis is to forego emissions trading and taxes, and stick with fixed emissions limits. As noted above, that choice forfeits the cost savings and more stringent standards made possible by market-based incentives, which may well be more important than the possible perversities around the hormetic optimum. A second option is to constrain the market by limiting bunching (acquiring extra allowances) and draining (selling allowances), such as by forbidding buyers to buy more than a certain amount, or sellers to sell more than a certain amount, or anyone to emit more than some maximum and minimum. Those options may be helpful but also inhibit the market and curtail its cost savings. They are partial moves toward fixed emissions limits. A third option is to segment the market geographically so as to prevent trades that cause bunching and draining in certain places. For example, if the community near source A is especially vulnerable, the policy could allow emissions trading but prohibit sales to A. Or the acid rain trading program could restrict sales to sources whose emissions would be deposited in the sensitive Adirondacks (as New York State tried to do, by forbidding LILCo from selling allowances to upwind sources). These approaches also inhibit transactions and forego cost savings. Thus in each of these cases, the gains from avoiding hormetic perversities must be weighed against the costs of using the less cost-effective design.[9]

The fundamental reason that emissions trading and taxes confront problems under hormesis is not, however, a problem with trading or incentives. It is the focus on emissions as the target of the policy. Hormesis is vexing because it reveals that emissions and exposure are not proportionate to risk. And as Hammitt notes, even without hormesis, if emissions are not always equally proportionate to exposure — if emissions translate into different exposures at different sources — then emissions trading and taxes will not outperform command-and-control policies as well as they do when emissions yield equal exposures everywhere.

The lesson is that emissions are only an intermediate stage in the production of risk. Policy can operate at several different control points, including the inputs to the firm (e.g. fuels and materials), the firm's internal processes and technologies (e.g. combustion methods, scrubbers), the firm's outputs such as emissions (e.g. emissions limits, allowances or taxes), the ambient levels of the pollutant (e.g. workplace ambient standards or state-level NAAQS or CO2 levels), exposure (e.g. worker protective equipment), and ultimately risk (e.g. tort liability or workers' compensation). In practice we can see examples of all of these approaches. Much air and water pollution policy has focused on the firm's technology, emissions, or ambient levels. And tradable allowance policies have often focused on emissions (although the lead phasedown involved tradable gasoline content credits applied upstream to refiners — an arrangement that helped avoid any local emissions effects).

But if emissions are not closely related to exposure and in turn to risk, or are sometimes inversely related to risk because of hormesis, then policy should turn to controlling risk rather than emissions. As Hammitt notes briefly, "A more complicated system, in which the tax or quantity of permits required per unit of exposure varies across subpopulations in proportion to the marginal benefits of reducing exposure could provide superior outcomes but would be more complicated to develop."[10]

One approach would be a tax-subsidy regime which taxes emissions (or exposure) above eM, and subsidizes them below eM.[11] A second approach is to trade risk allowances, not emissions allowances.[12] "Risk bubbles" have been advocated within firms and across site-specific projects. For example, Paul Portney has suggested that "regulated entities, public or private, should be allowed to relax pollution controls at one point and install new protective measures elsewhere, subject to a demonstration that an overall improvement in health or environmental quality will take place as a result of the change."[13] In Richard Stewart's formulation, "Under risk bubbles, a facility operates under an aggregate, multi-media 'umbrella' of residuals limitations for the facility as a whole. This 'risk cap' creates economic incentives because higher levels of discharges by the facility of a given residual in a given medium carry an opportunity cost in the form of the resources that must be devoted to reducing other residuals in order to stay within the cap. In effect, the cap creates an implicit internal residuals trading market."[14] Thus a plant could relax air pollution controls at one place in exchange for achieving an extra reduction in water pollution emissions elsewhere in the plant or at another plant in the area, subject to EPA approval of the net reduction in risk. EPA and the plant owners would negotiate the bubbling arrangement,[15] and the plant owners would bear the burden of proving that their water pollution reductions are at least equal to the air pollution increase in terms of environmental risk. EPA has begun to use risk bubbles to allow net risk improvements at individual firms through Project XL.[16] Others have advocated letting parties responsible for Superfund cleanups negotiate agreements with host municipalities and states in which all would agree to spend some of the funds on other measures that would reduce risks even more, such as air pollution control or city street lighting or vaccines for children.[17] Frank Cross advocates setting an aggregate risk limit (called a "risk

cup") and then requiring new sources of additional risk to purchase offsetting reductions in risk from existing sources of risk.[18] The Clean Air Act's provisions for air toxics offsets (section 112(g)) and CFC-substitutes (section 612) both authorize interpollutant trading based on achieving reductions in overall risk.

The concept of risk bubbles, like emissions bubbles, represents an initial project-based version of locational flexibility. And just as the experience with emissions bubbles under EPA's bubble, netting and offsets policies in the late 1970s and early 1980s became the basis for the more formal and more successful emissions allowance trading policies for lead, CFCs and SO2 in the 1980s and 1990s, so too can risk bubbles build the groundwork for formal risk allowance trading. Under formal risk allowance trading, each source would be allocated allowances to cause a limited number of units of harm, significantly reducing harm from the unregulated level. Then each source, perhaps assisted by the government, would calculate the effects of its activities on risk and apply its allowances accordingly. Low-cost risk abaters would reduce risks further and sell allowances; high-cost risk abaters would buy allowances.

Risk bubbles and risk allowance trading would overcome the problem of bunching, hotspots, and hormetic perversity — and perhaps even unfairness — because they would measure results in risk rather than in emissions. Thus if bunching emissions at location A would increase risk, such as to a sensitive subpopulation, that would require additional risk allowances, and sources would have an incentive to reduce those risks. If reducing emissions below the hormetic nadir eM would increase risk, that would require additional risk allowances (not fewer emissions allowances). To be sure, a firm could purchase additional risk allowances, increasing the risk it poses. But as noted above, it is unlikely or inconceivable that a firm would purchase costly allowances to emit or cause risk at levels higher than it would have done before being regulated. Thus the creation of a system of risk allowances, even if initially set at the current level of risk in the aggregate (preventing increases in risk), is unlikely to yield significant increases at specific firms. And if the risk allowances are issued at a lower level than currently obtains, net reductions could be anticipated in most places, with buyers increasing above the standard but not above the pre-regulation level.

In addition, risk bubbles and trading represent a vehicle for integrated multimedia pollution control,[19] overcoming the problems of cross-media shifts of pollution and risk-risk tradeoffs across substances, activities and fragmented narrow regulatory programs that have plagued our risk regulation regime.[20] Provided all significant risks created by the firm are counted, the risk bubble or allowance system would remove any incentive to shift away from one regulated activity or substance or medium into another; the firm would be accountable for every unit of risk it creates, and would have incentives to reduce its overall risk in the most cost-effective ways possible. "[T]he facility is forced to internalize the environmental risks posed by its operations and take them into account when evaluating the regulatory feasibility and the economic profitability of changes in technologies and production. Any increase in one residual must be paid for with reductions of others."[21] It would also stimulate a wave of new innovation in risk-reduction opportunities previously ignored under the current fragmented regulatory system.

Risk allowance trading is easy to parody and pillory as selling the right to murder. But this epithet is misplaced. First, risk allowance trading would not authorize intentional assaults or any other forms of risk creation that are already illegal. Criminal law prohibitions would always supersede regulatory risk allowances. Second, risk allowance trading would be used to sharply reduce risks from existing levels, protecting society against ills not yet addressed, and at lower cost than fixed limits on risk. Third, it would solve the problem of exposure to more sensitive subpopulations and the perversities of emissions trading under hormesis, thereby protecting society even better than emissions limits, taxes and trading systems do. Fourth, it would reinforce and improve on the tort law system, which provides ex post financial liability remedies for risk, by adding an ex ante quantity limit on risk – a more protective and credible regime, set at levels determined by comprehensive analysis of societal risks, without the high transaction costs, variabilities and unpredictabilities of litigation.

At the same time, risk bubbles and risk allowance trading would have to satisfy important criteria. They demand information about risk levels imposed by different activities, and some means of making those risks commensurable across different substances, activities, populations, and health endpoints. Stewart notes: "Each such facility is given an individual risk quota; a maximum level of risk that it may impose on the environment as a result of the residuals that it generates. The foundation of the 'risk bubble' system is a set of risk indices assigned to each residual and measured in some common risk unit. A common risk unit may, for example, be fixed as a level of the environmental hazard created by a discharge of one pound of a reference pollutant, the effects of which on human health and the environment are well known. The relative risk measure of other regulated pollutants are determined and assigned in relation to the baseline pollutant. An individual facility is free to select the levels of specific residuals that it generates as long as the aggregate risk imposed on human health and the environment by all of the total discharge does not exceed the facility's maximum risk quota."[22]

This will not be easy. The creation of "a comprehensive system for indexing the relative environmental risks of different residuals ... presents a problem of serious scientific difficulty."[23] We would need better information on emissions, associated exposures, and associated risks for a host of substances and activities, including their synergistic or offsetting effects in concert. Mechanisms,

dose-response functions, exposures, and variability across populations would need to be better understood and measured. Many uncertainties in current science would need to be ameliorated. Diverse risks would need to be translated into some version of commensurability — some set of multi-attribute risk metrics that can enable transparent accounting and review. And improved monitoring, perhaps by third parties as well as by government, would be needed to ensure that firms' risks are accurately counted, reported and checked against their allowances.

Yet these challenges are not unique to the risk bubble/trading approach; they apply equally to the current regime of fragmented command-and-control and emission trading systems. Not indexing is tantamount to indexing arbitrarily. "The relative stringency of command requirements for control of different residuals discharged into different media represents a de facto index of relative risk and regulatory priority. The virtue of a risk bubble approach is that it makes the indexing issue explicit and focuses scientific research and policy work on improving such values and dealing explicitly and systematically with risk versus risk and benefit versus benefit tradeoffs."[24] Moreover, commensurability across risks is an endogenous function of our investment in measuring and understanding, rather than an inescapable extrinsic barrier.[25] The familiar complaint that someone is inappropriately "comparing apples and oranges" actually disproves itself: people regularly do compare apples and oranges when choosing what to buy at the grocery store; the statement that apples are different from oranges itself constructs the comparison on which a choice could be made (such as based on flavor, color, nutrient content, price, and other attributes); and comparing the difference between "apples and oranges" to the difference between the other two things that the speaker is arguing cannot be compared (such as two risks) is itself a comparison that proves its feasibility.[26] Indices of the relative impacts of different pollutants can be constructed to enable transparent comparisons across pollutants contributing to the same risk and, potentially, different risks contributing to overall harm.[27]

Likewise, inadequate monitoring impairs technology-based and emissions-based regimes as well as risk-based regimes. Who knows if the scrubber is turned on, or operating effectively? Investments in better monitoring are needed in all regulatory systems. The risk bubble/trading approach would direct incentives toward (private sector) investment in better monitoring of the activities and attributes that actually contribute to harm. Dan Esty argues that although risk bubbles "depend on a greater degree of information than is generally currently available ... Today, however, with advanced pollution detection and tracking equipment, the technical dimension of the risk bubble problem is becoming increasingly tractable."[28] Likewise, Donald Elliott and Gail Charnley stress the practicality of designing effective risk bubbles and trading systems, if investments are made in the science of monitoring, measurement and comparison.[29] Regulatory regimes can be designed to reward industry's investments in improved monitoring. All of these challenges pose the cost of new information against the value of new information. If health and environmental risks are important to our society, then investing in improved methods of monitoring, measurement and comparison can yield large net gains.

The legal regime for such a risk-based approach, especially risk allowance trading across firms, has yet to be designed. Project XL was unable to stimulate widespread risk bubbling within firms largely because of current legal requirements for specific controls that such risk bubbles would have had to overcome; even if EPA had agreed to waive those requirements (in return for superior risk reductions), citizens' groups could sue to enforce them. Thus firms were hesitant to agree to risk reductions in some areas when EPA could not promise relief from other rules that did little to reduce risk. New legislation would be needed to authorize full risk bubbles with intra-firm flexibility. And risk allowance trading across firms would require major new legislation, likely with special attention to whether risk allowance trading would prevent or exacerbate hotspots. A potentially greater obstacle is the motivation of interest groups and members of Congress to maintain bureaucratic control and claim credit for separate initiatives by keeping the environmental regulatory system fragmented into different laws and committees; integration may make social sense but not political sense. Gradual experience with risk bubbling could be gained through initiatives like Project XL, through practice under Clean Air sections 112(g) and 612 noted above, and through multimedia permitting. EPA could begin developing the transparent indices and monitoring systems. Industry could be encouraged to finance this work by offering enhanced abatement credit in return for better monitoring of emissions, exposure and risk.[30] In addition, there is a move afoot in the current Congress to pass new "three pollutant" air quality legislation, covering SOx, NOx, and mercury, with a possibility of adding CO2 as well. This legislation would authorize some interpollutant trading (e.g. SOx-NOx), which would build experience with the monitoring, risk assessment, and commensurability needed for risk allowance trading.

F. Conclusion

Hammitt is right to raise the possible problems posed for emissions trading or taxes by hormesis, but as this comment has illustrated, the problem is not as serious as might be thought. For emissions trading or taxes to be undermined by hormesis, the necessary conditions involve not just hormesis but also local emissions effects, a level of protection set at or near the hormetic minimum-effects level, and a pattern of selling and buying by sources along the dose-response curve in a direction that poses a net increase in harm. Fairness concerns about hotspots are distinct from hormetic effects. And to the extent that hormesis does pose

problems for emissions trading and taxes, an eventual shift toward risk trading and taxes is the superior route.

At the same time, several issues of interest remain unaddressed. One question is how a hormetic dose-response function and a more complex emissions-exposure function would interact. Another is what an empirical review of the actual performance of fixed emissions limits and tradable emissions allowances in the past, illuminated by evidence about the dose-response function, would reveal about whether hormesis and emissions trading have conflicted in practice. A third is whether hormesis would have implications for the choice between taxes (prices) versus allowance trading (quantities) — which it might, given the importance of differently sloped marginal benefit functions for this choice.[31] A fourth is whether these issues bear on different methods of allocating allowances to sources. Further analysis and empirical study of dose-response functions and policy instruments could help improve the selection of instruments to match the complexities of real environmental problems.

FOOTNOTE

* I thank James Hammitt for helpful comments on an earlier draft.

REFERENCES

1.On hormesis and regulation in general, see Edward J. Calabrese & Linda Baldwin, U-shaped Dose-responses in Biology, Toxicology, and Public Health, 22 Annual Review of Public Health 15-33 (2001); Frank B. Cross, Legal Implications of Hormesis, 20 Human and Experimental Toxicology 122-128 (2001) (and related comments in this symposium issue).

2. James K. Hammitt, Economic Implications of Hormesis, BELLE –12:1-36 (2003).

3. See Jonathan B. Wiener, Global Environmental Regulation: Instrument Choice in Legal Context, 108 Yale Law Journal 677-800 (1999).

4. Hammitt, BELLE (2003), cited above (p.18)

5. See Stephen M. Johnson, Economics v. Equity: Do Market-Based Environmental Reforms Exacerbate Environmental Injustice?, 56 Wash. & Lee L. Rev. 111, 125- 26 (1999).

6. Hammitt, BELLE (2003), cited above, Figures 1 & 2 (p. 23)

7. Hammitt, BELLE (2003), cited above (p.17)

8. See Richard B. Stewart and Jonathan B. Wiener, Reconstructing Climate Policy (Washington DC: AEI Press, 2003).

9. For a helpful analysis of these options and a creative proposal, see Jonathan Remy Nash & Richard L. Revesz, "Markets and Geography: Designing Marketable Permit Schemes to Control Local and Regional Pollutants," 28 Ecology L. Q. 596 (2001).

10. Hammitt, BELLE (2003), cited above (p.18)

11. See Marc J. Roberts and Michael Spence, "Effluent Charges and Licenses under Uncertainty," Journal of Public Economics, 5: 193-208 (1976).

12. See Paul R. Portney, "Reforming Environmental Regulation: Three Modest Proposals," Issues in Science and Technology (Winter 1988) pp.74-81, and "Reforming Environmental Regulation: Three Modest Proposals," Columbia Journal of Environmental Law 13: 201, 207-09 (1988).

13. Portney, Columbia J. Envtl. L., cited above, at 207.

14. Richard B. Stewart, "A New Generation of Environmental Regulation?" 29 Capital University Law Review 21, 96 (2001). See also id. at 116-122 (analyzing risk bubbles in detail).

15. See William F. Pedersen, "Contracting With The Regulated For Better Regulations," 53 Administrative Law Review 1067 (Fall 2001) (advocating government-industry contracts that apply risk bubbles to improve risk protection while lowering costs).

16. See Stewart (2001), cited above, at 64-66 (analyzing Project XL agreements that authorize risk bubbling); Daniel A. Farber, "Rethinking Regulatory Reform after American Trucking," 23 Pace Law Review 43, 70 (Winter 2002).

17. See Harvard Group on Risk Management Reform, "Reforming Risk Regulation: Achieving More Protection at Less Cost," 1 Human & Ecological Risk Assessment 183-206 (1995).

18. See Frank B. Cross, Incorporating Hormesis in Risk Regulation, 30 Envtl. L. Rep. 10778 (September 2000).

19. See Lakshman Guruswamy, "Integrated Pollution Control: The Expanding Matrix," 22 Environmental Law 77-90 (1992).

20. See John D. Graham and Jonathan B. Wiener, Risk vs. Risk: Tradeoffs in Protecting Health and the Environment (Harvard Univ. Press, 1995).

21. Stewart (2001), cited above, at 117.

22. Stewart (2001), cited above, at 117.

23. Stewart (2001), cited above, at 119.

24. Stewart (2001), cited above, at 119.

25. "For risks that seem very 'dissimilar,' risks cannot (yet) be measured on a unidimensional scale, and the exercise of informed value judgments becomes all the more central. ... But it is chiefly our lack of methods of comparison ... that makes these risks seem 'dissimilar' or noncomparable, not an inherent incommensurability. As we improve methods of risk analysis, the idea of calculating the 'net risk" of a risk portfolio, or the change in net risk due to a risk tradeoff, may become more meaningful." Graham & Wiener, Confronting Risk Tradeoffs, in Risk vs. Risk, supra, at 33.

26. See Hiram E. Chodosh, Comparing Comparisons: In

Search of Methodology, 84 IOWA L. REV. 1025, 1061-63 (1999) (arguing that "apples and oranges" are regularly compared and can be compared).

27. Indices to compare multiple pollutants contributing to a single risk have been successfully developed and used in the treaties addressing CFCs (ozone depletion potential or ODP) and GHGs (global warming potential or GWP). Indices to compare different risks with each other can draw on willingness to pay/accept approaches (using either revealed or stated preferences). See W. Kip Viscusi, Wes Magat & Joel Huber, "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Tradeoffs for Chronic Bronchitis," 21 J. Envtl. Econ. & Mgt. 32-51 (1991). Short of full monetization, such indices can be constructed using multiattribute decision theory, by combining different factors into an index through additive weighting. See Jonathan Baron, Thinking and Deciding 341-48 (2d ed. 1994); Ralph L. Keeney & Howard Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs (1976); John W. Payne et al., The Adaptive Decision Maker 24-25 (1993); Detlof von Winterfeldt & Ward Edwards, Decision Analysis and Behavioral Research 259-77 (1986).

28. Daniel C. Esty, "Next Generation Environmental Law: a Response to Richard Stewart," 29 Capital University Law Review 183, 196 (2001) (citations omitted).

29. See E. Donald Elliott & Gail Charnley, "Toward Bigger Bubbles: Why Interpollutant and Interrisk Trading Are Good Ideas and How We Get There from Here," Forum Applied Research & Public Policy 48 (Winter 1998). Similarly: "On the one hand, the more broadly the measures of equivalent compliance are defined, the greater the opportunities for flexibility, improved performance, and cost savings. Thus, a comprehensive risk bubble that would permit one kind of environmental risk to be traded off against another would potentially provide the greatest benefit. On the other hand, the difficulties of measuring "equivalency" become greater the more broadly the concept is defined. Thus, it is inherently easier to equate one pound of sulfur dioxide released in one part of a factory with an equivalent pound of the identical chemical released somewhere else in the same facility. However, as common metrics are developed for comparison purposes, dissimilar environmental risks will be able to be traded off more broadly against one another."

E. Donald Elliott, "Toward Ecological Law and Policy," in Thinking Ecologically: The Next Generation of Environmental Policy 184 (Marian R. Chertow & Daniel C. Esty, eds., 1997).

30. See Jonathan B. Wiener, "Solving the Precautionary Paradox: Policy Approaches to Improve Measurement of Greenhouse Gas Sources and Sinks," in J. van Ham et al., eds., Non-CO2 Greenhouse Gases (Dordrecht Netherlands: Kluwer Academic Publishers, 1994), pp. 527-531.

31. See Martin Weitzman, "Prices versus Quantities," 41 Rev. Econ. Stud. 477 (1974).

ECONOMIC IMPLICATIONS OF HORMESIS: RESPONSE TO COMMENTATORS

James K. Hammitt
Center for Risk Analysis
Department of Health Policy and Management and
Department of Environmental Health
Harvard School of Public Health
718 Huntington Ave., Boston, MA 02115
Phone: 617-432-4030
Fax: 617-432-0190
E-mail: jkh@harvard.edu

1. Introduction

I thank the expert commentators for clarifying some of the issues I addressed, extending the analysis of some of these issues, and introducing other important issues to the discussion. While the commentators addressed many issues, I will organize my response around two general themes: complexity of the regulatory approach and the use of economic-incentive instruments.

2. Optimal Complexity of Law and Regulation

As all the commentators note, regulation of substances with hormetic exposure-response functions requires more information than regulation of substances with linear no-threshold exposure-response functions. Information is required on the shape of the exposure-response function and on the levels of individual exposures, in order to determine whether reductions in exposure will be beneficial or harmful to health. Rather than simply comparing health risks with the costs of controlling exposure, hormesis requires regulators to balance potential adverse and beneficial health effects against one another as well as against control costs.

While Morgenstern notes that the recognition of beneficial and adverse health effects of environmental regulation is not new, Griffiths points out that environmental laws in the United States (as elsewhere) are often written in ways that are not supportive of maximizing the difference between benefits and costs. Many laws address only one risk at a time and direct regulators to limit exposures, implicitly assuming that smaller exposures will reduce harm. Such laws provide only a crude tool for promoting health and welfare but have the virtue of being relatively simple (in concept, if not always in application).

In contrast, Wiener discusses the possibility of using more comprehensive and nuanced regulatory approaches including tradable risk allowances. In principle, more comprehensive regulatory approaches can provide greater efficiency, in the sense of maximizing health benefits net of controls costs, but at the cost of greater complexity of design and potentially greater need for information.

Because actual and ideal exposure levels may differ among individuals, Smith and Evans note that approaches which encourage individual adaptation and control of exposure may be superior to government regulations that impose a common standard. When individual adaptation is reasonable, the conflicts in choosing a best single level described in Section 4 of my paper can be alleviated. But several factors limit the potential for individualized control of exposure to many substances of regulatory concern. First, it may be much more difficult or costly for individuals to reduce exposure to environmental pollutants widely dispersed through the environment than for the firms or other entities that release these pollutants to restrict their emissions. Second, individuals may be rationally ignorant about the health effects of many substances, as the benefit of learning about these substances and subsequently altering exposure to them may not exceed the cost of learning. Third, when relying on intuition rather than on more careful analysis, humans make systematic errors in evaluating and interpreting probabilities (Kahneman et al., 1982) and in other aspects of risk assessment and decision making (Margolis, 1996; Sunstein, 2002). For these reasons, it can be more efficient for individuals to delegate responsibility to government to investigate potential health effects and to adopt policies to regulate exposure to many environmental substances.

The best combination of government regulation and individual adaptation may vary across contexts, depending in part on individuals' ability to determine and alter their personal exposures. As noted by Smith and Evans, methods to provide information about exposure and its effects and to permit or encourage individualized responses may be superior to methods that attempt to achieve a common exposure level.

3. Economic-Incentive Instruments for Regulation

In Section 5 of my paper, I suggest that economic-incentive instruments (which are often thought to provide more health protection at lower cost than command-and-control instruments) may be less suitable when the substance to be controlled has a hormetic rather than a linear no-threshold exposure-response function. The intuition is that economic-incentive instruments do not provide the control over the spatial distribution of emissions that can be achieved by command-and-control instruments and, with a hormetic exposure-response function, the health effects of the regulation depend on the interaction of the induced change in exposure with individuals' initial exposure

levels. When emissions rather than exposures are regulated, environmental fate and transport processes influence the population distribution of exposure that results from any spatial pattern of emission reduction. I also note that more complicated instruments that differentiate "across subpopulations in proportion to the marginal benefits of reducing exposure could provide superior outcomes." Morgenstern also notes that "hot spots" may make simple pollution taxes and tradable permits less attractive, but suggests that hybrid approaches that combine economic incentives with command-and-control mechanisms may be more desirable.

Wiener extends my discussion by offering a comprehensive analysis of the effects of hormesis and local effects on the relative merits of economic-incentive and command-and-control instruments. He clarifies that the relative merits of simple economic-incentive and command-and-control instruments depend on the extent to which the distribution of exposure across the population is influenced by the spatial distribution of releases to the environment. If exposures are independent of the spatial distribution of releases, whether the exposure-response function is characterized by a hormetic or other shape is unimportant. It is only when the distribution of exposures depends on the spatial distribution of sources and the exposure-response function differs from the linear no-threshold case that economic incentives can reallocate total releases in a way that affects health adversely. In this case, the spatial reallocation can also affect health beneficially. Wiener notes that, even if the economic-incentive mechanism leads to adverse changes in the spatial distribution of emissions compared with an alternative command-and-control instrument, the additional emission reductions that can be achieved with economic-incentive instruments (because they offer a lower control cost) may more than offset the adverse spatial effect.

Wiener also provides an example of a more complicated instrument—risk-allowance trading—that incorporates the differential effects of emissions from alternative sources and thus overcomes the weaknesses of simple tradable permit schemes. A system of tradable risk allowances naturally requires more information about the health effects of emissions at alternative sources than does a simpler permit system, but no more than is required by an optimal command-and-control system. Like other economic-incentive instruments, a tradable risk-allowance system does not require the detailed information on source-specific control costs that are required under an optimal command-and-control system.

4. Conclusion

Hormesis is typically defined in terms of a single endpoint which responds in opposite directions to low and high exposures (e.g., Calabrese and Baldwin, 2002). As Griffiths notes, when one is concerned with human wellbeing, the restriction to a single endpoint is unimportant and many of the examples that are most prominent in policy discussion involve beneficial effects on some endpoints and adverse effects on others. Smith and Evans identify hormetic-like responses in three non-health contexts, including people's preferences for proximity to agricultural fields, for crowding at beaches and other recreational sites, and for the degree of challenge associated with rock-climbing, skiing, and whitewater-rafting sites.

The net benefits of regulatory stringency may also exhibit hormetic-like features. As illustrated by Figure 1 of my article, as exposures to an environmental pollutant are reduced from their uncontrolled level, the sum of the health damages and control costs typically decreases. With additional exposure reductions, the total harm begins to increase. Although the figure is drawn so that regulated exposures always produce less total harm than unregulated exposures, it is also possible for stringent regulations to be more harmful than no regulation, if the total cost of eliminating exposure exceeds the total harm from unregulated exposure.

Under a more general definition that includes beneficial and adverse effects on different endpoints, hormetic-like responses are ubiquitous, as evidenced by aphorisms like "too much of a good thing" and "moderation in all things." The existence of hormetic-like responses poses no fundamental challenge to economic analysis. Indeed, the economic perspective focuses on the need to balance competing effects and anticipates that good solutions are more likely to be found somewhere in the middle of the set of possible actions rather than at an extreme. As a result, the economic perspective is especially congenial to the existence of hormesis and hormetic-like effects.

REFERENCES

Calabrese EJ and Baldwin LA. 2002. Defining hormesis. *Biological Effects of Low Level Exposures* 10(2):25-30.

Kahneman D, Slovic P, and Tversky T. 1982. *Judgment under uncertainty: heuristics and biases*. Cambridge University Press, New York.

Margolis, H. 1996. *Dealing with risk: why the public and the experts disagree on environmental issues*. University of Chicago Press, Chicago.

Sunstein, CR. 2002. *Risk and reason: safety, law, and the environment*. Cambridge University Press, New York.