Josh Cohen <cohenj@hsph.harvard.edu>
05 28'2002 04:47:46 PM

Record Type: Record

To: John F. Morrall III/OMB/EOP@EOP

cc:

Subject: Comment on Fed Register notice of March 28, 2002...

Dear Mr. Morrall,

Attached, please find a letter commenting on OIRA's Federal Register notice posted on March 28, 2002 (vol. 67, no. 60, pp. 15013 to 15045).

Sincerely,

Joshua Cohen

 - Weinstein1996.pdf

 - Evans1994.pdf

 - Russell1996.pdf

 - Siegel1996.pdf

 - Thompson1997.pdf

 - 2002 05 24 Comment to OIRA.doc

---

****

Josh Cohen, Ph.D.
Senior Research Associate
Harvard Center for Risk Analysis
718 Huntington Avenue
Boston, MA 02115

Tel: 617-432-0394
Fax: 617-432-0190
email: cohenj@hsph.harvard.edu

Harvard Center for Risk Analysis
718 Huntington Avenue
Boston, MA 02115

May 24,2002

John Morrall
Office of Information and Regulatory Affairs
Office of Management and Budget
NEOB, Room 10235
725 17" Street, NW
Washington, DC 20503

Dear Mr. Morrall,

This letter expresses our support for some of the methodological approaches outlined in the report published by your office in the Federal Register on March 28,2002 (Vol. 67, No. 60, pp. 15014 to 15045) and suggests additional methodologies that we believe OIRA should consider in its efforts to evaluate federal regulations.

In general, we believe that economic assessments of regulations should use approaches, such as benefit-cost analysis and cost-effectiveness analysis, that account for both anticipated benefits and costs. The technical issues associated with these methodologies have been extensively studied. A good example of a set of standard practices was developed by the Panel on Cost-Effectiveness in Health and Medicine convened by the U.S. Public Health Service in the 1990s. Although the recommendations were developed in the context of medical decision making, they are generally applicable to other policy contexts where health risks are a key outcome. Key features of the Panel's recommendations include the use of quality-adjusted life years (QALYs) as a measure of health effects in cost-effectiveness analyses, and discounting of future health and economic consequences. QALYs are similar to the measure known as disability-adjusted life years (DALYs), but they differ in that, unlike DALYs, they give equal weight to health improvements at all ages. We have included three articles summarize the Panel's recommendations (Russell *et al.*, 1996; Weinstein *et al.*, 1996; Siegel *et al.*, 1996), as well as an editorial from a recent issue of the Journal of the American Medical Association that supports the use of cost-per-QALY analyses to guide policy decisions (Mark, 2002).

A part of considering both costs and benefits is the valid treatment of the uncertainty inherent in estimating the probability and magnitude of various outcomes. The field of Decision Science has demonstrated that rational decision making depends on fully characterizing the range of plausible values for uncertain quantities rather than, for example, depending exclusively on "worst case" estimates. The enclosed paper by Evans *et al.* (1994) describes how such a distribution can be developed when so-called "objective" information is not available.

Finally, we urge OIRA to apply economic evaluation techniques in assessing which information should be gathered to improve decision making. The enclosed paper by Thompson and Evans (1997) describes how "value of information" techniques can be used to identify which information can be expected to best improve the ability of policy makers to identify economically efficient policies. Value of information techniques can also help to distinguish between situations in which it is best to gather additional information before choosing how to act, and those

situations in which the cost of additional information – in terms of either the time needed to gather it or its financial cost – is too great to justify putting off action.

Sincerely,

Joshua T. Cohen, Ph.D.
Senior Research Associate
Harvard Center for **Risk** Analysis
Harvard School of Public Health

John S. Evans, S.D.
Senior Lecturer on Environmental Science
Department of Environmental Health
Harvard School of Public Health

James K. Hammitt, Ph.D.
Associate Professor of Economics and Decision Sciences
Department of Health Policy and Management
Harvard School of Public Health

Karen M. Kuntz, Sc.D.
Associate Professor of Health Decision Science
Departments of Health Policy and Management and Biostatistics
Harvard School of Public Health

Peter Neumann, Sc.D.
Assistant Professor of Policy and Decision Sciences
Department of Health Policy and Management
Harvard School of Public Health

Milton C. Weinstein, Ph.D.
Henry J. Kaiser Professor of Health Policy and Management
Departments of Health Policy and Management and Biostatistics
Harvard School of Public Health

List of enclosed/attached papers:

Evans, **J.S.,** Graham, J.D., Gray, G.M., and Sielken, R.L. 1994. **A** distributional approach to characterizing low-dose cancer risk. ***Risk Analysis.*** 14(1):25-34.

Russell, L.B., Gold, M.R., Siegel, J.E., Daniels, N., and Weinstein, M.C. 1996. The role of cost-effectiveness analysis in health and medicine. *JAMA*. 276(14):1172-1177.

Siegel, J.E., Weinstein, M.C., Russell, L.B., and Gold, M.R. 1996. Recommendations for reporting cost-effectiveness analyses. *JAMA*. 276(16):1339-1341.

Thompson, K.M., and Evans, **J.S.** 1997. The value of improved national exposure information for perchloroethylene(Perc): **A** case study for dry cleaners. Risk Analysis. 17(2):253-256.

Weinstein, M.C., Siegel, J.E., Gold, M.R., Kamlet, M.S., and Russell, L.B. 1996. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. ***JAMA.*** 276(15): 1253-1258.

Reprinted from JAMA © The Journal of the American Medical Association October 16, 1996 Volume 276

Consensus Statement

Recommendations of the Panel on Cost-Effectiveness in Health and Medicine

Milton C. Weinstein, PhD; Joanna E. Siegel, ScD; Marthe R. Gold, MD, MPH; Mark S. Kamlet, PhD; Louise B. Russell, PhD; for the Panel on Cost-Effectiveness in Health and Medicine

Objective.—To develop consensus-based recommendations for the conduct of cost-effectiveness analysis (CEA). This article, the second in a 3-part series, describes the basis for recommendations constituting the reference case analysis, the set of practices developed to guide CEAs that inform societal resource allocation decisions, and the content of these recommendations.

Participants.—The Panel on Cost-Effectiveness in Health and Medicine, a nonfederal panel with expertise in CEA, clinical medicine, ethics, and health outcomes measurement, was convened by the US Public Health Service (PHS).

Evidence.—The panel reviewed the theoretical foundations of CEA, current practices, and alternative methods used in analyses. Recommendations were developed on the basis of theory where possible, but tempered by ethical and pragmatic considerations, as well as the needs of users.

Consensus Process.—The panel developed recommendations through 2½ years of discussions. Comments on preliminary drafts prepared by panel working groups were solicited from federal government methodologists, health agency officials, and academic methodologists.

Conclusions.—The panel's methodological recommendations address (1) components belonging in the numerator and denominator of a cost-effectiveness (C/E) ratio; (2) measuring resource use in the numerator of a C/E ratio; (3) valuing health consequences in the denominator of a C/E ratio; (4) estimating effectiveness of interventions; (5) incorporating time preference and discounting; and (6) handling uncertainty. Recommendations are subject to the "rule of reason," balancing the burden engendered by a practice with its importance to a study. If researchers follow a standard set of methods in CEA, the quality and comparability of studies, and their ultimate utility, can be much improved.

JAMA. 1996;276:1253-1258

COST-EFFECTIVENESS analysis (CEA) has emerged as a basic tool in the evaluation of health care practices. Despite widespread application, there remain disparities in the methods that investigators employ.[1] Some of these disparities can be traced to a misunderstanding of the principles of CEA, while others reflect divergent views on key methodological choices. For example, an investigator who fails to account for important negative side effects of a therapy in estimating effectiveness is making a clear error, while investigators who include or exclude the financial costs of lost productivity that accompany illness are reflecting different views of how productivity should be accounted for in a CEA.

The divergence of methods used to conduct CEA interferes with the ability of decision makers charged with resource allocation to make appropriate comparisons of cost-effectiveness (C/E) ratios across programs. As described in the first article of this series,[2] this concern about lack of standardization has led the Panel on Cost-Effectiveness in Health and Medicine to develop a set of recommendations for the practice of CEA that can serve as a point of reference for investigators who seek comparability with other analyses in the literature. The panel refers to this set of methodological practices as the reference case.

The reference case will not address all types of questions regarding the cost-effectiveness of interventions. In some cases, depending on the goals of the analysis, the author may prefer to highlight an analysis based on a slightly different set of principles, or one based on quite different assumptions. In the interest of comparability, however, we urge that the reference case set of assumptions and practices be included in every CEA that is designed to permit broad comparisons across interventions or that might be used for this purpose.

The recommendations outlined here, together with others that provide more detailed methodological guidance, are expanded in the full report of the panel.[3] While this article focuses on the reference case recommendations, we also describe a few recommendations that are intended to improve the conduct of analyses but that are not explicitly incorporated within the reference case.

RATIONALES FOR RECOMMENDATIONS

Reference case analyses are intended to inform resource allocation decisions and, as described in the first article of this series, are conducted from the societal perspective for this reason.[2] Specific rec-

From the Departments of Health Policy and Management (Dr Weinstein) and Maternal and Child Health (Dr Siegel), Harvard School of Public Health, Boston, Mass; Office of Disease Prevention and Health Promotion, US Public Health Service, Washington, DC (Dr Gold); Heinz School, Carnegie Mellon University, Pittsburgh, Pa (Dr Kamlet); and the Institute for Health, Health Care Policy, and Aging Research, Rutgers University, New Brunswick, NJ (Dr Russell).

A complete list of the Panel on Cost-Effectiveness in Health and Medicine membership and staff appears at the end of this article.

Corresponding author: Marthe Gold, MD, MPH, Office of Disease Prevention and Health Promotion, 200 Independence Ave SW, Room 738G, Washington, DC 20201.

ommendations for conducting CEAs from the societal perspective were based on a number of considerations, including economic and decision theory; consistency in the accounting of costs and consequences; ethical concerns; pragmatic concerns; and needs of users of analyses. In some instances, where neither theory nor these other considerations led to a clear choice, the panel recommended a conventional practice for the sake of consistency across studies.

Where possible, recommendations were based on theoretical considerations in order to provide a defensible, logical, and consistent framework for methodological choices. At one level, CEA can be based solely on the mathematical theory of "optimization."[4] In that formulation, the decision maker is free to select any objective to maximize (eg, life-years or quality-adjusted life-years [QALYs]) and to specify the particular resource constraint under which allocations must be made (eg, national health care costs). However, this framework alone provides no guidance on key issues that arise from a societal perspective, such as which costs to include in the analysis, how to measure costs, and whose values to incorporate into the definition of health consequences. Therefore, drawing from recent literature suggesting a link between CEA and welfare economic theory,[5] the panel relied on economic theory for many of the recommendations for the reference case. In addition to economic theory, principles of decision theory were invoked to define the basis for individual preferences.

Some recommendations were dictated by the need to maintain a logical accounting of costs and health effects. For example, theoretical considerations, along with a basic presumption about the definition of the C/E ratio,[2] led to a clear decision concerning the placement of all health effects in the denominator of the C/E ratio. Then, the "accounting" principle that no cost or effect should be counted twice disallowed the inclusion of health effects—even in monetary form—in the numerator of the C/E ratio.

The implications of welfare economics were often modified in the interest of producing recommendations that were both pragmatic and ethically acceptable. The practical need to obtain data on health outcomes, utilization of services and unit costs, and weights for health-related quality-of-life (HRQL) states led to such compromises. For example, while medical prices are not an exact reflection of the true value of resources, pragmatism suggests that prices be used to approximate costs except where distortions are likely to be significant and important to the analysis.

Ethical considerations sometimes tempered recommendations based on economic theory or were decisive in choices among alternatives. Most fundamentally, the decision to use QALYs as the effectiveness measure reflects the ethical stance that QALYs accruing to different people or at different stages in life should be valued equally, even though welfare economics implies that health benefits should be weighted by willingness to pay.

Where theory did not offer a clear choice, the panel based some recommendations simply on the need for a clear convention to which analysts would adhere in the reference case. In some cases, the recommendation was somewhat arbitrary. For example, the choice of a standard time discount rate, while guided by theory and data, is fixed by the need for a standard practice.

Finally, needs of the potential users of CEAs influenced several recommendations, playing a particularly great role in the panel's recommendations for the reporting of CEAs as described in the third article of this series.[6] They also entered into the recommendations regarding the evidence of effectiveness used in analyses and the treatment of uncertainty. For example, sensitivity analyses, which explore the implications of alternative assumptions and data, are often recommended so that decision makers can gain confidence in the conclusions of an analysis or identify areas for further investigation.

While some of the reference case recommendations address common errors in the practice of CEA, many more represent the panel's view of the best among several defensible choices. The task of developing CEA standards is analogous to the formulation of the consumer price index (CPI), which is used to adjust for inflation based on the prices of a typical market basket of goods and services. The choice of what items go into that market basket, and how they are weighted, reflects judgments made by the Bureau of Labor Statistics and its advisors. Legitimate opposing views exist.[7] However, there is an implied consensus that the CPI will be used so that industry, government, and consumers can have a shared understanding of the inflation rate. Indeed, flows of resources, such as the level of Social Security payments, depend in part on the CPI. While this country does not base policy directly on C/E ratios as it does with the CPI, it is important for many decision makers to be able to rely on a dependable yardstick for measuring cost-effectiveness of health services.

RECOMMENDATIONS

The panel's recommendations fall largely into 8 categories: (1) the nature and limits of CEA and of the reference case; (2) components belonging in the numerator and the denominator of a C/E ratio; (3) measuring terms in the numerator of a C/E ratio (costs); (4) valuing the health consequences in the denominator of a C/E ratio; (5) estimating effectiveness of interventions; (6) time preference and discounting; (7) handling uncertainty in CEA; and (8) reporting guidelines. The first group of recommendations, regarding the nature and limits of CEA, has been described in the first article of this series.[2] The last group, regarding reporting guidelines, is the subject of the third article.[6] Additional recommendations regarding research to develop improved data for CEA and improved methods are described in the full report of the panel.[3] This article summarizes the remaining 6 categories of recommendations.

Components Belonging in the Numerator and the Denominator of a C/E Ratio

Cost-effectiveness analysis rests on the proposition that a decision maker wishes to select programs so as to maximize some desired objective subject to a resource constraint. In practice, CEA in health care has been based on the premise that health benefits are the objective that societal decision makers wish to maximize, subject to a constraint on health care resources. This formulation leads directly to the construction of a C/E ratio in which the net expenditure of health care resources (a monetary measure) goes in the numerator and the net improvement in health (a nonmonetary measure) goes in the denominator.

Unfortunately, however, this definition is incomplete. It leaves open to question whether certain costs and consequences should be thought of as health care costs or savings (numerator), or health decrements or improvements (denominator), and it completely ignores nonhealth costs and effects, such as those associated with economic productivity, the environment, or education. Therefore, if analyses are to handle such issues consistently, the choice between numerator (resource impact) and denominator (HRQL impact) must follow an established convention. In any case, the societal perspective dictates that all important impacts on human health and on resources must be included somewhere, either in the numerator or the denominator. With this principle in mind, the panel reached the following recommendations regarding the distinction between costs and health consequences.

By convention, the denominator of a C/E ratio is reserved for the improvement in health associated with an intervention. Thus, effects of an intervention on length of life and on morbidity, including the full value of HRQL to patients, should be incorporated in the denomina-

tor. In order to avoid double counting, monetary values for lost life-years should not be imputed in CEA and should not be included in the numerator of the C/E ratio. For a reference case analysis, HRQL should be captured by a measure that, at a minimum, implicitly incorporates the effects of morbidity on productive time and leisure. When instruments used to measure health states are silent concerning the consideration of lost income, we assume that financial effects have been considered by the respondent and that it is therefore not necessary to account for these effects in the numerator. Preferences for health states ideally should be elicited using health status measures that explicitly invite respondents to consider the full range of impacts of the health status change, including loss of income and leisure activities.

Currently, some instruments used to measure health states explicitly exclude consideration of lost income. Moreover, methods that measure changes in HRQL in monetary terms, such as contingent valuation (willingness to pay), have also been used. While these approaches are technically valid, a CEA in which these practices were used would not constitute a reference case analysis.

The numerator of a C/E ratio captures changes in resource use associated with an intervention. The major categories of resource use that should be included are costs of health care services; costs of patient time expended for the intervention; costs associated with caregiving (paid or unpaid); other costs associated with illness, such as child care and travel expenses; economic costs borne by employers, other employees, and the rest of society, including so-called friction costs associated with absenteeism and employee turnover[8]; and costs associated with nonhealth impacts of the intervention, such as on the educational system, the criminal justice system, or the environment.

The handling of patient time in CEA presents challenges and is the focus of several reference case recommendations. Time spent by individuals seeking health care or undergoing an intervention is a component of the intervention, and thus it should be valued in monetary terms and incorporated into the numerator of the C/E ratio. Time spent sick (morbidity time) is part of the health outcome measured in the denominator of the CEA, as described above. In some instances (eg, when recuperating from surgery), time could be categorized either as morbidity time (valued in the denominator) or as an input to the intervention itself (costed out in the numerator); as a general rule, this time should be considered as morbidity time. These recommendations are not based on any fundamental theoretical consideration, but are made for consistency across reference case analyses.

Measuring Terms in the Numerator of a C/E Ratio (Costs)

A change in the use of a resource caused by a health intervention should be valued at its opportunity cost, which is the value the resource could have produced if it were spent in its best available alternative use. In economics, this principle is the basis for valuing resource flows in society.

Several implications arise from the opportunity cost principle. First, it is the difference in resource use between an intervention and the intervention with which it is being compared that should be included in the numerator of the C/E ratio. That is, costs should reflect the marginal or incremental resources consumed or saved, rather than total resources. Fixed costs—costs unaffected by the level of implementation of an intervention—should generally be excluded from consideration. However, resource costs should be measured from a long-term perspective, which implies that many costs that may be fixed in the short run (such as most of what is usually considered overhead in the financial accounts of hospitals and other health care providers) are in fact variable in the long run and should be included in CEAs.

Direct measurement of opportunity costs is difficult and often impossible. To the extent that market prices of healthcare inputs reflect opportunity costs, they provide an appropriate means for valuing changes in resources. According to economic theory, prices in competitive markets reflect opportunity costs of resources. However, when prices do not adequately reflect opportunity costs because of market distortions, they should be adjusted appropriately. Examples of adjustments commonly used in CEAs include the use of ratios of cost to charge (RCCs) to adjust hospital prices, the use of management accounting systems to estimate costs, and the use of third-party payments to providers in lieu of fees to reflect provider opportunity costs. When substantial bias is present in prices and adjustment is not feasible, the panel recommends that more suitable proxies for opportunity costs be considered, including the possibility of conducting "microcosting" studies within provider organizations. (Such studies collect information on the range of inputs to a service, such as the nursing care, supplies, and ancillary services constituting a day of hospital care.)

Costs should be measured in constant dollars, that is, in the dollars of a fixed year. When the original data are for different years, the effect of price inflation must be removed, either by inflating the data from an earlier year to the chosen year or by deflating the data from a later year. Depending on whether the resources being valued are more representative of goods and services in the economy at large or in the medical care sector, either the CPI or its medical care component(s) is suitable for inflation adjustment in CEAs.

Transfer payments (such as cash transfers from taxpayers to welfare recipients) associated with a health intervention redistribute resources from one individual to another. While administrative costs associated with such transfers could be included in the numerator of a C/E ratio, the transfers themselves should not be since, by definition, their impact on the transferor and the recipient cancel out from the societal perspective.

Time costs for individuals in the labor force should generally be valued by the wage rate as an acceptable measure of opportunity cost of time. The reference case recommendation is to use wages corresponding to the age and gender composition of the target population. However, group-specific wages may influence the conclusions of an analysis in ethically problematic ways. For example, a policymaker might object to having the wage differential between men and women reflected in the results of a CEA. In these instances, sensitivity analyses should be conducted to explore the specific nature of this influence. Because of such ethical concerns, and because of practical problems in obtaining data on wages by characteristics other than age and gender, the panel does not recommend using wages specific to target groups defined by race, ethnicity, or other characteristics.

Wage rates generally do not adequately reflect the value of time for persons engaged primarily in leisure—such as retired persons—or in activities for which they are not compensated—such as household activities. Average age- and gender-specific wages among persons in the labor force may be applied to approximate the opportunity cost of time for persons of similar age and gender not in the labor force.

Should health care costs that result solely from the fact that a successfully treated patient lives longer be attributed to the health intervention? Which future costs should be included in a CEA? For example, a cost-effective analysis of antihypertensive therapy found that excluding noncardiovascular disease costs in future years would reduce the C/E ratio by 5% to 20%, with the greatest impacts on the ratios for younger population groups.[9] To clarify the issues, we define 5 categories of induced costs that may or may not be germane in a CEA. These are (1) costs for intervention-related diseases incurred in years of life that would have been lived anyway; (2) costs for unrelated diseases

that are incurred in years of life that would have been lived anyway; (3) health care costs for related diseases that ensue in years of life added (or subtracted) as a result of the intervention, (4) health care costs for unrelated diseases that occur in years of life added (or subtracted) by the intervention, and (5) nonhealth care costs typified by commodities such as food and shelter that occur in years of life added (or subtracted) by the intervention.

The handling of some of these categories of costs is uncontroversial. Costs of related diseases in the original life span clearly must be included in the analysis. For example, costs and savings associated with treatment of strokes and myocardial infarctions must be included in analyses of hypertension programs. Costs of treating adverse effects of treatment must be included as well. On the other hand, because unrelated health and nonhealth care costs occurring throughout the expected years lived without the intervention would cancel from the incremental cost calculation in the numerator of the C/E ratio, these may be omitted from the analysis. It may actually be preferable to omit these unrelated costs because their measurement may add to error in the estimation of costs with and without the intervention.

Costs for intervention-related diseases that occur in added years of life are typically included in CEAs. For example, if a fatal myocardial infarction is delayed 5 years by a coronary bypass operation or a cholesterol-lowering regimen, the costs of treating coronary events ensuing throughout the 5 years should be, and usually are, included. Similarly, costs of an ongoing therapy throughout added years of life, such as lifelong antihypertensive treatment and its medication side effects, are always included.

Costs of diseases unrelated to the intervention and ensuing as a result of added years of life have been the source of more debate.[4,5,10-12] Difficulties with the choice to include or exclude them are illustrated by the example of a cholesterol-lowering intervention. The analyst might decide to exclude all unrelated costs occurring in years of life gained because of the program. In this case, costs of illnesses such as arthritis and Alzheimer disease would be excluded. However, age-specific "background" costs of coronary heart disease—that is, the level of disease that would occur among people who are not candidates for the intervention—are also "unrelated" to the intervention and should also be excluded. To neglect to do so would provide an uneven playing field for comparisons of interventions affecting different diseases: life-prolonging heart disease interventions would be encumbered with all future costs of heart disease even though they only target an excess risk, while suicide prevention programs would not. To avoid the practical and conceptual problems in disengaging the "related" and "unrelated" elements of costs for "related" diseases, it would be preferable to include all these costs. However, this choice would impose a burden on the analyst frequently not warranted by the importance of future costs.

Because there are unresolved theoretical and empirical questions and because health care costs in added years of life are typically small compared with the other costs in an analysis, the panel concluded that the reference case may either include or exclude these costs. Whenever the investigator has reason to believe that inclusion or exclusion of future health care costs may make a significant difference to the analysis, a sensitivity analysis should be performed to assess the effect on the C/E ratio.

We now consider nonhealth care costs in added life-years. Although there is no precedent in CEA for including these costs, one could reasonably argue that if health care costs in added years can be included, future expenditures on food, clothing, and shelter should also be included. The theoretical answer is that the net economic burden of survivors on the rest of society (consumption minus productivity) should indeed be included as a cost. However, if these nonhealth care costs are truly "unrelated," then their consistent inclusion or exclusion would only add or subtract a constant from the C/E ratio.[5] Whether nonhealth care costs are in fact "unrelated," or at least approximately so, is an unresolved empirical question. Nonetheless, on the assumption that these costs can reasonably be considered to be unrelated and to avoid placing an unnecessary burden on the analyst, the panel does not recommend including future nonhealth care costs in reference case analyses.

Valuing the Health Consequences in the Denominator of a C/E Ratio

As discussed in the first article of this series,[2] a reference case analysis should measure health effectiveness in terms of QALYs. These QALYs incorporate changes in survival and changes in HRQL by weighting years of life to reflect the value of the HRQL during each year.[13,14] In order to be consistent with the QALY construct, the quality weights must be measured by or transformed into the interval scale on which optimal health has a value of 1 and death has a value of 0.

The weights used in QALYs should be based on a health-state classification system that reflects health-related domains (attributes) that are important for the particular analysis. In order to qualify as a reference case analysis, the CEA should use a generic health-state classification, that is, a classification that applies broadly across diseases and conditions. Disease-specific health-state classifications are appropriate for a reference case analysis provided that they are designed to be mapped onto or embedded within a generic system. Some examples of commonly used health-state classification systems that may be suitable for CEA include the Health Utilities Index,[14,15] the EuroQol,[16] the Quality of Well-Being Scale,[17] and the Years of Healthy Life measure.[18] The Health Utilities Index, for example, consists of 8 domains (vision, hearing, speech, dexterity, mobility, cognition, emotion, and pain), each of which is classified into either 5 or 6 levels. Each combination of levels is assigned a weight, using a formula based on multiattribute utility theory and a community preference survey.[13,14]

The weights used in QALYs should be based on preferences for health states. In a reference case analysis, these weights should be based on community preferences, rather than those of patients, providers, or investigators. The rationale for community preferences has been described in the first article of this series.[2] Health status scales that are not preference weighted, such as the Medical Outcomes Study Short-Form Health Survey (SF-36),[19] are not suitable for CEA in their present form. Use of patient preferences to value health states is acceptable in a reference case analysis only when adequate information is unavailable regarding community preferences.

The weights assigned to health states should be interval scaled; that is, the method of measurement should be one in which the ratio of differences between values is meaningful. (By analogy, the Fahrenheit, Celsius, and Kelvin scales are equivalent and appropriate measures of temperature, because the intervals between degrees reflect meaningful differences in temperature.) According to decision theory, preference weights obtained from standard gamble questions and, under certain conditions, time trade-off questions satisfy the interval property. At the same time, psychometric research suggests that rating scales can produce interval data. These claims are mutually inconsistent, since there is apparently not a linear relationship between rating scales and standard gambles or time trade-offs.[20,21]

It remains an open question whether standard gambles, time trade-offs, rating scales, or other measures such as person trade-offs[22] produce the closest approximation to the type of interval-scaled weights needed for QALYs. For example, some research suggests that respondents may introduce distortions

into responses to utility questions such as the standard gamble, compromising their theoretical desirability. Therefore, the panel does not recommend one source of weights over the others; for purposes of the reference case, preference weights can come from measurement systems that rely on any of these techniques. The discrepancies associated with different measurement strategies pose a problem for standardization that will be important to address in future research.

To date, most CEAs using QALYs have assumed that a year of life gained by an intervention is valued at 1.0 QALY. In fact, people are rarely in the state of optimal health that a full QALY implies; the use of a value of 1.0 for years that life is prolonged will therefore overstate an intervention's effectiveness and underestimate the true C/E ratio. The panel recommends that, when calculating QALYs gained from a life-extending intervention, estimates of age- and sex-specific HRQL should be applied to the years of extended life—even if the intervention itself has no effect on HRQL. Similarly, when estimating QALYs gained by ameliorating disease symptoms, a return to average rather than optimal HRQL should be credited. It should be noted that this use of average quality of life in reference case analyses means that studies using ratios of cost per year (unadjusted) of life saved will not be comparable to reference case results.

Sociodemographic characteristics, such as age, sex, or race, are associated with HRQL. When the QALYs produced by an intervention vary as a result of these sociodemographic differences, reference case results are affected in ways that may be ethically problematic. For example, an intervention that extends the lives of 80-year-olds may appear less cost-effective than an equally effective intervention applied to 20-year-olds, not only because fewer years are gained, but also because the average quality of those years is less. In these instances, sensitivity analyses should be conducted to indicate explicitly how results are affected.

Estimating Effectiveness of Interventions

The quality and validity of a CEA depend crucially on the quality of the underlying data that describe the effectiveness of interventions and the course of illness without intervention. Data may be obtained from primary data collection efforts specifically intended to inform the CEA or from secondary data sources. The appropriateness of various sources of data will depend on the purpose of a CEA. The consequences of misestimation of cost-effectiveness may be regarded as more serious by some decision makers than by others. For example, the Food and Drug Administration might desire a greater level of certainty in distinguishing the cost-effectiveness of very similar drugs for the purpose of reviewing marketing claims than a formulary manager might demand in adopting a new drug.

For the purpose of a reference case analysis, acceptable data for estimation of effectiveness may come from a variety of sources: randomized controlled trials, observational studies, uncontrolled experiments, or descriptive series. The analyst should select outcome probabilities from the best-designed (and least-biased) sources that are relevant to the question and population under study. There are often trade-offs between the internal validity of data (optimized in randomized trials) and their external validity in actual practice. Meta-analysis and other synthesis methods can be used when no single study has sufficient power to detect effects or when studies produce conflicting results. Expert judgment should be used only as placeholders where no adequate empirical data exist, or when the parameter of interest plays only a minor role in the analysis.

Modeling is a valid and necessary scientific procedure for estimating effectiveness for CEA. Typically, data from randomized trials are combined with observational data and public health statistics in models that are used to estimate changes in life expectancy and quality-adjusted life expectancy. Models may incorporate features such as logistic regression to estimate incidence of disease or death contingent on risk factors; Bayesian analysis to estimate posttest probabilities of disease from data on sensitivity, specificity, and prevalence; and life-table analysis to estimate life expectancy from survival curves. Models include population and cohort models, deterministic and probabilistic models, decision analysis and state-transition models. Because of limited time horizons and selected study populations in clinical studies, failure to use models to extrapolate from primary data can lead to greater error than the models themselves would introduce. Models should be used as complements to, not substitutes for, direct primary or secondary empirical evaluation of effectiveness. Readers are referred to the full report of the panel for more discussion of the roles of particular types of data and models in CEA.[3]

Time Preference and Discounting

Interest rates reflect people's preference for having money and material goods sooner rather than later. Similarly, people value health outcomes that occur in different time periods differently. In CEA, time preference for resources is reflected by discounting future costs to present value. Discounting the value of future expenditures requires that health effects experienced in the future also be discounted at the same rate. This conclusion is based on the observation that people have opportunities to exchange money for health, and vice versa, throughout their lives. Failure to discount health effects will lead to inconsistent choices over time; for example, it will appear that delaying investments will always result in a program's becoming more cost-effective. For this reason and based on other evidence and considerations outlined in its full report,[3] the panel recommends that costs and health outcomes occurring during different time periods should be discounted to their present value and that they should be discounted at the same rate.

Although a wide variety of discount rates are used in the literature and can be defended, a convention is needed to achieve consistency across analyses. Theoretical considerations suggest that the real (inflation-adjusted) discount rate should be based on time preference, the difference in value people assign to events occurring in the present vs the future. Further, economic theory suggests that time preference is reflected in the rate of return on riskless, long-term securities. Empirical evidence is consistent with this rate's being in the vicinity of 3% per annum (net of inflation). Direct evidence on time preferences for health outcomes is also consistent with a discount rate of 3%.

The panel therefore recommends the use of a real, 3% discount rate in the reference case. Before discounting, all costs should be adjusted for inflation. Because many published CEAs have used a discount rate of 5%, future analyses should include sensitivity analysis using 5% as well as other rates in the range of 0% to 7%. The discount rate should be reviewed and possibly revised periodically, to reflect important changes in economic conditions. To ensure that analyses will remain comparable, however, both 3% and 5% should continue to be used for at least the next 10 years.

Handling Uncertainty in CEA

Cost-effectiveness analyses are subject to uncertainty with regard to estimates of effectiveness, the course of illness, HRQL consequences and preferences, and health care utilization and costs. Users of analysis need information on the degree to which CEA conclusions might change with changes in assumptions or values.

Sensitivity analysis is an appropriate tool with which to respond to this need.

The simplest method, which should be used in all CEAs, is a univariate (1-way) sensitivity analysis, in which estimates or assumptions are changed one at a time. These establish where uncertainty or lack of agreement about some key parameter's value could have substantial impact on the conclusion of a CEA. They suggest areas where efforts to obtain additional data might be warranted in terms of impact on decisions and, conversely, areas where additional precision would be unlikely to change results. For example, the estimated cost-effectiveness of thrombolytic therapy in older patients with suspected myocardial infarction was shown to be stable when the efficacy of therapy, the prevalence of myocardial infarction, or the incidence of stroke was varied, thus suggesting that further research to evaluate the risks and benefits of streptokinase in this age group was not warranted.[23]

One-way sensitivity analyses understate the overall uncertainty in the C/E ratio; therefore, analysts should also conduct multivariate (multiway) sensitivity analyses, in which several estimates or assumptions are changed at the same time, for important parameters. If possible, a reasonable confidence interval or credible interval for the C/E ratio should be estimated based either on statistical methods or on simulation. The value of multivariate sensitivity analysis is greatest when there is reason to believe that estimation errors for key parameters are correlated, for example, if studies that overestimated the effectiveness of thrombolysis also tended to underestimate the associated risk of stroke. Several methods for performing such statistical analyses and simulations are described in the panel's full report.[3]

CONCLUSION

The panel recognizes that many of the recommended methodological approaches are not broadly practiced at present, and there may be gaps in the availability of data to satisfy the criteria for the reference case. For example, there is no broad consensus on which health-state classification systems are suitable for CEA, and values of community-based weights for some systems are not publicly available. The ability to weight years of life by population averages of community preference weights is limited by the lack of appropriate (age- and sex-specific) population data for some systems. In the area of costs, there are no well-accepted methods for determining time costs for individuals outside the labor force, and few good-quality data on resource use, reflecting costs rather than charges and clearly applicable to the populations under study, are readily available to analysts.

The intention of these recommendations is to move the field of CEA closer to standardization in the near term where possible and to identify desirable practice where optimal methods are not currently feasible. All of these recommendations for the reference case, and for CEAs in general, are subject to a "rule of reason." When a parameter estimate or an element of the analysis is unlikely to have an appreciable effect on the result, then it may be acceptable to use shortcuts to obtain them; expert opinion may be used to assess them; or they may be excluded from the analysis altogether. Examples may include the weights assigned to short-term and mild impairments of HRQL, costs of unrelated health care in added years of life, or the incidence or costs of minor side effects of treatments. The rule of reason applies if the cost of obtaining more precise estimates of the parameter in question would exceed the value of achieving more precision in the final cost-effectiveness result. However, the burden is on the analyst to justify suboptimal methods of parameter estimation or exclusion of effects from an analysis.

If researchers endeavor to follow a standard set of methods in CEA and to obtain the required inputs for their studies, much will have been accomplished toward improving the utility of this form of analysis. It is hoped that the recommendations contained here will stimulate rapid progress toward availability of the necessary data and tools, so that the practice of CEA can soon become as established as many other forms of scientific inquiry.

The Panel on Cost-Effectiveness in Health and Medicine membership and staff: Norman Daniels, PhD; Dennis G. Fryback, PhD; Alan M. Garber, MD, PhD; Marthe R. Gold, MD, MPH; David C. Hadorn, MD; Mark S. Kamlet, PhD; Joseph Lipscomb, PhD; Bryan R. Luce, PhD; Jeanne S. Mandelblatt, MD, MPH; Willard G. Manning, Jr, PhD; Donald L. Patrick, PhD; Louise B. Russell, PhD; Joanna E. Siegel, ScD; George W. Torrance, PhD; Milton C. Weinstein, PhD.

We thank Kristine I. McCoy, MPH, for coordination, research, and editorial assistance in association with this project.

References

1. Udvarhelyi S, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? *Ann Intern Med.* 1992;116:238-244.
2. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC, for the Panel on Cost-Effectiveness in Health and Medicine. The role of cost-effectiveness analysis in health and medicine. *JAMA.* 1996; 276:1172-1177.
3. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine.* New York, NY: Oxford University Press; 1996.
4. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med.* 1977;296:716-721.
5. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *J Health Econ.* In press.
6. Siegel JE, Weinstein MC, Russell LB, Gold MR, for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations for reporting cost-effectiveness analyses. *JAMA.* In press.
7. Newhouse JP. Medical care costs: how much welfare loss? *J Econ Perspect.* 1992;6:3-21.
8. Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. *J Health Econ.* 1995;14:171-189.
9. Weinstein MC, Stason WB. *Hypertension: A Policy Perspective.* Cambridge, Mass: Harvard University Press; 1976.
10. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* Oxford, England: Oxford University Press; 1987.
11. Russell LB. *Is Prevention Better Than Cure?* Washington, DC: Brookings Institution; 1986.
12. Meltzer DO. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ.* In press.
13. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care.* 1989;5:559-575.
14. Patrick DL, Erickson P. *Health Status and Health Policy: Allocating Resources to Health Care.* New York, NY: Oxford University Press; 1993.
15. Torrance GW, Boyle MH, Horwood SP. Application of multiattribute utility theory to measure social preferences for health states. *Operations Res.* 1982;30:1043-1069.
16. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16:199.
17. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res.* 1988;23:203-235.
18. Erickson P, Wilson R, Shannon I. *Years of Healthy Life.* Hyattsville, Md: Centers for Disease Control and Prevention, National Center for Health Statistics; 1995. Dept of Health and Human Services publication (PHS) 95-1237.
19. Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey, I: conceptual framework and item selection. *Med Care.* 1992;30:473-483.
20. Torrance GW. Social preference for health states: an empirical evaluation of three measurement techniques. *Socio-economic Plann Sci.* 1976; 10:129-136.
21. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes: comparison of assessment methods. *Med Decis Making.* 1984;4:315-329.
22. Nord E. Methods for quality adjustment of life years. *Soc Sci Med.* 1992;34:559-569.
23. Krumholz HM, Pasternak RC, Weinstein MC, et al. Cost effectiveness of thrombolytic therapy with streptokinase in elderly patients with suspected acute myocardial infarction. *N Engl J Med.* 1992;327:7-13.

Printed and Published in the United States of America

Risk Analysis, Vol. 14, No. 1, 1994

A Distributional Approach to Characterizing Low-Dose Cancer Risk

John S. Evans,[1] John D. Graham,[2,4] George M. Gray,[2] and Robert L. Sielken, Jr.[3]

Received October 26, 1992; revised June 14, 1993

Since cancer risk at very low doses cannot be directly measured in humans or animals, mathematical extrapolation models and scientific judgment are required. This article demonstrates a probabilistic approach to carcinogen risk assessment that employs probability trees, subjective probabilities, and standard bootstrapping procedures. The probabilistic approach is applied to the carcinogenic risk of formaldehyde in environmental and occupational settings. Sensitivity analyses illustrate conditional estimates of risk for each path in the probability tree. Fundamental mechanistic uncertainties are characterized. A strength of the analysis is the explicit treatment of alternative beliefs about pharmacokinetics and pharmacodynamics. The resulting probability distributions on cancer risk are compared with the point estimates reported by federal agencies. Limitations of the approach are discussed as well as future research directions.

KEY WORDS: Distributional analysis; probability tree; carcinogen risk assessment.

1. INTRODUCTION

Risk managers are increasingly asking hard questions about the scientific basis of quantitative risk estimates.[1] They want to know how much confidence can be placed in the methods used to calculate risk. They also want to know about major points of scientific agreement and disagreement, and how differences in scientific opinion are reflected in risk assessment.

Risk managers have good reasons for desiring more information about which estimates of cancer risk are most likely or least likely to be correct. They see this type of scientific uncertainty as a factor to consider when setting priorities among potential targets for risk management decisions and for determining the proper degree of stringency in protective regulations. Moreover, indications of uncertainty can help target priorities for scientific research, which may increase confidence in future risk assessments. Finally, reporting the relative likelihood of alternative estimates of risk can assist risk managers in efforts to make honest characterizations of risk to journalists and the public.[2]

While many risk assessments acknowledge uncertainty, quantitative indications of the extent of uncertainty rarely reach the desks of risk managers. For example, the statement that "excess cancer risk may be as low as zero" does not distinguish a 90% probability of zero risk from a 1% probability of zero risk. Moreover, the phrase "plausible upper bound" does not define how likely it is that the true risk exceeds the plausible upper bound (and by how much!). Risk managers need to be provided quantitative statements about the likelihood that various estimates of risk are correct.[3] Fortunately, a technical consensus is emerging that risk analysts can and should do a better job of quantifying the degree of scientific uncertainty in their estimates of risk.[4]

In this article, we demonstrate a probabilistic ap-

[1] Interdisciplinary Programs in Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115.
[2] Center for Risk Analysis, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115.
[3] Sielken, Inc., Suite 230, 3833 Texas Avenue, Bryan, Texas 77802.
[4] To whom all correspondence should be addressed.

0272-4332/94/0200-0025$07.00/1

proach to characterizing the likelihood of alternative values for low-dose cancer risk, one of the most contentious issues in risk assessment. Our approach is rooted in Bayesian concepts of subjective probability[5] (which see strengths of belief on the part of scientific experts as legitimate measures of likelihood) and is widely used in business, medicine, and engineering.[6] Here, we illustrate the distributional method without addressing the critical question of how "experts" should be selected. In previous and ongoing work, various procedures for selecting experts are being examined.[7]

Our distributional approach builds on other applications of decision analysis to environmental control problems.[8] Formaldehyde (HCHO) was selected for application of the approach because it is a chemical that has been subjected to several risk assessments,[9] and because it has a rich yet controversial experimental database that raises interesting challenges for uncertainty analysis.[10]

For illustrative purposes, we address the same analytic questions that EPA and OSHA analysts have addressed in recent risk assessments: How many additional cases of cancer result each year from 240,000,000 Americans inhaling average outdoor concentrations of approximately 2.28 parts per billion of formaldehyde? The EPA reported an "upper bound" point estimate of 124 excess cancers per year.[11] For OSHA the question was how much excess cancer risk is incurred by a worker who inhales 0.75 parts per million of formaldehyde for a working lifetime? OSHA constructed two point estimates: 1.6 in 1,000,000 and 2.0 in 1000.[12]

2. A PROBABILITY-TREE APPROACH WITH ILLUSTRATIVE WEIGHTS

The most direct method for characterizing the likelihood of alternative values of carcinogenic potency is to elicit from appropriate scientists their subjective probability density functions on dose-specific potency.[13] Although this direct method has been used to assess the noncarcinogenic health risks of lead and sulfate pollution,[14] it has not been applied to chemical carcinogenesis.

While the direct-elicitation method may be worth exploring, we rejected it in this case. In light of the fragmentation in scientific expertise on cancer potency, we decided to decompose cancer risk into component parts before obtaining subjective probabilities. This allows scientists to focus thinking on well-structured questions that may be closer to their knowledge base.[15] This article explores cancer risk characterization through a strategy of decomposition and recomposition.

For analytical purposes, the inputs to the carcinogenic potency of HCHO are summarized as a six-level probability tree. The proposed tree, in simplified form, is reproduced in Fig. 1. A more complete appendix, which portrays the full tree and associated probabilistic weights that have been used in our illustrative calculations, is available from the authors upon request.

The weights in the probability tree are crucial to the distributional analysis. The weights for the different alternatives at each level of the probability tree are nonnegative and sum to one. The weight for a complete path through a probability tree is obtained by multiplying the weights assigned to each alternative along the path. Like the probabilistic weights, each path weight is nonnegative and the sum of all path weights must equal one. Equal weights imply that the alternatives have equal likelihood of being correct, given current scientific evidence. If an alternative has been omitted from the tree, it has implicitly been assigned a probabilistic weight of zero.

In this article, where the probabilistic approach is illustrated, the weights have been assigned based on our reading of the scientific literature and have not been elicited from experts in toxicology, epidemiology, biostatistics, pathology, biochemistry, and cancer biology. Hence, the weights should be considered illustrative of the kinds of information that need to be elicited from the relevant scientific communities if the distributional approach is to be fully implemented. Reasonable scientists might disagree with the weights reported here.

In level one of the tree in Fig. 1, a probabilistic judgment must be made about whether HCHO has the potential to cause cancer in people. This judgment corresponds to the concept of "carcinogenic hazard," as reflected in carcinogen classification schemes. It is a judgment that traditionally is made independent of dose, which means that we are saying that there is an X percent chance that HCHO could cause cancer in some people under at least some conditions of exposure. We have assigned a probability of 0.8 to the event that HCHO is capable of causing cancer in humans, and a probability of 0.2 that HCHO is not capable of causing human cancer, which is arguably consistent with EPA's decision to classify formaldehyde as a (B1) "probable human carcinogen."[16]

In the second level of the tree, various mechanistic hypotheses about chemical carcinogenesis are proposed and assigned probabilistic weights. This is perhaps the most critical question in carcinogen risk assessment. We

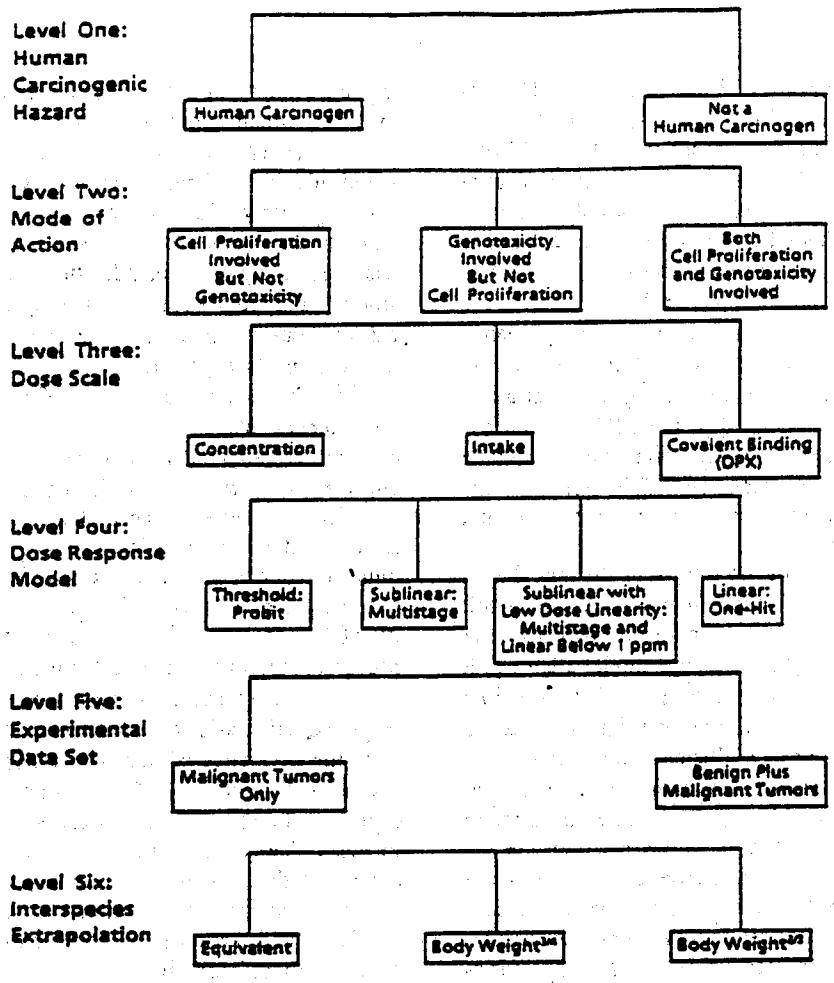


Fig. 1. The simplified tree for formaldehyde: the alternatives for the six major levels of the dose-response assessment.

assume that HCHO could cause cancer in the respiratory system by inducing chronic cell proliferation, by direct genotoxicity, or by a combination of cell proliferation and genotoxicity. A probability of 0.8 is assigned to the event the formaldehyde carcinogenicity is caused solely by chronic cell proliferation, on the basis of mechanistic research reported by scientists at the Chemical Industry Institute of Toxicology (CIIT).[17] The remaining probability, which turns out to be 0.195, is assigned to the event that both cell proliferation and genotoxicity play a role in formaldehyde carcinogenesis, since HCHO is known to be mutagenic.[18] A slight probability, say 0.005, is assigned to the possibility that direct genotoxicity explains all of formaldehyde's carcinogenic activity. If this level of the tree were devised by cancer biologists and toxicologists, it might include more branches, numerous subtrees, and possibilities for subtle biological interactions.

The third level of the probability tree considers three possible ideas about what is the relevant dose metric: the traditional measure of administered concentration (in ppm of HCHO), a measure of HCHO intake,[19] and a measure of covalent binding of formaldehyde to DNA in the respiratory system in rats and monkeys, as measured by scientists at CIIT.[20] The choice of dose scale reflects a judgment about the true relationship between the amount of HCHO administered and the amount that interacts with the target cells in the respiratory system. The traditional approach assumes that delivered dose is proportional to administered concentration, even at low

administered concentrations.[21] In contrast, CIIT's data suggest that the covalent binding of HCHO to DNA declines rapidly as administered doses are reduced.[22] Ideally, the branches here should reflect different pharmacokinetic models rather than simply different sources of data. A complication we do not consider is the biology of dose pattern (e.g., peaks vs. averages or continuous vs. intermittent exposures).

In our illustrative calculations, we assign probability weights of 0.1, 0.3, and 0.6 to the three dose scales, respectively. Although it has been suggested that all of the weight should be assigned to the CIIT data on covalent binding in rats and monkeys, some questions have been raised about the relevance of CIIT's short-term experiments to assessment of risks from chronic human exposures to HCHO.[23]

In the fourth level of the tree, four distinct shapes for the dose–responsive function are presented. The probit model reflects threshold-like behavior in dose–response curves. The unconstrained multistage model allows (but does not compel) linear dose–response behavior at low doses, and for this data set, the best fitting model is sublinear. The constrained multistage model compels a linear term at low-doses while allowing for curvature at high doses. And the one-hit model assumes linearity in tumor response from zero dose to doses within the experimental range. (Supralinear curves can also be added to the tree, although they are not plausible in the case of HCHO). For each data set and model choice, parameter uncertainty is estimated using standard bootstrapping methods, which account for the limited number of animals at risk in the CIIT bioassay of HCHO.

Bootstrapping is an approach based on Monte Carlo simulations for assessing the uncertainty in parameter estimations due to the inherent variability of small samples. Simulation techniques were used to generate sets of hypothetical data from the bioassay. Each data set consisted of four data points [i.e., values of r (the number of responders) and n (the number of animals at risk)] corresponding to the four dose groups in the Kerns *et al.*'s bioassay. The simulated values of r_d were drawn from binomial probability functions with parameters π_d (the true probability of response at dose d) and n_d (the number of animals at risk at dose d). Estimates of π_d were based on the observed responses in the Kerns' bioassay. Uncertainty in the parameters of the four dose–response models of interest is assessed by fitting the model to each of the key potential data sets and observing the variability in the parameter estimates.

Our analysis relies on the tumor incidence data from the CIIT bioassay, since it is the largest and most widely reviewed long-term inhalation bioassay of HCHO.[24] A key uncertainty in evaluating the tumor incidence data in rats is whether to count only squamous cell carcinomas of the nasal cavity, or whether to also count polypoid adenomas (benign tumors) as relevant to human carcinogenicity. The incidence of polypoid adenomas exhibits an inverse dose–response relationship, which has been a topic of considerable interpretative discussion.[25] Polypoid adenomas may be a precursor of adenocarcinomas, a rare malignant tumor that was observed in the high-dose groups in Kerns *et al.*,[24] Albert *et al.*,[26] and Sellakumar *et al.*[26] On the other hand, a more recent inhalation study of HCHO in rats did not observe any polypoid adenomas or adenocarcinomas that appeared to be related to HCHO exposure.[27]

The fifth level of the probability tree has two distinct branches: one that includes polypoid adenomas in tumor incidence counts and one that excludes them. For purposes of probabilistic analysis, we assign a probability of 0.2 to the event that polypoid adenomas are a precursor to formaldehyde-induced adenocarcinomas and relevant to human carcinogenicity. In the dose–response calculations performed in this branch of the tree, the incidence of rats with polypoid adenomas is added to the incidence of squamous cell carcinomas at each dose level before dose–response modeling is performed. The remaining 0.8 probability is assigned to the event that polypoid adenomas are not related to HCHO exposure. In this branch of the tree, only the incidence of squamous cell carcinomas are included in dose–response modeling.

Since humans and rodents may not be equally sensitive to equal delivered doses of HCHO, the sixth level of the probability tree allows for three interspecies scaling options. If the dose scale represents the biologically effective dose, there is no need for interspecies adjustments in pharmacokinetics. The three scaling options reported here are intended to represent differences in pharmacodynamics between rodents and people. The most straightforward option assumes that rodents and people are equally sensitive to the same delivered dose of HCHO, whatever the proper measure may be. The other options assume that human responses scale roughly according to body weight to the three-quarters and two-thirds powers, respectively.

When weights are assigned to branches in the tree, their values are typically *dependent* on previous levels of the tree. For example, probabilistic weights assigned to each dose–response model, which are available upon request, are varied depending upon the assumed dose scale and biological mechanism. For example, when the traditional dose scale (ppm) is combined with the hypothesis that formaldehyde acts exclusively through cell

proliferation, the probit model is assigned the dominant probabilistic weight because it captures the desired "threshold-like" behavior in the dose–response curve. When the traditional dose scale (ppm) is combined with the hypothesis that both cell proliferation and genotoxicity are involved, the multistage models are assumed to be appropriate since they allow for a low-dose linear component that may be caused by genotoxicity. When covalent binding is considered the proper dose scale, it is assumed that some of the presumed nonlinearity between administered dose and tumor incidence has been accounted for. In this situation, more of the probabilistic weights is assigned to the linear models. By carefully considering the interactions of various levels of the tree, the analyst can avoid "double counting" the same information twice at different points in the tree.

3. PRODUCING A PROBABILITY DISTRIBUTION ON CARCIONOGENIC POTENCY

When computing a point estimate of the carcinogenic potency of HCHO, the risk assessor can report numerous values depending upon the selection of assumptions, data, and models. The probability tree in Fig. 1 illustrates the analytical options available in hazard identification, mechanism of action, dose scaling, dose-response modeling, counting tumors, and interspecies scaling. Implicit in the tree are 432 estimates of HCHO's potency (many of which are zero in this case). While risk assessors could report 432 cancer estimates for HCHO, risk managers would reasonably ask for some statement about which estimates are most likely and least likely to be correct.[28]

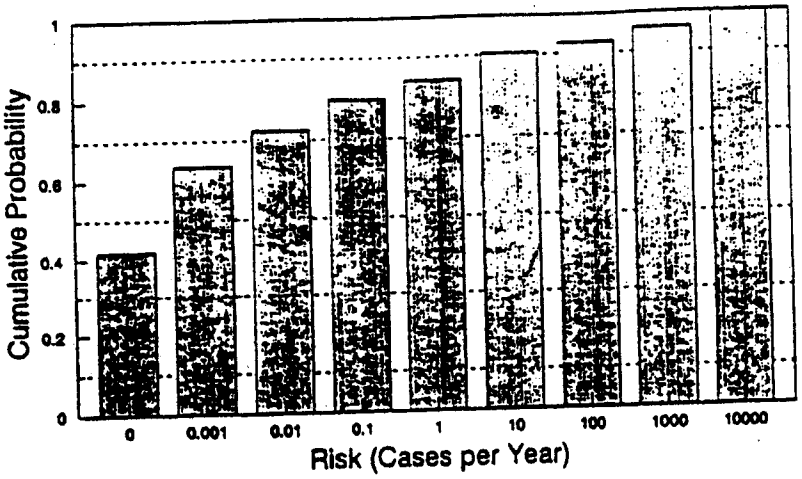
Our distributional approach combines these possibilities into a single probability distribution. Figure 2 reports a probability density histogram for cancer incidence using EPA's point estimate that 240,000,000 people are exposed continuously (24 hr/day) to average outdoor formaldehyde concentrations of 2.28 parts per billion of air. Figure 3 is a probability distribution on the incremental cancer risk associated with a worker inhaling 0.75 ppm of formaldehyde for a working lifetime. We used OSHA's deterministic assumptions that a worker breathes this level of HCHO for 8 hr per day, 5 days per week, 50 weeks per year, for 40 years in a 70-year lifetime. This exposure scenario corresponds to a lifetime average daily HCHO dose of approximately 0.1 ppm.

The results reported here reflect the weighted frequencies of the probability tree outcomes and the bootstrap distributions of risk estimates. Note that traditional "parameter uncertainty" (arising from the limited number of animals in the HCHO animal test) has been combined with "model uncertainty" (e.g., ignorance about shape of the dose–response curve below the experimental range) in the results reported here. In other words, each pathway in the probability tree has its own final risk distribution that arises from "its bootstrap," and its overall density is determined by the cumulative weight of the associated pathway in the tree. Since it was assumed that the current state of scientific knowledge supports some paths in the probability tree more than others, the numerical results with stronger scientific support exhibit larger probabilities. Thus, the overall probability distribution would inform the risk manager of the relative likelihood, given the current weight of the evidence, that incremental cancer risk is equal to various values.

4. SUMMARY STATISTICS AND COMPARISON WITH ESTIMATES REPORTED BY EPA AND OSHA

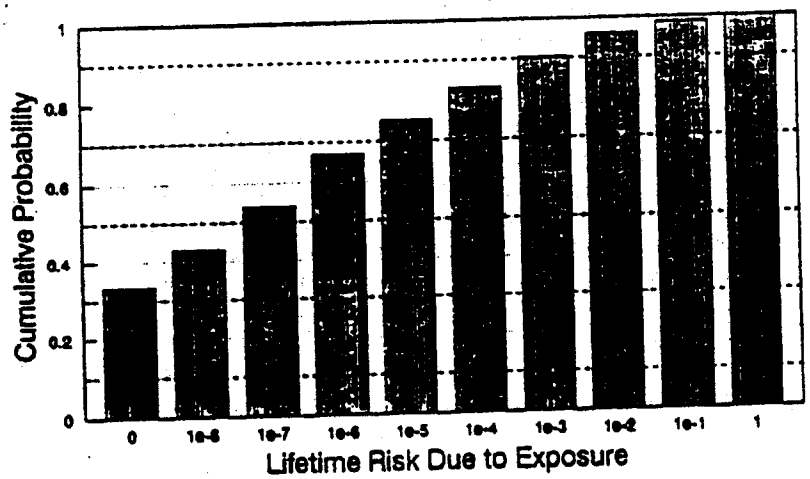
EPA reported an upper bound estimate that outdoor air concentrations of formaldehyde are associated with 124 excess cancers per year. It is interesting to consider where this point estimate falls in the probability distribution that we have constructed. Table I reports two sets of summary statistics. The first set includes various percentiles of the distribution, with an emphasis on the percentiles in the right-hand tail of the distribution. The second set reports the minimum value, the mean, and the maximum value of the distribution.

In addition to our major point, that point estimates of risk conceal the wide distributions reported here, several comments are worth noting. First, EPA's point estimate is considerably smaller than the maximum value of the distribution. In this respect, one can raise questions about whether EPA's estimate is large enough to be considered an upper bound on the true yet unknown risk.[29] Second, the most likely (modal) value of the distribution is essentially zero. While EPA guidelines acknowledge that true risk from low-level carcinogenic exposures may be as low as zero, our analysis suggests that some measures of the central tendency of the distribution on excess risk at 2.28 ppb are zero. While the median value is not zero, it is very close to zero. Third, not all values between zero and 124 are considered equally likely. Most of the likelihood falls at zero or near zero. Fourth, the mean of this distribution is far above the median and mode because the distribution has a long



At average outdoor concentration of 2.28 ppb.
Value at zero includes weights on risks < 0.001

Fig. 2. Cancer risk from ambient formaldehyde. Nationwide incidence attributable to exposure.



At concentration of 0.75 ppm.
Value at zero includes weight on risks <= 1e-8

Fig. 3. Cancer risk from workplace formaldehyde. Lifetime risk to exposed worker.

right-hand tail. Moreover, the mean is more than twice as large as EPA's point estimate.

The summary statistics for the occupational exposure scenario are reported in Table II, using individual risk rather than population risk as the measure of risk. The mean of the distribution, 4 in 1000, is larger than OSHA's 1-in-1000 benchmark of significant risk. On the other hand, the most likely and median values of the distribution are zero and 4 in 100,000,000 respectively. Note that the two point estimates reported by OSHA are both within the probability distribution of values.

Choosing a single summary statistic to represent any complex probability distribution is fraught with danger and implicit value judgments. For example, some analysts have expressed reservations about reporting only the upper bound of a risk distribution,[30] while others have questioned the wisdom of reporting only the most likely (modal) estimate of risk.[31] Even the expected-

Table I. Nationwide Annual Cancer Incidence Attributable to Exposure to 2.28 PPB of Formaldehyde

Summary statistics Percentiles		
5th	=	0
10th	=	0
25th	=	0
50th	=	9×10^{-6}
75th	=	0.05
90th	=	8
94.22nd	=	124
95th	=	220
99th	=	7800
Minimum	=	0
Mean	=	280
Maximum	=	11000

Table II. Lifetime Added Risk for a Worker Exposed to 0.75 ppm of Formaldehyde

Summary statistics Percentiles		
5th	=	0
10th	=	0
25th	=	0
50th	=	4×10^{-8}
75th	=	9×10^{-6}
90th	=	0.001
95th	=	0.003
99th	=	0.1
Minimum	=	0
Mean	=	0.004
Maximum	=	0.13

value of a risk distribution, which is a useful summary statistic in many regulatory contexts,[32] is not always an optimal summary statistic.[33]

5. STRENGTHS OF THE PROBABILISTIC APPROACH

Current methods of cancer potency assessment have been criticized because they censor large amounts of scientific information and neglect key sources of uncertainty. A major advantage of the probabilistic approach is that all of the available scientific evidence and theories can be incorporated into the assessment of carcinogenic potency.[34] Moreover, partial evidence and new biological theories can be considered in the distributional approach, although their impact on the distribution of cancer potencies is weighted by their assessed likelihood of reflecting the truth.

As new scientific data and understanding emerge, the probability distributions on cancer potency can be revised to reflect the new state of evidence. New evidence can impact the probability distribution on potency by changing the structure of the probability tree, the probability weights in the tree, or both. Unlike current methods of cancer risk assessment, which may discourage scientific research by placing excessive weight on default assumptions, the distributional approach would provide incentives for generating new scientific information, since the probability distribution on cancer risk would presumably change in response to new information.

The distributional approach might be criticized on the grounds that it relies on subjective scientific judgments to provide probabilistic weights. This is not a *relative* disadvantage because current methods of cancer potency assessment also rely on subjective scientific judgments (however hidden and unstated they may be). For example, numerous technical judgments are made about which data sets to use, what scaling factors to employ, and which dose–response models to fit. Some of these judgments are informed by risk assessment guidelines but scientific judgments are necessary to devise the guidelines and apply the guidelines responsibly to specific chemicals.

The major analytical difference is that the probabilistic approach requires scientific judgments in the form of explicit, subjective probabilities. A significant body of knowledge exists on the value of eliciting subjective probability distributions from scientists.[35] One of the major findings is that experts tend to exhibit overconfidence in their subjective probability assessments. Decision analysts have demonstrated that overconfidence can be reduced (but not eliminated) through intensive probability training.[36] While reliance on judgmental probabilities may understate the full degree of scientific uncertainty, the distributional approach conveys uncertainty more explicitly than does the point estimate that is currently in widespread use. It should also be remembered that the expert judgments used with current procedures to produce a point estimate also suffer from overconfidence, in a way that is less transparent to readers and decision-makers.

6. LIMITATIONS AND FUTURE DIRECTIONS

There are several conceptual and theoretical problems with the approach that we have demonstrated. While some of these problems are quite thorny and cannot be addressed fully here, we highlight several for the reader's reflection.

First, the combination of judgmental probabilities about model uncertainty with parameter uncertainties (based on bootstrapping) may cause some concern, particularly among scientists schooled in classical statistics. To a strict classicist, it is meaningless to report a probability distribution on an unknown population parameter that can, in reality, assume only one value. We believe no technical problem exists from a Bayesian perspective, where probabilities are applied to unknowns based on informed judgment. In future work, we intend to develop methods of presentation that distinguish model from parameter uncertainties.

Second, and perhaps more troubling to some, is the possibility that each level in the probability tree does not contain a collectively exhaustive set of possibilities. For example, all the possible mechanisms of formaldehyde-induced carcinogenicity may not have been identified. If this is the case, the results we have reported are misleading, since plausible states of the world have been neglected. Obviously, we have no way of assessing mechanistic hypotheses that have not been postulated, but they can be added to the tree as new insights emerge.

Third, some of the levels of the probability tree can be questioned on the grounds that they do not seem to define events in the sense of mathematical probability theory. The first level of the tree, which asks whether formaldehyde is a human carcinogen, is perhaps most susceptible to this criticism. Without considerable refinement, this level of the tree would probably not pass a "clairvoyance test," which is a device used by decision analysts to determine whether a probability statement is well defined. A better approach, which we are applying in current work on chloroform, might be to allow for the possibility of noncarcinogenicity in the interspecies extrapolation level of the tree. In any event, the risk assessment should allow for the possibility that, even though an agent causes cancer in rodents, it does not do so in people.

A final conceptual concern is that some definitional questions, such as the difference between a benign and a malignant tumor, should not be assessed with a judgmental probability. The treatment of benign tumors is important in this example, since the long right-hand tails in our results reflect the inclusion of benign tumors in dose-response estimation. If benign tumors are considered simply a different health effect, then there are no grounds for combining them with cancers in a probabilistic assessment (unless the utility losses of the two effects are similar). We have included them because some scientists believe that benign tumors in rodents are relevant to assessing the *carcinogenic* potency of formaldehyde in humans. In earlier work we heard this message from a qualified panel of pathologists and cancer biologists.[37] While there may be other ways to incorporate this insight into our probability tree, we believe the approach we have taken is defensible.

In addition to these conceptual and theoretical limitations, we have not fully addressed the scientific richness of the formaldehyde example. This article extends our earlier, more simplified assessments of formaldehyde's carcinogenicity,[38] but still more analysis is needed to account for the growing body of evidence on formaldehyde. For example, the human data on HCHO need to be analyzed to determine whether rodent-based estimates of risk are consistent with the findings of occupational epidemiology. Subtle differences in the rat and monkey data also need to be considered. The growing database on the role of cell proliferation in HCHO carcinogenicity should also be modeled more rigorously using, for example, a biologically based model. Such work is underway.[39] And the relative importance of endogenous vs. exogenous sources of HCHO needs to be explored by modelers.[40] In a comprehensive assessment of HCHO, sensitivity analysis of these and other critical issues should be reported.

There are a broader set of issues that could be addressed with the probabilistic approach that we have not explored here. In the future we intend to address questions such as heterogeneity in human sensitivity to cancer, variations in exposure, and the proper weighting of animal and human evidence.

In ongoing work, we are beginning to perform formal probability elicitations on carcinogenicity questions with toxicologists, pathologists, cancer biologists, and epidemiologists to demonstrate how their insights can be incorporated into risk assessments. A project on the carcinogenicity of the chemical chloroform is now in progress. The project includes construction of a probability tree in consultation with experts, training of experts in probability assessment, and elicitation of subjective probability weights relevant to the assessment of chloroform's carcinogenic potency. It is hoped that this work, along with related work by other decision analysts, will increase confidence in the distributional approach to carcinogen risk assessment. In the years ahead, it should be feasible to provide risk managers with quantitative

indications of the probability that alternative cancer potency values reflect the truth, although only breakthroughs in mechanistic science can ultimately eliminate the uncertainty.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support provided by the American Industrial Health Council and the Risk Science Institute. John Evans also received partial support from the U.S. Environmental Protection Agency's Cooperative Agreement with Harvard on Environmental Health and Public Policy. Robert Sielken's previous work on HCHO was supported by the Formaldehyde Institute. Helpful comments and suggestions were provided by John Bailar, Robert Barnard, John Brauman, Joshua Cohen, Anthony Cox, Adam Finkel, Neil Hawkins, James Swenberg, and Tammy Tengs.

REFERENCES

1. F. Henry Habicht II, "Guidance on Risk Characterization for Risk Managers and Risk Assessors," (Office of the Administrator, U.S. Environmental Protection Agency, Washington, D.C., February 26, 1992).
2. Report of an Ad Hoc Study Group on Risk Assessment Presentation, *Presentation of Risk Assessments of Carcinogens* (American Industrial Health Council, Washington, D.C., July 1989).
3. Adam M. Finkel, *Confronting Uncertainty in Risk Management: A Guide for Decision-Makers* (Center for Risk Management, Resources for the Future, Washington, D.C., January 1990).
4. M. Granger Morgan and Max Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Risk and Policy Analysis* (New York, Cambridge University Press, 1990); Kenneth T. Bogen, *Uncertainty in Environmental Health Risk Assessment* (New York, Garland Publishing Inc., 1990); Dale Hattis, "Three Candidate 'Laws' of Uncertainty Analysis," *Risk Analysis* 10, (1990).
5. Bruno de Finetti, *Theory of Probability* (Wiley, New York, 1974); Howard Raiffa, *Decision Analysis: Introductory Lectures on Choice under Uncertainty* (Addison-Wesley, Reading, Massachusetts, 1968).
6. Max Henrion, "Assessing Probabilities: A Review," in *Expert Judgments for Policy Analysis* (Brookhaven, Upton, New York, 1978).
7. J. D. Graham, N. C. Hawkins, and M. J. Roberts, "Expert Scientific Judgment in Quantitative Risk Assessment," *Banbury Report Number 31: Carcinogen Risk Assessment: New Directions in the Qualitative and Quantitative Aspects* (Cold Spring Harbor Laboratory, New York, 1988), pp. 231–244; N. C. Hawkins and J. D. Graham, "Expert Scientific Judgment and Cancer Risk Assessment: A Pilot Study of Pharmacokinetic Data," *Risk Analysis* 8, 611–621 (1988).
8. Warner North, "Representing Uncertainty Explicitly in Environmental Risk Management: An Example from Acid Deposition Policy Analysis" (Proceedings of the Specialty Conference on Environmental Risk Management, *Is Analysis Useful?* Air Pollution Control Association, Chicago, Illinois, April 1986); Harvey R. Richmond, "Development of Probabilistic Health Risk Assessment for National Ambient Air Quality Standards" (Presented at APCA International Specialty Conference on Regulatory Approaches for Control of Air Pollution, Atlanta, Georgia, February 17–20, 1987); R. G. Whitfield, T. S. Wallsten, R. L. Winkler, and H. M. Richmond, *Assessing the Risk of Chronic Lung Injury Attributable to Ozone Exposure*, (Argonne National Laboratory, Argonne, Illinois, August 1990); James K. Hammit, "Subjective-Probability-Based Scenarios for Uncertain Input Parameters: Stratospheric Ozone Depletion," *Risk Analysis* 10, 93–102 (1990); R. H. Harris and D. E. Burmaster, "Restoring Science to Superfund Risk Assessment," *Toxics Law Reporter* 25, 1318–1323 (1992); K. M. Thompson, D. E. Burmaster, and E. Crouch, "Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments," *Risk Analysis* 12, 53–64 (1992).
9. K. G. Brown, "Risk Assessment of Laboratory Rats and Mice Chronically Exposed to Formaldehyde Vapors," *Risk Analysis* 5, 171 (1985); T. B. Starr, "Quantitative Cancer Risk Estimation for Formaldehyde," *Risk Analysis* 10, 85–91 (1990).
10. John D. Graham, Laura Green, and Marc Roberts, *In Search of Safety: Chemicals and Cancer Risk* (Cambridge, Massachusetts, Harvard University Press, 1988).
11. *Cancer Risk from Outdoor Exposure to Air Toxics* (U.S. Environmental Protection Agency, Washington, D.C., EPA-450/89, 1990).
12. U.S. Department of Labor, Occupational Safety and Health Administration, "Occupational Exposure to Formaldehyde," *Federal Register* 57, 22305, footnote 7 (1992).
13. M. Granger Morgan, Max Henrion, and Samuel C. Morris, *Expert Judgments for Policy Analysis* (Brookhaven National Laboratory, Upton, New York, BNL 51358, 1979).
14. R. G. Whitfield and T. S. Wallsten, "A Risk Assessment for Selected Lead-Induced Health Effects: An Example of a General Methodology," *Risk Analysis*, 9, 197–207 (1989).
15. E. J. Bonano, S. C. Hora, R. L. Keeney, and D. von Winterfeldt, *Elicitation and Use of Expert Judgment in Performance Assessment for High-Level Radioactive Waste Repositories* (Sandia National Laboratories, Albuquerque, New Mexico, SAND 89–1821, May 1990).
16. S. E. Spedden and P. B. Ryan, "Probabilistic Connotations of Carcinogenic Hazard Classifications: Analysis of Survey Data for Anchoring Effects," *Risk Analysis* 12, 535–541 (1992).
17. T. M. Monticello, F. J. Miller, J. A. Swenberg, T. B. Starr, and J. E. Gibson, "Association of Enhanced Cell Proliferation and Nasal Cancer in Rats Exposed to Formaldehyde," *The Toxicologist*, 12, 1017 (1992).
18. V. S. Goldmacher and W. G. Thilly, "Formaldehyde is Mutagenic for Cultured Human Cells," *Mutation Research* 116, 417–422 (1983).
19. R. L. Sielken, "The Capabilities, Sensitivity, Pitfalls, and Future of Quantitative Risk Assessment," in *Environmental Health Risks: Assessment and Management* (R. Stephen McCall ed.) (University of Waterloo Press, California, 1989), pp. 95–131.
20. M. Casanova, K. T. Morgan, W. H. Steinhagen, J. L. Everitt, J. A. Popp, and H. Heck, "Covalent Binding of Inhaled Formaldehyde to DNA in the Respiratory Tract of Rhesus Monkeys: Pharmacokinetics, Rat-to-Monkey Interspecies Scaling, and Extrapolation to Man," *Fundamental and Applied Toxicology* 17, 409–428 (1991).
21. D. G. Hoel, N. L. Kaplan, and M. W. Anderson, "Implication of Nonlinear Kinetics on Risk Estimation in Carcinogenesis," *Science* 219, 1032–1037 (1983).
22. T. B. Starr and R. D. Buck, "The Importance of Delivered Dose in Estimating Low-Dose Cancer Risk from Inhalation Exposure to Formaldehyde," *Fundamental and Applied Toxicology* 4, 740–753 (1984). M. Casanova and H. Heck, "The Impact of DNA-Protein Cross-Linking Studies on Quantitative Risk Assessments of Formaldehyde," *CIIT Activities* 11, 1–6 (1991).
23. Report of the Science Advisory Board, *Formaldehyde Risk Assessment Update* (U.S. EPA, Washington, D.C., September 1992).
24. W. D. Kerns, K. L. Pavkov, D. J. Donofrio, and J. A. Swenberg,

"Carcinogenicity of Formaldehyde in Rats and Mice After Long-term Inhalation Exposure," *Cancer Research* 43, 4382–4392 (1983).
25. Report on the Consensus Conference on Formaldehyde, *Environmental Health Perspectives* 58, 334 pp.; J. A. Swenberg and C. J. Boreiko, "Appropriateness of Polypoid Adenoma Data for Quantitative Risk Assessment" (Chemical Industry Institute of Toxicology, Research Triangle Park, North Carolina, 1985).
26. R. E. Albert, A. R. Sellakumar, S. Laskin, M. Kuschner, N. Nelson, and C. A. Snyder, "Gaseous Formaldehyde and Hydrogen Chloride Induction of Nasal Cancer in the Rat," *Journal of the National Cancer Institute* 68, 597–603 (1982); A. R. Sellakumar, C. A. Snyder, J. J. Solomon, and R. E. Albert, "Carcinogenicity of Formaldehyde and Hydrogen Chloride in Rats," *Toxicology and Applied Pharmacology* 81, 401–406 (1985).
27. M. Tobe, T. Kaneko, Y. Uchida, E. Kamata, Y. Ogawa, Y. Ikeda, and M. Saito, "Studies of the Inhalation Toxicity of Formaldehyde" (National Sanitary and Medical Laboratory Service (Japan), 1985), p. 1294.
28. Robert L. Sielken, "Decision Analysis and Quantitative Risk Characterization: Information Analysis-Based Risk Characterization" (Chemical Manufacturers Assn., Washington, D.C., 1990).
29. Adam M. Finkel, "Is Risk Assessment Really Too Conservative? Revising the Revisionists," *Columbia Journal of Environmental Law* 14, 427–467 (1989).
30. Office of Management and Budget, "Current Regulatory Issues in Risk Assessment and Risk Management," *Regulatory Program of the U.S. Government* (Washington, D.C., 1990–91), pp. 13–26.
31. John S. Evans, John D. Graham, George M. Gray, Adrienne Hollis, Barry Ryan, Andrew Smith, Mark Smith, and Allison Taylor, "Summary of Workshop to Review an OMB Report on Regulatory Risk Assessment and Management," *Risk: Issues in Health and Safety* 3, 71–83 (1992).
32. R. J. Zeckhauser and W. K. Viscusi, "Risk Within Reason," *Science* 238, 559–564 (1990).
33. Adam M. Finkel, "Edifying Presentation of Risk Estimates: Not As Easy As It Seems," *Journal of Policy Analysis and Management* 10, 296–303 (1991).
34. R. L. Sielken, "Useful Tools for Evaluating and Presenting More Science in Quantitative Cancer Risk Assessments," *Toxic Substances Journal* 9, 353–404 (1989).
35. H. Otway and D. von Winterfledt, "Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls," *Risk Analysis* 12, 83–94 (1992).
36. D. Kahneman, P. Slovic and A. Tversky (eds.), *Judgment Under Uncertainty; Heuristics and Biases* (New York, Cambridge University Press, 1982).
37. S. K. Wolff, N. C. Hawkins, S. M. Kennedy, and J. D. Graham, "Selecting Experimental Data for Use in Quantitative Risk Assessment: An Expert Judgment Approach," *Toxicology and Industrial Health* 6, 275–291 (1990).
38. J. D. Graham and G. M. Gray, "Air Toxics: Characterizing the Risks," in *Toxic Air Pollutants from Mobile Sources* (Air and Waste Management Association, Pittsburgh, Pennsylvania, 1992), pp. 43–52.
39. R. B. Conolly and M. E. Andersen, "Biologically Based Pharmacodynamic Models: Tools for Toxicological Research and Risk Assessment," *Annual Review of Pharmacology and Toxicology* 31, 503–523 (1991).
40. Tito C. Cascieri and John J. Clary, "Formaldehyde: Oral Toxicity Assessment," *Comments in Toxicology* 4, 295–304 (1992).

Reprinted from JAMA © The Journal of the American Medical Association October 9, 1996 Volume 276 Copyright 1996, American Medical Association

Consensus Statement

The Role of Cost-effectiveness Analysis in Health and Medicine

Louise B. Russell, PhD; Marthe R. Gold, MD, MPH; Joanna E. Siegel, ScD; Norman Daniels, PhD; Milton C. Weinstein, PhD; for the Panel on Cost-Effectiveness in Health and Medicine

Objective.—To develop consensus-based recommendations guiding the conduct of cost-effectiveness analysis (CEA) to improve the comparability and quality of studies. The recommendations apply to analyses intended to inform the allocation of health care resources across a broad range of conditions and interventions. This article, first in a 3-part series, discusses how this goal affects the conduct and use of analyses. The remaining articles will outline methodological and reporting recommendations, respectively.

Participants.—The Panel on Cost-Effectiveness in Health and Medicine, a nonfederal panel with expertise in CEA, clinical medicine, ethics, and health outcomes measurement, was convened by the US Public Health Service (PHS).

Evidence.—The panel reviewed the theoretical foundations of CEA, current practices, and alternative procedures for measuring and assigning values to resource use and health outcomes.

Consensus Process.—The panel met 11 times during 2½ years with PHS staff and methodologists from federal agencies. Working groups brought issues and preliminary recommendations to the full panel for discussion. Draft recommendations were circulated to outside experts and the federal agencies prior to finalization.

Conclusions.—The panel's recommendations define a "reference case" cost-effectiveness analysis, a standard set of methods to serve as a point of comparison across studies. The reference case analysis is conducted from the societal perspective and accounts for benefits, harms, and costs to all parties. Although CEA does not reflect every element of importance in health care decisions, the information it provides is critical to informing decisions about the allocation of health care resources.

JAMA. 1996;276:1172-1177

From the Institute for Health, Health Care Policy, and Aging Research, Rutgers University, New Brunswick, NJ (Dr Russell); Office of Disease Prevention and Health Promotion, US Public Health Service, Washington, DC (Dr Gold); School of Public Health, Harvard University, Cambridge, Mass (Drs Siegel and Weinstein); and the Department of Philosophy, Tufts University, Medford, Mass (Dr Daniels).

A complete list of the Panel on Cost-Effectiveness in Health and Medicine membership and staff appears at the end of this article.

Corresponding author: Marthe Gold, MD, MPH, Office of Disease Prevention and Health Promotion, 200 Independence Ave SW, Room 738G, Washington, DC 20201.

THE HEIGHTENED awareness in recent years of the need to live within budgets in the health care sector would seem to create the perfect climate for cost-effectiveness analysis (CEA). Cost-effectiveness analysis is a method for evaluating the health outcomes and resource costs of health interventions. Its central function is to show the relative value of alternative interventions for improving health. Analyses provide information that can help decision makers in a variety of settings weigh alternatives and decide which best serve their programmatic needs. Yet CEA is rarely used to inform decisions about health services in the United States.

The case of Oregon illustrates the problems that arise when policymakers attempt to use CEA. Faced with providing medical care to its low-income population within the constraints of its Medicaid budget, the Oregon Health Services Commission initially tried to set priorities for covered services using CEA. In 1990 to 1991, the commission generated a list of "condition-treatment" pairs ordered by their cost-effectiveness ratios. The list was withdrawn after public criticism of its counterintuitive ranking of some interventions.

For editorial comment see p 1180.

Oregon's effort provoked criticism at the time and has continued to be the subject of debate about the role of CEA.[1-5] Some have observed that cost-effectiveness ratios do not adequately reflect important issues, such as distributive justice and competing values outside of health. The commission subsequently adopted a process that included cost-effectiveness as 1 of 13 factors on which the ranking of services was based, including equity, "benefits many," and community compassion. The role of CEA was to supplement these qualitative factors by providing standardized, quantitative estimates of the likely cost per unit of health benefit for each intervention.

But, CEA presents problems even in

this more limited role. The method can be difficult to follow, and results are often presented in a way that impedes rather than facilitates understanding. Studies vary widely in the health effects and costs included and in the way these are valued and combined, so that studies of the same intervention can produce very different cost-effectiveness ratios; potential users may be confused and suspicious that CEA can be manipulated to support almost any conclusion.[6-8] Finally, although the CEA literature has grown enormously in recent years,[9] many interventions have not been evaluated.

In 1993, the US Public Health Service (PHS) convened the Panel on Cost-Effectiveness in Health and Medicine to address these problems. The panel, a nonfederal group of 13 scientists and scholars with expertise in CEA, clinical medicine, ethics, and health outcomes measurement, met 11 times during 2½ years to review the state of the field and develop recommendations to improve the quality and comparability of studies. Comparability is essential if CEA is to help decision makers evaluate tradeoffs and choose among alternatives. The panel was charged with developing recommendations that would provide a framework for consistent practice across conditions and interventions—preventive, therapeutic, rehabilitative, and public health.

The panel's focus on policy decisions and resource allocation at a broad level reflected the increasing attention directed to CEA by federal health agencies. The Centers for Disease Control and Prevention, in addition to their own studies, have developed courses and guides to train staff and are working with states interested in CEA. The National Institutes of Health are collecting cost data in addition to outcome data in a few trials. The Agency for Health Care Policy and Research has begun to use CEAs in developing practice guidelines and in the technology assessments it conducts for the Health Care Financing Administration. The Food and Drug Administration has been involved in an intense review of its regulatory role with respect to the marketing of drugs on the basis of cost-effectiveness claims.

In a survey conducted before the first meeting, panel members identified methodologically challenging or contentious areas: the fundamental foundation for CEA; how best to frame an analysis (plan the approach to the analysis and outline the study's main elements); the perspective of an analysis; types of costs and data sources for costs; measuring and valuing outcomes; components of cost-effectiveness ratios (what should go in the numerator and the denominator); time frame for analysis; discounting; ethical issues, especially distributive implications; appropriate comparator programs; statistical issues; and the overriding issue of standardization of methods.

The panel organized itself into 9 working groups: the role of CEA; theoretical foundations; the framing of analyses; measuring effectiveness; valuing health outcomes; measuring costs; discounting; evaluating uncertainty; and reporting. At meetings, which included PHS staff and representatives of other federal agencies, the working groups developed issues for panel discussion. Panel members and staff then drafted chapters on each subject, outlining areas of agreement and disagreement and proposing recommendations. Recommendations were debated until consensus was reached, or, in a few cases, until it was apparent it would not be. Chapters were revised to set out the arguments supporting (or preventing) a consensus-based recommendation. Throughout the process, experts from federal agencies and the academic community critically reviewed and helped shape the work.

This article is the first of 3 that summarize the panel's discussions and recommendations, which are presented more completely in its full report.[10] Here we introduce the reference case, the panel's proposed mechanism for improving comparability. The panel's work is aimed at both analysts and users of CEAs. For analysts, the recommendations describe why comparability is important and how to achieve it. For users, they offer a guide to the evaluation and use of CEAs and should ultimately make CEA easier to use and more useful.

THE REFERENCE CASE

Cost-effectiveness analysis is a method for evaluating the outcomes and costs of interventions designed to improve health. The results are usually summarized in cost-effectiveness ratios that demonstrate the cost of achieving a unit of health effect (eg, the cost per year of life gained) for diverse types of patients and for variations of the intervention. In a cost-effectiveness ratio, changes in health due to an intervention, compared with a specific alternative, are captured in the denominator; and changes in resource use, compared with the same alternative and valued in monetary terms, are captured in the numerator.

It is common practice to define a base case that incorporates data and methods the analyst thinks best represent the interventions and choices under consideration. The analyst then conducts a series of sensitivity analyses using different data or methods to test the robustness of the results. Base-case results and selected sensitivity analyses are presented. Within a given study, alternatives must be analyzed in comparable fashion so that the cost-effectiveness ratios reflect true differences in costs and health effects.

No single study can provide all the information needed to compare health services across a broad range of conditions and interventions. But, if individual studies adhered to a common standard, they could collectively provide the necessary results and do so more efficiently than a comprehensive analysis undertaken specifically for the purpose. Thus, comparability across studies is crucial for evaluating the broad allocation of health care resources.

To promote comparability of CEAs while leaving analysts free to address issues specific to a particular problem, the panel proposes that studies include, either as the base case or in addition to it, a reference case. The reference case is defined by a standard set of methods and assumptions. It includes a set of standard results: the reference case results. While an investigator might also present results based on different methods and assumptions to serve the other purposes of the analysis, the reference case serves as a point of comparison across studies. It should be included whenever the CEA is intended to contribute to decisions about the broad allocation of health care resources.

To build the reference case on a solid foundation, the panel reviewed the theoretical roots and practical applications of CEA. The goal was to develop complete, consistent, and theoretically grounded recommendations that were sufficiently tractable and detailed to provide practical guidance for analysts. The second and third articles in this series summarize the recommendations that define the reference case and the reasoning behind them.[11,12]

Use of the reference case would address the problems with CEA in 3 ways. First, by setting standards for the costs and health effects that should be included and the ways in which they should be valued, the reference case offers analysts and users a benchmark that allows them to evaluate the quality of a study and determine whether its results can be compared with other studies. Second, the reference case includes recommendations for reporting results, described in the third article in this series, designed to make it easier for users to see what was done and how the results compare with those of other studies. Third, as analysts begin to include the reference case in their results, they will contribute to a growing pool of studies that can be compared.

The need to standardize CEA has been recognized for at least a decade,[12-16] but the resources to tackle this difficult task with sufficient time to work through the issues and develop appropriate recommendations had not previously been available in the United States. Earlier attempts at standards were very general, apparently in the belief that 1 set of standards could serve all analyses. The panel recognized that recommendations must be tailored to the kind of decision analyses are intended to inform; as noted, it focused on decisions that involve evaluating a broad range of interventions that can apply to widely different kinds of people and conditions. The panel's recommendations for the reference case are much more comprehensive and detailed than previous efforts in order to provide guidance on all issues that determine comparability across studies. Founded on state-of-the-art thinking in the field, they represent a reasonable consensus that can move the field forward.

PERSPECTIVE

When health maintenance organizations (HMOs) or government programs evaluate an intervention, they consider the costs they will experience in providing or paying for it. Costs incurred by patients or others, such as for outpatient medication or home care after hospital discharge, may be irrelevant from their perspective. They may also disregard some outcomes. For example, it may matter little to the HMO or government program how soon patients return to work after an illness, although it may matter a great deal to individuals, their employers, or the government agency responsible for disability payments.

While the use of a particular perspective (eg, HMO, employer, government program, or individual) is appropriate for informing decisions from that perspective, studies based on different perspectives are not comparable. The perspective for the reference case is of overarching importance because it reflects the type of decisions the analysis is intended to inform and determines which costs and health effects go in the cost-effectiveness ratio and how the costs and effects are valued. To serve the goal of facilitating comparisons across interventions and patient groups, the panel recommends the societal perspective for the reference case.

In a CEA conducted from the societal perspective, the analyst considers everyone affected by the intervention, and all health effects and costs that flow from it are counted, regardless of who would experience them. Health effects include both benefits and harms, even when these occur in people who are not the intended recipients of the intervention. Resource costs include all resources used, whether or not money changes hands. Using CEA to inform allocation of health resources accepts the existence of a limit on health spending; when the societal perspective is adopted, the analysis acknowledges the value of competing uses for society's resources.

Programs to reduce the incidence of neural tube defects by increasing the folic acid intake of pregnant women demonstrate the breadth of the societal perspective. Adding folic acid to cereal grains would allow the improvement of women's nutritional status without effort and would be helpful for those with inadequate access to medical care and, thus, to counseling about nutrition. Everyone who buys cereal-grain products pays the cost. But fortification puts older people at risk because it masks pernicious anemia, which, untreated, can cause neurological problems. A CEA conducted from the societal perspective would include the harms to the elderly as well as the benefits to infants and all related costs.

Because the societal perspective includes all costs and health effects, it does not necessarily show employers, HMOs, insurers, government programs, or individuals exactly what they want to know to make choices best suited to their interests. An analysis might suggest that, from the societal viewpoint, it would be better to cover exercise programs for older people than coronary artery bypass surgery for risk groups for which effectiveness of bypass surgery is very low and cost per year of life gained very high. If individuals in those risk groups and their clinicians were to evaluate the same interventions strictly from the patients' perspective, they would count only costs patients would pay out-of-pocket, ignoring substantial sums paid by insurers or public programs. Bypass surgery might appear desirable from this perspective even if its benefits were exceedingly small.

The societal perspective represents the public interest rather than that of any group. It is compatible with the traditional principle that decisions affecting people with differing interests are more likely to be fair if they are made by those who will not gain or lose from them. Many philosophers, operating from diverse perspectives, have suggested a thought experiment to show why individuals and groups might accept the societal perspective even when it does not perfectly represent their interests. The experiment is to imagine that we are viewing the world before our birth (*ex ante*) and to ask what type of world we would like it to be.[17-24] From that vantage point, we would not yet know which health problems we would experience, only that there was some possibility that we might develop any of them, and we might well then prefer a system in which decisions about health interventions reflected the seriousness of the condition and the ability of alternative interventions to improve it without reference to individuals, budgets, or special interests. We would not wish to have any health problem neglected entirely because that neglect would affect us if we developed the problem. And we would want areas other than health care to receive resources so that our other needs and aspirations could be met. The panel's choice of the societal perspective is based on this ethical framework.

The societal perspective is also a pragmatic choice, exactly because it does not represent the viewpoint of any particular group. Instead, it provides a benchmark against which to assess results from other perspectives. Only the societal perspective never counts as a gain what is another party's loss. If an employer adopts an intervention that reduces the employer's health insurance costs but increases costs for Medicare, or if a public health intervention improves the health of 1 group but causes unwanted side effects for another, the societal perspective includes both changes. No perspective has a stronger claim to be the basis for comparability across studies.

EXAMPLE OF IMPLICATIONS OF THE SOCIETAL PERSPECTIVE: VALUING HEALTH EFFECTS

For the reference case, the measure of health effect must be comparable across interventions and conditions and capable of capturing the impact of interventions with different effects. Life-years gained, often used in CEAs done from the societal perspective, are an important metric, but give little credit to interventions that primarily improve quality of life (eg, cataract surgery) and fail to account for adverse effects. Over the past 2 decades, investigators have developed measurement strategies that permit calculation of the health-related quality of life (HRQL) associated with conditions or interventions. These HRQL measures classify people into health states defined on a continuum, from least desirable to most desirable, in terms of some or all of the following: physical function, psychological function, social role function, perceptions of health, and symptoms. A number of these measurement systems are preference based and capture people's values for states of health. In general, health states are scaled from 0 (dead) to 1 (optimal health); however,

states worse than death can be accounted for by assigning them negative scores.[25]

Some methods for collecting preferences involve asking patients or members of the public to locate their preferences for health states directly on the 0 to 1 continuum using techniques such as standard gamble, time trade-off, and category rating.[26,27] Another approach measures preferences indirectly, relying on health classification systems such as the Health Utilities Index,[28] the Quality of Well-Being Scale,[29] or the EuroQol[30] that include premeasured preferences for defined health states. These premeasured values come in part from direct measurement of some preferences (using standard gamble, time trade-off, or category rating) and in part from application to those preferences of multiattribute utility theory[31] or statistical inference[32] to fill in values not measured directly. In health classification systems, premeasured values are based on results of community surveys.

Although assessments of criterion validity, where one measure is considered the gold standard against which all others are judged, are impossible since, by definition, there is no set of preferences that is correct for all people, many of these methods have shown good reliability and sensitivity to changes in clinical conditions, and convergent validity between methods.[26] These techniques are reviewed in the panel's report.[10]

A preference-based system accomplishes 2 important tasks for CEA. First, it makes it possible to combine length of time health states are experienced with quality of that time to create a summary measure: quality-adjusted life-years (QALYs). Quality-adjusted life-years provide a common metric for recording effects of different interventions.[27,33-36] Second, since the purpose of investing in health is to make people better-off, it seems appropriate to let them be the judge of what constitutes better or worse outcomes and of the relative magnitudes of health effects. The welfare economic foundations of CEA, which assign primacy to individual preferences, are compatible with this view.

For these and other reasons described in its full report, the panel recommended QALYs as the measure of health effect for the reference case. Given this decision, a difficult issue arises: whose preferences should be used for the reference case?

The choice is between patient preferences and those of a representative community sample. Patient preferences are values that people experiencing a condition assign to their own health. Community preferences are values assigned by representatives of the general population, which contains people with disabilities or chronic illnesses in proportion to the prevalence of their condition. While some studies support the idea that both groups assign similar values to the same states,[26,30,37-41] others suggest that people experiencing a condition attach higher values to the associated health states.[42-45] Analysts have used patient[46] as well as community preferences[32,33] to calculate QALYs for use in CEAs, but to date there are no studies comparing the impact of using one or the other in the same analysis.

Many investigators use patient preferences because they believe that people experiencing the health states have the most accurate appreciation of their conditions, that it is ethically appropriate to solicit information from those directly affected when evaluating interventions for a condition, and that community preferences discriminate against people who are disabled or ill. In a CEA designed to allow individualization in the choice between 2 medications for a particular illness such as arthritis, where subtle side effects might be important to capture, patient preferences are appropriate. For the reference case, however, a logical extension of the societal perspective and the *ex ante* position embedded within it is that the best articulation of society's preferences comes from a representative community sample.

However, aware of the issues of discrimination raised by Oregon's use of community preferences,[5] the panel carefully examined the implications of endorsing community preferences for reference case analyses because of the possibility that those with disabilities or chronic illnesses assign higher scores to their health states.

Consider an intervention that can be used for many people, some of whom may have disabilities—suppose coronary artery bypass graft surgery is targeted to everyone with ischemic chest pain—and suppose that persons who are paraplegic score a health state that includes wheelchair-dependent mobility higher than does the general public. To calculate QALYs using community preferences, an investigator would use the scores for states with and without chest pain assigned by a representative sample of the community. All who undergo coronary artery bypass graft surgery, including paraplegic persons, will be credited with this gain. Thus, the cost-effectiveness of interventions directed at conditions unrelated to the disability would be the same for the disabled and the general population, as long as subgroups defined according to comorbidities are not analyzed separately.

Subgroups are, however, often analyzed separately when the evidence about health outcomes or cost suggest significant differences. Using community weights for outcomes in these cases might assign fewer QALYs to a subgroup than if the subgroup's own preferences were used. The panel recommended that when there are important differences in preferences among subgroups, analysts should conduct sensitivity analyses to show the impact of differences.

When cost-effectiveness of treatment to cure or prevent paraplegia is evaluated, community preferences will always yield as many QALYs, or more, than calculations based on patient preferences. If a state with limited mobility is rated lower by the general public than by persons in that state, more QALYs are gained by relieving it, and the cost-effectiveness ratio for the intervention is more favorable. However, for the occasional case of a lifesaving intervention aimed specifically at the disabled or ill, use of community preferences when preferences differ assigns fewer QALYs to the intervention than would use of patient preferences; this is because the difference between the community's value for the state and 0 (death) is less than the difference for the group itself.

In the great majority of instances, the panel found that use of community preferences in the reference case would not discriminate against the disabled. Where it might, the panel recommended sensitivity analyses be used to explore the influence of community vs patient preferences on the cost-effectiveness ratio, to let decision makers know whether that influence was significant. As noted, the *ex ante* position provides compelling theoretical reasons for using community preferences in an analysis done from the societal perspective. Also, although persons with disability and illness may adjust successfully to their conditions, a more compelling goal from the societal perspective is to avoid disability and promote full function in all domains of health. Finally, from a practical standpoint, standardizing practice requires that the source of preferences be consistent. For all these reasons, the panel endorsed the community as the source of preferences for the reference case.

It bears repeating here that reference case analyses are intended to inform decisions at the level of broad resource allocation and may provide little guidance about optimal bedside management of individuals. There the preferences and conditions of individual patients may point to decisions different from those supported by a reference case CEA.

CEA AS AN AID TO DECISION MAKING

Choices involve ethical issues, and the choices made in defining the reference case express certain ethical points of view. By counting all costs and health effects, the societal perspective reflects the public interest, not the interest of any group in society. Using the preferences of a community sample to value health states incorporates community values into the decision-making process. Other choices with ethical implications, such as the discount rate or use of wages to value people's time, are discussed in the second article. The panel recognized these implications and that some issues are left unresolved by these choices and by the methods available to implement them.

Quality-adjusted life-years provide another noteworthy example. Summing them involves the assumption that all QALYs are equal, no matter who gains them or when during the life span. This implies that it makes no difference whether QALYs benefit people in good or poor health and that 2 therapies that produce equal numbers of QALYs—one perhaps in the form of small benefits for many people, the other in the form of large benefits for a few—are equal in value. Intuition and research suggest that deviations from this assumption can be important.[47,48] Decision makers (and the general public) might wish to give preference to those in poor health, because of their greater need, or to the intervention that provides large benefits for a few, exactly because it made such a large difference for those few. Thus, although QALYs have the advantage that they count changes in quality as well as quantity of life, they do not as currently defined, and perhaps never can, perfectly reflect everything about health that matters to people.

Values outside of health care, which often influence choices about health services, cannot be quantified in CEA. As an example, individuals' right to privacy has blocked compulsory testing for the human immunodeficiency virus except in special situations like the military, although diagnosis and treatment can be delayed.

Some nonhealth benefits or harms could be captured by cost-benefit analysis (CBA), which values all effects in monetary terms. The panel accepted the majority view in the health care sector that monetizing health introduces ethical issues avoided by use of health-specific measures. However, in its full report, the panel built bridges to CBA and acknowledged that opportunities to compare interventions can be lost by relying solely on CEA. It noted that CBA, CEA, and cost-consequence analysis (in which costs and effects are calculated, but not aggregated into QALYs or cost-effectiveness ratios) are complementary forms of analysis; the use of one does not preclude the use of others in a study.

No method of making decisions about health care resources allocation provides a complete procedure for resolving ethical issues. Whether decision makers use criteria like medical necessity, expected benefit, standards of evidence, CEA, or CBA, the issues of fairness, feasibility, and values are not completely captured by the analysis and must be weighed against factors that are. A CEA, however, offers more complete information than these other methods about the size and composition of health effects and costs.

That CEA does not reflect all trade-offs of potential relevance has implications for how it should be used in decision making. The methodology of CEA is constructed to serve a straightforward goal: to identify interventions that produce the most health with the resources available. In the textbook explanation of the method, once cost-effectiveness ratios have been computed using a comparable method, the decision maker ranks interventions from lowest to highest cost per QALY and, starting with the lowest cost-effectiveness ratio, selects interventions in order until the budget has been spent. The chosen interventions produce the greatest number of QALYs possible with that budget.

But real-world decisions must balance health against other goals—fair access to services, help for those worst off, and values outside health affected by health decisions. Thus, it is seldom appropriate to apply CEA mechanically. The panel recommended that CEA be used as an aid to decision makers who must weigh the information it provides in the context of these other values.

The best CEAs available suggest that the current allocation of health care resources falls well short of producing the most health possible.[49-51] Some interventions are applied extensively, in ways that produce little health for large expenditures, while other interventions do not receive enough resources. Use of CEA in the decision-making process could contribute to improvements in the effective use of resources.

To allow decision makers to evaluate trade-offs carefully, CEAs should present not only cost-effectiveness ratios, but also background information about the elements that make up costs and effects: kinds and magnitudes of costs, who is helped or harmed, and how much they are affected. The detailed information can help decision makers evaluate the trade-offs between those elements of the decision that are well captured by CEA and those that are not.

Comparability is the foundation on which the usefulness of a method for choosing among alternatives must be built. Differences in cost-effectiveness should reflect true differences among interventions, not unnecessary differences in method. The panel achieved consensus on a detailed set of recommendations designed to promote comparability of CEAs. We believe these recommendations can do much to overcome problems that have interfered with past use of CEA.

The Panel on Cost-Effectiveness in Health and Medicine membership and staff: Norman Daniels, PhD; Dennis G. Fryback, PhD; Alan M. Garber, MD, PhD; Marthe R. Gold, MD, MPH; David C. Hadorn, MD; Mark S. Kamlet, PhD; Joseph Lipscomb, PhD; Bryan R. Luce, PhD; Jeanne S. Mandelblatt, MD, MPH; Willard G. Manning, Jr, PhD; Donald L. Patrick, PhD; Louise B. Russell, PhD; Joanna E. Siegel, ScD; George W. Torrance, PhD; Milton C. Weinstein, PhD.

We thank Kristine I. McCoy, MPH, for coordination, research, and editorial assistance in association with this project.

References

1. Hadorn DC. Setting health care priorities in Oregon: cost-effectiveness meets the Rule of Rescue. *JAMA.* 1991;265:2218-2225.
2. Eddy DM. Oregon's methods: did cost-effectiveness analysis fail? *JAMA.* 1991;266:2135-2141.
3. Klevit HD, Bates AC, Castanares T, Kirk EP, Sipes-Metzler PR, Wopat R. Prioritization of health care services: a progress report by the Oregon Health Services Commission. *Arch Intern Med.* 1991;151:912-916.
4. Brown LD. The national politics of Oregon's rationing plan. *Health Aff (Millwood).* 1991;10:29-51.
5. Office of Technology Assessment. *Evaluation of the Oregon Medicaid Proposal.* Washington, DC: US Congress; 1991. Office of Technology Assessment publication OTA-H-531.
6. Kassirer JP, Angell M. The Journal's policy on cost-effectiveness analyses. *N Engl J Med.* 1994; 331:669-670.
7. Evans RG. Manufacturing consensus, marketing truth: guidelines for economic evaluation. *Ann Intern Med.* 1995;123:59-60.
8. Brown ML, Fintor L. Cost-effectiveness of breast cancer screening: preliminary results of a systematic review of the literature. *Breast Cancer Res Treat.* 1993;25:113-118.
9. Elixhauser A, Luce BR, Taylor WR, Reblando J. Health care CBA/CEA: an update on the growth and composition of the literature. *Med Care.* 1993; 31(suppl):JS1-JS11.
10. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine.* New York, NY: Oxford University Press; 1996.
11. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB, for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA.* In press.
12. Siegel JE, Weinstein MC, Russell LB, Gold MR, for the Panel on Cost-Effectiveness in Health and Medicine. Reporting cost-effectiveness analyses. *JAMA.* In press.
13. Russell LB. *Is Prevention Better Than Cure?* Washington, DC: Brookings Institution; 1986.
14. Task Force on Principles for Economic Analysis of Health Care Technology. Economic analysis

of health care technology: a report on principles. *Ann Intern Med.* 1995;123:61-70.
15. Henry C. Economic analysis as an aid to subsidisation decisions: the development of Australia's guidelines for pharmaceuticals. *PharmacoEconomics.* 1992;1:54-67.
16. Canadian Coordinating Office of Health Technology Assessment (CCOHTA). *Guidelines for Economic Evaluation of Pharmaceuticals.* Ottawa, Ontario: Canadian Coordinating Office of Health Technology Assessment; 1994.
17. Harsanyi JC. Cardinal utility in welfare economics and in the theory of risk-taking. *J Polit Econ.* 1953;61:434-435.
18. Harsanyi JC. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J Polit Econ.* 1955;63:309-321.
19. Rawls J. *A Theory of Justice.* Cambridge, Mass: Harvard University Press; 1971.
20. Daniels N. *Just Health Care.* Cambridge, Mass: Harvard University Press; 1985.
21. Daniels N. *Am I My Parent's Keeper? An Essay on Justice Between the Young and the Old.* New York, NY: Oxford University Press; 1988.
22. Menzel PT. *Strong Medicine.* New York, NY: Oxford University Press; 1992.
23. Dworkin R. What is equity? 2: equality of resources. *Philos Public Aff.* 1981;10:283-345.
24. Eddy DM. The individual vs society: resolving the conflict. *JAMA.* 1991;265:2405-2406.
25. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. *Med Decis Making.* 1994; 14:9-18.
26. Froberg DG, Kane RL. Methodology for measuring health-state preferences, III: population and context effects. *J Clin Epidemiol.* 1989;42:585-592.
27. Patrick DL, Erickson P. *Health Status and Health Policy: Allocating Resources to Health Care.* New York, NY: Oxford University Press; 1993.
28. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes.* New York, NY: Oxford University Press; 1987.
29. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res.* 1988;23:203-235.
30. The EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16:199-208.
31. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Operations Res.* 1982;30:1043-1069.
32. Kaplan RM, Bush JW, Berry CC. The reliability, stability, and generalizability of a health status index. In: *Proceedings of the Annual Meeting of the Social Statistics Section, 704-9; San Diego, Calif.* Washington, DC: American Statistical Association; 1978.
33. Boyle MH, Torrance GW, Sinclair JC, Horwood SP. Economic evaluation of neonatal-intensive care of very-low-birth-weight infants. *N Engl J Med.* 1983;308:1330-1337.
34. Oldridge N, Furlong W, Feeny D, et al. Economic evaluation of cardiac rehabilitation soon after acute myocardial infarction. *Am J Cardiol.* 1993; 72:154-161.
35. Spilker B. *Quality of Life Assessment in Clinical Trials.* New York, NY: Raven Press; 1990.
36. Weinstein MC, Stason WB. *Hypertension: A Policy Perspective.* Cambridge, Mass: Harvard University Press; 1976.
37. Balaban DJ, Sagi PC, Goldfarb NI, Nettler S. Weights for scoring the quality of well-being instrument among rheumatoid arthritics: a comparison to the general population weights. *Med Care.* 1986;24:973-980.
38. Llewellyn-Thomas HA, Sutherland JH, Tibshirana R, Ciampi A, Till JE, Boyd NF. Describing health states: methodologic issues in obtaining values for health states. *Med Care.* 1984;22:543-552.
39. Boyle MH, Torrance GW. Developing multiattribute health indexes. *Med Care.* 1984;22:1045-1057.
40. Patrick DL, Sittampalam Y, Somerville S, Carter W, Bergner M. A cross-cultural comparison of health status values. *Am J Public Health.* 1985; 75:1402-1407.
41. Hadorn DC, Uebersax J. Large-scale outcome evaluation: how should quality of life be measured? I: calibration of a brief questionnaire and a search for preference subgroups. *J Clin Epidemiol.* 1995; 48:607-618.
42. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. *J Chronic Dis.* 1978;31:697-704.
43. Najman J, Levine S. Evaluating the impact of medical care and technologies on the quality of life: a review and critique. *Soc Sci Med.* 1981;15: 107-115.
44. Epstein AM, Hall JA, Tognetti J, Son JH, Conant L. Using proxies to evaluate quality of life. *Med Care.* 1989;27(suppl):S91-S98.
45. Slevin ML, Stubbs L, Plant HJ, et al. Attitude to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. *BMJ.* 1990;300:1458-1460.
46. Nease RF Jr, Kneeland T, O'Connor GT, et al, for the Ischemic Heart Disease Patient Outcomes Research Team. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical practice guidelines. *JAMA.* 1995;273:1185-1190.
47. Harris J. QALYfying the value of life. *J Med Ethics.* 1987;13:117-123.
48. Daniels N. Rationing fairly: programmatic considerations. *Bioethics.* 1993;7:223-233.
49. Russell LB. Too much for too few: what cost-effectiveness tells us about prevention. In: Skelton WD, Osterweis M, eds. *Promoting Community Health: The Role of the Academic Health Center.* Washington, DC: Association of Academic Health Centers; 1993.
50. Russell LB. *Educated Guesses: Making Policy About Medical Screening Tests.* Berkeley, Calif: University of California Press and Milbank Memorial Fund; 1994.
51. Russell LB, Siegel JE, Daniels N, Gold MR, Luce BR, Mandelblatt JS. Cost-effectiveness analysis as a guide to resource allocation in health: roles and limitations. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-effectiveness in Health and Medicine.* New York, NY: Oxford University Press; 1996.

Printed and Published in the United States of America

Reprinted from JAMA © The Journal of the American Medical Association October 23/30, 1996 Volume 276

Consensus Statement

Recommendations for Reporting Cost-effectiveness Analyses

Joanna E. Siegel, ScD; Milton C. Weinstein, PhD; Louise B. Russell, PhD; Marthe R. Gold, MD, MPH; for the Panel on Cost-Effectiveness in Health and Medicine

Objective.—This article, the third in a 3-part series, describes recommendations for the reporting of cost-effective analyses (CEAs) intended to improve the quality and accessibility of CEA reports.

Participants.—The Panel on Cost-Effectiveness in Health and Medicine, a nonfederal panel with expertise in CEA, clinical medicine, ethics, and health outcomes measurement, convened by the US Public Health Service.

Evidence.—The panel reviewed the theoretical foundations of CEA, current practices, alternative methods, published critiques of CEAs, and criticisms of general CEA methods and reporting practices.

Consensus Process.—The panel developed recommendations through 2½ years of discussions. Comments on preliminary drafts were solicited from federal government methodologists, health agency officials, and academic methodologists.

Conclusions.—These recommendations are proposed to enhance the transparency of study methods, assist analysts in providing complete information, and facilitate the presentation of comparable cost-effectiveness results across studies. Adherence to reporting conventions and attention to providing information required to understand and interpret study results will improve the relevance and accessibility of CEAs.

JAMA. 1996;276:1339-1341

DECISION MAKERS who use cost-effectiveness analyses (CEAs) to inform decisions for a range of interventions may consult multiple studies. Standardization of methods is essential for valid comparisons of cost-effectiveness (C/E) ratios. As described,[1-3] the Panel on Cost-Effectiveness in Health and Medicine was convened to recommend standards for CEA. To serve as the point of comparison across studies, a standard set of assumptions and practices referred to as the reference case was developed by the panel.

For optimal use, CEA results must also be reported in a standard way. Differences among reports complicate users' ability to find, interpret, and adapt information. Studies of published CEAs have found widespread inadequate reporting.[4-6] Previous direction on reporting has been too general to provide adequate guidance,[7-11] and recent guidelines targeting subsets of CEAs are not intended for the majority of analyses.[12-14] Our recommendations for standard reporting of reference case analyses are presented herein.

From the Departments of Maternal and Child Health (Dr Siegel) and Health Policy and Management (Dr Weinstein), Harvard School of Public Health, Boston, Mass; the Institute for Health, Health Care Policy, and Aging Research, Rutgers University, New Brunswick, NJ (Dr Russell); and the Office of Disease Prevention and Health Promotion, US Public Health Service, Washington, DC (Dr Gold).

A complete list of the Panel on Cost-Effectiveness in Health and Medicine membership and staff appears at the end of this article.

Corresponding author: Marthe Gold, MD, MPH, Office of Disease Prevention and Health Promotion, 200 Independence Ave SW, Room 738G, Washington, DC 20201.

THE REFERENCE CASE ANALYSIS AND THE JOURNAL REPORT

Reports of cost-effectiveness should allow determination of whether the results can be juxtaposed with those of other CEAs. Elements to include in a journal report are discussed below and summarized in the checklist (Table).

Framework of the CEA

The reporting of study framework explains the motivation for the research, including research objectives, and outlines the study design.[3] An explicit statement of the analysis perspective, which, for the reference case will be societal, is essential. The quality and appropriateness of the results depend on whether costs and outcomes are derived consistently from the stated viewpoint, yet many studies (75%-82%) fail to identify perspective.[4-6]

To assess an intervention's impact, CEA describes and contrasts costs and outcomes of a course of events expected to occur with the intervention and costs and outcomes of a comparator course of events without the intervention. Program elements of interventions and comparators that will define the analysis should be outlined, such as site, target population, and frequency of an intervention. Descriptors of the target population may include demographic, socioeconomic, behavioral, and clinical characteristics.

Analysis boundaries should be described, explaining the extent to which relevant benefits and harms are included. For example, does an analysis of a smoking-cessation program for pregnant women include effects on eventual health and health care utilization of the fetus? On that of siblings? The time frame

Checklist for Reporting the Reference Case Cost-effectiveness Analysis*

- **Framework**
  - Background of the problem
  - General framing and design of the analysis
  - Target population for intervention
  - Other program descriptors (eg, care setting, model of delivery, timing of intervention)
  - Description of comparator programs
  - Boundaries of the analysis
  - Time horizon
  - Statement of the perspective of the analysis
- **Data and Methods**
  - Description of event pathway
  - Identification of outcomes of interest in analysis
  - Description of model used
  - Modeling assumptions
  - Diagram of event pathway (model)
  - Software used
  - Complete description of estimates of effectiveness, resource use, unit costs, health states, and quality-of-life weights and their sources
  - Methods for obtaining estimates of effectiveness, costs, and preferences
  - Critique of data quality
  - Statement of year of costs
  - Statement of method used to adjust costs for inflation
  - Statement of type of currency
  - Source and methods for obtaining expert judgment
  - Statement of discount rates
- **Results**
  - Results of model validation
  - Reference case results (discounted at 3% and undiscounted): total costs and effectiveness, incremental costs and effectiveness, and incremental cost-effectiveness ratios
  - Results of sensitivity analyses
  - Other estimates of uncertainty, if available
  - Graphical representation of cost-effectiveness results
  - Aggregate cost and effectiveness information
  - Disaggregated results, as relevant
  - Secondary analyses using 5% discount rate
  - Other secondary analyses, as relevant
- **Discussion**
  - Summary of reference case results
  - Summary of sensitivity of results to assumptions and uncertainties in the analysis
  - Discussion of analysis assumptions having important ethical implications
  - Limitations of the study
  - Relevance of study results for specific policy questions or decisions
  - Results of related cost-effectiveness analyses
  - Distributive implications of an intervention

*Modified from Gold et al.[1]

should be indicated and, for the reference case, should be long enough to capture all significant benefits, harms, and costs.

Data and Methods Section

A description of the event pathway, which details progression of the disease, events associated with the intervention, and disease- and treatment-related events following intervention, provides a useful overview. Typically, CEAs rely on a model that may be used to integrate secondary data, extend an analysis beyond primary data, or define a pattern from data. The models used should be described, including supporting assumptions.

Details about estimates of effectiveness, costs, and preference weights in the analysis should be provided. A review of CEAs using clinical trial data found that few contained enough information concerning trials to form an opinion about their adequacy,[5] eg, only about a third reported participation and drop out rates, comparability of treatment and nontreatment groups, and whether the analysis was based on results from those entering or completing the trial.

The health states used for valuing health-related quality of life should be identified, listing the states if a published system (eg, Health Utilities Index) is not used. The population from which preference weights were obtained and the method for eliciting weights (eg, time trade-off, rating scale) should be identified. For cost data, method for inflation adjustment, type of currency, and year of costs should be indicated. Sources of data on health care utilization (physical units) and unit costs, and methods for measuring and valuing time costs are also important.

Controversies associated with data used in the analysis should be discussed, including disputes regarding effectiveness or published discrepancies in estimates of cost, effectiveness, or health outcomes. Related assumptions in the analysis should be justified.

The methods section should also indicate the discount rate used for costs and health effects and describe sensitivity analyses performed.

The Results Section

If a simulation model has been used, validation tests of the model are generally reported first. These tests demonstrate that the model produces verifiable information, increasing confidence in results that are not directly verifiable.

The results of the CEA follow, including results of the base-case analysis—the analysis using the data and methods the analyst thinks best characterize the choices under consideration—and sensitivity analyses. If the primary purpose of the analysis is to inform broad resource allocation decisions, the base case will be the reference case analysis.[1] If serving a different purpose, the analysis should report reference case results in addition to those of the base case. Reference case results should be clearly identified.

A basic set of reference case results includes the following: total costs and effectiveness, incremental costs and effectiveness, and incremental cost-effectiveness (C/E) ratios, with discount rates of 3% (reference case) and 5%, the latter for comparison with past analyses. The totals give a sense of the magnitudes of cost and effect. Providing intermediate steps (incremental costs and effects) allows verification of methods. Undiscounted results are also often of interest and should be reported. Presentation of results in tables facilitates identification of findings.

The basic results should be supported by discussion and analytical detail on components of costs and effects. We emphasize that CEA is an aid to decision making rather than a decision-making process; information about components can be important in decisions. The reference case outcome measure is quality-adjusted life-years (QALYs), but reporting number of life-years gained (unadjusted for quality of life) in addition is usually warranted, as it distinguishes lifesaving from quality-enhancing effects and allows consideration of alternative quality-of-life schemes. The benefits, harms, and costs to different groups should be presented for evaluation of distributive issues.

Of note, C/E ratios described in the basic results are incremental, comparing each intervention with the next most effective option. Average ratios, calculated as a program's cost divided by its effectiveness, do not reflect availability of options with intermediate cost and effectiveness. That incremental ratios provide the appropriate measure of cost-effectiveness is not controversial.[15-17] Nonetheless, many studies present only average results.[4]

Cost-effectiveness ratios should not appear for options less effective but more costly than an alternative (dominated) or for those less effective but more costly than some combination of 2 other options (dominated by extended dominance).[15,16] By excluding these options, C/E ratios for efficient program options can be correctly determined. However, dominated programs can be viable policy options even though not efficient; total costs and effectiveness can be used to calculate C/E ratios for dominated options when needed.

Key sensitivity analyses, which test alternative data and assumptions, should follow the results, providing an indication of their stability. These include tests on important components of costs, and outcome probabilities and values, as well as models used. Discount rates of 3% and 5% will have been reported, but additional sensitivity analysis of the discount rate may be conducted if the C/E ratio is likely to be sensitive to the rate used. If future health care costs of unrelated diseases are relevant, the effect of sensitivity analysis on them should be described.

Reporting of sensitivity analyses to illuminate important distributive issues is recommended whether or not their results affect the C/E ratio. For example, if distinct subgroup preferences are identified, if health-related quality of life is influenced by age, gender, or race, or if alternative methods of valuing time

would substantially influence time cost estimates, sensitivity analyses should be reported.

The Discussion Section

Quantitative results should be translated into a qualitative description of cost-effectiveness in the discussion section and placed in the context of the concerns motivating the analysis. Robustness of results, reflected in a summary of sensitivity analyses, should be described.

A discussion of study limitations is essential, assisting readers in determining the generalizability of results, likelihood of errors, and possible bias. For example, data on intervention effectiveness might disproportionately reflect studies of men. Results should be compared with those from other studies of the intervention to determine consistency of findings and with studies of related interventions to permit comparisons among alternative resource uses. A global set of options for addressing a problem should be discussed. Thus, a coronary bypass CEA should review cost-effectiveness of angioplasty and medical therapy, but diet modification and smoking cessation are also relevant.

Distributive implications may be of crucial importance to decision makers and may involve costs or savings from implementing an intervention, the distribution of health effects, or both. For example, pesticide use might improve population nutritional status but pose risks to farmworkers. Income transfers (welfare or disability payments), which are not costs from the societal perspective, may be of great concern in setting policy.

Ethical problems may arise from analysis assumptions. For example, if life-tables or health-related quality-of-life weights specific to the target population are used, results will be different than if population averages are used. Comment on the extent to which results are influenced by disparities among subgroups may be appropriate. Some interventions, such as prenatal genetic testing for markers of disease, raise ethical questions unrelated to cost-effectiveness. The discussion of these issues improves credibility and may avert simplistic use of CEA results.

Authors may be tempted to make statements concerning whether an intervention is cost-effective. Because cost-effectiveness is relative, absolute statements should be viewed with caution. Interventions are better described as being more or less cost-effective than others, except when actually cost saving.

Technical Addendum and Technical Report

Due to space constraints, a journal article rarely permits full reporting of a CEA. Reviewers may lack information needed to judge study quality. A technical addendum, submitted with the journal report and elaborating on methods, addresses this concern. The addendum could be published with the journal report. For readers requiring more information on a study, the panel recommends that a detailed technical report be made available on request.[3]

CONCLUSION

Standardizing the conduct of CEAs is designed to facilitate comparisons of interventions. However, unless the reporting of reference case results is also standardized, cost-effectiveness results will continue to be difficult to use. Our recommendations for standard reporting of reference case analyses are designed to promote greater relevance for CEA in informing decisions.

Panel on Cost-Effectiveness in Health and Medicine membership and staff: Norman Daniels, PhD; Dennis G. Fryback, PhD; Alan M. Garber, MD, PhD; Marthe R. Gold, MD, MPH; David C. Hadorn, MD; Mark S. Kamlet, PhD; Joseph Lipscomb, PhD; Bryan R. Luce, PhD; Jeanne S. Mandelblatt, MD, MPH; Willard G. Manning, Jr, PhD; Donald L. Patrick, PhD; Louise B. Russell, PhD; Joanna E. Siegel, ScD; George W. Torrance, PhD; Milton C. Weinstein, PhD.

We thank Kristine I. McCoy, MPH, for research, coordination, and editorial assistance in association with this project.

References

1. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC, for the Panel on Cost-Effectiveness in Health and Medicine. The role of cost-effectiveness analysis in health and medicine. *JAMA*. 1996; 276:1172-1177.
2. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB, for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276:1253-1258.
3. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
4. Udvarhelyi S, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? *Ann Intern Med*. 1992;116:238-244.
5. Mason J, Drummond M. Reporting guidelines for economic studies. *Health Econ*. 1995;4:85-94.
6. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care*. 1992;30:231-243.
7. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. New York, NY: Oxford University Press; 1987.
8. Shepard DS, Thompson MS. First principles of cost-effectiveness analysis in health. *Public Health Rep*. 1979;94:535-543.
9. Office of Technology Assessment. *The Implications of Cost-Effectiveness Analysis of Medical Technology, Congress of the United States, August 1980*. Washington, DC: US Government Printing Office; 1980.
10. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med*. 1977;296:716-721.
11. Task Force on Principles for Economic Analysis of Health Care Technology. Economic analysis of health care technology: a report on principles. *Ann Intern Med*. 1995;122:60-69.
12. Henry D. Economic analysis as an aid to subsidisation decisions: the development of Australia's guidelines for pharmaceuticals. *PharmacoEconomics*. 1992;1:54-67.
13. Commonwealth Department of Human Services and Health. *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee*. Canberra, Australia: Australian Government Publishing Service; 1995.
14. Canadian Coordinating Office for Health Technology Assessment. *Guidelines for Economic Evaluation of Pharmaceuticals: Canada*. Ottawa, Ontario: Canadian Coordinating Office for Health Technology Assessment; 1994.
15. Weinstein MC. Principles of cost-effective resource allocation in health care organizations. *Int J Technol Assess Health Care*. 1990;6:93-103.
16. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programmes. *Health Serv Res*. 1972;7:118-133.
17. Cantor SB. Cost-effectiveness analysis, extended dominance, and ethics: a quantitative assessment. *Med Decis Making*. 1994;14:259-265.

Printed and Published in the United States of America

Risk Analysis, Vol. 17, No. 2, 1997

217

The Value of Improved National Exposure Information for Perchloroethylene (Perc): A Case Study for Dry Cleaners

Kimberly M. Thompson[1] and John S. Evans[1]

Received September 8, 1995; revised July 10, 1996

Opportunities to improve our information about risk continue to arise and lead decision makers to indirectly address the issue of the value of improved information through resource allocation decisions. Statistical decision analysis techniques provide an analytical framework for valuing information explicitly in the context of regulatory decision making. This paper provides estimates the value of improved national estimates of perchloroethylene (perc) exposure from U.S. dry cleaners in the context of EPA's recently promulgated National Emissions Standard for Hazardous Air Pollutants (NESHAP) with emphasis on exposure information. Consistent with the NESHAP decision, we relied on EPA's technology and economic assessments. In this first cut analysis, estimates of the exposures of workers, consumers of dry cleaning services, and the general public are probabilistically characterized to reflect uncertainty about exposure and potency. We consider the net benefits of the different control options by assessing the associated changes in the total annual population risks and valuing them in monetary terms, with no constraints placed on maximum individual risks. The results suggest that the expected value of perfect information (EVPI) about potency exceeds the EVPI about exposure. Sensitivity analyses demonstrate how the choices of the valuation parameters and distributions used to characterize uncertainty in the model affect the estimates of the value of information.

KEY WORDS: Perchloroethylene; dry cleaning; uncertainty; value of information; decision analysis.

1. INTRODUCTION

In 1990, the Environmental Protection Agency's (EPA) Science Advisory Board (SAB) found that the EPA's ability to assess (and implicitly to manage) human health risks is limited by its lack of pertinent exposure data.[1] In response to this finding, the EPA has proposed to conduct studies like the multimillion dollar National Human Exposure Assessment Survey (NHEXAS) which is intended to provide national exposure information to improve risk management decisions.[2] This paper provides estimates the value of improved national estimates of perchloroethylene (perc) exposure from U.S. dry cleaners using statistical decision analysis techniques in the context of EPA's recently promulgated National Emissions Standard for Hazardous Air Pollutants (NESHAP) with emphasis on exposure information.[3] We selected perchloroethylene (perc) for this case study because it was one of the 12 substances that the EPA initially identified as being "key pollutants" to include in the NHEXAS.[4]

Perc is a volatile organic compound that may cause adverse health effects[5–9] ranging from acute CNS, liver, and kidney effects for exposure to relatively high levels of perc (over 678,000 $\mu g/m^3$ = 100 ppm), to recently reported induced color vision loss[10,11] and chronic kidney toxicity at lower levels.[12,13] However, at ambient and current occupational concentrations, carcinogenicity is a concern and it has been the primary basis for regulation.[14] The EPA currently places perc in the continuum between classes C (possible carcinogen) and B2

[1] Harvard School of Public Health, Harvard Center for Risk Analysis, 718 Huntington Ave., Boston, Massachusetts 02115.

0272-4332/97/0400-0253$12.50/1

Table I. Model Industry Profile (D-D = Dry-to-dry and T = Transfer machines)[20,21]

Machine size (load capacity in kg)/type	Annual output of cleaning at full capacity (kg/yr)	Total number of facilities	Number of uncontrolled facilities	Percent of uncontrolled
3.6/D-D	6800	3042	1615	53
11.3/D-D	28,400	2575	797	31
13.6/D-D	34,000	1691	625	37
15.9/D-D	39,700	7178	2114	29
20.4/D-D	51,100	2024	791	39
22.7/D-D	56,800	1438	638	44
27.2/D-D	68,100	91	39	43
45.4/D-D	113,500	1605	661	41
15.9/T	47,700	6524	3941	60
22.7/T	68,100	1095	439	40
45.4/T	136,200	726	182	25
Total		27,989	11,842	42

(probable carcinogen), and substantial amounts of uncertainty and controversy exist about perc's carcinogenicity[14,15] and potency.[6] This case considers perc exposure from dry cleaners because dry cleaners are the major users of perc.[14]

The approach we take is based on statistical decision analysis which provides an analytical framework for valuing information in the context of a decision. The approach is well-developed and has been previously applied in the context of numerous environmental health risk management decisions.[16-19] While our analysis is similar to previous applications, it differs because we focus on understanding the value of national exposure information from the perspective of three possible levels of decision making—the individual facilities, facilities with a particular machine type and size, and facilities with a particular machine type.

Consistent with the NESHAP decision, we relied on EPA's industry model and its technology and economic assessments described briefly in Section 2.[20,21] Section 2 also reviews the exposure and dose–response information available for perc and dry cleaners. Since the EPA's formal analysis did not include quantitative estimates of risk or social benefits of the regulation, we developed a risk model to characterize the total population risk that may be reduced by control of perc emissions from dry cleaners and an integrated model to value the benefits in monetary terms. These models, discussed in Section 3, rely on assessing the risk for a number of different exposed populations that may experience reduced exposures with control, and then combines these estimates the get total population risk related to the decision. In this first cut analysis, estimates of the exposures of workers, consumers of dry cleaning services, and the general public are probabilistically characterized to reflect uncertainty about exposure and potency. Based on these characterizations, estimates of the expected values of perfect information are discussed in Section 4 along with the results of sensitivity analyses. Section 5 provides a discussion of insights from the analysis and limitations to its application. In particular, we consider the implications of these results for the design of national exposure studies like the NHEXAS and how technological changes in the industry may change the analysis and results.

2. LITERATURE REVIEW

2.1. Industry Structure, Technology, and Motivation for Regulation

During the past two decades, the multibillion dollar dry cleaning industry has been regulated under a number of environmental statutes and has achieved a substantial amount of emission control.[14] With over 30,000 individual facilities, the industry is largely composed of small businesses and the technology used by the industry is highly variable. Since the 1950s, perc's popularity as a dry cleaning solvent has grown consistently, and in the early 1990s it was estimated that approximately 85–90% of all dry cleaners use perc.[14] While the sizes (and capacities) of machines vary substantially, two general types of machines are used. Transfer machines include a separate washer and dryer similar to domestic laundry machines. Dry-to-dry machines perform both washing and drying in a single unit. In order to characterize the industry, EPA developed an industry profile that classified the existing stock of perc machines into machine type and size categories as shown in Table I.

For any particular facility, decisions about emissions reduction are complex. They involve costs to the owner (which may or may not be passed on to the consumer and/or the workers) and potential benefits to the owner, workers, their families, consumers of dry cleaning services, and the public. Control decisions are further complicated by uncertainty associated with estimating the benefits of control, difficulties associated with valuing the benefits, facility-specific characteristics, local and state requirements, and competing opportunities to reduce risks associated with other types of exposure.

Ambient emissions account for a substantial fraction of perc losses from some dry cleaning machines and they are categorized into two types: vented emissions

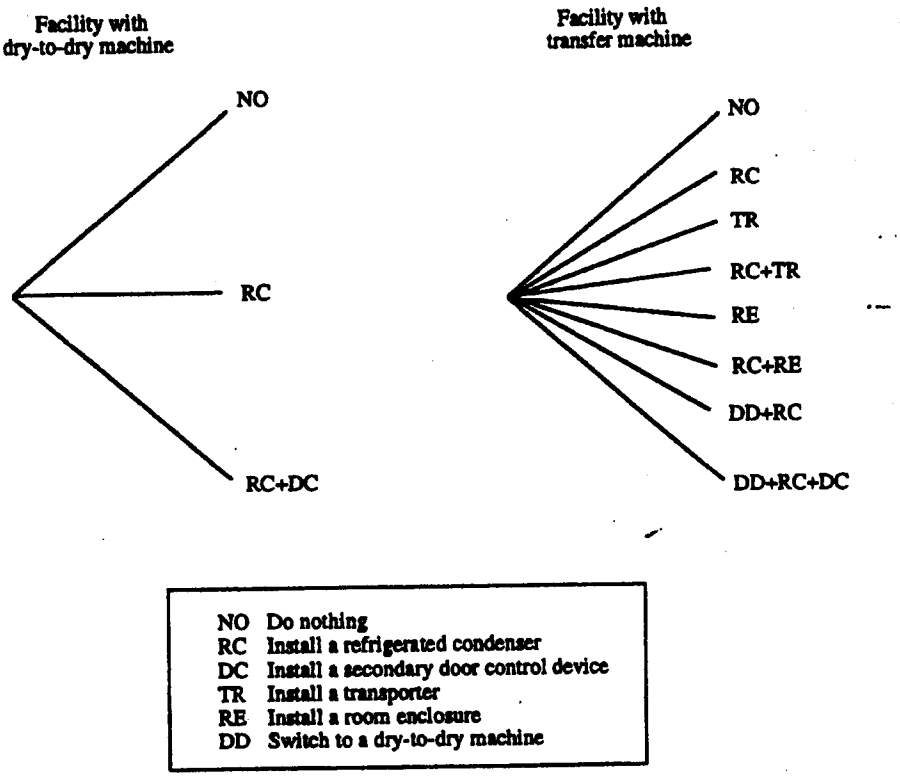


Fig. 1. Decision trees for two types of facilities.

and fugitive emissions. Vented emissions include all emissions that may be collected easily and passed through a control device. Fugitive emissions include perc losses from clothing transfer and leaky process equipment and other losses that are not easily collected.[20,21] Since the use of emissions control devices provides a potentially cost-effective method for reducing ambient emissions, exposure, and risk to perc from dry cleaners, the NESHAP considered requiring control devices for existing uncontrolled facilities (estimates of the numbers and percents of uncontrolled facilities are shown in the last two columns of Table I). The NESHAP control alternatives can be summarized using one of the two decision trees shown in Fig. 1 (which shows only the nondominated actions).

Two decision trees are required because different control options are applicable for the two types of machines, dry-to-dry or transfer, that a facility may use. Based on current information about the costs and performance of carbon adsorbers and refrigerated condensers (RCs), EPA assumed that refrigerated condenser technology dominates for the control of vented emissions (i.e., it costs less and is equally or more effective). Consistent with the NESHAP requirements, we implicitly assume that facilities currently using carbon adsorbers operate them properly, and consequently we only consider facilities with uncontrolled machines. As shown in Fig. 1, we assume that for facilities with uncontrolled dry-to-dry machines the decision is whether to do nothing (NO), to install a refrigerated condenser (RC), or to install both a refrigerated condenser and secondary door control (DC) to limit emissions when the machine door is opened (RC + DC).[20,21] We assume that the decision for facilities with transfer machines is whether to do nothing (NO), to install a refrigerated condenser (RC), to buy a transporter unit to reduce emissions associated with transferring clothes from the washer to the dryer (TR), to enclose the machine in a room (RE), to replace the transfer machine with a new dry-to-dry machine equipped with a refrigerated condenser and perhaps secondary controls on the door (DD + RC or DD + RC + DC), or some combination of these.[20,21] The current state-of-the-art technology in perc dry cleaning machines is the closed-loop (nonvented) dry-to-dry perc machine (DD + RC + DC), and an average-sized new machine costs on the order of $50,000 (capital costs only). We

Table II. Annual Control Costs ($1989/yr) for Machines (Same Categories as Table I) and Control Technologies for Interest Rates (i) of 2.5% and 5% (Not Including Solvent Credit)[20,21]

	Annual control costs with i=5%				
Machine size/type	RC	DC	TR	RE	DD+RC
3.6/D-D	7349	993			
11.3 to 27.2/D-D	1461	993			
45.4/D-D	1971	993			
15.9/T	1927	993	753	2076	5192
22.7/T	1927	993	753	2076	5760
45.4/T	2433	1041	807	2154	8071
	Annual control costs with i=2.5%**				
Machine size/type	RC	DC	TR	RE	DD+RC
3.6/D-D	7215	944			
11.3 to 27.2/D-D	1365	944			
45.4/D-D	1838	944			
15.9/T	1798	944	704	1883	4501
22.7/T	1798	944	704	1883	4955
45.4/T	2268	983	758	1961	7006

Table III. Emission Factors and Control Efficiencies for Different Machine Types and Control Options

	Dry-to-dry machines					
	Emissions factor (kg Perc/100 kg clothes)			Control efficiency		
Control option	Vented	Fugitive	Total	Vented	Fugitive	Total
NO	3.1	2.5	5.6	0	0	0
RC	0.2	2.5	2.7	0.95	0	0.52
RC+DC	0.0	2.5	2.5	0.99	0	0.55
	Transfer machines					
	Emissions factor (kg Perc/100 kg clothes)			Control efficiency		
Control option	Vented	Fugitive	Total	Vented	Fugitive	Total
NO	4.0	5.0	9.0	0	0	0
TR	4.0	3.1	7.1	0	0.38	0.21
RE	4.0	2.6	6.6	0	0.48	0.27
RC	0.6	5.0	5.6	0.85	0	0.38
RC+TR	0.6	3.1	3.7	0.85	0.38	0.59
RC+RE	0.6	2.6	3.2	0.85	0.48	0.64
DD+RC	0.2	2.5	2.7	0.95	0.5	0.70
DD+RC+DC	0.0	2.5	2.5	0.99	0.5	0.72

costs on the order of $50,000 (capital costs only). We have implicitly assumed that switching solvents, closing, and selling the facility have higher social costs than would be justified and consequently that they are not feasible options (the implications of this assumption are discussed in Section 5). Table II gives the estimated annual control costs of the different technologies for two interest rates (i = 2.5% and i = 5%),[22] and Table III gives the estimated control efficiencies.[20,21]

Although human exposure to perc emitted by dry cleaners could also be reduced by other types of control, these are not considered in this analysis. For example, while the exposure of consumers of dry-cleaned clothes to emissions of residual perc left on the clothes has been explored,[23–25] the industry and EPA have not identified effective control strategies for reducing these emissions from perc-cleaned clothes.[26] In addition, while the strategy of relocating dry cleaners out of residential buildings would clearly reduce the perc exposures reported in residential buildings,[27,28] the costs of such a strategy are probably best characterized on a case-by-case basis.

For this case study, we assume that the decision maker is an EPA regulator trying to establish a regulatory standard for the industry that minimizes total social costs (TSC):

$$TSC = CC + HC \quad (1)$$

where CC represents the control costs and HC represents the health costs associated with the total population cancer risks resulting from emissions of perc by the dry cleaner. We do not consider the noncancer health risks (associated with perc or its photochemical degradation products) or ecological risks because we assume that these are small. In addition, we did not put any constraints on maximum individual risks, although such constraints could be added to this type of analysis. The dilemma to adding such constraints in this case is that the maximum individual risks are experienced by workers, and regulation of worker exposure falls under the jurisdiction of the Occupational Health and Safety Administration (OSHA).

We identify three levels on which the regulator may wish to set standards depending on the amount of facility-specific information considered: (1) by facility—choosing the control option that minimizes social costs associated with each particular facility, (2) by machine category—choosing the control option that minimizes social costs associated with representative facilities for each machine type and size category (Table I), and (3) by machine type—choosing the control option that minimizes social costs associated with all facilities using a particular machine type (dry-to-dry and transfer). Thus, we assume that a regulator who desires optimal decisions at the facility level is concerned with aspects of each facility that distinguish it from other facilities with respect to risk (e.g., the location of the facility and the levels of exposure experienced by workers in the facility). In contrast, we assume that a regulator deciding by machine category or type would focus on the aggregate

risk from the industry and not on variability associated with individual facilities.

2.2. Exposure Studies

A systematic national survey that fully characterizes perc exposure and risk among the U.S. population has not been conducted. Nonetheless, everyone is exposed to some level of perc via inhalation and several studies provide insight about the magnitude of such exposures. The EPA's Total Exposure Assessment Methodology (TEAM) Study results suggest that indoor and outdoor samples collected at the participants' homes are similar (with indoor levels slightly higher), and that exposure outside the home (associated with personal activities) may be substantial.[29,30] Many studies have provided ambient measurements of perc which collectively suggest that daily average concentrations of perc vary spatially and temporally and that ambient concentrations of perc in urban and suburban areas are typically several times higher than those in rural areas.[31] Perc measurements in remote areas reflect the global transport of perc that occurs because of its slow destruction by reaction with hydroxyl radicals in the upper troposphere.[32] Two studies provide insight into exposure to perc from water.[33,34]

While these studies are informative for assessing total exposure to perc, they do not provide insight about exposure to perc specifically from dry cleaners. More importantly, they do not support the development of predictive models that are required to assess the costs and benefits of control for dry cleaning facilities. However, a number of dry cleaning-specific exposure studies have been performed that provide information relevant for this case. These studies suggest that exposure to perc from dry cleaners may occur in a number of different ways. First, workers at dry cleaning facilities (both machine operators and others) are exposed to perc emitted into the facility from dry cleaning machines and storage vessels and during the processing of cleaned clothes.[35–42] Second, everyone is exposed to ambient emissions of perc from a dry cleaner because perc has a long lifetime in the atmosphere, although exposures to individuals who live in the vicinity of dry cleaning shops are the highest.[41,43,44] Third, customers who use dry cleaning services are exposed to perc when they enter a dry cleaning facility to drop off or pick up clothing and when they transport, wear, and store the clothing in their homes.[23–25,29,41] Fourth, the families of dry cleaning workers may be exposed to perc brought home by the worker,[45–47] possibly by the workers exhaling perc stored in their bodies.[48] Fifth, people may be exposed to perc while using laundries that offer coin-operated dry cleaning services.[41] Finally, exposure to perc may occur from contaminated waste water and/or solid waste which is transported away from the facility. Although limited industrywide information exists about the amounts and locations of perc contaminated waste associated with dry cleaning facilities, currently most of the industry's waste water and solid waste are treated as hazardous waste.[14]

2.3. Dose–Response Studies

Epidemiological evidence of perc's carcinogenicity is largely inconclusive,[6,7] although various studies suggest excess mortality from leukemia, bladder, lung, cervix, kidney, skin, esophageal, and/or colon cancer.[49–56] Mutagenicity assays and genotoxicity studies have been inconclusive as well, and the mechanism by which perc may cause cancer is unknown. Although only a relatively small percent of perc is cleared from the body by metabolism, metabolites of perc are believed to be responsible for its potential carcinogenicity and toxic effects.[5–9]

The concern about perc's carcinogenicity is based on the results of animal bioassays and mechanistic studies. Two bioassays[57,58] showed dose–response trends for liver tumors in mice. The survival of the rats used in the National Cancer Institute (NCI) bioassay[57] was inadequate to support any conclusions, but the rats showed some evidence of chronic respiratory and kidney disorders. Rats in the National Toxicity Program (NTP) bioassay[58] showed a statistically significant dose–response trend for leukemias based on NTP historical data for background leukemia responses. However, these results are weakened by the fact that the laboratory that conducted this bioassay has a substantially higher background of leukemia responses than most.[59] Three other limited bioassays[60–62] did not show statistically significant evidence of increased tumors in the test animals. Based on the NCI and NTP bioassays, the EPA derived a unit risk for perc of 5.8×10^{-7} $(\mu g/m^3)^{-1}$ or approximately 1×10^{-2} (mg metabolized/kg/d)$^{-1}$.[8]

The State of California discussed and demonstrated the large amount of uncertainty that exists about the status of perc as a human carcinogen and about its carcinogenic potency.[6] California considered the implications of making different assumptions in its derivation of a point estimate of unit risk. Specifically, California derived 144 point estimates using different NCI and NTP animal bioassay datasets, different physiologically-based pharmacokinetic (PBPK) models, different inter-

species scaling factors, and both administered and metabolized doses. It was found that the 95% upper confidence limit of the linear parameter of the "linearized" multistage model (q_1^*) ranged from 0.0050 to 0.42 (mg metabolized/kg/d)$^{-1}$ and from 0.00019 to 0.085 (mg administered/kg/d)$^{-1}$, which gave an 84-fold range in metabolized dose and a 447-fold range in administered dose. This three order of magnitude range in q_1^* is remarkable, and it provided a basis for California to conclude that "Clearly, then, the [PBPK] model, study, route of exposure, and tumor type taken to calculate the appropriate dose in dose–response assessment is [sic] a significant factor in cancer-risk extrapolation for [perc]" (Ref. 6, App. B, pp. 5–32). McKone and Bogen (1992)[34] developed a distribution for perc's potency that reflects uncertainty in the choice of animal bioassay dataset and interspecies scaling factors for estimating potency given (1) the assumed physiologically-based pharmacokinetic (PBPK) model for estimating metabolism, and (2) the multistage model for extrapolating from high animal to low human doses. It implicitly places equal weights on different bioassay results and a probability of 20% on the possibility that perc is not a carcinogen (i.e., that it has a potency of zero). The distribution gives expected and 95th-percentile potencies of approximately 0.1 and 0.4 (mg metabolized/kg/d)$^{-1}$, respectively.[34]

Several sources of uncertainty may not be fully captured by the California analysis or in the potency distribution given by McKone and Bogen (1992).[34] These include the apparent incompatibility of some of the available datasets used to calibrate the PBPK models with other equally plausible calibration datasets,[63] differences in assumptions made by different PBPK models,[63,64,65] and the issue of selecting the appropriate dose surrogate.[66] In addition, no one has seriously considered the use of other dose–response models for perc (e.g., threshold models)[67] or attempted to elicit a dose–response distribution from experts (e.g., see Ref 68).

3. MODELS

3.1. Risk Model

Based on the literature review, we develop a risk model that relies on assessing the risk for a number of different exposed populations and then combining these to estimate the total risk as shown in Eq. (2):

$$\tilde{R} = \sum_{p=1}^{P} \tilde{R}_p = \tilde{\beta} \cdot f \sum_{p=1}^{P} \tilde{M}_p \cdot \tilde{E}_p \cdot \tilde{Y}_p \qquad (2)$$

where $\tilde{R}$ = the total annual population risk (cases/yr), $\tilde{R}_p$ = the total annual population risk for exposed population p (cases/yr), $\tilde{\beta}$ = the potency (cases/person)·(mg metabolized/kg/d)$^{-1}$, f = a conversion factor accounting for the typical breathing rate (20 m^3/d) and body weight (70 kg) of an exposed individual, and 0.001 mg/μg, $\tilde{M}_p$ = the fraction which is metabolized for population p (mg metabolized/mg inhaled), $\tilde{E}_p$ = annual population exposure (people·μg/m^3/yr), and $\tilde{Y}_p$ = model uncertainty factor (dimensionless).

We considered four different exposed populations for this case study: (1) workers, (2) families of worker (referred to as families), (3) people who consume dry cleaning services (referred to as consumers), and (4) people exposed to ambient emissions (referred to as the public) because some of the control strategies have the potential to reduce their exposures and consequently lead to health benefits. Workers are exposed to the highest concentrations (on the order of 10,000–100,000 μg/m^3). Further analysis suggested that the risks to families were very small compared to the risks for workers, and consequently we assume that the benefits associated with reducing risks to families are negligible. Consumers were considered because in aggregate they can account for a substantial population exposure even though each individual spends relatively little time in the facility picking up and dropping off clothes. The public was included in the model because of its potentially significant aggregate exposure and because most of the control alternatives under consideration were primarily intended to reduce ambient emissions and to limit exposure of the public.

For all of the exposed populations except the public, we further characterize $\tilde{E}_p$ according to Eq. (3):

$$\tilde{E}_p = \tilde{N}_p \cdot \tilde{F}_p \cdot \tilde{C}_p \qquad (3)$$

where $\tilde{N}_p$ = the aggregate number of exposure events in population p, which is the number of people times the annual number of exposure events per capita (people·events/yr), $\tilde{F}_p$ = the fraction of time spent (as an annual average) exposed per event (events)$^{-1}$, and $\tilde{C}_p$ = the average perc concentration during the period of exposure (μg/m^3).

To estimate $\tilde{E}_{public}$, we evaluated two components. One accounts for the effects on the population near the source ($\tilde{E}_{local}$) and the other accounts for the fact that perc may be transported globally and affect people in other parts of the world ($\tilde{E}_{global}$). Because of differences in pop-

ulation density and prevailing meteorological conditions that depend on the location of the facility, there are differences in exposure to ambient emissions associated with dry cleaners with similar ambient emission rates (Q). To estimate $\tilde{E}_{local}/Q$ for a facility, we used Version II of the EPA's Human Exposure Model (HEM).[69] This version of the HEM combines a sector-averaged Gaussian plume dispersion model and meteorological observations with population data to estimate the population exposure from a source. The model finds the centroids of census blocks (the finest resolution available with the Census data), estimates the concentration at each centroid, multiplies that concentration by the 1990 Census population in the block, and aggregates the results for all of the blocks.[2] The location of the facility is a key input into the HEM because it determines the appropriate meteorological conditions and population data. We use an additional term ($\tilde{Y}_{local}$) to characterize the uncertainty in annual average concentrations predicted by Gaussian dispersion models.

To estimate $\tilde{E}_{global}/Q$, we use a simple one-compartment model, described in Appendix A, which accounts for dilution and transport of perc with the global circulation, and includes uncertainty about perc's lifetime in the atmosphere (L) assuming that perc is only destroyed by photochemically reacting with OH radicals. We include uncertainty about the model in the atmospheric lifetime term. Since in both of these models exposures are proportional to the emission rate (Q), they can be expressed as:

$$E_{public} = Q \cdot \left(\frac{\tilde{E}_{local}}{Q} \cdot \tilde{Y}_{local} + \frac{\tilde{E}_{global}}{Q} \right) \qquad (4)$$

3.2. Integrated Decision Model

Given the risk model, the decision maker's objective is to choose the control option, k, which minimizes the total social cost (TSC), i.e., the sum of control costs and health costs to the various exposed populations:

$$\tilde{TSC}_k = CC_k - S_k + \tilde{\beta} \cdot f \cdot V \sum_{p=1}^{P} \tilde{M}_p \cdot \tilde{E}_{p,k} \cdot \tilde{Y}_p \qquad (5)$$

where CC_k is the annualized control cost of option k(\$/yr), S_k is the credit for solvent savings (\$/yr), V indicates the value of preventing a statistical case of cancer (\$/case), and $\tilde{E}_{p,k}$ is the annual population exposure for population p for option k (people·μg/m³).

The solvent savings for option k (S_k) are estimated using Eq. (6):

$$S_k = PC \cdot (Q_{NO} - Q_k) \qquad (6)$$

where PC = cost of perc (\$/kg) = 0.683,[22] Q_{NO} = emissions associated with doing nothing (baseline emissions) (kg/yr), and Q_k = emissions associated with option k (kg/yr).

We further assume that $\tilde{E}_{p,k}$ is related to the baseline population exposure ($\tilde{E}_{p,NO}$) and can be modeled as:

$$\tilde{E}_{p,k} = \tilde{E}_{p,NO} \cdot (1 - e_{p,k}) \qquad (7)$$

where $e_{p,k}$ is efficiency of control option k for reducing the exposure of population p. For $\tilde{E}_{public,k}$ we use the total control efficiencies, while for $\tilde{E}_{workers,k}$ and $\tilde{E}_{consumers,k}$ we use the control efficiencies for fugitive emissions (given in Table III) as explained below.

3.3. Input Selection

The distributions and point estimates we used as inputs to the risk model (summarized in Table IV) reflect our judgment about the existing information. The uncertain parameters include: perc's carcinogenic potency, $\tilde{\beta}$; the fraction of inhaled perc which is metabolized at high doses, $\tilde{M}_{high}$, and low doses, $\tilde{M}_{low}$; the lifetime of perc in the atmosphere, $\tilde{L}$; the uncertainty in estimates of annual average population exposure derived from the EPA's HEM, $\tilde{Y}_{local}$; the variability of local population exposure per unit emissions, $\tilde{E}_{local}/Q$; the levels of exposure experienced by operators, $\tilde{C}_{operators}$, and nonoperators, $\tilde{C}_{nonoperators}$; the annual average fraction of time workers spend in a facility, $\tilde{F}_{workers}$; the annual average fraction of time consumers spend in a facility, $\tilde{F}_{consumers}$; and the aggregate number of consumer trips made to a facility per year unit output, $\tilde{N}_{consumers}/O$.

For some of the parameters, characterization of uncertainty depends on the level of the decision. For facility-specific decisions, the relevant exposures are: the exposures of the workers in the facility, the exposures of the customers who use the facility, the exposures of members of the public living nearby, and the exposure from global dispersion of the perc emitted. For machine type or category decisions, the exposures of interest are: the average exposure of workers in a facility of this type (and size), the average exposure of consumers visiting a facility of this type (and size), the average exposure of people in communities near dry cleaners, and the average global exposure.

[2] Since the HEM is based on a sector averaged Gaussian plume model and it does not take into consideration the presence of complex terrain (e.g., mountain ranges), it may not give good predictions in such areas.

Table IV. Input Distributions

Parameter	Exposed population	Decision level: F=facility C=category M=machine	Distribution or point estimate: lognormal (median,gsd); uniform (min,max); triangular (min,max, most likely)	Source or basis
$\tilde{\beta}$	All	F, C, M	Empirical	Ref. 34
$\tilde{M}_{low}$	Public, consumers	F, C, M	Uniform(0.038,0.46)	Ref. 34
$\tilde{M}_{high}$	Workers	F, C, M	Uniform(0.0033, 0.0067)	Ref. 7
$\tilde{L}$	Public	F, C, M	Uniform(1,9)	Appendix A
$\tilde{E}_{total}/Q$	Public	F	Lognormal (0.53,2.2)	Ref. 69, mean of distribution
		C, M	0.74	
$\tilde{Y}_{local}$	Public	F	Lognormal(1,2)	Ref. 72
$\tilde{Y}_{local}$	Public	C, M	Lognormal(1,1.4)	Ref. 72
$\tilde{C}_{operators}$	Workers	F	Lognormal(2.0×10^3,2.3)	Ref. 35
$\tilde{C}_{operators}$	Workers	C, M	Lognormal(2.0×10^3,1.04)	Ref. 35
$\tilde{C}_{nonoperators}$	Workers, consumers	F	Lognormal(4.8×10^4,4.3)	Ref. 35
$\tilde{C}_{nonoperators}$	Workers, consumers	C, M	Lognormal(4.8×10^4,1.3)	Ref. 35
$\tilde{F}_{workers}$	Workers	F	Uniform(0.2,0.3)	Local dry cleaners
		C, M	0.25	
$N_{customers}/Q$	Consumers	F	Uniform (0.57,069)	Ref. 75
		C, M	0.63	
$\tilde{F}_{consumers}$	Consumers	F	Triangular(3.8×10^{-4}, 1.9×10^{-3}, 9.6×10^{-6})	Local dry cleaners
		C, M	9.6×10^{-3}	
Y global, Y workers, Y consumers	All	F, C, M	1	
$N_{operators}$	Workers	F, C, M	1	
$N_{nonoperators}$	Workers	F, C, M	5	

For all exposed populations, we use the distribution for $\tilde{\beta}$ for metabolized doses given by McKone and Bogen (1992).[34] We used two distributions for $\tilde{M}_p$: (1) one for low (ambient) exposures, and (2) one for higher (occupational) exposures. This distinction was made because at higher exposures, lower fractions of the inhaled perc are metabolized than at lower exposures. However, during their brief visits to a dry cleaner, we assume that consumers continue to metabolize perc at the higher rate of metabolism characteristic of low exposure.

For a facility at any specified location, the value of $\tilde{E}_{local}/Q$ can be readily obtained using the EPA's HEM. Consequently, if the locations of all uncontrolled facilities are known, then a distribution reflecting the site-to-site variability in population exposure from ambient emissions can be characterized by multiple runs of the HEM. To characterize the site-to-site variability in exposure to ambient emissions, we ran the EPA's HEM using the locations of 100 dry cleaners for an annulus around the source ranging from 20 m to 50 km. The locations were randomly selected from a representative list of 600 U.S. perc dry cleaners provided by a national trade association. The average from this distribution of 100 site-specific values ($\overline{E_{local}}$) was used to characterize the mean population exposure per unit of ambient emissions of perc for machine type and category decisions.

Some investigators have used a "factor of two" to describe uncertainty in annual average concentrations predicted by Gaussian dispersion models[70,71] based on validation studies conducted under ideal conditions (e.g., flat terrain, within 10 km downwind).[72] Under less ideal conditions (e.g., complex terrain or meteorology and urban releases), other investigators have estimated errors for annual average concentrations in excess of tenfold.[73] For facility-specific decisions, we assume that the HEM estimates the average concentration reasonably

well for the very large averaging areas and times appropriate to this analysis and account for model uncertainty using a lognormal distribution for $\tilde{Y}_{local}$ with a median of 1 and a geometric standard deviation of 2. We recharacterize the uncertainty in $\tilde{Y}_{local}$ for the machine type or category decisions based on the assumption that random errors associated with individual facilities will average out nationally and consequently uncertainty in the national average exposure would be somewhat less (we use a geometric standard deviation of $\sqrt{2}$).

Most studies of worker exposure have found differences that depend on both job type (machine operator vs. nonoperator) and machine type (dry-to-dry vs. transfer) (see Table I of Ref. 48 for a summary). These data indicate that the exposures of transfer machine operators are typically two times greater than dry-to-dry machine operators, and that exposures of machine operators are typically almost 2 times greater than those of nonoperators. Consistent with a preliminary EPA analysis,[74] we assumed that worker exposures were unchanged by the introduction of a control device (a carbon adsorber and/or refrigerated condenser) and consequently that workers would not benefit from such control. This assumption implies that worker exposure is primarily related to fugitive emissions, and it is further supported by the recent measurements of worker exposure in facilities using controlled state-of-the-art machines.[41]

By far, the largest study of worker exposure in the dry cleaning industry is the personal exposure study conducted by the International Fabricare Institute (IFI), although Ref. 35 does not provide adequate information to generate distributions of worker exposure. We requested data from the IFI in order to generate distributions for worker exposure and to explore the issue of whether control would reduce worker exposure. Because these data were stratified according to machine type, job type, and whether or not the facility reported using a vapor control device, we were able to perform some preliminary analyses to test this assumption. Although the data are cross-sectional, not controlled for facility size, and otherwise limited, they were consistent with the assumption that vapor control did not change worker exposure (*t*-tests were not significant). Consequently, we stratified the data only for job type and machine type and fit lognormal distributions for $\tilde{C}_{operators}$ and $\tilde{C}_{nonoperators}$ for facilities with transfer machines (see Fig. 2 for histograms of the data).[3] For facility-specific decisions, we characterize uncertainty based on the means and standard deviations of the data, and for machine type and category decisions we use the standard errors of the mean. We encourage the dry cleaning trade associations to perform and publish analyses of their data and to further test these assumptions.

Based on discussions with local cleaners, we estimate that the annual average fraction of time spent working in a facility, $\tilde{F}_{workers}$, ranges between 0.2 and 0.3 (approximately 40 hr/week). We use the mean of this distribution for machine type and category decisions.

In order to estimate $\tilde{N}_{consumers}/O$ we used the results of a national solvents usage survey which provided a frequency distribution of commercial dry cleaning use.[75] We assumed that people who dropped off clothes less frequently than once a week made a separate trip to pick-up clothes, and that a person drops off an average of 2 kg of clothing per trip.[4] These assumptions imply that a cleaner processes 3341 kg of clothes per 100 people served and that 0.63 people-trips are made to the dry cleaner per kg of dry cleaning output (O). To characterize uncertainty about this relationship for facility-specific decisions, we assume that there may be a 10% error in either direction. We use the mean of the distribution for machine type and category decisions. Although it is possible that there may be a correlation between the frequency of visits and the amount of clothes dropped off, we are unaware of any data to suggest such a correlation and we do not account for the possibility here.

To estimate time-weighted average exposure from visiting a commercial cleaner, we assume that consumers entering the facility are exposed to perc at the same level as nonoperators in the facility. We were unable to find any data which discussed the average amount of time which people spend in the dry cleaner when they drop off and pick up clothes. Consequently, based on discussions with local dry cleaners, we assumed that $\tilde{F}_{consumers}$ for facility-specific decisions would be between 2 and 10 minutes (which is converted to a fraction by dividing by the number of minutes in a year). We use a point estimate of 5 minutes for machine type and category decisions as shown in Table IV.

4. DECISION, VALUE OF INFORMATION, AND SENSITIVITY ANALYSES

4.1. Method

The best control alternative for the decision maker to choose given no additional information is known as

[3] We were unable to reject the null hypothesis that the distributions were lognormal based on the Shapiro-Wilks statistic ($p = 0.18$, for both).

[4] A two-piece man's suit and a three piece woman's suit each weigh about 1 kg.

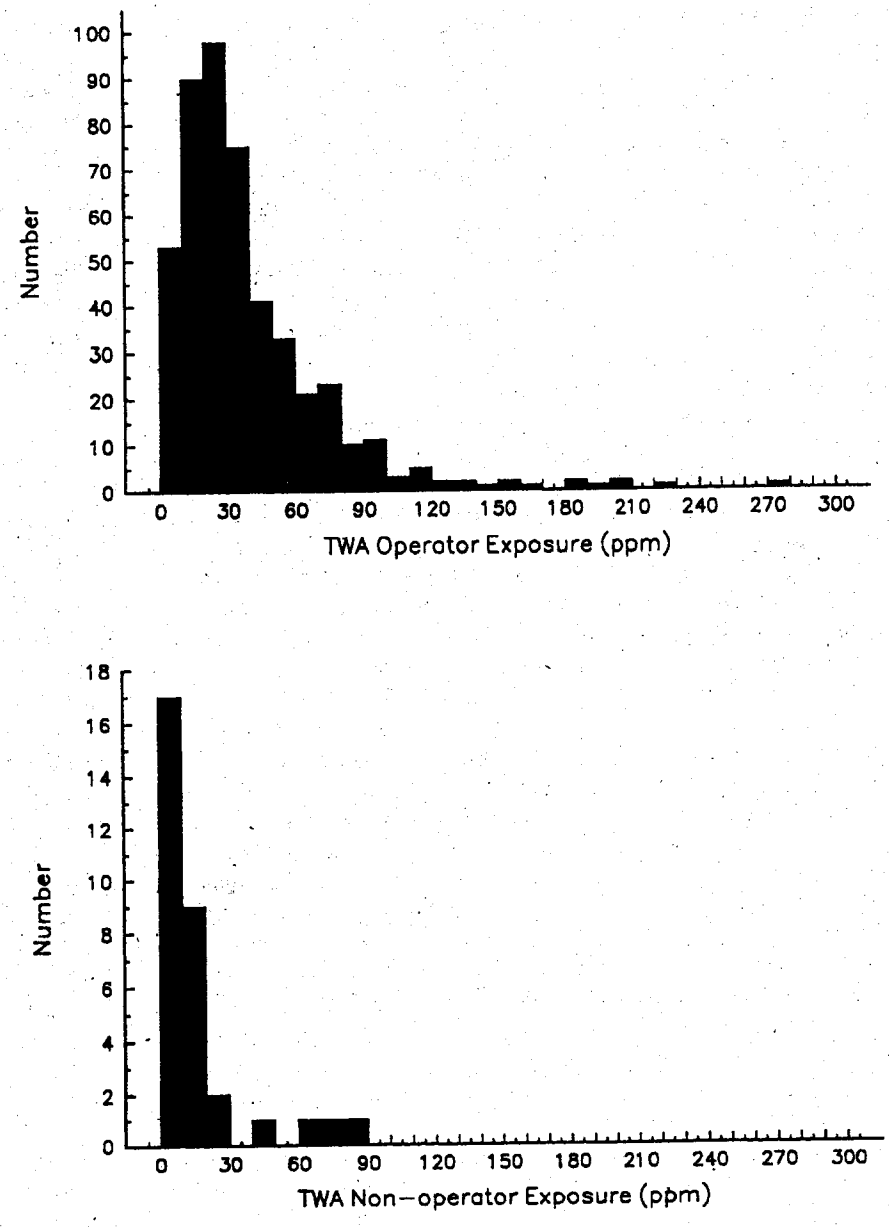


Fig. 2. Histograms of time-weighted-average (TWA) exposures for operators and non-operators at facilities with transfer machines.

the *a priori* k (denoted k^*). Due to uncertainty about the actual risk, the total social costs for any control alternative k are uncertain and are represented by a distribution (denoted $TS\tilde{C}_k$). For an expected value decision maker, by definition, the best alternative is the one with the lowest expected total social cost. The expected value of perfect information (EVPI) is the most that a decision maker should be willing to pay to resolve all of the uncertainty about risk prior to making a decision. This can be expressed symbolically as:

$$\text{EVPI} = \int_{R_{p,NO}} (TS\tilde{C}_k^* - TS\tilde{C}_k^{**}(R_{p,NO}))\, dR_{p,NO} \quad (9)$$

where k^{**} is the control alternative that minimizes $TSC(R_{p,NO})$ for any specific value of $R_{p,NO}$. In most cases, it is not possible to completely resolve the uncertainty in risk. However, there may be tests available that will

provide information about one or more of the components of the risk model. Estimating the expected value of perfect information for components of the risk equation provides an upper bound on the value of the information provided by the test. For any tests that appear to be worthwhile, the next step is to characterize the imperfect nature of the test and to assess the value of the imperfect information.

In order to assess the value of better exposure information for the entire industry, we perform the decision analysis for representative dry cleaners from each of the 11 size and machine type categories defined by the EPA[20,21] and then aggregate the results.[5] We used simulation techniques for all of our analyses using the methodology described by Thompson and Graham (1996).[76] Simulations were performed using SAS version 6.0.9 (SAS Institute, Inc., Cary, NC) to estimate the expected values of perfect information about risk and about several components of the risk equation. In the absence of an accepted policy indicating social preferences for discounting and valuing life, we computed the control costs based on an interest rate of 5% and value health benefits using V = \$3 million as a base case. In addition, we performed sensitivity analysis that vary V between \$1 and \$10 million for two different interest rates (i = 2.5% and 5%) based on recommendations made in Ref. 77 and 78, although recent studies of public preferences suggest the ranges may be larger.[79,80]

4.2. Results

The decision and value of information analyses for a single facility illustrate the levels of risk at issue and the effects of the different control alternatives. Figure 3 shows the total social costs for the different control alternatives as functions of the risk to the public for a facility with a 15.9 kg dry-to-dry machine located such that the population exposure from ambient emissions equals the mean of E_{local}. The best *a priori* decision is to do nothing. However, the expected value of the risk (at the 73rd percentile of the risk distribution) falls very close to the breakpoint (which occurs at the 74th percentile of the risk distribution) where the optimal decision switches from doing nothing to installing a refrigerated condenser. Consequently, there is value associated with obtaining information prior to making a decision. The expected value of perfect information (EVPI) for this facility is approximately \$370/yr.[6] Table V provides estimates of the value of perfect information about various components of risk for this facility and a summary of the contributions to variance in risk (CV). Since the uncertainty in potency ($\tilde{\beta}$) and metabolism ($\tilde{M}_{\text{low}}$) combined account for over 95% of the variability in risk ($\tilde{R}$), it is not surprising that the value of better information about these inputs exceeds that for better exposure information.

For a facility with a 15.9 kg transfer machine located such that the population exposure from ambient emissions equals the mean of E_{local}, Table VI provides the expected total social costs for the different control alternatives based on the expected values of the risk distributions for the different exposed populations. Given our assumptions, the population risks experienced by the consumers are an order of magnitude higher than those experienced by workers and the public, in part reflecting the order of magnitude difference between $\tilde{M}_{\text{low}}$ and $\tilde{M}_{\text{high}}$. The best *a priori* decision is to install a room enclosure and a refrigerated condenser, but the expected value of risk falls close to the breakpoint between installing the room enclosure alone and installing both the room enclosure and a refrigerated condenser. As shown in Table V, completely resolving uncertainty would be worth approximately \$1430/yr. The EVPI for potency ($\tilde{\beta}$) and metabolism ($\tilde{M}$) combined of approximately \$1210 is almost twice the EVPI for perfect exposure information of approximately \$740/yr. Much of the value of perfect exposure information comes from resolving uncertainty about $\tilde{C}_{\text{nonoperators}}$, which is both relatively uncertain and used to estimate the risks to both workers and consumers.

Performing this analysis for all of the categories, Fig. 4 provides the EVPIs and the EVPIs for exposure (denoted EVPEI) for a representative machine (assuming $\tilde{E}_{\text{local}} = \overline{E_{\text{local}}}$) in each category. These estimates are multiplied by the number of facilities in each category to estimate national EVPIs. The 3.6 kg and 45.4 kg dry-to-dry machine categories are not included in Fig. 4 because there is no value associated with collecting additional information for these facilities. The peaked trends in the EVPI and EVPEI for the dry-to-dry machines are indicative of the nature of the decisions made for the different categories. For the first three dry-to-dry machine categories the best *a priori* decision is to do

[5] EPA had 14 categories, but the control strategies for the three categories in the industrial sector were not considered in our analysis because there was a net savings associated with control from saving solvent and consequently the decision to control dominated.

[6] An estimate of the present value of the EVPI can be obtained by multiplying by a factor of 20 (i.e., assuming a constant stream of benefits at an interest rate of 5%.) Given our approach of aggregating from the facility level up to the national level, consideration must be made of reducing the numbers of uncontrolled facilities as facilities adopt control measures.

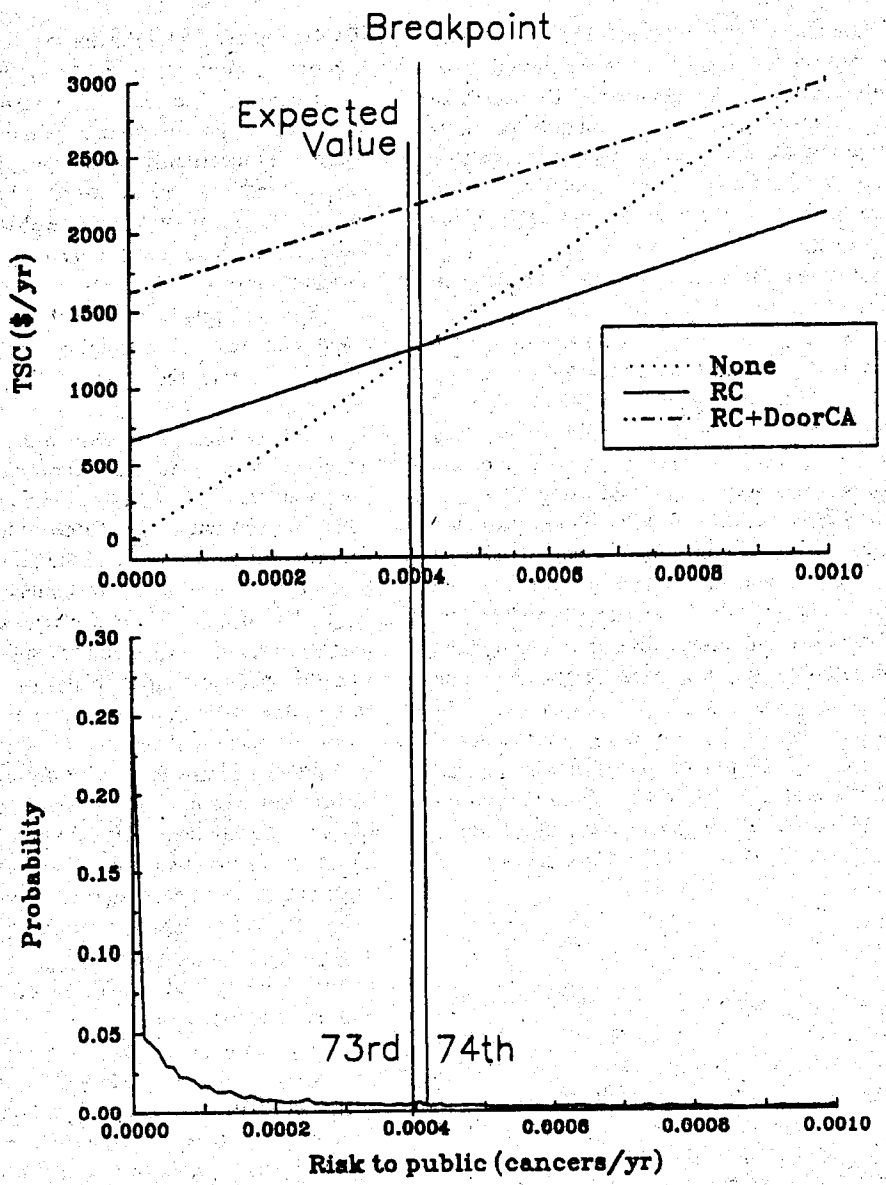


Fig. 3. Total social costs for the 15.9 kg dry-to-dry machine as a function of risk ($\bar{R}$).

nothing, while for the last three categories, the best *a priori* decision is to install a refrigerated condenser. When the expected value of $\bar{R}_{public}$ becomes relatively closer to the breakpoint between these two options, as it does for the 15.9 kg and the 20.4 kg dry-to-dry machines, information becomes relatively more valuable. The best *a priori* decision for the transfer machines is to require installation of both a room enclosure and refrigerated condenser. For decisions based on machine type only, the best *a priori* decision is to require nothing in facilities with dry-to-dry machines, and to require refrigerated condensers and room enclosures in facilities with transfer machines.

Nationally, we estimate that the EVPI about risk for the industry is approximately $4 million/yr for decisions made by category or machine type, and approximately $8 million/yr for decisions made by facility (as shown in Table VII). Information for facilities with transfer machines accounts for approximately 85% of these totals. The EVPIs about potency and metabolism

Table V. Contributions to Variance in Risk (CV) and EVPIs for Components of Risk for a Facility with a 15.9 kg Machine (Base Case)

	Machine type			
	15.9 kg dry-to-dry		15.9 kg transfer	
Random variable	CV	EVPI	CV	EVPI
$\tilde{R}$	100%	370	100%	1430
$\tilde{E}$	3%	170	10%	740
$\tilde{\beta}M$	97%	360	90%	1210
$\tilde{\beta}$	93%	340	88%	1080
$\tilde{L}$	1%	130	<1%	160
$\tilde{Y}_{local}$	2%	150	<1%	170
$\tilde{C}_{nonoperators}$			9%	640
$\tilde{F}_{nonoperators}$			<1%	100
$\tilde{N}_{nonoperators}$			<1%	120
$\tilde{F}_{workers}$			<1%	120
$\tilde{C}_{operators}$			<1%	120

First parameter resolved takes a big bite of VOI b/c it moves the optimal decision away from status-quo

combined ($\beta\tilde{M}$) is approximately \$4 million/yr for decisions made by category or machine type, and approximately \$7 million/yr for decisions made by facility. These values are much larger than the EVPIs for exposure (denoted the EVPEIs) which are approximately half a million dollars per year for decisions made by category or machine type, and increase to approximately \$3.5 million/yr for decisions made by facility.

Looking at the EVPIs for specific input to the risk model, we can consider which types of additional exposure information may be worth collecting. For decisions made by machine type or category, the most valuable exposure information appears to be information about the lifetime of perc in the atmosphere ($\tilde{L}$). For decisions made by facility, the most valuable exposure information appears to be information about the levels of exposure to nonoperators ($\tilde{C}_{nonoperators}$) in each facility that operates a transfer machine. This information, worth approximately \$2.8 million/yr, has a high value because it heavily influences whether a particular facility should install fugitive emissions control equipment to reduce the levels of exposure within the facility. Perfect information about other exposure variables—including $\tilde{L}$, $\tilde{Y}_{local}$, $\tilde{E}_{local}$, and $\tilde{F}_{consumers}$—may each be approximately a few hundred thousand dollars per year.

The EVPI and EVPEI are very sensitive to the choices of V and i (shown in Fig. 5). For facility-specific decisions (a) the EVPI ranges from \$3 to \$12 million/yr and the EVPEI varies from \$1 to \$3.6 million/yr. For decisions based on category (b), the EVPI ranges from \$1 to \$8 million/yr and the EVPEI ranges from \$0 to \$2 million/yr. For decisions based on the machine type (c), the EVPI ranges from \$1 to \$13 million/yr and the EVPEI ranges from \$0 to \$2 million/yr. In addition, if the decision maker chooses to ignore uncertainty in dose–response and use the EPA's unit risk estimate for perc, EVPIs for the industry drop to zero.

5. DISCUSSION

Since this analysis suggests that the EVPI for cancer risks due to perc exposure from dry cleaners may be on the order of several hundred thousand to several million dollars per year, it may be worthwhile to ask whether a study like the NHEXAS could collect the types of exposure information that we have identified as being valuable. In this case, the most valuable information for a decision maker is related to the dose–response parameters and particularly about the potency of perc (which the NHEXAS is unlikely to resolve).

The value of exposure information depends heavily on the level on which decisions are made. The contrast between the levels is clearly demonstrated in their relative values of exposure information for non-operators in facilities which use transfer machines. Decisions at all levels produce some value associated with collecting information about $\tilde{Y}_{local}$ and $\tilde{L}$. Information about the ability of the HEM to accurately predict aggregate population exposure ($\tilde{Y}_{local}$) may be worth collecting depending on the amount of uncertainty it resolves. In addition, the value estimated in this case represents a lower bound on the value of generally validating the HEM if such a validation study would reduce the uncertainty in exposure estimates for other similar types of sources. Better information about perc's lifetime in the atmosphere ($\tilde{L}$) and implicitly about the model we have used to characterize the global benefits of control may be worthwhile. However, this value depends on our assumption that the same V applies to the entire global population. For some parts of the global population, at the margin, there are far cheaper ways of reducing risks in other parts of the world than in the U.S. Nonetheless, the value of this information for this case may underestimate the true value of knowing perc's lifetime in the atmosphere if other decision makers would benefit from such information.

The regulator establishing facility-specific standards would find some other types of exposure information valuable. In particular, information about the time consumers spend in a dry cleaners per visit ($\tilde{F}_{consumers}$) appears to be worth its cost. This type of information and

Table VI. Estimates of the Expected Total Social Costs for the Different Control Alternatives for a Facility with a 15.9 kg Transfer Machine (Base Case)

Option (k)	Control costs (CC_k) ($/yr)	Solvent credit (S_k) ($/yr)	Expected value of risk (R_k) (cases/yr)[a] Consumers	Public	Workers	Total	Expected total social cost (TSC_k) ($/yr)
NO	0	0	5.2×10^{-3}	7.9×10^{-4}	5.6×10^{-4}	6.6×10^{-3}	19,700
RC	1927	1108	5.2×10^{-3}	4.9×10^{-4}	5.6×10^{-4}	6.2×10^{-3}	19,600
TR	753	619	3.2×10^{-3}	6.2×10^{-4}	3.5×10^{-4}	4.2×10^{-3}	12,700
RC+TR	2680	1727	3.2×10^{-3}	3.2×10^{-4}	3.5×10^{-4}	3.9×10^{-3}	12,600
RE	2076	782	2.7×10^{-3}	5.8×10^{-4}	2.9×10^{-4}	3.6×10^{-3}	12,000
RC+RE	4003	1890	2.7×10^{-3}	2.8×10^{-4}	2.9×10^{-4}	3.3×10^{-3}	12,000
DD+RC	5192	2052	2.6×10^{-3}	2.4×10^{-4}	2.8×10^{-4}	3.1×10^{-3}	12,500
DD+RC+DC	6185	2118	2.6×10^{-3}	2.2×10^{-4}	2.8×10^{-4}	3.1×10^{-3}	13,300

[a] Expected values of risk are based on the following expected values of the risk distributions in the absence of further control for the different exposed populations ($\tilde{R}_{p,s}$): $R_{public} = 7.9\times10^{-4}$, $R_{workers} = 5.6\times10^{-4}$, and $R_{consumers} = 5.2\times10^{-3}$.

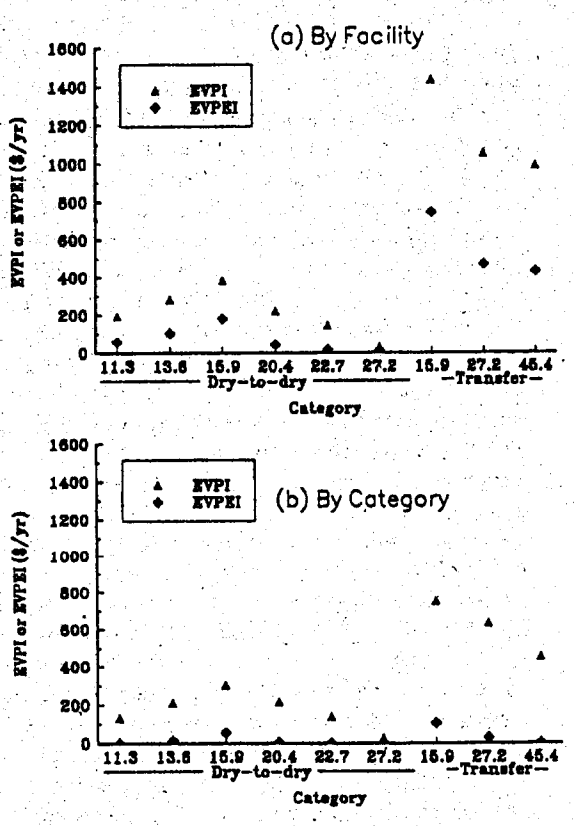


Fig. 4. EVPIs and EVPEIs for the different machines for decisions made by facility (a) and by category (b).

information about the number of trips to dry cleaners ($\tilde{N}_{consumers}$) could be collected in a time-activity study done as a part of the NHEXAS and would also be useful

Table VII. Estimates of the EVPIs for Components of Risk ($MM/yr) for the Base Case by Different Decision Levels

Random variable	By facility	By category	By machine type
$\tilde{R}$	7.6	4.4	4.0
$\tilde{E}$	3.6	0.6	0.2
$\tilde{\beta}M$	6.5	4.3	3.8
$\tilde{\beta}$	5.8	4.0	3.3
$\tilde{L}$	1.0	0.4	0.07
$\tilde{Y}_{home}$	1.1	0.23	0.03
$\tilde{C}_{consumers}$	2.8	0.01	0.001
$\tilde{F}_{consumers}$	0.4	0	0
$\tilde{N}_{consumers}$	0.5	0	0
$\tilde{F}_{workers}$	0.5	0	0
$\tilde{C}_{operators}$	0.5	0	0

to validate the assumption of independence used in this model. If the regulator did not know the locations of the uncontrolled facilities prior to conducting its analysis, then resolving this uncertainty would be important (i.e., it is worth on the order of several hundred thousand dollars per year). Though it may be surprising that information about exposure of operators is not of value, it is consistent with our approach that focuses on reductions in population risk as opposed to the risks of highly-exposed individuals.

Several possible extensions or refinements of this first-cut analysis are possible. First, while this analysis is based on our interpretations of the available information, it may be worthwhile to critically evaluate the default assumptions we made about uncertainty in cur-

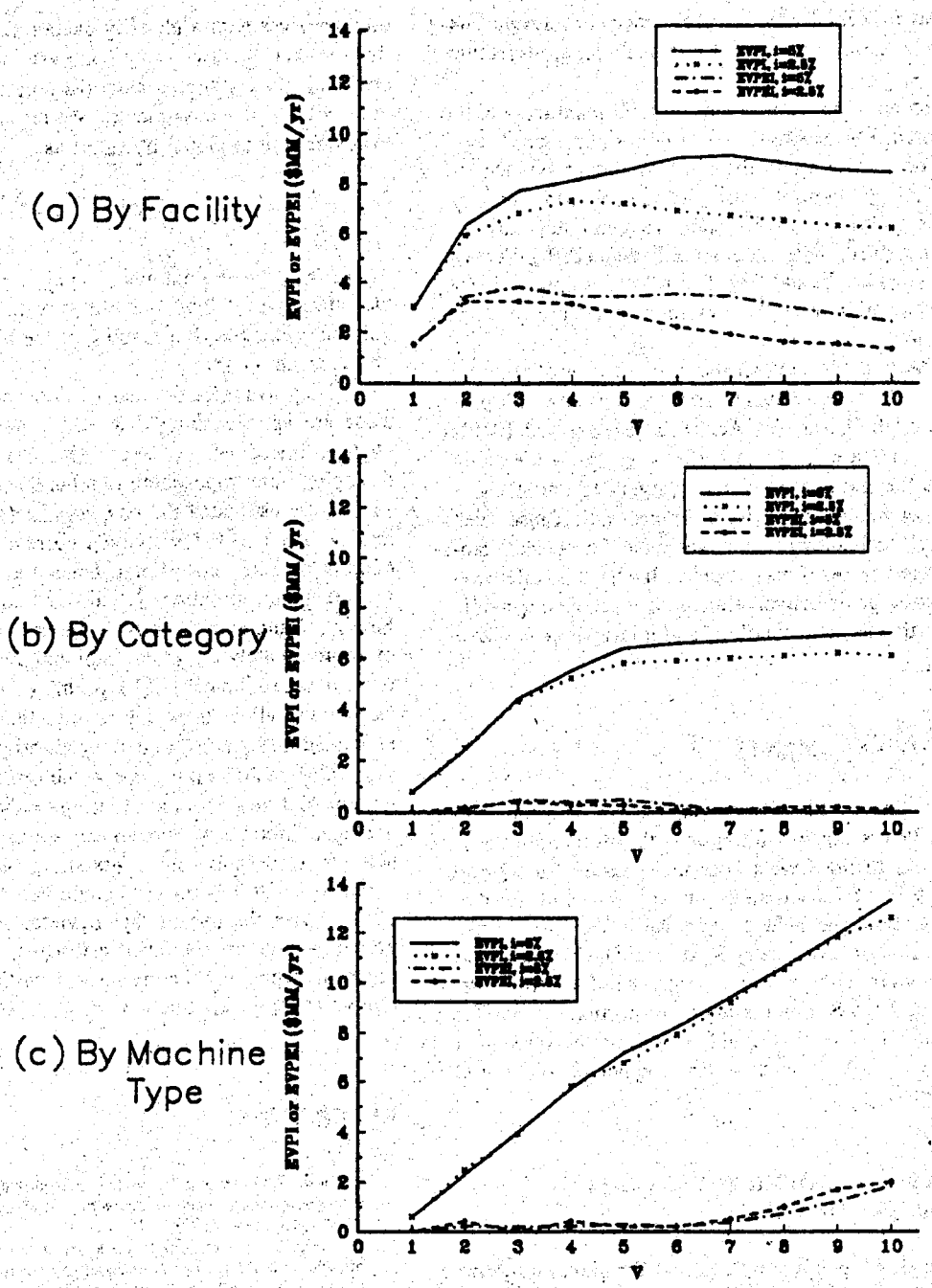


Fig. 5. Changes in the EVPI and EVPEI as a function of V and l from for decisions made by facility (a), category (b), and machine type (c).

rent estimates of exposure and potency, and to consult with relevant experts to obtain more accurate distributions, costs, and valuations for the analysis. This may be particularly true for some variables like the $\tilde{C}_{\text{monoperations}}$ which has a relatively high value of information, or $\tilde{Y}_{\text{local}}$ which changes the EVPI for exposure because it increases the expected value of $\tilde{R}_{\text{public}}$ and makes installation of the refrigerated condenser technology relatively

more favorable. Clearly, the preferences of relevant risk managers should be used for actual decision-making purposes.

Second, other facility-specific information could be considered. For example, we have not considered variability or uncertainty in the cleaning output, control efficiencies, or control costs, although different dry cleaners operate at different capacities and they may actually incur very different costs and obtain different control efficiencies. In addition, facilities operating delivery services or drive throughs may lower exposures of consumers and some workers.

EPA will soon revisit the issue of controlling perc emissions from dry cleaners under the residual risk provisions of the Clean Air Act. The decision and EVPIs at that point are likely to differ from this case study because the industry is shifting largely to the state-of-the-art nonvented dry-to-dry machines and because "wet cleaning" technology (using water for "dry clean" fabrics instead of perc) may soon be highly cost-effective. Thus, this type of analysis should be considered dynamic and iterative as the technology and information available change.

ACKNOWLEDGMENTS

We thank Mike Dusetzina, George Duggan, and Warren Peters, for providing us with information about and access to the EPA's Human Exposure Model and for many helpful discussions and George Smith for discussions about the industry and the NESHAP. Thanks also to John Graham, Barry Ryan, and David Paltiel and two reviewers for providing many helpful comments. We thank the U.S. Environmental Protection Agency for supporting this work through Cooperative Agreement CR818090-01-0, Program in Environmental Health and Public Policy.

APPENDIX A. MODEL FOR GLOBAL TRANSPORT OF PERC

We developed a simple model for global transport of perchloroethylene (Perc) based on a model which considered global transport of radon.[81] The model considers the region between 15 to 55 degrees N latitude as a pipe and assumes that air moves predominantly from West to East and that the importance of diffusion is negligibly small relative to advection. We assume that Perc is removed only by photochemical reaction with OH radicals and that the process can be characterized as a first-order process with a globally-averaged rate constant for this region (L). Based on our analysis, we found that we could adequately approximate the pipe with a single box with volume V. Consequently, we model the population exposure due to global transport as:

$$E_{\text{global}} = \frac{N_{\text{global}} \cdot Q}{V \cdot \bar{L}} = \frac{2.5 \cdot Q}{\bar{L}} \qquad \text{(A1)}$$

where N_{global} = the number of people in the box, Q = the emissions of Perc into the box in kg/yr, V = the volume of the box (km³), and $\bar{L}$ = the globally-averaged rate constant (1/yr).[7]

Using available population data, we estimated that there are approximately 3.8×10^9 people in the box (N_{global}). Based on the assumption that mixing occurs throughout the troposphere to a height of approximately 11 km, we estimated the box volume (V) to be approximately 1.6×10^9 km³. Given current worldwide production of Perc, we estimated that Perc emissions into the box were approximately 3.6×10^8 kg/yr (Q) in 1990. Finally, Wang *et al.*[82] provide an estimate of L for Perc of approximately 2.5/yr and a global average concentration of approximately 0.14 μg/m³ (20 ppt) for 1990. Combining all of these inputs into the box model, we estimated a global average concentration of 0.09 μg/m³ (13 ppt), which is very close to the estimated value. We assume that their is some uncertainty about the globally-averaged lifetime of Perc in the atmosphere. We characterize uncertainty in L assuming that it could be between 1 and 9/yr (i.e., we could be off by as much as 50%). Since the use of this L yields a wide range of global average concentration estimates, we felt it unnecessary to separately incorporate a term for model uncertainty. Thus, we assume that $Y_{\text{global}} = 1$.

REFERENCES

1. Reducing Risk, Setting Priorities and Strategies for Environmental Protection, Environmental Protection Agency, SAB-EC-90-021, Washington, D.C. (1990).
2. T. A. Burke, J. Creason, P. B. Ryan, and M. Schwab, An Approach to the Design of the National Human Exposure Assessment Survey (NHEXAS): A Consolidated Report, Environmental Protection Agency Draft Report, Office of Health Research, Washington, D.C. (1992).
3. EPA, National Emission Standards for Hazardous Air Pollutants for Source Categories: Perchloroethylene Dry Cleaning Facilities, Final Rule, Federal Register 58:49354, Environmental Protection Agency, Washington, D.C., September 22 (1993).

[7] Since kg/km³ are equivalent to μg/m³, the units of E_{global} are people-μg/m³.

4. Memorandum and Attachments Regarding Request for Comments on Recommended List of "Key" NHEXAS Pollutants from Elizabeth Bryan to NHEXAS Workshop Participants, June 10 (1992).
5. Toxicological Profile for Tetrachloroethylene, Department of Health and Human Services, Agency for Toxic Substances and Disease Registry, ATSDR/TP-92/18, Atlanta, GA (1993).
6. Proposed Identification of Perchloroethylene as a Toxic Air Contaminant Technical Support Documents (August 1991) and Revisions (April 1992), California Air Resources Board and the Department of Health Services, Sacramento, CA (1991).
7. Health Assessment Document for Tetrachloroethylene (Perchloroethylene), Final Report, Environmental Protection Agency, EPA/600/8-82/005F, Washington, D.C. (1985).
8. Addendum to the Health Assessment Document for Tetrachloroethylene (Perchloroethylene), Updated Carcinogenicity Assessment, Environmental Protection Agency, EPA/600/8-82/005FA, Washington, D.C. (1986).
9. Response to Issues and Data Submissions on the Carcinogenicity of Tetrachloroethylene (Perchloroethylene), Environmental Protection Agency, Office of Research and Development, EPA/600/6-91/002F, Washington, D.C. (1991).
10. A. Cavalleri, F. Gobba, M. Paltrinieri, G. Fantuzzi, E. Righi, and G. Aggazzotti, "Perchloroethylene Exposure Can Induce Colour Vision Loss," *Neurosci. Lett.* 179, 162–166 (1994).
11. D. Echeverria, R. F. White, and C. Sampio, "A Behavioral Evaluation of PCE Exposure in Patients and Dry Cleaners: A Possible Relationship Between Clinical and Preclinical Effects," *J. Occup. Environ. Med.* 37(6), 667–680 (1995).
12. A. Mutti, R. Alinovi, E. Bergamaschi, C. Biagini, S. Cavazzini, I. Franchini, R. R. Lauwerys, A. M. Bernard, H. Roels, and E. Gelpi, "Nephropathies and Exposure to Perchloroethylene in Dry Cleaners," *Lancet* 25(340), 189–193 (1992).
13. R. G. Price, S. A. Taylor, E. Crutcher, E. Bergamaschi, I. Franchini, and A. D. Mackie, "The Assay of Laminin Fragments in Serum and Urine as an Indicator of Renal Damage Induced by Toxins," *Toxicol. Lett.* 77, 313–318 (1995).
14. K. M. Thompson, "Cleaning up Dry Cleaners," in J. D. Graham and J. K. Hartwell (eds.) *The Greening of Industry: A Risk Management Approach* Chap. 4 (Harvard University Press, Cambridge, MA, 1997).
15. E. Drye, "Perchloroethylene," in J. D. Graham (ed.), *Harnessing Science for Environmental Regulation*, Chap. 6 (Praeger, New York, 1991), pp. 97–125.
16. G. L. Campbell, D. Cohan, and D. W. North, The Application of Decision Analysis to Toxic Substances: Proposed Methodology and Two Case Studies, Prepared for the Environmental Protection Agency, Office of Pesticides and Toxic Substances, NTIS PB82-249103 (1982).
17. A. M. Finkel and J. S. Evans, "Evaluating the Benefits of Uncertainty Reduction in Environmental Health Risk Management," *JAPCA* 37(10), 1164–1171 (1987).
18. J. K. Hammitt and J. A. K. Cave, *Research Planning for Food Safety, A Value-of-Information Approach* (Rand, Santa Monica, CA, 1991).
19. A. C. Taylor, J. S. Evans, and T. E. McKone, "The Value of Animal Test Information in Environmental Control Decisions," *Risk Anal.* 13(4), 403–412 (1993).
20. Dry Cleaning Facilities—Background Information for Proposed Standards, Draft EIS, Environmental Protection Agency, EPA-450/3-91-020a, Washington, D.C. (1991).
21. Economic Impact Analysis of Regulatory Controls in the Dry Cleaning Industry, Final Report, Environmental Protection Agency EPA-450/3-91-021, US EPA Office of Air Quality Planning and Standards, Washington, D.C. (1991).
22. Memorandum Regarding Quality Assurance of Dry Cleaning Cost-Effectiveness Tables from Kim Kepford (Radian Corporation) to George Smith, October 23 (1992).
23. K. W. Thomas, E. D. Pellizzari, R. L. Perritt, and W. C. Nelson, "Effect of Dry-Cleaned Clothes on Tetrachloroethylene Levels in Indoor Air, Personal Air, and Breath for Residents of Several New Jersey Homes," *J. Exp. Anal. Environ. Epidemiol.* 1(4), 475–490 (1991).
24. B. A. Tichenor, L. E. Sparks, M. D. Jackson, Z. Guo, M. A. Mason, C. M. Plunket, and S. A. Rasor, "Emissions of Perchloroethylene from Dry Cleaned Fabrics," *Atm. Environ.* 24A(5), 1219–1229 (1990).
25. T. Kawauchi and K. Nishiyama, "Residual Tetrachloroethylene in Dry-Cleaned Clothes," *Environ. Res.* 48(2), 296–301 (1989).
26. EPA, Proceedings: International Roundtable on Pollution Prevention and Control in the Dry Cleaning Industry, Environmental Protection Agency, EPA/774/R-92/002, Washington, D.C. (1992).
27. J. S. Schreiber *et al.*, New York State Departments of Health (Bureau of Toxic Substances) and Environmental Conservation (Division of Air Resources), October, "*Investigation of Indoor Air Contamination in Residences Above Dry Cleaners*," *Risk Anal.* 13(3), 335–344 (1991).
28. Upstairs, Downstairs: Perchloroethylene in the Air in Apartments Above New York City Dry Cleaners, Consumers Union, New York, NY, October (1995).
29. L. A. Wallace, "The Total Exposure Assessment Methodology (TEAM) Study: An Analysis of Exposures, Sources, and Risk Associated with Four Volatile Organic Compounds," *J. Am. Coll. Toxicol.* 8(5), 883–895 (1989).
30. L. Wallace, W. Nelson, R. Ziegenfus, E. Pellizzari, I. Michael, R. Whitmore, H. Zelon, T. Hartwell, R. Perritt, and D. Westerdahl, "The Los Angeles TEAM Study, Personal Exposures, Indoor–Outdoor Concentrations, and Breath Concentrations of 25 Volatile Organic Compounds," *J. Exp. Anal. Environ. Epidemiol.* 1(2), 157–192 (1991).
31. J. J. Shah and E. K. Heyerdahl, National Ambient Volatile Organic Compounds (VOCs) Database Update, Environmental Protection Agency, Office of Research and Development, EPA/600/3-88/010a (PB88-195631), Washington, D.C. (1988).
32. H. B. Singh, "Halogens in the Atmospheric Environment," in H. Singh (ed.), *Composition, Chemistry and Climate of the Atmosphere* Van Nostrand Reinhold Publishers, New York, (1995).
33. R. C. Cothern, W. A. Coniglio, and W. L. Marcus, Techniques for the Assessment of the Carcinogenic Risk to the U.S. Population Due to Exposure from Selected Volatile Organic Compounds from Drinking Water via the Ingestion, Inhalation, and Dermal Routes, Environmental Protection Agency, Office of Drinking Water EPA 570/9-85-001 (PB84 213941), Washington, D.C. (1985).
34. T. E. McKone and K. T. Bogen, Uncertainties in Health-Risk Assessment, An Integrated Case Study Based on Tetrachloroethylene in California Groundwater, *Regulat. Toxicol. Pharmicol.* 15, 86–103 (1992).
35. "Monitoring Solvent Vapors in Drycleaning Plants, International Fabricare Institute," *Focus Drycleaning* 14(3), July (1990).
36. H. Ludwig, M. Meister, D. Roberts, and C. Cox, "Worker's Exposure to Perchloroethylene in the Commercial Dry-Cleaning Industry," *Am. Ind. Hyg. Assoc. J.* 44(8), 600–605 (1983).
37. B. L. Materna, "Occupational Exposure to Perchloroethylene in the Dry Cleaning Industry," *Am. Ind. Hyg. Assoc. J.* 46(5), 268–273 (1985).
38. D. Solet, T. G. Robins, and C. Sampaio, "Perchloroethylene Exposure Assessment Among Dry Cleaning Workers," *Am. Ind. Hyg. Assoc. J.* 51, 566–574 (1990).
39. M. X. Petreas, S. M. Rappaport, B. L. Materna, and D. M. Rempel, "Mixed-Exhaled Air Measurements to Assess Exposure to Tetrachloroethylene in Dry Cleaners," *J. Exp. Anal. Environ. Epidemiol.* 1(Suppl.): 25–39 (1992).
40. G. M. Toutonghi, Characterization of Perchloroethylene Exposure in Commercial Dry Cleaning Workers with Biological Monitoring, M.S. thesis, University of Washington, Department of Environmental Health (1993).
41. D. J. Moschandreas and D. S. O'Dea, "Measurement of Perchloroethylene Indoor Air Levels Caused by Fugitive Emissions from

Unvented Dry-to-Dry Dry Cleaning Units," *JAWMA* **45**, 111–115 (1995).
42. S. Howie, Ambient Perchloroethylene Levels Inside Coin-Operated Laundries with Dry-Cleaning Machines on the Premises, Environmental Protection Agency, Environmental Monitoring Systems Laboratory, EPA 600/4-82-032, NTIS PB82- 230947, Washington, D.C. (1981).
43. J. Bumgarner *et al.*, Coin-Operated Dry Cleaner Study, Memorandum prepared by the US EPA Environmental Monitoring Systems Laboratory (as cited by Ref. 16) (n.d.).
44. Cancer Risk from Outdoor Exposure to Air Toxics, Vols. I and II, Final Report, Environmental Protection Agency, Office of Air Quality Planning and Standards, EPA-450/1-90-004a and b, Washington, D.C. (1990).
45. L. Wallace, E. Pellizzari, L. Sheldon, T. Hartwell, R. Perritt, and H. Zelon, Exposures of Dry Cleaning Workers to Tetrachloroethylene and Other Volatile Organic Compounds, Measurements in Air, Water, Breath, Blood, and Urine, Presented at the Annual Conference of the International Society for Exposure Analysis and Environmental Epidemiology, Atlanta, GA, November 15–18 (1991).
46. G. Aggazzotti, G. Fantuzzi, G. Predieri, E. Righi, and S. Moscardelli, "Indoor Exposure to Perchloroethylene (PCE) in Individuals Living with Dry-Cleaning Workers," *Sci. Total Environ.* **156**, 133–137 (1994).
47. G. Aggazzotti, G. Fantuzzi, E. Righi, G. Predieri, F. M. Gobba, M. Paltrinieri, and A. Cavalleri, "Occupational and Environmental Exposure to Perchloroethylene (PCE) in Dry Cleaners and Their Family Members," *Arch. Environ. Health* **49**(6), 487–493 (1994).
48. K. M. Thompson and J. S. Evans, "Workers' Breath as a Source of Perchloroethylene (Perc) in the Home," *J Exp. Anal. Environ. Epidemiol.* 3(4), 417–430 (1993).
49. A. Blair, P. Decoufle, and D. Grauman, "Causes of Death Among Laundry and Dry Cleaning Workers," *Am. J. Pub. Health* **69**, 508–511 (1979).
50. S. D. Kaplan, Dry Cleaning Workers Exposed to Perchloroethylene. A Retrospective Cohort Mortality Study. NIOSH, Department of Health and Human Services Contract No. 210-77-0094, Cincinnati, OH (1980).
51. R. M. Katz and D. Jowett, "Female Laundry and Dry-Cleaning Workers in Wisconsin. A Mortality Analysis," *Am. J. Pub. Health* **71**, 305–307 (1981).
52. R. S. Lin and I. I. Kessler, "A Multifactorial Model for Pancreatic Cancer in Man," *JAMA* **245**, 147–152 (1981).
53. R. W. Duh and N. R. Asal, "Mortality Among Laundry and Dry-Cleaning Workers in Oklahoma," *Am. J. Pub. Health* **74**(11), 1278–1280 (1984).
54. D. P. Brown and S. D. Kaplan, "Retrospective Cohort Mortality Study of Dry Cleaning Workers Using Perchloroethylene," *J. Occup. Med.* **29**, 535–541 (1987).
55. A. Aschengrau, D. Ozonoff, C. Paulu, P. Coogan, R. Vezina, T. Heeren, and Y. Zhang, "Cancer Risk and Tetrachloroethylene-Contaminated Drinking Water in Massachusetts," *Arch. Environ. Health* **48**(5), 284–292 (1993).
56. A. M. Ruder, E. M. Ward, and D. P. Brown, "Cancer Mortality in Female and Male Dry-Cleaning Workers," *J. Occup. Med.* **36**(8), 867–874 (1994).
57. Bioassay of Tetrachloroethylene for Possible Carcinogenicity, Department of Health, Education, and Welfare, National Institutes of Health, DHEW Pub. No. (NIH) 77–813, Bethesda, MD (1977).
58. Toxicology and Carcinogenesis of Tetrachloroethylene (Perchloroethylene) (CAS No. 127-18-4) in F344/N Rats and B6C3F1 Mice (Inhalation Studies), Natl. Toxicol. Program Tech. Rep. Ser. 311, Bethesda, MD (1986).
59. Health Effects Assessment of Perchloroethylene, Review of the Office of Research and Development's Draft Document, "Response to Issues and Data Submissions on the Carcinogenicity of Perchloroethylene (EPA/600/6-91/002A)" by the Environmental Health Committee, Environmental Protection Agency Science Advisory Board (A-101), EPA-SAB-EHC-91-013, Washington, D.C. (1991).
60. L. W. Rampy, J. F. Quast, M. F. Balmer, B. K. I. Leong, and P. J. Gehring, Results of a Long-Term Inhalation Toxicity Study on Rats of a Perchloroethylene (Tetrachloroethylene) Formation, Midland, MI: Toxicology Research Laboratory, Dow Chemical, Co.,
61. J. C. Theiss, G. D. Stoner, M. D. Shimkin, and E. K. Weisburger, "Tests for Carcinogenicity of Organic Contaminants of United States Drinking Waters by Pulmonary Tumor Response in Strain A Mice," *Cancer Res.* **37**, 2717–2720 (1977).
62. B. L. Van Duuren, M. Goldschmidt, G. Loewengart, C. Smith, S. Melchione, I. Seidman, and D. Roth, "Carcinogenicity of Halogenated Olefinic and Alipathic Hydrocarbons in Mice," *J. Natl. Cancer Inst.* **63**, 1433–1439 (1979).
63. D. Hattis, P. White, and P. Koch, "Uncertainties in Pharmacokinetic Modeling for Perchloroethylene. II. Comparison of Model Predictions with Data for a Variety of Different Parameters," *Risk Anal.* 13(6), 599–610 (1993).
64. D. Hattis, P. White, L. Marmorstein, and P. Koch, "Uncertainties in Pharmacokinetic Modeling for Perchloroethylene. I. Comparison of Model Structure, Parameters and Predictions for Low-Dose Metabolism Rates for Models Derived by Different Authors," *Risk Anal.* 10(3), 449–457 (1990).
65. D. Hattis and R. Goble, "Expected Values of Projected Cancer Risks from Putative Genetically-Acting Agents," *Risk Anal.* 11(3), 359–363 (1991).
66. D. Farrar, B. Allen, K. Crump, and A. Shipp, "Evaluation of Uncertainty in Input Parameters of Pharmacokinetic Models and the Resulting Uncertainty in Output," *Toxicol. Lett.* **49**, 371–385 (1989).
67. F. Y. Bois, L. Zeise, and T. N. Tozer, "Precision and Sensitivity of Pharmacokinetic Models for Cancer Risk Assessment: Tetrachloroethylene in Mice, Rats, and Humans," *Toxicol. Appl. Pharmacol.* **102**, 300–315 (1990).
68. J. S. Evans, G. M. Gray, R. I. Sielken, A. E. Smith, C. Valdez-Flores, and J. D. Graham, "Use of Probabilistic Expert Judgment in Uncertainty Analysis of Carcinogenic Potency," *Regulat. Toxicol. Pharmacol.* **20**, 15–36 (1994).
69. HEM-II User's Guide, US Environmental Protection Agency, Office of Air Quality Planning and Standards, EPA/450/3-91-0010, Research Triangle Park, NC (1991).
70. N. C. Hawkins, "Conservatism in Maximally Exposed Individual (MEI) Predictive Exposure Assessments: A First-Cut Analysis," *Regulat. Toxicol. Pharmicol.* **14**, 107–117 (1991).
71. A. C. Taylor, Assessing Uncertainty in the Estimation of Environmental Exposure and Cancer Potency, Sc.D. dissertation, Harvard University (1992).
72. J. F. Bowers, A. J. Anderson, and A. H. Huber, An Evaluation Study of the Industrial Source Complex (ISC) Dispersion Model, Presented at the 74th Annual Meeting of the Air Pollution Control Association, Paper #81-20.4, Philadelphia, PA, June 21–26 (1981).
73. C. W. Miller and L. M. Hively, "A Review of Validation Studies for the Gaussian Plume Atmospheric Dispersion Model," *Nucl. Safety* **28**(4), 522–531 (1987).
74. EPA, Options for Regulating Perchloroethylene Emissions in the Dry Cleaning Industry: A Cost-Benefit Analysis. Draft Report, US Environmental Protection Agency, Washington, D.C. January (unpublished draft on file with the authors) (1988).
75. Westat, Household Solvents Products: A National Usage Survey, Final Report, Prepared for the US EPA Exposure Evaluation Division of the Office of Toxic Substances Under Contract Number 68-02-4243, July (1987).
76. K. M. Thompson and J. D. Graham, "Going Beyond the Single Number: Using Probabilistic Risk Assessment to Improve Risk

Management," *Human Ecol. Risk Assess.* 2(4), 1068–1084 (1996).
77. W. K. Viscusi, *Fatal Tradeoffs: Public and Private Responsibilities for Risk* (Oxford University Press, 1992).
78. *Valuing Health Risks, Costs, and Benefits for Environmental Decision Making* (National Academy Press, Washington, D.C., 1990).
79. M. L. Cropper, H. S. Aydede, and P. R. Portney, "Preferences for Lifesaving Programs: How the Public Discounts Time and Age," *J. Risk Uncert.* 8, 243–265 (1994).
80. G. Tolley, D. Kenkel, and R. Fabian, *Valuing Health for Policy: An Economic Appraisal* (University of Chicago Press, Chicago, IL, 1994).
81. E. K. Turekian, Y. Nozaki, and L. K. Benninger, "Geochemistry of Atmospheric Radon and Radon Products," *Ann. Rev. Earth. Planet Sci.* 5, 227–255 (1977).
82. C. L. L. Wang, D. R. Blake, and F. S. Rowland, "Seasonal Variations in the Atmospheric Distribution of a Reactive Chlorine Compound, Tetrachloroethylene ($CCl_2 = CCl_2$)," *Geophys. Res. Lett.* 22(9), 1097–1100 (1995).