NICHOLAS ASHFORD

*Massachusetts Institute of Technology, Cambridge, Massachusetts*

# Compliance costs: the neglected issue

**Research in the USA over the last 20 years suggests there is strong evidence that the major benefits of technological innovation implemented in response to stringent health and safety standards are usually neglected in cost-benefit calculations.**

Regulation of worker health and safety is acknowledged to result in health benefits to workers and economic costs to employers. The latter are sometimes shared by workers and consumers in the form of lower wages or salary increases and higher prices. However, the history of OSH regulation in the United States over the last twenty years reveals that this simplified view of regulation neglects the important role that technological innovation plays in:

- reducing the actual costs of compliance with a new regulation to a fraction of pre-promulgation estimates
- yielding a benefit in terms of savings in material, water and energy costs, and
- changing the nature of process and product technology, resulting in reduced environmental damage and its associated costs and compliance burden.

The US Office of Technology Assessment (OTA) investigated the technology-forcing aspects of standards promulgated by the US Occupational Safety and Health Administration (OSHA) over the last twenty years and found that:

- technological innovation usually resulted from stringent regulation, and
- traditional CBA performed prior to a standard's implementation failed to anticipate significant economic benefits accruing to the innovating industrial firm.

## CHANGING ECONOMIC THEORIES

The reductionist version of neoclassical economic theory predicts that since health, safety and environmental regulations impose non-productive investment by industry on pollution control, regulation can only be a burden on innovation, and hence on economic growth, because of the diversion of resources away from R&D. A more modern view is the so-called Porter Hypothesis proposed in 1991, which argues that regulations may actually stimulate growth and competitiveness [1]. In fact that suggestion, and the empirical evidence that supports the hypothesis, goes back to a series of publications from researchers at MIT beginning twelve years earlier [2,3,4,5] although Porter does not seem to be conscious of it [6].

Porter bases his hypothesis on the notion that firms that first address environmental problems by developing technological responses to control pollution gain a market advantage by doing so – the so-called 'first mover' advantage [7]. One could describe the Porter Hypothesis as having a weak and a strong form. Porter himself actually discusses only the weak form which says that regulation, properly designed, can cause the (regulated) firm to undertake innovations that not only reduce pollution – which is a hallmark of production inefficiency – but also save on materials, water, and energy costs. This can occur because the firm, at any point in time, is suboptimal. If the firm is first to move by complying in a clever way, other firms will later have to rush to comply – and do so in a less thoughtful and more expensive way. Thus, there are 'learning curve' advantages to being first and early. What is missing from Porter's analysis are details about the process of innovation, how change actually occurs in industrial firms, what kinds of firms are likely to come up with what kinds of technical responses, and how very stringent regulation can confer competitive advantage beyond what he calls 'innovation offsets'. While Porter stresses the importance of going beyond the 'static model' of compliance responses, he is in fact talking mostly about modest or incremental innovation in reducing pollution.

> *Traditional methods over-estimate the costs of protecting workers and under-estimate the health and safety benefits achievable by adopting superior technologies* ”

The strong form of the Porter Hypothesis was first proposed by Ashford and his colleagues at MIT after years of cross-country and US-based studies that showed that stringent regulation could cause dramatic changes in technology, often by new firms or entrants displacing the dominant technologies. These changes go beyond traditional end-of-pipe pollution control, and result in more fundamental changes in the production process encompassing pollution prevention/cleaner production. The former results in modest reduction of pollution; the latter often brings about more significant, even radical, technological change, which is often accompanied by savings in energy, water and material resources, as well as reducing pollution at source.

The replacement of dominant technologies by new entrants, rather than incremental change by existing technology providers, has been the source of the most important radical innovations over this century. It stands to reason that any strong change in market conditions – be it sudden factor cost changes, new opportunities from new consumer or societal demands, an energy crisis, or demanding regulation – could stimulate significant innovation. Porter did not actually discuss the strong form of the hypothesis, which is by far the more important and interesting effect. Paradoxically, pollution prevention/cleaner production initiatives that do not also include concerns for worker health and safety can result in either media-shifting from environment to workplace, or problem-shifting from gradual expected pollution to sudden and accidental releases of chemicals, sometimes resulting in catastrophic accidents[8].

In contrast to the situation in the US, regulation-induced radical technological change in the environmental area has not been observed as much in Europe [9] where a more cooperative style of government-industry interaction often prevails. This cooperative style has tended to reduce the stringency of European regulations [10]. In the US, there is ample evidence that regulation – if properly designed and implemented – can prompt the kind of fundamental technological change that can significantly reduce human and environmental exposure to toxic substances in the workplace, in the environment and in consumer products [11]. Prior work at the Massachusetts Institute of Technology developed models to explain the effects of regulation on fundamental technological change, based on empirical evidence over two decades [2,3,4,5]. The particulars of this model – the nature of the regulatory stimulus, the characteristics of the responding industrial sectors, and the resulting implications of the model for explaining technological responses to regulation and for designing innovative regulatory strategies – are discussed elsewhere[12].

## COST-BENEFIT ANALYSIS EXAMINED

During the past two decades, cost-benefit has become the dominant method used by many policy-makers to evaluate government intervention in the areas of health, safety and environment. As conceived in theory, cost-benefit analysis:

- enumerates all possible consequences, both positive and negative, that might arise in response to the implementation of a candidate government policy
- estimates the probability of each consequence occurring
- estimates the benefit or loss to society should each occur, expressed in monetary terms
- computes the expected social benefit or loss from each possible consequence by multiplying the amount of the associated benefit or loss by its probability of occurrence; and
- computes the net expected social benefit or loss associated with the government policy by summing over the various possible consequences. The reference point for these calculations is the state of the economy in the absence of government policy, termed the 'baseline'.

The mechanics of constructing a cost-benefit analysis can be seen with reference to Table 1, which presents a relatively disaggregated matrix of the various positive and negative consequences of a government policy – such as regulation – for a variety of actors. The consequences are separated into economic, health and safety, and environmental effects, and those affected are organized into policy-relevant groups of actors, such as firms, workers, consumers and 'others'. Initially, the consequences are represented in their natural units:

- economic effects are expressed in monetary units
- health and safety effects are expressed in mortality and morbidity terms; and
- environmental effects are expressed in damage to ecosystems, etc.

All of the consequences of a candidate policy (or regulation) are described fully in terms of the times during which they occur. What traditional CBA does is translate all of these consequences into 'equivalent' monetary units (since a dollar in an earlier time period could be invested to earn interest over time) by discounting each to present value and aggregating them into a single dollar value intended to express the net social effect of the government policy.

**Table 1: Matrix of policy consequences for different actors**

| Group | Economic effects | Health/safety effects | Environmental effects |
|---|---|---|---|
| Producers | $C_\$$ | | |
| Workers | $C_\$$ | $B_{H/S}$ | |
| Consumers | $C_\$$ | $B_{H/S}$ | |
| Others | $C_\$$ | $B_{H/S}$ | $B_{Environment}$ |

The cost-benefit calculation can be expressed in simple mathematical terms by the following equation:

$$V = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (B_{ij} - C_{ij})}{(1 + r)^{i}}$$

where $B_{ij}$ and $C_{ij}$ are the $j^{th}$ type of policy benefit and cost, respectively, in the $i^{th}$ year after the policy is introduced, and B and C are expressed in monetary units; r is the appropriate discount rate; and V is the (discounted) present value of the policy.

Elsewhere the author has argued that health, safety and environmental benefits should be treated differently to costs in computing their present value [13]. One approach would allow for discounting of non-monetisable benefits, but at a lower discount rate. This approach can be defended in terms of a belief that certain amenities, such as health, become more valuable relative to other goods in this society as time passes and the standard of living improves. The following relationship would separate the factors affecting the present value of health impairment prevented in year n:

$$B_n = \frac{B(1 + Œ)^{n}}{(1 + r)^{n}}$$

where:
B = metric, expressed in person-years of health impairment prevented in any one year,
Œ = the subsequent annual fractional increase in societal value of health impairment prevented
and r = annual discount rate.

For small values of r and Œ this is equivalent to:

$$\frac{B}{(1 + r - Œ)^{n}}$$

Thus the 'effective discount rate' (r - Œ), or time rate of preference, will be less than the discount rate used for monetary benefit or cost calculations. Note that, in principle, if the society's valuation of health benefits increases rapidly over time, the effective discount rate for benefits could even be negative! Thus, instead of the traditional cost-benefit approach which is biased against interventions requiring cost expenditures early and yielding benefits later (such as is the case with chronic disease), this treatment makes long-term investments in health much more attractive.

The increasing concern with global environmental effects that occur far into the future, such as climate change, are a striking example of a societal judgement that the future may count more than the present. Here, the application of traditional positive discount rates are not likely to justify precautionary actions needed to prevent future catastrophic or otherwise irreversible harm[14].

When there is only one policy option, cost-benefit analysis dictates that option should be implemented only if its anticipated net social effect is positive. In general, however, numerous policies or sets of policies are possible. Each policy can be differentiated according to the various features that it comprises – type of policy instrument, policy level or stringency, firms covered, etc. In this situation, according to the cost-benefit criterion, the policy with the largest expected net social benefit, when compared to the baseline, should be implemented.

As a decision-making tool, CBA offers several compelling advantages:

- First, it clarifies choices among alternatives by evaluating consequences in a systematic and rational manner
- Second, it professes to foster an open and fair policy-making process by making explicit the estimates of costs and benefits and the assumptions on which those estimates are based
- Third, by expressing all of the gains and losses in monetary terms, discounted to their present value, it permits the total impact of a policy to be summarised using a common metric and represented by a single dollar amount.

As a practical matter, however, CBA possesses several serious limitations. The ensuing dissection of CBA is not intended to suggest a wholesale rejection of the technique, but to caution against the uncritical application of an imperfect methodology and the unqualified acceptance of its results[15].

## PROBLEMS IN ESTIMATING PUBLIC POLICY BENEFITS

The benefits of a specific government policy concerning occupational health and safety are generally the reduced social costs associated with a decrease in the number (or severity) of job-related injuries and illnesses, where the decrease is brought about by the policy in question. Prominent examples of policy benefits include reductions in medical expenses, productivity losses, physical disability, pain and suffering, and loss of life. Estimation of the policy benefits in CBA is a formidable task because it is difficult to predict the reduced risk of injury and disease and to monetise the associated benefits.

There are many problems in trying to determine the effects of a government policy on the incidence of job-related injuries and disease. The baseline occupational risks may not be scientifically established. In most cases, the precise relationship between exposure and disease is simply not known. Estimating the effects of the policy on worker exposure levels may also be rather uncertain, depending as it does on assumptions about company and worker behaviour, as well as on technical production relationships.

Additionally, many of the benefits of government policy, such as reductions in physical disability, pain and suffering and loss of life, have no clearly defined economic value (as compared to the market prices established for labour and medical services). The traditional methods of monetising these benefits – surveys and market studies – have to a large extent been unsuccessful. Interviews and questionnaires asking individuals what they would be willing to pay for a stated reduction in risk have inherent limitations, since answers to hypothetical questions have been shown to be poor indicators of a person's behavior. Imputing the value of risk reduction from an individual's market behavior is also a seriously flawed approach[16]. Individual actions are normally undertaken for a variety of reasons, and it is difficult to isolate what portion is motivated by a desire to reduce the risk of bodily impairment, pain and suffering, or a premature death. Furthermore, consumers are rarely well informed about the risks confronting them and have a well-documented history of being unable to process the risk information at their disposal in an expected manner[17,18,19]. As a result, the assumption of economic efficiency underlying attempts to value risks from consumer market decisions is untenable in practice.

The job market itself is the place where policy analysts have most frequently turned to derive the value of a reduction in risk. Recall that, according to economic theory, the risk-compensating wage premium (hazard pay) represents the workers' valuation of job risk. However, the same job market imperfections that produce a socially-excessive level of workplace risk and create a need for government intervention also undermine the usefulness of the risk premium as a measure of the worker's risk valuation. For example, job-related diseases that the worker does not know about will not be reflected in the wage premium for risk. Moreover, workers may have difficulty in understanding risk information. In theory, they are just as likely to over-react as under-react to hazard information. In practice, worker risk perception appears to be dominated by an "it-can't-happen-to-me" attitude[20]. This results in known risks being understated and therefore undervalued. Another job market defect, externalities, causes the observed wage premium for risk to measure only the *worker's* valuation of an incremental risk, but not the value family members, friends and other interested parties attach to the risk. Furthermore, models of the risk-compensating wage differential assume a perfectly-competitive job market; violation of this assumption means that the resulting estimates will 'misinterpret' the true wage premium for risk. This is a particularly serious problem since there may be no way to adjust the estimates to correct for the mis-specification.

## PROBLEMS IN ESTIMATING PUBLIC POLICY (INTERVENTION) COSTS

Although the costs imposed by a government policy seem rather easy to identify and to express in economic terms, they are usually no more certain or reliable than the benefits. One reason is that policy analysts rarely have access to detailed, independent information about actual, and potential, production relationships and associated costs in an industry. Instead, they must depend to a large extent on industry-provided data to develop estimates of the costs to industry of complying with the public policy. Since higher compliance costs make a policy less attractive, industries adversely affected by the policy may choose to inflate their reported compliance costs.

In addition, compliance cost estimates often fail to take three significant factors into account:

- economies of scale, which reflect the fact that an increase in the production of compliance technology often reduces unit costs
- the ability of industry to learn over time to comply more cost-effectively, i.e. what management scientists refer to as the learning curve, and
- compliance costs based on present technological capabilities ignore the role played by technological innovation in reducing those costs[21].

The last factor is particularly crucial. A recent retrospective analysis of eight OSHA regulations issued between 1974 and 1989 by the OTA concluded that the agency's estimates of economic impacts systematically and significantly over-estimated compliance costs by ignoring the innovative response of industry to the enacted standards[22]. Five of these regulations addressed toxic substances and are discussed in detail below.

## THE OTA STUDY

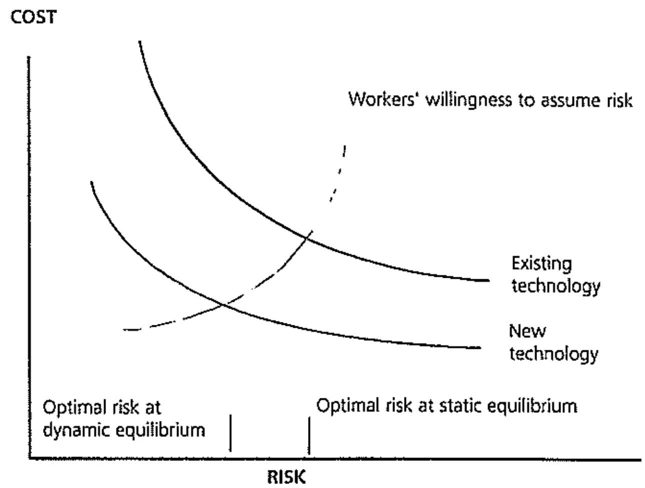
The results of the OTA study are summarized in Table 2 for the five health standards that were investigated. The study concluded that:

*"OSHA's current economic and technological feasibility analyses devote little attention to the potential of advanced or emerging technologies to yield technically and economically superior methods for achieving reductions in workplace hazards... Opportunities are missed to harness leading-edge or innovative production technologies (including input substitution, process redesign, or product reformulation) to society's collective advantage, and to achieve greater worker protection with technologically and economically superior means.*

*"Intelligently-directed effort can yield hazard control options – attributes that would, no doubt, enhance (for regulated industries and their workforces) the 'win-win' character of OSHA's compliance requirements in many cases and support the achievement of greater hazard reduction."*

There is thus overwhelming and convincing evidence that failure to include technological innovation in assessing the costs and benefits of workplace regulation renders cost-benefit analysis of minimum use in efforts to protect workers from occupational health hazards. Estimates using traditional approaches over-estimate the costs of protecting workers and under-estimate the health benefits achievable by developing or adopting superior technologies.

Thus, reliance on traditional cost-benefit analysis leads to a suboptimal level of regulation and hence a suboptimal level of occupational health and safety[23]. This of course, is one reason for the enthusiasm in some quarters for using CBA to determine the acceptability of health and safety regulation.




Optimal risks at static and dynamic equilibrium.

### Table 2: OTA case histories

- **Vinyl chloride**
  (1) significant process (polymerization) innovation
  (2) final compliance costs were one quarter of pre-promulgation best estimates
- **Cotton dust**
  (1) aggressive re-tooling of entire production process yielding significant productivity improvements
  (2) final compliance costs were one third of pre-promulgation best estimates
- **Lead (secondary smelters)**
  (1) lack of enforcement, plus exercise of the option of paid medical removal of workers, yielded a small fraction of predicted compliance costs
- **Ethylene oxide**
  (1) because of fear of legal liability, the hospital industry chose to replace existing sterilisation equipment with innovative technology, thereby reducing the ambient level well below the required standard
  (2) engineering costs were, however, about the same as predicted, although new technology was adopted
- **Formaldehyde (metal foundries)**
  (1) significant innovation by formaldehyde resin suppliers
  (2) costs were a half of pre-promulgation best estimates

Source: *Gauging Control Technology and Regulatory Impacts in Occupational Safety and Health: An Appraisal of OSHA's Analytic Approach.* Washington, DC: U.S. Congress, Office of Technology Assessment (OTA-ENV-635); September 1995.

REFERENCES:

1 Porter, Michael E., "America's Green Strategy," *Scientific American*, April 1991, p 168.

2 Ashford, N. A., Heaton, G. R., and Priest, W. C., "Environmental, Health and Safety Regulations and Technological Innovation," in *Technological Innovation for a Dynamic Economy*, C. T. Hill and J. M. Utterback (eds.), Pergamon Press, Inc., NY, 1979, pp. 161-221.

3 Ashford, N.A. and Heaton, G. R., "Regulation and Technological Innovation in the Chemical Industry," *Law and Contemporary Problems*, Duke University School of Law, Volume 46, Number 3, Summer 1983, pp. 109-157.

4 Ashford, N. A., Ayers, C., and Stone, R. F., "Using Regulation to Change the Market for Innovation," *Harvard Environmental Law Review*, Volume 9, Number 2, Summer 1985, pp. 419-466.

5 Ashford, N. A. "Understanding Technological Responses of Industrial Firms to Environmental Problems: Implications for Government Policy," in *Environmental Strategies for Industry: International Perspectives on Research Needs and Policy Implications*, K. Fischer and J. Schot (eds.), Island Press, Washington, DC, 1993, pp 277-307.

6 Jaffe, A., Peterson, S., Portney, P., and Stavins, R., "Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us?" *Journal of Economic Literature*, Vol. 33, March 1995, 132-163 (page 154).

7 Nehrt, Chad, "Maintainability of First Mover Advantages When Environmental Regulations Differ Between Countries" in *Academy of Management Review*, 1998, 23(1):77-97.

8 Ashford, N.A., "Industrial Safety: The Neglected Issue in Industrial Ecology" in the Special Issue on Industrial Ecology, Ashford, N. A. and Côté, R. P. (Eds), *Journal of Cleaner Production*, 1997. 5(_), pp 115-121.

9 Kemp, René "Environmental Regulation and Innovation: Key Issues and Questions for Research" in *The Impact of EU Regulation on Innovation of European Industry*, Hemmelskamp, J. and Leone, F. (Eds.), Joint Research Center, the European Union, Seville, 1998, pp 12 et seq.

10 Gouldson, Andrew and Murphy, Joseph, *Regulatory Realities: The Implementation and Impact of Industrial Environmental Regulation*, Earthscan Publications, London 1998.

11 Strasser, Kurt, "Cleaner Technology, Pollution Prevention, and Environmental Regulation" in *Fordham Environmental Law Journal* 1997, 9(1).

12 See notes 3-5. See also Ashford, N. A. "An Innovation-Based Strategy for the Environment," in *Worst Things First? The Debate Over Risk-based National Environmental Priorities*, A. M. Finkel and D. Golding (eds.), Resources for the Future, Washington, DC, 1994, pp. 275-314.

13 Ashford, N. A. "Alternatives to Cost-Benefit Analysis in Regulatory Decisions," *Annals of the New York Academy of Sciences*, Volume 363, April 30, 1981, pp. 129-137. See also Ashford, N. A. and Caldart, C. C. "Economic Issues in Occupational Health and Safety", Chapter 5 in Technology, Law and the Working Environment, Revised Edition, Island Press, 1996, 641 pages.

14 Ashford , N. A. , "A Conceptual Framework For The Use of the Precautionary Principle in Environmental Law," in Raffensperger, C. and Tickner, J. (Eds.) *Protecting Public Health and the Environment: Implementing the Precautionary Principle*, Island Press, 1999.

15 An alternative approach to cost-benefit analysis, trade-off analysis, has been suggested. In this approach, like cost-benefit analysis, the matrix depicted in Table 1 is a starting point for the analysis. However, in trade-off analysis, the matrix elements are expressed in their natural units, such as monetary costs, mortality and morbidity statistics, etc. No attempt is made to place a monetary value on health, safety and environmental effects. The time period in which each effect is experienced is fully described, but the health, safety, and environmental effects are not necessarily discounted. Trade-offs between worker health and costs to producers, consumers and others are made in a transparent manner by the politically-accountable decision-maker. Thus *accountability*, rather than accounting, is fostered. See Ashford, N. A. and Caldart, C. C. "Economic Issues in Occupational Health and Safety", Chapter 5 in *Technology, Law and the Working Environment*, Revised Edition, Island Press, 1996, 641 pages.

16 Fischer, G. W., "Willingness to pay for probabilistic improvements in functional health status: A psychological perspective." In: Mushkin, S. J and Dunlop, D. W., eds. *Health: What Is It Worth? Measures of Health Benefits*. New York: Pergamon Press; 1979.

17 Tversky. A. and Kahneman, D., "Judgment under uncertainty: Heuristics and biases." *Science*. 1974;185:458-468.

18 Fischhoff, B. "Cost benefit analysis and the art of motorcycle maintenance". *Policy Sciences*. 1977;8:177-202.

19 Machina, M. J., "Choice under uncertainty: Problems solved and unsolved." *Journal of Economic Perspectives*. Summer 1987;1:121-154.

20 Ashford, N. A., *Crisis in the Workplace: Occupational Disease and Injury*. Cambridge, Massachusetts: MIT Press; 1976, 579 pages.

21 See note 4.

22 *Gauging Control Technology and Regulatory Impacts in Occupational Safety and Health An Appraisal of OSHA's Analytic Approach*. Washington, DC: U.S. Congress, Office of Technology Assessment (OTA-ENV-635); September 1995.

23 Full application of cost-benefit analysis dictates that solutions to worker health and safety problems be "economically efficient", i.e., that intervention is justified until the marginal benefits of further protection no longer equal marginal costs. A slightly less stringent condition – often expressed in guidelines from the US Office of Management and Budget – is that the total benefits of regulation at least exceed the costs. Using the costs of *existing* technology in this calculus leads to static efficiency. Instead, if the costs of *new* technology were factored into the calculus, a different, *dynamic* efficiency would be achieved, with a greater level of worker protection at lower cost, leading to a "win-win" situation (see Ashford and Caldart, note 15). See Figure 1 for a graphical representation of dynamic versus static efficiency.