BENEFIT-COST ANALYSIS IN PUBLIC HEALTH

Lester B. Lave\textsuperscript{1} and Satish V. Joshi\textsuperscript{2}

\textsuperscript{1}Graduate School of Industrial Administration and Heinz School of Public Policy, \textsuperscript{2}PhD Student, Heinz School of Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

KEY WORDS: benefit-cost analysis, cost-benefit analysis, cost-effectiveness, evaluation, monetizing

ABSTRACT

This chapter gives an overview of benefit-cost analysis (BCA) and similar evaluation tools, such as cost-effectiveness analysis and technology-based standards, in the context of public health. We describe these evaluation tools, how they are used, their shortcomings, and how they should be interpreted. As with other powerful tools, they are subject to misuse and misinterpretation, even by professionals doing the analysis.

INTRODUCTION

The rising costs of health care and a general desire to lower government expenditures have led to a close scrutiny of public health programs. For more than 20 years, public health officials have had to justify many programs by an analysis of their social benefits and costs. Although benefit-cost analysis (BCA) may often require extensive data collection and analysis, public health professionals should welcome the use of this tool. It is helpful in making public health programs more efficient and effective while providing evidence to convince decision makers that these programs are worthy of support. However, BCA is not a panacea. Public health professionals have much to gain from knowing what the tool is and is not, and how the results should be interpreted.

One of the first official acts of President Reagan was to promulgate Executive Order 12291 requiring agencies to conduct a BCA of all major rules (those imposing costs on the economy of at least $100 million). Requiring this analysis was not peculiar to President Reagan. Rather, some form of White House review of the benefits and costs of regulation had been required since President Nixon. President Clinton modified the executive order slightly, but
the essential parts remain in force. Just before the 1994 elections, Congress came close to enacting legislation requiring BCA. The Republicans made BCA a part of the "Contract with America" and the House of Representatives moved quickly after the election to pass legislation.

The 1995 House and Senate hearings on legislation to require BCA of major rules had many environmentalists and public health professionals testifying against the legislation (27). Both Carol Browner, Administrator of EPA, and Donna Shalala, Secretary of Health and Human Services, expressed vigorous disagreement with the notion of requiring BCA. They seemed to regard the proposed legislation as a means of slowing or stopping new government regulation rather than as a means of improving the programs and isolating those programs that should be rejected. We agree that requiring BCA could impede regulation, particularly if existing programs must be evaluated and no new resources are available, but we praise BCA for the good it can do for public health programs. However, to reap these benefits, the community must have a good understanding of what BCA is, what it is not, and how it can be helpful in improving the quality of public health decisions.

THE BENEFITS OF BENEFIT-COST ANALYSIS

Although they do not agree on the precise interpretation of BCA, practitioners agree that their tool is useful for posing the right questions and collecting and analyzing the relevant data. They disagree on whether the tool should dictate decisions, but agree that where decisions involve the health of many individuals or millions of dollars, the BCA framework can inform decision makers. However, as shown below, BCA can be expensive and time-consuming. The depth of analysis, and even the decision as to whether to do an analysis, must depend on the value of the information to be gleaned and the importance of the decision. The tool must be used with discretion and flexibility.

BCA should be thought of as a framework to improve the quality of decisions. For example, myriad screening programs have been suggested for breast, prostate, colon, skin, and other cancers; stringent regulations exist for abating air and water pollution, occupational exposure to toxicants; and control of drinking water quality and pesticide residues in food. Each of these programs is potentially expensive but each also offers large potential health benefits. Should society require that every screening and environmental quality program be funded? Should all programs shown to be efficacious be funded? Should society offer first the programs that are most cost-effective, i.e. those giving the greatest health improvement per dollar expended? Should programs be offered only if their social benefits exceed their social costs?
A tool to help policy makers decide among programs is likely to be most helpful if it examines the following issues:

1. **Problem definition and statement of objectives:** What is the problem we are trying to solve? Often people have different perceptions of the problem and thus come to very different conclusions about the desirability of proposed interventions. For example, is the problem excessive deaths from breast cancer? Is it excessive incidence? Or is it a lack of concern for women’s health problems? These three definitions would give rise to very different interventions, from better screening and treatment to prevention programs to broader attention to women’s health problems.

2. **Identification of all reasonable means to accomplish the stated objectives:** What is the full panoply of interventions? For example, the goal of fewer breast cancer deaths could be accomplished by improved treatment, by improved screening, or by primary prevention. Resources would be spent in quite different ways in each of these interventions. Tools are needed to compare the efficacy and costs of each program.

3. **Analysis of the benefits and costs of each alternative:** The benefits and costs must be quantified and, if reasonably possible, be translated into dollars. Where the translation is controversial, the analysis must be careful to summarize the effects in terms of a multidimensional array. If the goal is fewer breast cancer deaths, a variety of interventions could improve treatment at increasing costs, with an extreme case being establishing tertiary medical care centers in every community to treat breast cancers. How many deaths would be prevented by each type of intervention to improve treatment? What is the estimated cost of each intervention? What are the life expectancy and the quality of life of women who have been treated? What is the quality of life during treatment? Analogous questions should be asked for early detection programs: How many deaths would be prevented? What is the increase in life expectancy? What is the quality of life for each group—particularly women who are incorrectly diagnosed as having breast cancer or as not having breast cancer? How should the inconvenience and discomfort of screening be weighed against the increased life-expectancy of the small proportion who are helped? There might be significant psychological and medical costs associated with a false positive, requiring a woman to undergo further tests and undergo the anxiety of thinking that she may have breast cancer. Primary prevention has the virtue of requiring fewer women to undergo treatment. However, prevention might require a change in diet or other habits that could be costly to the individual, and might lower the quality of life.

4. **Use of all relevant data:** The analysis should emphasize a systematic, analytic approach that uses all relevant data and helps to structure a research agenda for gathering the crucial missing data. For example, at what age should
breast cancer screening start? How often should women be screened? How should information on family and personal history of breast cancer be used?

5. Perspective and values used in the analysis: We suggest using a social perspective, but this is not uniquely defined. An analyst who is 25, black, and was raised in Harlem probably has a different perception of the social perspective than an analyst who is 65, white, and was raised in a small town in Washington. To be most helpful, the analysis should consider a range of important perspectives and values, corresponding to the range of most groups in society.

6. The time pattern of benefits and costs: Programs generally involve expenditures and benefits that extend over many years. A dollar spent now is more costly than one spent a decade from now, if only because the dollar could be put into the bank to grow in value. Similarly, benefits a decade from now are worth less than benefits today. For example, suppose that one program would prevent 100 breast cancer deaths each year starting now; a second program has the same costs and prevents 101 deaths each year, but there is a ten-year lag before any deaths are prevented. Which program is preferred?

7. Analysis of uncertainties: Each of the previous steps is filled with uncertainties, from quantification to discounting. The analysis should examine the uncertainties explicitly to discover which are most important for the conclusions. For example, what is the detection rate and increase in life expectancy associated with screening women aged 40–49 for breast cancer? We present below the range of outcomes from various studies of mammography. A sensitivity analysis is needed to account for the likely range of outcomes.

8. Interpretation of the results: Neither the facts nor the analysis “speak for themselves”; the analyst must interpret these for the reader.

These eight steps define an approach to gathering information and performing an analysis to inform social decisions (adapted from Reference 17). They can be thought of as providing a framework for asking questions, collecting data, and deciding which questions are most important. It is hard to quarrel with these steps for evaluating large-scale public health programs, such as removing toxic substances from drinking water or screening for breast cancer: The price is large and the health implications are potentially large. Good decision making requires a careful analysis using all steps.

For major public health decisions, such as setting drinking water standards or urging people to get more exercise, the resources required for the analysis are small compared to the social costs and benefits of selecting the best program. An elaborate analysis is not appropriate to decide the operating hours
for a public health clinic. While that decision is important, an elaborate analysis would be far more expensive than the value of the information gained.

Even when the decision does not justify an elaborate analysis, these steps provide a good template for thinking through the problem. What is the problem and what are the goals? What are the benefits and costs of alternative actions? What do available data show about these benefits and costs?

Strictly speaking, these steps constitute a BCA only if all the effects are quantified and then monetized. The benefits and costs that occur over a period of time must be brought to the present so that they are fully commensurate with current benefits and costs. These steps are necessary since a BCA requires calculating the present dollar value of benefits, the present dollar value of costs, and the net benefit of each program.

The Reagan Executive Order requires an agency to choose the alternative with the greatest net benefit, unless this is precluded by law. This requirement meant that failure to quantify or to monetize a benefit or cost implied that the benefit or cost was ignored. For example, if the psychological costs of a false positive on the breast cancer screen were not quantified and evaluated in dollar terms, they were ignored in the subsequent decision making. As shown below, there is no realistic hope of quantifying and monetizing all important benefits and costs. Unquantified or unmoneitized benefits and costs that are omitted from the analysis will bias decision makers. If the quantification or monetization is biased or done badly, decision makers will be misled.

A different problem is the distribution of benefits and costs: Who pays and who is benefitted? For public health programs, most of the costs are borne by individuals other than those who reap the benefits. The implied income redistribution can be an important attribute of a program. There is no realistic attempt to account for these distributional issues in the analysis.

Thus, a BCA emphasizes systematic analysis and use of data. No one should claim that BCA will find a social optimum. "Once benefit-cost analysis is understood as a process meant to yield information rather than to make decisions, practitioners of benefit-cost analysis need not take sides in controversies over the nature of justice" (21).

ALTERNATIVE FRAMEWORKS AND CRITERIA

BCA is one of several decision frameworks that Congress has written into legislation. Others include: (a) no risk, (b) risk-risk, (c) technology-based standards, (d) risk-benefit analysis, and (e) cost-effectiveness analysis (11). (a) No risk framework is exemplified by the Delaney Clause that forbids the Food and Drug Administration from permitting carcinogens to be added to food, no matter how small the risk or how large the benefit. When Congress adopts the
no risk framework, it is assuming that benefits are trivially small and the consequences of permitting the action are horrendous. 

(b) Risk-risk analysis is used by the FDA to regulate drugs and medical devices. Drugs and X rays have some risk of harming the patient; thus the FDA trades off the possible risk of the intervention against the possible risk of no intervention. The FDA does not account for the cost of treatment and other factors and does not compare one type of treatment to another, e.g. pharmaceuticals vs surgery for coronary artery disease. 

(c) Technology-based standards are frequently used for regulation of industrial emissions to air or water; agencies are instructed to choose a control technology with no explicit consideration of costs and health or other benefits. 

(d) Risk-benefit analysis attempts to weigh the risks of a technology or intervention against the benefits of that technology or intervention. Costs and other issues are not considered. 

(e) Cost-effectiveness analysis attempts to achieve a specified goal, e.g. lowering the infant mortality rate by 20%, at the least cost. 

(f) Benefit-cost analysis attempts to identify, quantify, monetize, and then compare all of the social benefits with all of the social costs.

These frameworks, (a)–(f), are arrayed in rough order with respect to the breadth of considerations, tradeoffs considered, and thus efficiency. The frameworks are also in rough order in terms of the amount of data and analysis required. For example, cost-effectiveness analysis requires that the various costs be combined into a total cost but does not require analysis of the benefits; the framework does not require that benefits and costs be compared (25a). Despite these limitations, cost-effectiveness will result in the socially efficient outcome if the goal chosen happens to be the efficient one, the one that would be selected by a benefit-cost analysis.

Each framework makes implicit assumptions about ethics. For example, the “no risk” framework assumes that protecting human health is the only goal; no tradeoffs are relevant concerning such other attributes as health vs private or public consumption. The no risk framework is simplistic in not recognizing more complicated implications, such as the risk-risk tradeoffs of the second framework. The risk-risk framework continues to focus only on health. Technology-based standards rest upon a naïve assumption that the technology drives the solution and that other considerations, including health and consumption, are either not important or are handled in an obvious fashion. Risk-benefit analysis examines the health risks and general benefits of a technology. It assumes that other types of risk are not relevant. Finally, cost-effectiveness is based on the premise that efficiency is the most important attribute, that choosing the desired goal is simple or not possible analytically. Although Congress could be forthright in declaring its goals, more often it expresses its values more subtly by choosing one of these decision frameworks.
DIFFICULTIES IN BENEFIT-COST ANALYSIS

A host of difficulties impede BCA. The most important difficulty is quantifying the effects of an intervention. How much additional life expectancy, on average, will result from screening women aged 40 to 49 years for breast cancer? Screening detects breast tumors earlier than they would be detected without screening. The increase in the survival rate comparing those screened with those not screened is due, in part, to detecting the disease earlier than if the woman were symptomatic. Meanwhile, some abnormalities will be detected and treated that may not have progressed to breast cancer in the absence of treatment. Women with these abnormalities bias upward the survival rate attributed to screening. In addition, these women suffer psychological trauma, a decrease in the quality of life, and may even have their lives shortened as an adverse effect of the mistaken treatment of these abnormalities.

Whether the analyst is trying to estimate the effect of breast cancer screening, suspended particles in the air, trihalomethanes in drinking water, pesticide residues in food, or occupational exposures to toxic substances, quantifying all the health effects of an intervention is difficult. We emphasize that the difficulty is inherent, not a problem caused by appealing to BCA. Sensible decisions cannot be made without good estimates of the health benefits of the intervention. Criticism of BCA tends to focus on the monetization of these effects, rather than on the magnitude of effects. This criticism is misguided insofar as the real problem is the quantification of the benefits of intervention.

Monetizing the benefits of health improvements does add controversy to the analysis. What is the dollar value of adding a year of life expectancy to the population of 50–59 year old women? Some studies look at decisions that people currently make, such as taking risky jobs or purchasing less safe products, such as cars without air bags or cars that are smaller and lighter (28). Few people are skeptical of using market prices for the supplies used in a public health clinic. There is less comfort in using market prices in other areas, such as the risk premium in the pay of underground coal miners who may have little alternative employment.

The way in which economists value most goods and services not traded in the market is via surveys that ask individuals how much they would be willing to pay to get something they desire or how much they would have to be paid to give it up (contingent valuation). This technique has been used for valuing an additional asthma attack, other kinds of morbidity, and the rate of time preference in saving lives (5, 26). People are asked how much they would be willing to pay to lower their chance of being killed this year in a highway crash by one chance in 100,000. In many cases, the answers can be compared to actual purchasing decisions that people make, such as buying cars with air bags or antilock brakes or buying larger, safer cars (28).
In some cases, monetization can be avoided by concluding that the health benefits of the intervention are small or nonexistent. No further analysis is needed. When the intervention does produce substantial benefits, they will need to be monetized in a BCA. This monetization cannot be done with mathematical precision. However, in many cases, the uncertainty about the valuation estimate is smaller than the uncertainty in quantifying the health benefits. For example, there is considerable uncertainty about how much women would be willing to pay to lower their chance of dying of breast cancer by one chance in 1000 per year. However, there is still greater uncertainty concerning the reduction in relative risk as a result of an annual screening program in 40–49 year-old women.

Since quantification and valuation are fraught with uncertainty, the question is whether the estimates help or mislead government officials. Miller has no patience with this question: “If we have given government officials the power to impose costs on the activities of private citizens or otherwise control them, we cannot say they cannot be trusted to weigh benefits and costs in determining what they do. If regulatory statutes allow discretion, the enforcing agency would be irresponsible—and in my view unfair—not to weight the benefits and costs of its policies” (16).

BCA is filled with uncertainties concerning the health benefits of the intervention, their monetization, the costs of the intervention, and other aspects of the analysis. To be helpful to decision makers, the analyst should be careful to present the data and assumptions clearly. The analyst should undertake a sensitivity analysis and state clearly which of the results are most sensitive to the uncertainties in the data and monetization.

**IMPLEMENTATION DIFFICULTIES**

Whatever the theoretical difficulties, they are dwarfed by the difficulties of implementing BCA. For example, the tool has been criticized as biased and as serving the incumbent politicians. However, “any technique employed in the political process may be distorted to suit parochial ends and particular interest groups. Cost-benefit analysis can be an advocacy weapon, and it can be used by knaves. But it does not create knaves, and to some extent it may even police their behavior. The critical question is whether it is more or less subject to manipulation than alternative decision processes. Our claim is that its ultimate grounding in analytic disciplines affords some protection” (13).

In practice, BCA is an unimaginative, bookkeeper’s activity. The tool seems to encourage a narrow analysis. All too often economists miss the primary benefits (or costs) in doing the analysis. The conceptual difficulties are the grist for academic theorists. Reality is more bleak. In a few cases catching national attention, BCAs are careful and complete (16a). Even these analyses
are subject to the limitations described above. More generally, BCA is done by apprentice analysts without the time or resources for careful analysis (8). Still, a conceptually correct BCA that is done quickly can be helpful, even in a limited way. Conversely, analyses that are not conceptually correct can mislead. For example, a novice analyst might count the jobs created by an air pollution control program as a social benefit; in fact they are a social cost. The benefit consists of clear air; program expenditures needed to achieve clear air are a cost, not a benefit.

In practice, what decision makers learn from BCA comes from the executive summary. But no one- or two-page summary can indicate the range of uncertainties and other qualifications that a decision maker must know to use the analysis intelligently. Furthermore, reading the report might detect fundamental conceptual flaws in the analysis, something the summary is unlikely to indicate.

CASE STUDIES

To complement the theoretical discussion above of the strengths and weaknesses of BCA, we illustrate the tool by applying it to three case studies.

A: Regulating Coke Oven Emissions

Coke making is the first step in steel making. It involves heating coal to a high temperature in an airtight chamber. The impurities are volatilized, leaving almost pure carbon. The volatilized coke oven gas (COG) is a complex mixture of over 10,000 different chemicals in the form of gases, vapors, and particles. Modern coke ovens collect the COG produced in coke making and further process it or use it as a fuel. However, some of the COG is released to the environment during charging, discharging, and cooling operations, and due to leakages through imperfect seals around doors and pipes (fugitive emissions). Among the diseases associated with COG are lung and urinary tract cancers, skin tumors, bronchitis, and emphysema. There is strong scientific evidence, based on epidemiological surveys, animal studies, and chemical analyses, that coke oven emissions are carcinogenic at exposure levels experienced by coke oven workers. Since these emissions pose health hazards for both the workers and the populace living around coke ovens, they have been regulated by OSHA and the EPA.

BCA can be very useful in evaluating these regulations. However, estimating the benefits of regulating industrial emissions such as COG requires data on current emissions rates, prospective control technologies and the extent to which they reduce emissions, the ambient concentrations associated with different emission rates, populations exposed and the exposure rates, the dose-
response rates for the relevant diseases and populations, and the timing and intensity of various health effects. Estimating costs requires information on costs and disruptions, including productivity changes, of each relevant control technology.

Here we outline the procedures used by EPA in developing coke oven regulations to highlight some of the subjective decisions and assumptions often necessary in carrying out a practical BCA, which may not be apparent from the executive summaries and how systematic biases in these can misinform and mislead (see References 12, 23 for more detailed analyses).

To estimate the uncontrolled emission rates, the EPA study assumed that coke ovens operate at full capacity 90% of the time. Three control options were considered that were estimated to cost $7 million, $19 million and $81 million per year of operations, and reduce annual coke oven emissions to 450 Mg, 420 Mg, and 100 Mg, respectively, from base emissions of 720 Mg. EPA chose to use analytic diffusion models instead of actual site measurements to estimate ambient concentrations due to these emissions. Generic air dispersion models were used to estimate the ambient air concentrations in an area within a radius of 50 km from each source. EPA also developed some site-specific dispersion models based on local meteorological conditions but did not use them in their final analyses. To estimate lifetime exposure to COG, EPA assumed that individuals are exposed 24 h per day, 365 days per year for 70 years to these ambient concentrations.

Similarly, EPA used a series of conservative procedures to arrive at the estimates of cancer risk from these exposures. NIOSH and the steel industry conducted an elaborate ongoing epidemiological study that discovered and quantified the excess cancer deaths for coke oven workers (14, 15, 22). One major limitation of these epidemiological studies was that they collected no data on individual exposure levels or cigarette smoking habits. EPA risk analysis assumed that coke oven workers had smoking habits similar to other steel workers and that exposure levels were the same as those observed in coke plants in the 1960s, both of which are likely to overstate risks. The next step was to estimate the cancer risks to the population that is exposed to much lower doses than were the coke oven workers for whom data were available. The EPA study chose to extrapolate using the linear functional form, which gave the greatest estimated risk level. However, the quadratic functional form fitted the data best and gave much lower estimates for disease incidence. EPA followed its standard procedure of using the linear dose-response relationship and the upper bound of the 95% confidence interval, instead of the maximum likelihood estimates on the parameters to calculate unit risks of exposure. The number of cancer deaths averted was calculated based on the population living within the 50-km radius. Under these assumptions, control option 1 was estimated to save 2.6 lives per year at a cost of $2.7 million/life saved; option 2,
an additional 0.3 lives at $41 million/life saved; and option 3, 3.1 lives at $26 million/premature death averted. Using maximum likelihood estimates for the linear dose-response relationship or the quadratic dose-response relationship would have lowered the estimated number of deaths by at least factors of 10 and 19; thus the cost per premature death averted would be $27–51 million for option 1, $410–800 million for option 2, and $260–500 million for option 3 (12).

The process of cascading conservative estimates at each stage led to risk estimates that are orders of magnitude greater than estimates based on more realistic or "best" estimates. The result is regulation calling for much greater control—and greater expense. Lave & Leonard (12) argue that true cancer risk to people living around coke ovens may be almost zero and the most likely cost per life saved is in the order of $270–$4100 million. Reuter & Steger (23) studied geographical patterns of cancer incidence around coke ovens in Allegheny County and found no evidence of increased cancer risk due to closeness to coke ovens.

**B: Mammography Screening for Women under Age 50 Years**

The effectiveness of mammography screening in reducing breast cancer mortality in women older than 50 years is well established. In contrast, mammography screening for women under age 50 years has been the subject of a continuing debate among medical professionals. Twelve medical groups, including the American Cancer Society (ACS), American Medical Association (AMA), National Cancer Institute (NCI), and American College of Radiology, recommend mammography at 1–2 year intervals for women in the age group 40–49 years, whereas the US Preventive Services Task Force, the Canadian Task Force on Periodic Health Examination, the International Union Against Cancer (UICC), and the American College of Gynecology and Obstetrics declined to recommend mammography for women with average risk in this age group (3).

The issue is important. Estimates of new cases of breast cancer found each year in this age group vary from 18,300 (7) to 29,000 (6). There are about 17 million women in this age group. Assuming 25% coverage and a cost of $100/screen, the annual cost of screening at one-year intervals would be about $425 million.

The difference in recommendations concerning screening may be due to differences in goals. Possible goals of early detection, preventing premature deaths, improving the well-being of women, appear to be similar, but actually differ significantly. The American Cancer Society appears to focus on the first goal: "American Cancer society feels that the value of screening transcends cost .... As long as mammography represents the only method by which breast
cancers can be diagnosed in a subclinical stage, adequate screening opportunities must be provided to women as a whole” (6). The implicit assumption is that early detection will improve survival rates and well-being. Kerlikowske et al focus on the second goal when they state: “The goal of screening mammography is to reduce mortality from breast cancer” (10). For this goal, improving treatment after detection is an alternative to screening for early detection. Costs are not important in either goal. ACS explicitly rejects consideration of costs, whereas Kerlikowske et al do not comment on costs. The third goal focuses on improving well-being, which suggests an examination of primary prevention through changes in lifestyle, such as diet, instead of focusing only on early detection (secondary prevention). Indeed, changing lifestyle might be more cost-effective than screening.

Estimating the benefits of mammography screening requires quantifying the following linkages: the relationship between prevalence and incidence of breast cancer by age group and by other risk factors such as genetics, personal habits, and medical history; the relationship between timing of screening procedures and proportion of cancers detected by various combinations of procedures; additional risks due to screening procedures; lead time gained through screening and the effectiveness of available therapy in terms of reductions in morbidity and mortality as a function of time of detection. The outcomes of the screening program include reduction in premature deaths due to early detection; gains/losses in quality of life due to changes in morbidity levels; the extra years of life gained; premature deaths due to screening radiations; and unnecessary biopsies in the case of false positive screenings. Quantifying the (financial and other) costs of screening requires consideration of the costs of false positives and additional cancers due to screening radiation.

Breast cancer is a relatively well studied problem. However, evidence on the most important health outcome, i.e. effectiveness of routine mammography screening in reducing mortality in this age group, is inconclusive. Table 1 summarizes the estimated relative risks of mortality for the screened with respect to unscreened populations, obtained in major studies of breast cancer.

Kerlikowske et al (10) find that the overall summary relative risk of cancer deaths in the screened and unscreened populations arrived at by the meta-analyses of all the nine studies was 0.93 (95% CI, 0.76 to 1.13), and for randomized controlled trials was 0.92 (95% CI, 0.75 to 1.13). They also find that the studies that had 10-12 years of follow-up had a lower summary RR than those that had only 7-9 years follow-up (0.83 and 1.02, respectively). Smart et al (24) conduct meta-analysis of a subset of the above trial results and estimate the relative risk to be 0.84 (95% CI, 0.69 to 1.02).

Screening appears to increase the risk of cancer death in the short run. In general, the RR declines with increased number of years of follow-up.
Table 1  Relative Risks (RR), odds ratios and 95% confidence intervals (CI) of screening mammography in women aged 40-49 years: results from major studies.\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Study</th>
<th>RR and CI</th>
<th>Years follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Edinburgh</td>
<td>0.98 (0.45–2.10)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.78 (0.46–1.51)</td>
<td>10</td>
</tr>
<tr>
<td>2. Malmo</td>
<td>1.29 (0.74–2.25)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.51 (0.22–1.17)</td>
<td>12</td>
</tr>
<tr>
<td>3. Kopparberg</td>
<td>0.76 (0.32–1.77)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.75 (0.41–1.36)</td>
<td>12</td>
</tr>
<tr>
<td>4. Ostergotland</td>
<td>1.03 (0.50–2.10)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1.28 (0.76–2.33)</td>
<td>12</td>
</tr>
<tr>
<td>5. Canadian I</td>
<td>1.36 (0.84–2.21)</td>
<td>7</td>
</tr>
<tr>
<td>6. HIP</td>
<td>0.81 (0.53–1.24)</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0.77 (0.50–1.16)</td>
<td>10</td>
</tr>
<tr>
<td>7. Stockholm</td>
<td>1.04 (0.53–2.05)</td>
<td>8</td>
</tr>
<tr>
<td>8. Gothenberg</td>
<td>0.73 (0.27–1.97)</td>
<td>7</td>
</tr>
<tr>
<td>9. Nijmegen</td>
<td>1.23 (0.31–4.81)</td>
<td>8</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Source: Reference 10, Table 2
\textsuperscript{b}All except Nijmegen are randomized control trials, while Nijmegen is a case-control study.

However, the changes in mortality risks are not statistically significant, even when results from many studies are combined to gain more statistical power. Kerlikowske et al (10) hypothesize that the key to increases in mortality in the short run, but better performance with longer follow-up, may be the timing of menopause. They suggest that screening begin at menopause, or at age 50 if menopause has not occurred. However, if the cancers occurring before age 50 are more aggressive, annual screening may not be frequent enough to detect the tumor early enough for effective intervention. If so, more frequent screening would be required. The higher frequency of screening must balance the risks of additional screening against the benefits of early detection (see below).

A detailed BCA would also require relative risk estimates at various time intervals of follow-up. Eddy et al (7) combine results from four short-term studies (BCDDP, Swedish, Nijmegen, and Florence) and one longer-term study (HIP in New York), using “confidence profile” method, and estimate that annual mammography for women aged 40–49 years would decrease mortality by 6% at five years, 14% at 10 years, 19% at 15 years, 22% at 20 years, and 26% at 30 years, and add an average of 3.5 years to the life of a woman destined to get cancer. However, they warn that there is considerable uncertainty about these estimates. They estimate that the savings in initial and terminal treatment costs due to early detection are very small (about 1%)
compared to the total costs of a screening program. They do not attempt to value the quality of the life in the additional years of life gained.

Screening is not risk free. Exposure to screening radiation may increase risk of cancer. Eddy et al (7) estimate the lifetime risk of cancer due to ten screening radiations at 1/25,000, and the natural risk of cancer in the age group 40–49 years at 128/10,000. As seen from Table 1, the reduction in mortality due to screening can vary from 2% to 20% of this risk, i.e. about 6/25,000 to 102/25,000. In other words, the risk induced by screening radiations can alone offset between 1.6% to 16% of the benefits of screening. Confirmatory testing of false positive screens may impose further avoidable risks, and there will be roughly 100 false positives for every cancer case detected (at 1% FP rate) or between 50 and 226 FPs per cancer death avoided (7). In addition to risks due to biopsies and costs of testing, there are nontrivial psychological costs of false positives.

In conclusion, in breast cancer screening in younger women, the main tradeoff is between the likely number and quality-of-life years saved, and the costs and risks of the screening program. The current evidence is too weak to draw firm conclusions. There is a good case to be made for putting more resources into primary prevention compared to screening.

C: Promotion of Exercise as a Preventive Health Measure

Of the three examples, defining the goals is most important for promoting exercise. The goals might be: (a) the reduction of mortality and morbidity, (b) reductions in health care expenses, (c) improvement of well-being, (d) improvement of appearance, or (e) reducing obesity and eliminating sedentary lifestyles. The choice of goals is likely to shape the type of intervention that is desired. If we assume that the goal is to improve longevity by reducing CHD deaths, we neglect reductions in morbidity, improvements in fitness and improved lifestyles; a BCA should account for all the outcomes.

The linkage between physical activity and longevity is as follows: Physical activity→physical fitness→higher quality life→low hypertensive-metabolic-arteriosclerotic disease (HMAD) risk→longer life (18, 19). However, these linkages are confounded by interrelationships between different types of physical activity (leisure time, occupational, and other chores), levels/intensities of activities, different types of fitness, (morphological, muscular, motor, cardiorespiratory, and metabolic), heredity, lifestyle (sleep, stress, smoking), personal attributes (age, sex, motivation), nutrition, physical environment (weather), social environment (marriage, children, pets, other social networks), and different physiological mechanisms through which increased fitness affects health outcomes in terms of mortality, morbidity, and wellness. Additional complications arise due to the dynamic nature of exercise programs and progression of CHD, and health risks due to exercise itself (adapted from Reference 2).
The problem of accurate and consistent measurement of these factors is not a trivial one. Hence it is almost impossible to design studies that control for all the factors and also establish causality.

So far, most research pertaining to the relationship between physical activity and CHD has been epidemiological; many studies are only suggestive because of small numbers of subjects or failure to control for important factors. In all studies, uncontrolled factors mean that the observed association may not be causal. The studies, especially the earlier ones, suffer from problems of self-selection, crude measures of exercise and CHD, dependence on self-reports, and inadequate control of confounding and covariate factors. Very few of the studies are RCTs (1, 18-20). The estimated summary relative risk of CHD associated with inactivity varies among studies, but generally ranges from 1.5 to 2.4 with a median of about 1.9 (20). Though it is well accepted that exercise improves fitness defined as capacity for work, little is known about how exercise and improved fitness protect against heart disease. Researchers have hypothesized that exercise affects CHD by preventing development of atherosclerosis or the accumulation of fatty deposits on blood vessels; or through secondary prevention by affecting blood pressure, HDL levels, blood coagulation, collateral blood flows to the heart, etc. However, the evidence is weak (4, 25).

As with any preventive activity, there are risks associated with exercise. They include increased risk of CHD events during and after exercise, and risk of muscular and skeletal injuries. These risks vary by age, sex, and exercise intensity. Other risks typically not considered include increased road and other accidents, dog bites, increased effect of urban pollution, heightened risks in conventional diseases if one exercises during illness, osteoporosis, cumulative damage to heart muscles due to exercise during infections, etc. Few data exist to quantify these risks. Also, CHD incidences attributable to exercise are likely to be underreported due to delayed effects, confusion with normal aches and pains of exercise, and deliberate concealing when they occur in commercial facilities.

The financial costs of exercise would include direct costs of equipment, facilities, counseling, initial medical evaluation, and medical care for injuries. Indirect costs include the value of time spent exercising and time lost to injury.

Inability to quantify the medical benefits and costs means that there are very few full-fledged BCAs of exercise programs. We could find one study that attempted a CEA of exercise as a health promotion activity. Hatziandreu et al (9) made several assumptions in their CEA to overcome data limitations. For example, they used the age-specific incidence rates of CHD from the Framingham study to calculate separate rates for exercisers and non-exercisers, by assuming that 10% of all men exercise, and that exercise reduces the rate of CHD by 50%. Similarly, they assume that people who are neutral toward
exercise value the time spent in exercise at the prevailing wage rate, those who dislike it at twice the wage rate, and those who like exercise as costless. They consider only the medical costs and lost wages due to increased risk of injury as the value of risks associated with exercise. They estimate the direct and indirect cost per CHD death avoided at $250,000.

In conclusion, in all these cases the main difficulty is in quantifying the health effects with adequate accuracy and establishing a prima facie case for the intervention. The issues of valuation, distributional justice, ethics, and achievement of social optimum, which have been the focus of much of the academic and political debate, are of secondary importance in many practical BCA in public health.

CONCLUSION

Benefit-cost analysis can be an extremely useful tool for public health professionals. It is a systematic, scientific approach to evaluation of programs; it uses analysis rather than emotion to inform difficult decisions. BCA focuses on the need to justify programs before recommending them to the public. It can serve to convince political leaders and the public of the social benefits of public health programs.

Despite the real advantages of BCA, one must take care that it is used properly and the results are not over-interpreted. For example, the Reagan Executive Order directing agencies to choose the alternative with the greatest net benefits makes no sense either in theory or practice. Public health professionals and political decision makers must be aware of the limitations of the tool and know how to interpret the analysis.

Literature Cited

mography screening in women under age 50 years. *JAMA* 259:1512–19

### Physical Activity, Fitness, and Health