On Pain

If you come to think about it, physical pain has many singularities. Of all human experiences it is, as long as it lasts, the most absorbing; and it is the only human experience which, when it comes to an end, automatically confers a real if not perhaps a very high kind of happiness. It is also the only experience this side of death which is by its nature solitary. But the oddest thing about it is that despite its intensity, despite its unequaled power over mind and body, when it is over you cannot really remember it at all.

My Aunt's Rhinoceros and
Other Reflections
Peter Fleming
Simon and Schuster, 1956.

Measuring the Effectiveness of Prevention: I

JUDITH R. LAVE AND LESTER B. LAVE

School of Urban and Public Affairs,
and Graduate School of Industrial Administration,
Carnegie-Mellon University

Introduction

The old adage, "An ounce of prevention is worth a pound of cure," is not a reasonable guide to the allocation of resources in the health care field. If it were, and we had preventive measures for all diseases, then health care would consist only of prevention; acute health care would only be a dim memory. Unfortunately, effective preventive techniques are lacking and the adage is misleading because it lumps into one category—prevention—a host of different kinds of program.

The basic public health programs of developing sewage disposal systems and potable water supplies surely returned more than sixteen units of health for each unit of prevention. Some current inoculation programs, such as those for measles and polio, have high payoffs. Incidence of measles continues to be a problem and bears an inverse relation to immunization. Cases of polio have almost disappeared; on the other hand, to discontinue immunization would almost surely result in serious recurrence. Vaccination against smallpox is now unnecessary because there is so little incidence of the disease that there is no reason to put up with the complications caused by the vaccination process. Other programs such as the screening for cancer of the cervix are controversial with respect to efficacy, periodicity, and the appropriate target population. Finally there are a series of programs, such as the installation of seat belts in automobiles and the treatment of asymptomatic hypertension, which are highly efficacious in theory, but, since they require patient compliance, they have proved to have low efficacy in practice.

Preventive health care thus encompasses a vast array of programs. Some of these are implemented within the traditional medical care system (inoculations); others have no close relationship with this system (seat belts); some require no patient action or deci-

sions (sewage disposal); others need the patient's complete cooperation (management of asymptomatic hypertension). Preventive programs have different levels of effectiveness, and different factors affect the level of a particular program's effectiveness both in theory and in practice.

As a society we could not possibly afford the cost of all the programs that might offer some potential of disease prevention. We have to decide, both collectively and individually, how many physician hours, drugs, patient time, and regulatory efforts to devote to each preventive program. The most effective prevention programs were, in general, implemented first. Through primary prevention it was possible to lower the infectious disease mortality and morbidity rates. We are now in the fortunate, but more difficult, position of choosing among programs which have much lower levels of effectiveness; the apparent lack of dramatic opportunities is one measure of how far we have come. Our higher incomes lead us to demand health even more than our grandparents did and so, even though current preventive programs do not have such dramatic effects, we are willing to pay more for each unit of prevention.

Many programs compete for our attention, and some of them are of doubtful value. In order to make intelligent choices about which programs to fund collectively (or to purchase as individuals) it is important to gain as much information as possible. In this paper we describe several of the aspects of program evaluation. We go on to describe some problems that appear to be specific to the evaluation of health care programs. Next we present some of the results of the extant literature on the evaluation of preventive services. We conclude with some recommendations.

Before we begin, we should point out that evaluation of acute health care programs is as critical as evaluation of preventive health care programs. The proper balance of acute and preventive services cannot be achieved without a careful, comparable analysis of both.

A Primer on Evaluation

There are four basic steps involved in any evaluation. (1) List the effects of the program (both beneficial and negative effects). (2) Relate the effects quantitatively to the program. (3) Translate the effects into some single metric. (This includes comparing inputs and outcomes both at present and over time.) (4) Compare the positive

and negative effects of the program (Dasgupta and Pearce, 1972; Haveman et al., 1973; Klarman, 1974).

Consider the Clean Air Act of 1967; the objective was to improve air quality (Lave and Seskin, 1970, in press; Stern, 1968; Environmental Protection Agency, 1972). The first step in an evaluation of the Clean Air Act involves listing its effects. The cost of implementing the act includes: the capital cost of the abatement equipment that must be constructed and added to existing and new equipment; the operating cost of the equipment (as well as any increased operating costs of old equipment); the cost of redesigning plants to enable them to obtain pollution control; and the cost of possible plant shutdowns and the resulting short-run unemployment due to the new emission standards. In addition, there is the cost of further research and development that must be done in order to achieve emission standards. The benefits we hope to derive from cleaner air are many. These include: lower morbidity and mortality rates; improved visibility; lower cleaning costs; longer lives for some materials; less damage to plants and animals; and a generally improved quality of life.

The second step in the evaluation is to relate the benefits and the costs of the programs to the degree of abatement and to quantify the relationships. Ideally one would want to know both how much it would cost to attain a given level of air quality (stage 1, costs versus emission levels, and stage 2, emission levels versus ambient air quality) and the benefits associated with each level of air quality. The air cost-quality relationship is difficult to ascertain but not so difficult as the air quality-benefit relationship. This is not only because there are many pollutants, but also because each pollutant has a different effect.

The third step is to translate the effects of the program into a single metric. The economists' metric of choice is the dollar. The costs of implementing the program can be readily monetized—in fact most of the costs of implementation, with the exception of the temporary increase in the unemployment rate and design cost, are already stated in dollars. Many of the benefits of the Clean Air Act, however, cannot be readily translated into this metric. We can estimate the benefits of lower cleaning costs and reduced damage to buildings by the dollars saved on cleaning and building repair. We can value the benefits of a decrease in visits to the physician and hospital episodes (as a result of improved health status) by the

decreased expenditure on these services. It is more difficult to monetize the value of improved visibility. Other benefits, such as improved health status and improved life expectancy, can be measured in dollars only by making controversial assumptions.

If we are content to stop with the listing of the effects as determined in step two, we can avoid this whole issue; this approach may be reasonable if only a single program is being considered. However, such an approach is not reasonable if many programs are being compared. In order to evaluate and compare many programs directed at similar problems, it is necessary to group at least some of the benefits into categories (for example, person years saved and disability days saved). Then it becomes possible to analyze the programs in terms of how much improvement is achieved in each health category for each dollar spent.

However, stopping short of measuring all benefits in the same metric is unsatisfactory. To compare the efficacy of a number of different kinds of programs, some method of aggregating the benefits, either an intuitive one or a formal one, must be used. Intuitive aggregation is inherently suspicious since individuals differ in their judgments; certainly a casual judgment is inappropriate for decisions involving lives and millions of dollars. We would not pretend that aggregating benefits is easy or without controversy, but we see no real substitute for attempting to translate all the effects into one common metric so that all can be compared.

As noted above, most economists prefer to use dollars as their common metric. We hesitate to use such language before a distinguished audience from the APHA: dollars somehow seem crude and incommensurate with human suffering, particularly when we think of weighing green pieces of paper against lives. However, dollars represent real resources that can be used to produce other health care programs, other kinds of public services, or private consumption goods such as food, clothing, or even entertainment. Perhaps a better way of phrasing the question is to ask whether resources should be used for a particular preventive program, for general medical care, for improving education, or for increasing consumption by the poor. The dollar is a good metric for such comparisons. This is not to say that we know how to translate all health benefits into dollars; that question (as discussed below) is the focus of much research.

Once the benefits and the costs have been determined and

perhaps monetized, the fourth step is to compare the benefits to the costs. At this point an assessment can be made as to whether an ongoing program (the Clean Air Act) is worthwhile at its current level or whether a potential program should be implemented and if so at what level. (Such an assessment can also be made at the end of the second step.) The decision maker must consider each step of the analysis very carefully. Given how complicated it is to perform a good evaluation, it is highly unlikely that perusal of a one-page summary would be sufficient.

From the amount of criticism directed at the various approaches used to monetize the benefits of health programs, one would think that the third step is the most difficult step. In actuality, the most difficult, as well as the most important, step of the four is the second. In order to relate the effects quantitatively to a program, one must first establish causality. For example, does cigarette smoking cause lung cancer (Sterling, 1975)? Does cigarette smoking cause cardiovascular disease? Does soft water cause heart disease? Do chemicals in the environment cause most of the cancers (National Academy of Sciences, 1974)? Do screening and early case finding lead to decreased morbidity and mortality rates? And in the case of the Clean Air Act, how much disease does chronic air pollution really cause?

A discussion of these causal links reads like a textbook on epidemiology or on the philosophy of science. If A and B are associated, there are four logical conclusions: (1) the effect merely arose by chance, (2) A causes B, (3) B causes A, and (4) there is some other factor or set of factors, C, which is the cause of both A and B. The fourth explanation, spurious correlation, is the most difficult to disprove. Whatever might be said about the relative merits of retrospective evaluation, prospective evaluation, or laboratory experiments, the only basis we have for judging causation is post hoc ergo propter hoc. However, one must be careful not to cry causality without a painstaking analysis of the data and with careful attention to alternative explanations (Hill, 1965; Blalock, 1964; Lave and Seskin, in press).

Another reason why the second step is difficult is that, even after causality has been ascertained, each effect must be quantitatively linked to the program. By the quantitative link, we mean the "dose-response" curve. For example, the Delaney amendment, which bans any food additives shown to be carcinogenic, makes no

sense to us for it does not consider the dose response curve. A food additive which can be shown to be carcinogenic in a dose a million times greater than would ever be consumed by man is to be given the same attention and response as an additive which is carcinogenic in doses which are commonly taken. The importance of the quantitative link is underscored by the outcry generated by the recent FDA ban on saccharin.

Biostatisticians have tended to concentrate on the first part of step two (the establishment of causality) and have neglected the second part (the estimation of the quantitative association). Both steps are important. Both parts are necessary for a complete evaluation.

Unresolved (Perhaps Unresolvable) Issues

Economists have directed their attention to some of these problems; other problems have been investigated by other disciplines. We begin by discussing a noneconomic problem.

1. The Ethics of Data Collection

An inherent problem in evaluating any preventive program is the potential lack of data. Theoretically, the clinical trial is the classic research design for obtaining such data but in practice there are often problems in implementing and sustaining a trial. If physicians suspect a technique is more (or less) effective than existing techniques, they tend to use it exclusively (or not at all). It is difficult to reconcile one's obligation to an individual patient with one's obligation to knowledge in favor of the latter. Since human health and welfare are at stake, ethical issues must arise in a clinical trial. The cost of running clinical trials is only one reason why more are not done.

Whatever the difficulties, clinical trials are essential. Current medical practice is filled with treatments of preference for which there is no more than a shred of evidence of superiority over alternative treatments. Worse, there have been treatments of preference later shown to be ineffective or even pernicious, such as radiation of the thyroid glands in children. This is not the place to discuss questions of informed consent or the obligation of the physician to do his best for each patient; however, there are many examples where shortsighted considerations or overzealous treatment has harmed

rather than helped patients. For example, universal screening for breast cancer using mammography has come under fire recently. Some experts have argued that the incidence of breast cancer in young women (not otherwise identified as high risk) is so low, and the additional detection power of mammography sufficiently small, that screening of young women winds up saving fewer lives than are destroyed by the screening test (through cancer caused by radiation from the test). A diagnostic tool useful for screening women with specific symptoms, or even useful for screening high risk women, may be pernicious when used for screening low risk women.

2. *The Value of Health*

We noted that if one is going to translate the health benefits of a preventive or acute medical program to a common metric, that metric will probably be dollars. To some it is evident that health is priceless and that we cannot put a value on it (Schelling, 1968; Hirshleifer et al., 1974). However, taken literally, this would mean that society should not cease to expend resources to improve health until there is absolutely no possibility that further expenditures would improve health. It is obvious, however, that we all act as if there are tradeoffs between health and other goods as we allocate our resources, privately and publicly. Some of us choose to engage in risky occupations, to participate in risky sports, or to drive cars less safe than others we might buy.

Determining a value for health, for a life or for a certain state of health, has been discussed since the turn of the century and has been the subject of extensive literature (Dublin and Lotka, 1930; Rice, 1966; Acton, 1973; Cooper and Rice, 1976; Thaler and Rosen, 1973; Berg, 1973; Mishan, 1971). At the moment we simply do not have good measures. For some restricted purposes the present value of wages may be a good approximation. However, there are situations where wages do not provide a good indicator. For example, excessive noise in the workplace leads to partial deafness with a latency period of twenty years. There are little or no lost wages or additional medical costs associated with the deafness, but, it seems evident that individuals would be willing to pay a great deal to avoid this disability. In such cases the value of wages plus medical costs is simply an irrelevant indicator of the value of health. In other cases, such as comparing benefit-cost ratios for programs directed at dif-

ferent demographic groups, the measure is completely inappropriate. Society doesn't agree with the implications of wage comparisons that show children and homemakers to be worth little and retired people to be worth nothing.

3. The Discount Rate

In any program, when the benefits or costs accrue over time they must be made commensurate and discounted back to a single period through the use of an interest rate. However, we observe many different interest rates at a point of time. Consumers face interest rates as high as 18 percent when they borrow and as low as 5 percent when they save. Corporations attempt to get returns of 30 percent before taxes (but have averaged only about 10 percent after taxes). When funds are to be spent on a public project, the relevant questions are: "Where do these funds come from?" and "What rate of return are we foregoing by investing in this project?" Since these questions are difficult to answer, economists have advocated using a range of discount rates, perhaps from 5 to 10 percent, in order to see whether the project evaluation is sensitive to the precise discount rate used (Dasgupta and Pearce, 1972; Klarman, 1974). Many projects will not be attractive at even the lowest discount rate and other projects will be attractive at even the highest discount rate.

Making benefits and costs commensurate in different periods is even more of a problem for health care projects. A consumer may discount future health states at a rate different from future consumption of goods and services. For example, a consumer who saves and has funds in a savings account which earns 5 percent would make consumption expenditures commensurate in different periods by using a 5 percent discount rate; that is, such a consumer could exchange consumption in one period for consumption in another by applying a 5 percent annual discount rate. At 5 percent interest, $1 is worth $3 twenty years hence. This means that the consumer would be willing to trade $3 worth of goods and services twenty years from now for $1 now or vice versa. However, such a consumer might regard disability or death twenty years from now as being much closer in value to disability or death today. It is not evident that future adverse health states would be discounted at the same rate. Some consumers would spend almost as much to protect themselves against adverse health outcomes that have a twenty-year latency

period as they would against immediate adverse health outcomes; other consumers would be willing to spend nothing to protect themselves against adverse outcomes that have a twenty-year latency period.

4. Volunteer Services

Two of the problems discussed above were related to the monetization of benefits; similar problems can arise in monetization of costs. In many health care activities, resources are provided at costs below market prices. In the extreme case, volunteer workers perform activities that would be costly to purchase. For evaluation purposes all resources should be valued at their opportunity cost. Thus, if a facility is given rent-free for a project, the correct value is the rent that the facility could have earned in an alternative use.

This problem occurs in a slightly different form when one program is piggybacked on another. For example, in an outpatient clinic the "slack time" of a nurse or a doctor could be used for a test or an inoculation. However, if an additional nurse or physician would have to be hired once the program changed from experimental to service status, then the cost of the latter ought to be used in determining the costs of the program rather than the fact that the enthusiasm of current participants means that no salary costs are incurred.

5. Errata

i. Evaluations are often filled with fallacies. One common fallacy is that the creation of new jobs, whether for deliverers of health care or construction workers on a new facility, is a benefit, not a cost. The wages should be counted as a cost since many programs compete for funds; because we cannot undertake all projects, we should do first those with the greatest benefit relative to cost.

ii. One must distinguish between a program's potential effectiveness and its probable effectiveness when widely implemented. Programs which are effective in clinical settings with highly trained personnel (or other controlled environments) may be less effective in the field. Consider potable water, sanitation, and seat belts. The full potential of improved water quality and sanitation can be achieved because success does not depend on individual compliance. In con-

trast, because the automobile seatbelt as a protection system relies entirely on individual compliance (only about 25 percent of automobile occupants have their seatbelts fastened), a potentially efficacious system is partially ineffective in practice.

iii. One must not assume that if approaches *A* through *Y* don't work and *Z* is the only remaining approach, that it must work. At the National Conference on Prevention in June 1975 there were some discussions that smacked of this fallacy. All participants agreed that if people followed healthful lifestyles, their health status would improve. If possible, it would be more effective to change the personal habits of individuals in order to prevent disease than to try to repair ill health with personal medical services. Some people then concluded that health education is the most effective means for health improvement. While we are mildly skeptical about the efficacy of health education, the point is that it is not logical to jump from a conclusion that health education could be effective to the conclusion that health education will be effective.

Is Evaluation Helpful?

Before we summarize some of the results of prior evaluations of preventive programs, we should again raise the fundamental question: is evaluation helpful?

The list of problems above appears formidable. Added to these should be the fact that in many instances not much information is available to a physician or to a public official faced with a decision, for example, to ban a suspected carcinogen. In recent years many suspected substances have been banned with hardly a shred of evidence of their harmfulness, much less sufficient information to conduct a careful evaluation. Furthermore, new treatments are often used before there is conclusive evidence of their efficacy.

There is an inherent difference between the task of a public health official or physician faced with an immediate health problem and the scientific questions of causality and efficacy. While we would not want to imply that decision makers should always wait for conclusive evidence on causality or efficacy (indeed, the cost of waiting might be prohibitive), we would also not want the scientific questions neglected. Prudence dictates that decision makers often act before conclusive evidence is accumulated; but scientific investigation should not stop with such a decision. Evaluation is neces-

sary in order to answer the scientific questions and thereby to determine if the original prudent decision was correct or should be revised.

Even if conclusive information is unavailable, the public health official and the physician (and the patient) can benefit from the evaluation. Listing the effects of the decision and attempting to relate each to the scope of the program is valuable, even when one can only guess at the facts. Even without data, many flights of fancy are grounded by correctly formulating the problem in terms of inputs, outcomes, alternative uses of the resources, and alternative risks.

The difficulties in evaluation should be viewed not as an apology that we can do no better, but rather as a goad to improve this technique and to apply it more widely. We already spend over $130 billion on health care each year. There is no substitute for posing the questions correctly and for getting the best possible answers, even if all of the questions cannot be answered satisfactorily. Systematic analysis is better than visceral reaction.

Some Results

We co-chaired a task force on the Economic Impact of Preventive Medicine which reported to the National Conference on Prevention in June 1975 (Lave et al., 1975). With the help of an excellent committee, we went through much of the literature evaluating preventive health care programs. It seems that 1974 and 1975 were the years for reviewing prevention; in 1974 the New York Academy of Medicine's annual conference was devoted to prevention and health (New York Academy of Medicine, 1975) and from October 5 through December 21, 1974, the *Lancet* published a series of articles on screening (for a summary, see Holland, 1974). All three reports are recommended. Below we summarize some of our results.

Overall Impression

There were few analyses in which the four proposed steps above had been followed. The major exceptions were the evaluations of the vaccination programs (Axnick et al., 1969; Witte, 1974) and the evaluation of the multiphasic program at Kaiser-Permanente in Oakland, California (Collen et al., 1973). (In these studies, the

benefits were monetized at the market wage rates.) Most analyses stopped at trying to ascertain if the proposed intervention had an effect; that is, changed the health status of the population at risk. The difficulty that investigators had in establishing this association reinforces our argument that the second step of the evaluation is the most difficult and most important. The costs of the programs were often not identified. In addition, it was rare that varying degrees of intervention had been applied so that one could estimate the "dose-response" curve.

Environmental Control

Although we did not review any of the work on the efficacy of the traditional public health kinds of measures, we have no doubt that they were effective. Indeed, in a cross-national comparison, Stewart (1971) found that funds allocated to public health programs account for a substantial variation in mortality rates across the countries and are more important than the number of physicians.

Air pollution and water pollution come closest in concept to the environmental problems faced at the turn of the century. There is a good deal of evidence that lowering air pollution levels would lead to a significant decrease in morbidity and mortality in urban areas (Lave and Seskin, in press; EPA, 1972; Amdur, 1976; Peskin and Seskin, 1975). The epidemiologic evidence has been reinforced by studies in toxicology and occupational medicine. Considering current costs of abating air pollution (and monetizing the health benefits at wage rates), it is estimated that the benefits are considerably greater than costs for enforcing the particulate and sulfur oxide standards mandated by the Clean Air Act. Abating air pollution is a high priority program.

Abatement of water pollution is a lower priority program since people need not expose themselves to polluted water (Peskin and Seskin, 1975). In contrast to air pollution, our exposure to the adverse effects of water pollution is lessened by modern water treatment plants in urban water supply systems. However, even here there is growing evidence that carcinogenic substances are present in many urban water systems. This situation warrants monitoring.

Healthful Lifestyles

Much is being written on the theory that healthy habits will lower

morbidity and mortality rates while smoking, excessive drinking, excess weight, poor nutrition habits, too little sleep, too little exercise, and so on, will raise these rates (Belloc, 1975; Palmore, 1970). The evidence suggests that a widespread change in health habits could be highly effective in improving the health of the nation. It is, however, a big step to go from identifying this problem to designing a program for its resolution. While we would argue that much work ought to go into innovative experimental programs, we believe such programs ought to be carefully evaluated before they are implemented on a large scale.

Immunization

Immunizing vaccines have been developed for a number of communicable diseases. Benefit-cost analyses of the immunization programs have been conducted primarily at the Center for Disease Control (CDC). To assess the immunization programs, it was necessary to consider the efficacy of the vaccine, the cost of administering the vaccine, the benefits from the reduced incidence of the disease, and the cost of side effects. If one uses wage rates to monetize the health benefits, the measles vaccine program appears to have a benefit-cost ratio of about 10:1 (Axnick et al., 1969). The benefit-cost ratio of a mumps vaccine program is significantly lower (because the complications of the disease are rarer), and the CDC concludes that the costs of a mumps vaccine would exceed the benefits if given by itself, but benefits would exceed costs if the vaccine were given as part of a regular examination or in conjunction with other vaccines (Witte, 1974). Rubella vaccine has not been evaluated, but there is evidence that the economic cost of the rubella epidemic of 1965–66 ran to $1.5 billion (Witte, 1974). Clearly the benefit-cost ratio of a combined rubella-measles-mumps vaccine is very high as the costs are only a little higher than for one alone.

Screening for Disease

Screening is an example of secondary prevention. The basic purpose of screening is to detect the presence of a disease before the onset of symptoms, to initiate early treatment, and thus to influence subsequent morbidity and mortality patterns.

Screening for phenylketonuria (PKU) was made almost universal in the United States before analyses were done to justify the

program. While there is evidence that the effectiveness of the screening programs could be enhanced (Holtzman, 1973) and there are substantial difficulties with the evaluations to date, there appears to be agreement that the benefits of screening for PKU exceed its costs.

A wide range of estimates have been published concerning the benefits of screening for cancer of the cervix utero. Evidence on the efficacy of screening is confounded by the fact that mortality due to cervical cancer has been decreasing in the unscreened as well as the screened population and because of uncertainty about the natural course of the disease. Furthermore, those women who are more likely to be examined on a regular basis, middle and upper income women, normally have a lower rate of cervical cancer than do lower income women. There remains some doubt among the experts about both the efficacy of screening in general and the frequency with which screening should take place even if it is efficacious.

The evidence of the efficacy of screening for breast cancer is much better than that for cervical cancer. The basic data on the effects of screening for breast cancer are from a large scale clinical trial on older women, carried out at the Health Insurance Program of Greater New York (HIP), a prepaid group practice plan. The data from that program indicate that the mortality rate of the group which was screened was lower than that of the group which was not. During that trial, the effectiveness of both palpation and mammography were assessed; palpation was more effective, although each technique missed some tumors. The cost of the program was not indicated. A similar trial has not been conducted for younger women.

The best data we have on the effects of annual multiphasic screening come from a large-scale clinical trial carried out over a ten-year time period at the Kaiser-Permanente Health Plan in Oakland, California (Collen et al., 1973). The results of that study suggest that the effectiveness of annual screening is ambiguous. There were significantly fewer deaths from two diseases for which it is believed that early treatment makes a difference (cancer of the colon and rectum and hypertensive cardiovascular disease) among the screened population. However, in examining the differential health status of different population groups, it was found that, on the one hand, the morbidity and mortality rates of men who were forty-five to fifty-four years old at the beginning of the trial were

lower for those who received the treatment (were screened) than for those who served as the controls, while on the other hand, the morbidity and mortality rates of women who were forty-five to fifty-four years old at the beginning of the trial were higher for those who received the treatment than for those who did not. No discernible differences in mortality and morbidity rates could be ascertained for the younger cohorts.

We can do no better than to summarize the conclusions of our task force report.

1. Health care programs must be evaluated if we are to realize the potential of medical care, prevent iatrogenic disease, and contain cost; in particular, cost effectiveness can be applied and is an invaluable aid in deciding whether a program should be required, financed through public funds, encouraged, or perhaps even prohibited. The results of such evaluations are of as much interest to consumers who must decide how much to spend on preventive services as they are to public decision makers.

2. Both therapeutic and preventive health care programs ought to be evaluated by equally critical techniques and judged by the same criteria.

3. Research and development must be accomplished to determine the efficacy and costs of proposed programs; in particular, large scale clinical trials, some of the most important techniques for developing evidence on efficacy, should be encouraged. If possible, trials should be set up to determine the effect of alternating levels of the treatment.

4. General health status is the primary measure of efficacy; improved methods of measuring health status must be developed and applied.

References

Acton, J.P. 1973. Evaluating Public Programs to Save Lives: The Cost of Heart Attacks. Santa Monica, California: The Rand Corporation (R-950-RE).

Amdur, M. 1976. Toxicological Guidelines for Research on Sulfur Oxides and Particulates. Presented at Fourth Symposium on Statistics and the Environment, National Academy of Sciences, Washington, D.C.

Axnick, W.; Shawall, S.; and Witte J. 1969. Benefits due to Immunization Against Measles. *Public Health Reports* 8 (August).

Belloc, N. 1975. Relationship of Health Practices and Mortality. *Preventive Medicine* 2:67–81.

Berg, R. 1973. Establishing the Values of Various Conditions of Life for a Health Status Index. In *Health Status Indexes*, edited by R. Berg. Chicago: Hospital Research and Educational Trust.

Blalock, H. Jr. 1964. *Causal Inferences in Nonexperimental Research.* Chapel Hill: University of North Carolina Press.

Collen, M., et al. 1973. Multiphasic Checkup Evaluation Study, *Preventive Medicine* 2.

Cooper, B.S., and Rice D.P. 1976. The Economic Cost of Illness Revisited. *Social Security Bulletin* (February).

Dasgupta, A. and Pearce, D. 1972. *Cost-Benefit Analysis*. New York: Macmillan.

Dublin, L. and Lotka, A. 1930. *The Money Value of a Man*. New York: Ronald Press.

Environmental Protection Agency. 1972. The Economics of Clean Air. Washington D.C.: Government Printing Office.

Haveman, R., et al. 1973. *Benefit-Cost and Policy Analysis*. Chicago: Aldine.

Hirshleifer, J.; Bergstrom, T.; and Rapoport, E. 1974. Applying Cost Benefit Concepts to Projects which alter Human Mortality. UCLA, School of Engineering and Applied Science (November).

Hill, A. 1965. The Environment and Diseases: Associations and Causation. Proceedings of the Royal Society of Medicine. Section on Occupational Medicine, 58.

Holland, W.W. 1974. Taking Stock. *The Lancet* (December 21).

Holtzman, W. 1973. Screening for Phenylketonuria and its Problems. International Congress Series No. 310, Proceedings of the Fourth International Conference, Vienna.

Klarman, H.E. 1974. Application of Cost-Benefit Analysis to the Health Services and the Special Case of Technologic Innovation. *International Journal of Health Services* 4: 325–352.

Lave, J. et al. 1976. Report of the Task Force on the Economic Impact of Preventive Medicine. *Preventive Medicine USA*. New York: Prodist.

Lave, L. and Seskin, E. 1970. Air Pollution and Human Health. *Science*: 169.

———. In press. *Air Pollution and Human Mortality*. Johns Hopkins University Press.

Mishan, E.J. 1971. Evaluation of Life and Limb, A Theoretical Approach. *Journal of Political Economy* (July-August).

National Academy of Sciences. 1974. *Geochemistry and the Environment.* Washington, D.C.

New York Academy of Medicine. 1975. Prevention and Health Maintenance Revisited. 1974 Annual Health Conference of the New York Academy of Medicine. *Bulletin of the New York Academy of Medicine* (January).

Palmore, E. 1970. Health Practices and Illness Among the Aged. *Gerontologist.*

Peskin, H. and Seskin, E. 1975. *Cost-Benefit Analysis and Water Pollution Policy.* Washington, D.C.: The Urban Institute.

Rice, D. 1966. *Estimating the Cost of Illness.* Public Health Service Publication, 947-6. Washington, D.C. (May).

Schelling, T. 1968. The Life You Save May Be Your Own. In *Problems in Public Expenditure Analyses*, edited by S. Chase. Washington: The Brookings Institution.

Sterling, T.D. 1975. A Critical Reassessment of the Evidence Bearing on Smoking as the Cause of Lung Cancer. *American Journal of Public Health* 65: 939–953.

Stern, A. 1968. *Air Pollution.* New York: Academic Press.

Stewart, C. 1971. Allocation of Resources to Health. *Journal of Human Resources* (Winter).

Thaler, R. and Rosen, S. 1973. The Value of Saving a Life: Evidence from the Labor Market. Working Paper, Department of Economics, University of Rochester.

Witte, J.J. 1974. Recent Advances in Public Health. *American Journal of Public Health* 64: 939.

This paper was presented at the American Public Health Association meetings, September 1976. The work was supported by the National Center for Health Services Research under grant 3 RO1 HS01529. Any errors and opinions are those of the authors.

Address reprint requests to: Dr. Judith Lave, Associate Professor of Economics and Urban Affairs, School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

Cura Aegrotorum

First and foremost, special care is taken of the sick, who are looked after in public hospitals. There are four at the city limits, just beyond the walls. These hospitals are so spaciously laid out as to be comparable to as many small towns. The intent of this design is twofold: first, that the sick, however variable their number, should never be crowded so close together as to cause discomfort and inconvenience; and second, that those who have a contagious disease may be isolated from the rest in order to prevent the spread of infection. These hospitals are very well appointed and equipped with everything conducive to health. Besides, such diligent treatment and constant attendance of expert physicians are given that, though no one is sent there against his will, hardly anybody in the whole city, when suffering from illness, would not rather be nursed there than at home.

Sir Thomas More's
Utopia, Book 2,
1516

Measuring the Effectiveness of Prevention: II

SAM SHAPIRO

Health Services Research and Development Center,
The John Hopkins Medical Institutions

A Framework

Over the past decade or two, there has been an explosion in the literature of the United States and other countries on preventive health measures. Almost every theoretical, methodological, and substantive aspect of primary prevention and secondary prevention[1] has come under close scrutiny and is treated both on a broad conceptual level and in terms of highly specific applications (Breslow, 1976). To a large extent, the critical appraisals have centered on existing and proposed measures for primary and secondary prevention closely related to the content of personal health services and dependent on an effective interface between the health care system and individuals in a target population.

A powerful current of thinking about prevention that is gaining momentum and has the potential for shifting attention away from the health care system is summarized succinctly in The Forward Plan for Health, FY 1978–82 (U.S. Dept. HEW, 1976).

> In the absence of a major scientific breakthrough (e.g., a cancer cure), further expansion of the Nation's health system is likely to produce only marginal increases in the overall health status of the American people. Obviously, we must continue efforts to correct the inequities and the maldistribution of services in the current system, but in the long run, the greater benefits are likely to accrue from efforts to improve the health habits of all Americans and the environment in which they live and work.

The basis for this statement and LaLonde's (1974) comprehensive document on the Canadian situation is an examination of the leading causes of death and disability, and the factors related to the

[1]Primary prevention defined in terms of interventions designed to prevent the occurrence of disease or injuries responsible for disability or death; secondary prevention in terms of interventions aimed at altering favorably the course of disease through early detection and treatment.

environment and "life styles" that epidemiologic and laboratory investigations have implicated as causes.

Part of the evidence consists of clearly defined relationships between an agent such as smoking, industrial wastes, or occupational hazards, and specific causes of death. Other evidence is more global, such as the association between combinations of certain health habits (smoking, drinking, physical activity, regularities of meals, hours of sleep, height-weight relationship) and life expectancy, reported in the cohort studies by Belloc and Breslow (1972) and Belloc (1973) (Table 1). The data are provocative and focus attention on the *possibility* that major improvements in health status might occur through alterations in life styles. They cannot be interpreted as a prediction that general adoption of the changes would result in the most favorable life expectancies observed in the study.

A number of the "good health habits" included in the study represent prudent prescriptions for reasonable daily living, and the data provide a stimulus for public education, but, as commented on by Breslow and others, it is important to recognize that the arduous task of testing for efficacy has not been obviated. Actually, the closer we examine the details of "life styles" and the more the call is heard for individuals to assume greater personal responsibility for their health, the more complex become the issues. Not only are there severe problems in modifying behavior to effect change in "life styles," but in some instances larger and more pervasive socioeconomic changes may be preconditions, a consideration that is also present in primary prevention of many specific diseases. Without losing sight of this fact, the judgment being made here is that it is necessary to select and test the efficacy of discrete preventive actions. And for this there are guideposts applicable to proposals that bear on "life style" or on health care processes.

Particularly relevant are (a) the commentaries by Cochrane (1972), Holland (1974), and Sackett (1975) on the need for rigorous research to test the efficacy of screening and treatment procedures, (b) the criteria that have been advanced for judging the state of readiness for widespread application of preventive health measures (Wilson and Jungren, 1968), and (c) the recognition that changing provider and consumer behavior is extraordinarily difficult (Somers, 1976; Bryant, 1976; Green et al., 1975). In effect, while prevention has an essential role for the present and future in achieving high levels of physical and emotional functioning, major invest-

TABLE 1
Average Remaining Lifetime Using Death Rates of Three Health Practice Groups

		Number of Health Practices[a]		
	Age	*0–3*	*4–5*	*6–7*
Men	45	21.63	28.15	33.08
	55	13.77	20.21	24.95
	65	10.61	13.71	17.41
	75	7.43	10.23	11.22
	85	6.47	5.82	5.04
Women	45	28.58	34.08	35.84
	55	20.02	25.11	27.83
	65	12.35	17.30	19.87
	75	8.63	11.70	12.50
	85	4.63	7.50	7.61

Source: Data abstracted from Table X, page 79, "Relationship of Health Practices and Mortality," Belloc, Nedra B., *Preventive Medicine* 2: 67–81 (1973), based on 5½ years of follow-up of persons covered in a household survey, Alameda County, California, 1965.

[a]Number of positive responses to the following items: usually sleep 7 or 8 hrs; eat breakfast almost every day; eat between meals once in a while, rarely, or never; weight for men between 5% under and 19.99% over desirable weight for height; weight for women, not more than 9.99% over desirable weight for height; often or sometimes engage in active sports, swim or take long walks, or often garden or do physical exercises; drink not more than four drinks at a time; never smoked cigarettes.

ments are required, often over a long period, to develop new knowledge and methods for applying effectively what is known.

Issues Affecting Selection of Methodologies

The discussion that follows poses a number of issues that require resolution in order to determine the nature and scope of investigations concerned with efficacy and application of preventive health measures. Illustrative material is drawn from research that has been completed or is in progress and from among controversies about prevention. Attention is focused on prevention in which the individual's participation is essential. This is designed to sharpen the discussion and does not represent a judgment about the relative value of prevention in this area compared with prevention related to occupational, industrial, and environmental hazards.

There may be disagreement about the rate at which contributions might be expected from research in one area or the other, or their relative effectiveness in lowering morbidity or mortality rates. However, there are problems in both areas that affect large groups

of people as well as small, highly circumscribed sub-groups. Decisions concerning allocation of resources for research should clearly not be determined by whether personal health services or environmental problems are involved, but rather by whether the problem is significant because of its prevalence, its disabling, impairing or lethal effect, and whether there is a reasonable basis and method for testing the efficacy of the proposed preventive action.

With respect to efficacy, perhaps the most difficult and consequential of the issues from a methodological and, at times, ethical standpoint, is how "hard" does the evidence have to be to accept the preventive measure. Here we are dealing with a diverse array of questions. Are there competing hypotheses about the etiology or epidemiology of the condition? Are there risks to the subjects involved? What is the scope of the medical, social, or economic changes that would be needed to implement a positive finding? How persuasive does the relationship between prevention and effect have to be for those who determine whether or not action is to be taken? How much of an effort is required to replicate the study? What are the consequences of a wrong conclusion for future inquiry (for example, an erroneous negative result could continue the search for a preventive measure unnecessarily and at high cost, and, on the other hand, an erroneous positive result could divert attention from alternatives for preventive action)? Are we testing whether to introduce a preventive measure or to withdraw one that has been accepted as efficacious but about which doubt has arisen?

Closely associated questions concern the level, nature, and scope of the effect. For some problems, it may be adequate to demonstrate that a nontrivial effect, however this may be defined, has been achieved. For other problems that require large resources in the implementation stage, it is imperative to establish that the intervention has a high probability of causing a change that reaches a level agreed to as important. The "nature of the effect" question brings us face to face with an even more complex issue. For example, are we dealing with a problem that requires measurement of multiple types of outcome that may have different social values, as would be the case if mortality were to decrease but protracted, severe morbidity were to increase? On another and equally fundamental level, are the indicators of health status widely enough accepted to provide a basis for policy and planning purposes? Of considerable importance are the current efforts to go beyond the conventional mortality, morbidity, and disability measures central

to the examples discussed later (Berg, 1973; Elinson, 1976). Their goal is to develop scales of social, mental, and physical functioning that provide sensitive measures of health status applicable in determining the effectiveness of health care programs. Clearly, these developments are relevant for the measurement of the value of proposals for prevention.

With respect to "scope of the effect," a major issue is whether the intervention may have a significant impact on one segment of the population and not on another. Our knowledge of the etiology of disease and the influence of biological, social, or demographic factors on the course of disease is so incomplete that often one cannot assume that the effect will be the same on all strata of the population.

A completely rational approach to the design of an inquiry into the efficacy of a preventive measure would consider, in advance, all of these questions and perhaps still others. This does not mean, of course, that all have equal weight or that there would be universal agreement about what can be set aside. Further, an overriding factor may be the limitation of options in selecting the methodology or in the size of sample or characteristics of the study group. A randomized trial may clearly be the best theoretical choice from every standpoint but because of the course of events or the nature of the intervention this design is precluded as illustrated by the first two studies considered below.

Measuring Effectiveness — In Practice

Maximizing the Utility of Available Data

For years, the case of determining whether the Pap smear and treatment of cancer of the cervix in situ results in reduced incidence of invasive cervical cancer or mortality has been cited with great justification as a lost opportunity for definitive study. A randomized trial started twenty-five to thirty years ago could long ago have settled the issue but instead, major screening programs were started in a few areas with provision to monitor changes that occurred in rates of in situ and invasive cervical cancer and in mortality. Reductions became apparent and the Pap smear achieved widespread acceptance as a preventive measure.

For a long time and until quite recently, comparisons between

cervical cancer mortality rates in areas that had aggressive screening programs and those that did not showed similar trends; challenges about the efficacy of the Pap smear were common (Cochrane and Holland, 1971). This is not the time to review in detail the evidence that has altered the outlook. However, a new analysis of the Canadian experience by geographic area indicates a relationship between the extensiveness of screening and mortality from cancer of the uterus (Table 2) which persists when a wide range of socioeconomic variables are introduced (Miller et al., 1976). Although the investigators point to limitations in the analysis, their conclusion is that the efficacy of the Pap smear is established and on this there will undoubtedly be great agreement. Further, decisions about allocation of resources for the use of the Pap smear will be made in the absence of a precise estimate of effectiveness derived from research, although simulation models provide useful working estimates (Knox, 1973). Today the more important considerations are the epidemiologic evidence on where the problem is still significant, the periodicity of and efficiency of alternatives for applying the procedure. The point behind this account is not to suggest that a randomized trial is a luxury but that force of circumstances may make it impossible and other means for reaching conclusions need to be examined.

Selection of an Alternative to Randomized Trial

Another instance in which large stakes are involved concerns regionalization for perinatal care. Professional opinion holds that there is a great potential for reducing infant and perinatal mortality rates and possibly reducing in children the number of seriously handicapping conditions that have their antecedents in the antenatal, intrapartum, and postpartum periods. The mechanism would be well coordinated systems of care that respond appropriately and in a timely way on the basis of risks identified throughout pregnancy and in the postpartum periods. This goes beyond the infant transport systems that are already in operation and involves creating new relationships among hospitals and physicians and, often, shifts in maternity services. In addition, new costs may well be generated.

TABLE 2
Fall in Mortality from Cancer of the Uterus 1960–72 Related to Screening in 1966. Three Year Average Age-Standardized Rates, Female Aged 30–64 Counties and Census Divisions, Provinces of Nova Scotia, Quebec, Ontario, Manitoba, and Alberta

Screening in 1966, rate per 1,000 women aged 20 or more	*Percentage fall in mortality 1960–62 — 1970–72*
Less than 24	−27.5
24–49	−23.6
50–99	16.8
100–249	23.8
250 or more	38.7

Source: "Mortality from Cancer of the Uterus in Canada and Its Relationship to Screening for Cancer of the Cervix." A.B. Miller, J. Lindsay, and G.B. Hill: *Int. J. Cancer* 17: 602–612, 1976.

Although there is agreement among many authorities that these appraisals are correct, progress towards regionalization has been slow. Aside from the usual difficulties of carrying out fundamental changes in the structure and process of care, it remains unsettled whether the changes will, in fact, produce large enough reductions in mortality or morbidity to justify the effort and modifications required. Strong evidence is needed on magnitude of change and on whether the change varies with socioeconomic characteristics of the population associated with health care behavior during pregnancy, nutritional status, age-parity and other risk factors. One could hypothesize why one group might benefit and another not and it would be essential for an evaluation of regionalization to address this question.

A program to accelerate the regionalization of perinatal care that utilizes periodic risk assessments for secondary prevention is being supported by the Robert Wood Johnson Foundation (Merkatz and Johnson, 1976). The eight perinatal centers receiving funds from the Foundation are aiming their efforts at well-defined populations that cover a wide spectrum of ethnic, socioeconomic, and geographic variables and medical care environments. The national evaluation has a number of components in its strategy that relate to process and outcome, the details of which cannot be dealt with in this presentation.

However, the change in the mortality rate is to be measured

through a quasi-experimental design that utilizes before and after observations for the demonstration and comparison areas. Also, a major objective is to assess the relative effectiveness of regionalization for different segments of the population. The evaluation includes a study at one year of age of the changes in developmental and other deficits. What has been conceptualized is a spectrum of reproductive casualties to which society may give different values, and a measurement process that attempts to assess the extent to which a program influences changes in defined categories of those casualties.

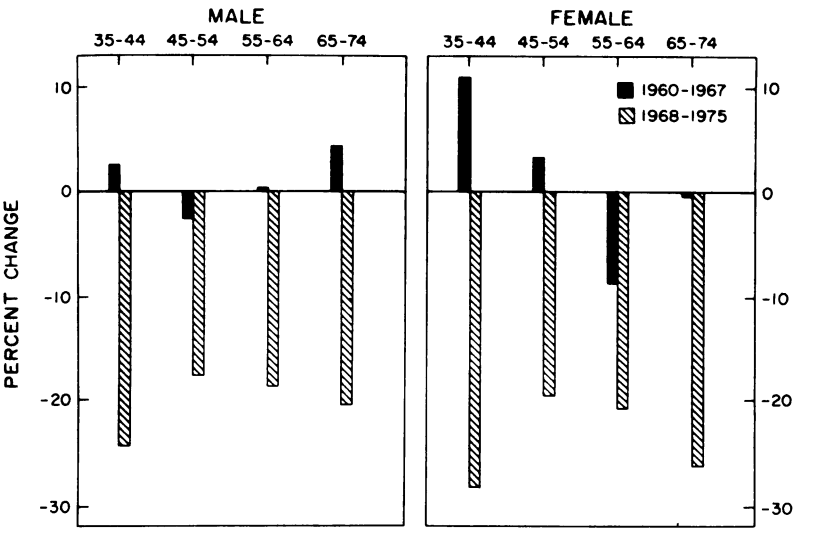
These are only two examples of serious attempts to measure the value of preventive measures. Neither utilizes the randomized clinical trial; one because of a missed opportunity to do so, the other because of the impossibility of introducing the intervention under such study conditions.

A Program of Randomized Trials

There are other interventions for primary and secondary prevention for which nothing short of a randomized trial would provide the needed information. A number of the trials are costly and entail many operational risks but they are aimed at health conditions experienced by large proportions of all segments of our society. In the aggregate they represent the most extensive clinical trials in populations of ambulatory patients that have been conducted in the U.S. As stated in the foreword to a briefing document prepared by National Heart, Lung, Blood Institute (1975) on clinical trials, "in spite of preliminary results obtained through use of new therapeutic modalities on small groups of patients, it is not axiomatic that projected benefits will be realized when findings are introduced into general medical practice." This rationale dictates certain requirements for a study of efficacy; for example, testing should produce measures of effect rather than a "Yes—No" answer; it should have the capability of providing data for major sub-groups of the population; it should be based on a rigorous methodology that would minimize controversy about accepting the results as valid; and attention should be given to problems of implementation.

The Multiple Risk Factor Intervention Trial (MRFIT) for the prevention of coronary heart disease among men and the Hypertension Detection and Follow-up Program (HDRP) for the reduc-

Percent Change in Death Rates from Ischemic Heart Disease by Age and Sex, 1960-1967 and 1968-1975



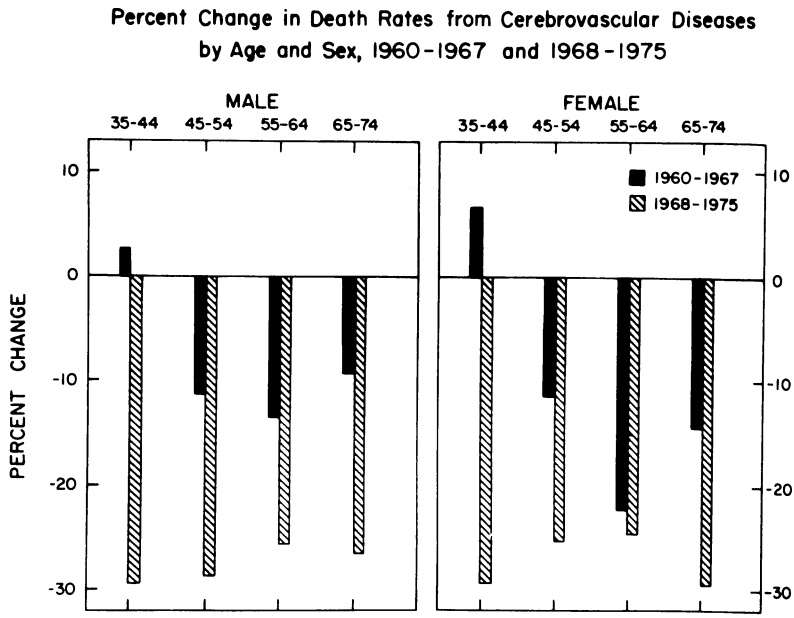
1968 and 1975 Death Rates per 100,000 (ICDA 410-413,420)

		35-44	45-54	55-64	65-74
Males	1975	70.5	292.3	779.1	1724.1
	1968	92.8	355.0	960.9	2157.4
Females	1975	16.4	70.9	258.9	789.1
	1968	22.8	87.6	319.4	1040.2

Source: Divison of Analysis, National Center for Health Statistics, Health Resources Administration, DHEW

FIG. 1.

tion of morbidity and mortality from hypertension among men and women are interesting cases. In MRFIT, prevention aims at a reduction in elevated lipids, elevated blood pressure, and cigarette smoking (JAMA, 1976). In HDRP, the interventions consist of carefully programmed antihypertensive therapies (Langford, 1976). Both trials are being conducted during periods of intensive campaigns in the general population to alter or control the risk factors involved. Whether or not these campaigns are responsible, the fact is that mortality from ischemic heart disease and cerebrovascular diseases has been decreasing at unprecedented rates among men and women (Figs. 1 and 2). From what is known through observational studies of the increased risks associated with the variables being addressed and from the VA hypertension therapeutic trials (Freis, 1967, 1970) the effect of the interventions should override these con-



1968 and 1975 Death Rates per 100,000 (ICDA 330-334, 430-438)

		35-44	45-54	55-64	65-74
Males	1975	11.6	33 9	108.0	363.1
	1968	16.4	47.7	145.8	496.1
Females	1975	11.8	30.7	77.1	2569
	1968	16.7	41.2	102.6	365 1

Source: Division of Analysis, National Center for Health Statistics, Health Resources Administration DHEW

FIG. 2.

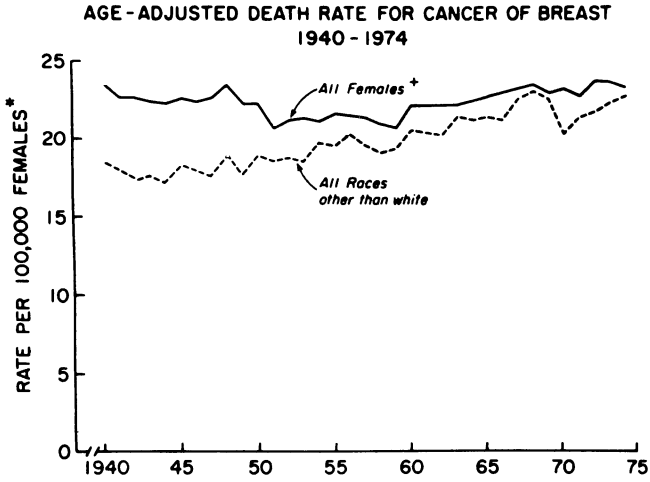
current changes. But the investigations carry a heavier burden in measuring change than originally contemplated and the experimented design (i.e., a randomized trial) assumes increased importance.

A Long Term Randomized Trial

The last investigation to be discussed here is the randomized clinical trial in the Health Insurance Plan of Greater New York to determine whether periodic screening with mammography and palpation results in reduced mortality from breast cancer (Shapiro, 1976). Some aspects of the experience are particularly relevant for this

panel in considering the efficacy of preventive actions. A serious disease, breast cancer is the target for secondary prevention. It is estimated that 7 percent of all women will at some time in their lives have a diagnosis of breast cancer, and a large proportion, about one out of four of all women, will have a breast biopsy. There has been no change in breast cancer mortality for almost 40 years (Fig. 3), and except for what the current adjuvant trials may produce, no benefits in prolonging life have been established for modifications in therapy that have taken place. From a study design standpoint, evaluation of screening's effectiveness in lowering breast cancer mortality requires gathering data over a long period (ten to fifteen years) during which incidence and treatment of the disease might change. Further reliable information must be obtained on the size of the benefit in view of the costs and complexities in making screening generally available. The project is still in the follow-up stages but it has already gone far in meeting its objectives. The picture that has emerged is a 30 percent reduction in mortality from breast cancer attributable to the screening program with all of the gains at ages 50 and over and none below 50 years of age (Table 3). Although the study did not provide for a randomized trial of the efficacy of mammography alone, case fatality rates for breast cancers detected in this way (palpation negative) are low (Table 4) and it is estimated that the inclusion of mammography accounts for about a third of the reduction in mortality.

The controversy that has erupted in the United States, about the extension of screening to the population at large, centers on the risks associated with radiation exposure from mammography and how to deal with the no benefit findings in the HIP study for women under 50 (Bailar, 1976). This is in addition to the issues of organization, costs of services, and consumer and professional acceptance that are faced when a major new program is contemplated, a set of problems the 27 NCI-ACS breast cancer detection demonstration projects was designed to address. A careful assessment of the radiation question using available estimates would seem to lead to the conclusion that the risks do not outweigh the benefits of using mammography for women over 50 years of age, but until new evidence appears there is no basis for routine screening of asymptomatic women or those at relatively low risk below age 50. More general reservations about screening for breast cancer with mammography and clinical examination do exist and new trials to



* RATES ARE ADJUSTED TO AGE DISTRIBUTION OF THE FEMALE POPULATION IN 1940.

† TREND LINE NOT SHOWN SEPARATELY FOR WHITE FEMALES; RATES ARE ALMOST IDENTICAL FOR "ALL" AND "WHITE" FEMALES.

NOTE: RATES PREPARED BY THE NATIONAL CENTER FOR HEALTH STATISTICS, PUBLIC HEALTH SERVICES, HEALTH RESOURCES ADMINISTRATION, DHEW.

FIG. 3.

TABLE 3
Breast Cancer Deaths by Age
Nine Years of Follow-up from Date of Entry

		Number of Deaths	
		Study	*Control*
Total		91	128
Age at Death:	40–49 years	17	17
	50–59	40	67
	60 and older	34	44
Age at Entry:	40–49 years	39[a]	48[a]
	50–59	40	60
	60–64	12	20
Age at Diagnosis:	40–49 years	30	27
	50–59	42	67
	60 and older	19	34

[a]Data by age at entry and age at diagnosis follow:

Age at Entry		Age at Diagnosis	
(40–49)	Total	40–49	50–59
Study	39	30	9
Control	48	27	21

TABLE 4
Cumulative Case Fatality Rates (per 100) Among Breast Cancer Cases[a] Detected on Screening by Modality

Modality[b]	*No. of Cases*	Years Following Diagnosis 7	8
Total	132	21.3	28.3 (4.0)[c]
Mammography only	44	9.2	14.4 (5.5)
Clinical only	59	20.3	31.8 (6.2)
Mammography and clinical	29	41.4	41.4 (9.2)
Cases positive on both modalities plus			
Mammography only cases	73	22.1	25.2 (5.2)
Clinical only cases	88	27.3	35.2 (5.2)

[a]Cases detected within five years after entry; follow-up through 12/31/75.
[b]Initial evidence for biopsy recommendation made independently by the two modalities.
[c]Numbers in parentheses are standard errors due to sampling.

test the cost-effectiveness of alternative approaches, including instruction for self-palpation, are being advanced.

From Research to Implementation

Thus far, the emphasis has been on measuring the effect of preventive measures. Most often, the issue that attracts attention when application is being considered is the costs or resources required with lesser attention being given to defining the problems of implementation and finding solutions. The question can be stated in a variety of ways; for example, what is the state of readiness among providers and the general public for accepting the results of an efficacy trial, and how rapidly and by what means can acceptance be achieved; are there changes needed in the structure, delivery and financing of care or in the skills of the providers of care or in their relationship to patients? Another formulation involves searching for effective strategiessfor diffusion of results; or for reaching target populations, linking the preventive measure to follow-up care, changing the health care and personal health behavior of the population, and assuring the accessibility and quality of the services. The importance of such issues seems obvious, in view of the fact that many of the health problems in our society are rooted in the chronic diseases or in other conditions for which the resolutions are complex.

All of this sharply demonstrates that it is a rocky course from a "hard" finding in an efficacy trial to taking action, and detours are more often than not encountered. There is no guarantee that results from rigorous research will be decisive in formulating policy or that a decision to move ahead will not be made even while a trial is under way. But this does not mean that such research on new proposals for prevention is nothing more than an academic exercise. Our past experience indicates that future challenges to a program should be anticipated, particularly if large resources and difficult changes are required and results become equivocal. The investment made in measuring the efficacy of the preventive measure will then be seen in retrospect as a wise, far-sighted action.

References

Bailar, J.C. 1976. Screening for Early Breast Cancer: Pros and Cons. *Cancer*, in press.

Belloc, N.B. 1973. Relationship of Health Practices and Mortality. *Preventive Medicine* 2: 67–81.

Belloc, N.B., and Breslow, L. 1972. Relationships of Physical Health Status and Health Practices. *Preventive Medicine* 1: 409–421 (August).

Berg, R.L., Chairman and Editor. 1973. Health Status Indexes. Proceedings of a conference conducted by Health Services Research. Chicago: Hospital Research and Educational Trust.

Breslow, L., Chairman. 1976. Task Force Report. Theory, Practice and Application of Prevention in Personal Health Services. In: *Preventive Medicine USA*. Sponsored by the John E. Fogarty International Center for Advanced Study in the Health Sciences, NIH and The American College of Preventive Medicine. New York: Prodist.

Bryant, J.H., Chairman. 1976. Task Force Report. Education and Training of Health Manpower for Prevention. In: *Preventive Medicine, USA*. Sponsored by the John E. Fogarty International Center for Advanced Study in the Health Sciences, NIH and the American College of Preventive Medicine. New York: Prodist.

Cochrane, A.L. 1972. *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust.

Cochrane, A.L., and Holland, W.W. 1971. Validation of Screening Procedures. *British Medical Bulletin* 27: 3–8.

Elinson, Jack, Editor. 1976. Sociomedical Health Indicators. *International Journal of Health Services* (3) Special issue.

Freis, E.D., Chairman. 1967. Veterans Administration Cooperative Study Group on Antihypertensive Agents: Effects of Treatment on Morbidity in Hypertension. I. Results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. *Journal of the American Medical Association* 202: 1028–1034.

———. Chairman. 1970. Veterans Administration Cooperative Study Group on Antihypertensive Agents: Effects of treatment on morbidity in hypertension. II. Results in patients with diastolic blood pressures averaging 90 through 114 mm Hg. *Journal of the American Medical Association* 213: 1143–1152.

Green, L.W.; Levine, D.M.; and Deeds, S. 1975. Clinical Trials of Health Education for Hypertensive Patients; Design and Baseline Data. *Preventive Medicine* 4 (December): 417–425.

Holland, W.W. 1974. Taking Stock. *The Lancet* 2: 1494–1497 (December 21).

Journal of the American Medical Association. 1976. Multiple Risk Factor Intervention Trial (MRFIT). National Study of Primary Prevention of Coronary Heart Disease. *JAMA* 235: 825–827.

Knox, E.G. 1973. A simulation system for screening procedures. In: *The Future—and Present Indicatives, Problems and Progress in Medical Care*, edited by G. McLachlan. Ninth Series. Oxford University Press for the Nuffield Provincial Hospitals Trust.

LaLonde, M. 1974. A New Perspective on the Health of Canadians; a working document. Ottawa (April).

Langford, H.G., Chairman. 1976. The Hypertension Detection and Follow-up Program: Hypertension Detection and Follow-up Program Cooperative Group. NHLI, NIH. *Preventive Medicine* 5: 207-215.

Merkatz, I.R. and Johnson, K.G. 1976. Regionalization of Perinatal Care for the United States. *Clinics in Perinatology* 3 (2): 271–276 (September).

Miller, A.B.; Lindsay, J.; and Hill, G.B. 1976. Mortality from Cancer of the Uterus in Canada and its Relationship to Screening for Cancer of the Cervix. *International Journal of Cancer* 17: 602–612.

National Heart, Lung, Blood Institute. 1975. Clinical Trials Briefing Document.

Sackett, D.L. 1975. Screening for Early Detection of Disease. To What Purpose? *Bulletin of the New York Academy of Medicine*, 51: 39–51 (January).

Shapiro, S. 1976. Current Observations from a Test of the Efficacy of Breast Cancer Screening and Their Implications. *Cancer*, in press.

Somers, A.R., Chairman. 1976. Task Force Report. Health Promotion and Consumer Health Education. In: *Preventive Medicine, USA*. Sponsored by the John E. Fogarty International Center for Advanced Study in the Health Sciences. NIH and the American College of Preventive Medicine. New York: Prodist.

U.S., Department of Health, Education, and Welfare, Public Health Service. 1976. *Forward Plan for Health FY 1978-82.*

Wilson, J.M.G., and Jungren, G. 1968. Principles and Practice of Screening for Disease. *Public Health Papers No. 34*, pp. 26–27. Geneva: World Health Organization.

This paper was presented at the American Public Health Association meetings, Miami Beach, Florida, September, 1976.

The Health Services Research and Development Center, at The Johns Hopkins Medical Institutions, Baltimore, Maryland, is supported in part through Grant No. HS-01964, National Center for Health Services Research.

Address reprint requests to: Sam Shapiro, Health Services Research and Development Center, The Johns Hopkins Medical Institutions, 624 North Broadway, Baltimore, Maryland 21205.

Measuring the Effectiveness of Prevention: III Commentary

MYRON E. WEGMAN

School of Public Health,
University of Michigan

It is obvious in the preceding papers that the techniques of measuring program effectiveness are multiple, interrelated, and not easily assessable themselves as methods. I shall try to use some practical illustrations to highlight some aspects that seem to me clear and some very murky.

My first point is that regardless of specific method, having done an evaluation once does not mean that it's done for all time. To be sure, the ever present need for the evaluation process is implicit in both the Lave and Shapiro papers but there is also a "time" or secular factor in evaluation. Situations do change, as in smallpox eradication, the very example cited by the Laves. Worldwide smallpox control (which they meant, I believe, rather than eradication, per se) has far different significance for the United States today than it did fifty years ago. As recently as 1947 millions of people were vaccinated in New York City in a period of a few weeks. The panic attending appearance of an imported case, with immediate contact cases, led to a glossing over of the benefit-risk ratio: there was city-wide vaccination, with neither time nor personnel available for the usual precautions. In the face of an epidemic the risk of any control procedure is minimized and evaluation of its effectiveness often neglected.

One cannot mention smallpox without noting the anticipation throughout the world that the last case anywhere has occurred. Stop and think about that. As a measure of effectiveness there can be nothing better, certainly, than the ultimate in prevention, no more cases at all—and, therefore, no need for further specific preventive measures. There is still an uneasy fear in many quarters that the virus is still hiding in an isolated village somewhere, and to answer that fear, intensive surveillance has been planned for at least two years.

When one speaks of eradication, in its specific sense, the tearing up of a disease by the roots so that it will never reappear, one thinks immediately of malaria, object of the first commitment to a worldwide eradication program. Assessment of results was built into the plan when first undertaken in the Americas in 1954 and worldwide in 1955, and was designed and scaled for the several stages of the eradication plan. There have been individual and isolated successes and failures but they add up to an overall result much less favorable than with smallpox. On the other hand, it is cogently argued that much of the reason that the malaria effort ran into difficulty was the inadequacy of personnel and resources from the outset. The moral commitment was there, but the willingness to mount an overwhelming attack from the beginning was not. One may well say that whole-hearted dedication cannot overcome half-hearted support.

Despite failure to reach the ultimate goal of eradication, the malaria experience is instructive in other ways as regards measuring effectiveness. Both previous papers discussed the need to distinguish between life-saving and health-saving. Shapiro wisely pointed out the relative futility of exchanging what may well be a temporary reduction in mortality for an increase in morbidity. This is particularly appropriate for malaria where another obvious measure of effectiveness is economic impact, favorable or unfavorable. There are some who suggest that reduction in both mortality and morbidity, as health conditions improve, may indeed be counter-productive.

One well known example is the thesis that measures to reduce infant mortality interfere with natural selection and survival of the fittest. This thesis conveniently forgets that the most easily preventable deaths in infancy are environmentally, not genetically, induced. I should like to cite two instances that may cast light on some of the issues involved.

One story occurred more than twenty years ago in a rural area of what was then British Guiana. The mayor of a village in the interior received the district health officer, his staff, and his international advisers with open arms when they came in with a program to clean up malaria. About a year later, when they came to check on progress, there were men armed with rifles to keep them away. A conference was held, under a white flag, at which the mayor said frankly, "We don't want any more of this public health nonsense.

We used to be very happy here. A year ago we had a population of 10,000 workers and only 5,000 jobs. Of our people 5,000 were always sick with malaria so there were always 5,000 jobs for the others. Everybody who could work was working. Then you came along and wiped out the malaria. We still have 10,000 workers and we still have only 5,000 jobs. But now there are always 5,000 unemployed and we have crime, robberies and murders that we never had before. It's all your fault!"

Obviously there are many, many other parts of the world where malaria control has brought economic benefit. I cite the People's Republic of China because it is more vivid in my memory. When I was in China in 1973 I learned that the malaria program had been given special importance because with chronic shortage of personnel, preventable illnesses were given top priority. Chairman Mao's first dictum in health was "stress prevention." I shall not go into the economics or the societal implications of a country of 800 million where retired people are urged to go back to work not just to supplement their pensions but because there are more jobs than people. Clearly, economic planning needs to be related to health planning as well as vice versa. To accept fatalistically that the existing situation necessarily limits the jobs that can be made available and that therefore enough illness or death should be allowed to exist in order not to have more people than jobs is just as unacceptable as to say that the way to control population is to let infant death rates stay high.

These illustrations underline the importance of evaluating the effect of any national communicable disease program on other countries, whether or not eradication is the goal. An attack on smallpox, or malaria, or tuberculosis, or measles, must be correlated, at least in part, with how the results affect and are affected by neighboring states. Effectiveness at one stage of development, national or worldwide, may be very different at a later stage.

In a general sense, the influence of changing times and conditions on health programs illustrates the interrelationship between quantity and quality, so dear to the dialectical philosophers. Quantitative change in health needs is taking place constantly as is the effectiveness of health procedures. At some point, the magnitude of these changes reaches the level where one may say a qualitative difference exists. At that point substantial change in procedures, immunization programs, or whatever, may be indicated.

A somewhat more homely illustration of the time and quantity factors occurred at Charity Hospital in New Orleans 25 years ago, in connection with an outbreak of diarrhea in newborn babies. An epidemiologic study showed that the disease pattern was unusual, in that occurrence was limited entirely to breast fed babies. Although further studies never did identify a specific microorganism the epidemic was clearly infectious in pattern. There was, moreover, clear epidemiologic association of the disease with the "preventive" use of boric acid solution to wipe off the mother's nipples each time before she breast fed her baby. Two of the bottles of boric acid solution on the ward were subsequently shown to be contaminated with common microorganisms. All we did was to stop the "preventive" action, refuse to use any of the insistently recommended antibiotics, and the epidemic ceased.

All of this reinforces for me the need for permanent, constant and iterative evaluation of ongoing practices.

I turn now to another kind of problem in assessing the effectiveness of preventive techniques. This illustration is also from the field of child health (in case there is any doubt, that's my original field) and raises the question of the success of the effort to reverse the seemingly inexorable worldwide change from breast feeding to bottle feeding.

In developed countries it is quite possible, although it requires some effort, to obtain the benefits from bottle feeding that come almost automatically from breast feeding. The major problem comes from putting the milk—whether cow's, goat's, human's, is unimportant—in a bottle. It is the preservation, transfer, and preparation that open all sorts of possibilities for things to go wrong. Even though environmental pollution is currently low in this country—I believe, however, that the Laves underestimated how much water pollution still exists—danger of contamination always exists where there are multiple steps in the delivery process. Besides other advantages, breast feeding—the direct producer-consumer system—is obviously unmatched in preventing difficulties of contamination and infection. The boric acid business I mentioned earlier was, after all, a "technical" intervention. To the extent breast feeding is diminished one is lessening the effectiveness of a potent preventive procedure.

All of the above is well known and one might say that in our milieu it is of minor importance. But remember that behavior in the

United States is an example for the world. And worldwide, in many of the developing countries, one of the greatest dangers to health is the steady and frightening decrease in breast feeding. Bottle feeding has become the fashion, the mode, what the "rich people" do, the thing to imitate.

Efforts by health authorities have been chiefly in the field of education and propaganda, with indifferent results, if one considers overall figures on proportion of infants breast fed. I often think wistfully—wouldn't it be great if we could get the enormous advertising machinery of the brassiere manufacturers on our side?

Incidentally, a curious and more hopeful reversal has taken place in recent years. An increasing number of younger, college-educated mothers are insisting on breast feeding, although sometimes, sad to say, they are faced with opposition from obstetrical and nursing staff. This increase came to my attention again recently when house officers in our pediatric service told me they guessed that, at age 3 months, 40 percent of the babies seen at University Hospital are breast fed and perhaps only 5 percent at the Wayne County Hospital.

The breast feeding situation is one example of attempts to change health behavior. Certainly, our behavior in regard to many aspects of our daily lives that affect health must be changed, if we are, all of us, to be better off. Yet quantifying the long range effects of health education on behavior is not readily achievable by the techniques we've heard of so far. One of the greatest difficulties in evaluating change in this area is, in my view, the wonderful tendency of us public health workers to equate activity with accomplishment. The very difficulty of the task, however, underlines the need to give more attention to the development of methods to assess the techniques of education, to measure differences between teaching and learning, to decide what practices are effective.

In contrast with dealing with the abstractness of behavior, the effectiveness of such specifics as "preventive" physical examination and laboratory procedures must also be measured. I am persuaded that early detection of treatable diseases and problems, for example, obesity, is very important. But for many years I have had serious doubts that most current routines of periodic examination and screening are the best way to accomplish this. What is far more logical, it seems to me, is to select specific screening procedures to be directed at specific population groups and to adapt the

periodicity to such factors as age, sex, social characteristics, special disease suseptibility, and the likelihood that something useful may be done as a result of the detection. This is the sort of targeted approach that came out of the conference on preventive medicine mentioned by Judith and Lester Lave. Evaluating the effectiveness of targeting rather than a dispersed, shotgun approach should be feasible.

A favorite bete noire of mine has to do with failure to choose clinical laboratory procedures wisely. Even when not related directly to preventive activities, carrying out these procedures affects both health care costs and effective use of available facilities. In the brief hospital ward supervision I undertake each year as professor of pediatrics I am concerned at the amount of unsolicited laboratory results presented to me. When I ask why a test was done the answer is usually either, "We thought you'd ask for it" (on the contrary I ask for justification for the test!), or that "It's cheaper." It apparently is cheaper to do a panel of twenty tests than to pick out the ones you need. A true evaluation might, however, ask how often inevitable laboratory errors in one of the unneeded procedures requires costly repetition. Beyond the fruitless multiplication of cost, incidentally decreasing availability of funds for preventive activity, such a wholesale and unanalytic approach pushes aside the more effective and more easily evaluated target program.

A new and rather frightening reason for proliferating laboratory work, which surely will influence definitions of effectiveness, is the "need" for defense against accusations of malpractice. If we are measuring effectiveness in reaching an objective and the objective is not related, and even possibly antithetical, to better health, we are all in trouble.

An even broader definition of prevention is implied in Shapiro's citation of regionalization in perinatal care. In the '30s and '40s we were organizing transportation systems to move premature infants to sophisticated neonatal centers. Effectiveness is increased when the transportation of the mother at risk is done before the baby is born, thus diminishing the risks for both.

Finally, a word about the conclusions and the consequent decisions that follow evaluation. I was impressed with Professor Shapiro's analysis of why one did not always have to have a randomized clinical trial to come to a sound and secure conclusion. The randomized planned trial has distinct advantages, but just because it

cannot be done for one reason or another, evaluation is not hopeless. Shapiro noted that there are times when one simply has to make a decision about starting a program before the results are all in and Judith and Lester Lave commented on the insecurity of conclusions when there are not enough data.

Yet public health people are often asked to decide whether to continue or institute a given preventive procedure when one simply cannot be certain if it is effective or not. Without diminishing by one whit the need to persevere in seeking the needed information, I remind you, in closing, of a remark by Alan Gregg, the late great vice president of the Rockefeller Foundation: "The essence of wisdom is the ability to make the right decision on inadequate evidence."

Address reprint requests to: Myron E. Wegman, School of Public Health, University of Michigan, Ann Arbor, Michigan 48104.

The editor has, for many years, indulged in the archaic pleasures of keeping a commonplace book. Literary gleanings, often serendipitously hit upon, are entered as they illuminate the workings of health and society. These apothegms and paragraphs are a useful reminder that wit and wisdom are rarely dulled by the dust of libraries.

We commend this device to our readers; it is a well-assessed, low-level technology. And the personal pleasures afforded are so readily shared.

Your own discoveries would be welcomed by the editor for inclusion in *Health and Society* as space permits.

D.P.W.