SCREENING FOR EARLY DETECTION OF DISEASE: TO WHAT PURPOSE?*

DAVID L. SACKETT, M.D., M.Sc. Epid.

Professor, Department of Clinical Epidemiology and Biostatistics
Professor, Department of Medicine
McMaster University Medical Center
Hamilton, Ontario
Canada

THIS topic is a timely one, and its timeliness takes on an additional perspective when we note the differing attitudes which prevail in the United States and Canada about the value of screening and the evidence required for evaluating its validity.

First, I propose to describe briefly and to distinguish between the several different motives for the performance of screening. I shall then focus upon the main topic of this conference: screening for the prevention of disability and untimely death. I shall make only passing reference to specific screening maneuvers and strategies, since other speakers will present up-to-date reviews in some of these special areas. I shall, instead, attempt to provide you with a series of strategies or yardsticks with which to assess more critically both today's evidence and that which will be presented to you in the future concerning the clinical efficacy of preventive screening. Thus, I shall attempt to introduce a way of thinking rather than a global conclusion about the value or uselessness of screening for prevention.

There are five different motivations for carrying out screening maneuvers. Only one of these is of central concern to this conference—the prevention of disability and untimely death—but it is important to identify the four others so that we can set them aside and avoid confusing further an area that is already complex. Although this may sound paradoxical, at least three of the five motivations involve maneuvers in which efficacy for the patient is simply not at issue.

The first of the five is screening for protection of an economic

*Presented in a panel, Prevention: The State of Knowledge, as part of the 1974 Annual Health Conference of the New York Academy of Medicine, *Prevention and Health Maintenance Revisited,* April 25 and 26, 1974.

This report was supported in part by Research Grant MA-5195 from the Medical Research Council of Canada and National Health Grants 606-22-12 and 606-1084-22 from Health and Welfare Canada, Ottawa, Ontario.

wager, and is typified by the life-insurance examination. The holder of the life-insurance policy is betting the company that he is going to die while the insurance company is betting him that he is going to live—at least until he pays his premiums. The screening is not performed in an effort to determine the state of the policyholder's health nor necessarily to improve it, but simply to guarantee that the company will win more bets than it loses.

The second motivation involves the protection of other citizens. It takes the form of several public-health measures for the control of infectious disease and the prevention of accidents: for example, the screening of crane operators in steel mills for seizure disorders and cardiovascular disease. Here again the primary motivation is not necessarily the benefit of the person examined; in this case it is the protection of those around him.

The third motivation for screening involves its provision as an alternative to personal health services, and results in the performance of a broad array of multiphasic health tests in lieu of an interview and examination by a clinician. This form of screening can be characterized in altruistic terms as the provision of some kind of health service to the poor in an effort to separate the well from the sick, or in a more cynical fashion as the buying-off of health needs with a computer and an SMA-60. Proponents of this approach suggest that the issue of efficacy is largely irrelevant in this case, and simply view this form of screening as a means for filling the gap between the demands of those who view medical care as a right and the ability of the existing system to meet that demand.[1] This opinion has extraordinarily little credibility outside of the United States.

The two remaining forms of screening are carried out in the expectation that they will, in fact, benefit the person being screened and must, therefore, be justified on the basis of evidence that they do more good than harm. In the first of these, the motivation is the acquisition of clinical baseline information about the patient at a time when he is asymptomatic and healthy for use at some later date when he may, with greater or lesser suddenness, become ill. The prototype here is the baseline electrocardiogram in the healthy middle-aged man. How comforting it is, as you arise from your bed in the middle of the night to evaluate his chest pain, to know that you have an earlier electrocardiogram for comparison purposes. I must admit that my faith in

the value of these clinical baselines has been shaken somewhat since I left a teaching post and returned to a general consulting practice in internal medicine. I am now reaching the tentative conclusion that, while these clinical baselines do have an effect upon the diagnostic label I use to identify a subsequent sudden illness, they appear to have very little effect upon my choice of either immediate or long-term therapy. This emerging realization, when added to the fact that these baseline data are often inaccessible in the middle of the night, has raised questions in my mind concerning the cost-effectiveness of this approach.

The final motivation for screening—that which has brought us to this conference—is screening for the detection of those conditions in which our subsequent clinical maneuvers will do the patient more good than harm. This is the focus of this conference: the early detection of those with disease or a predisposition to it in order to institute a series of clinical maneuvers which will have a favorable effect upon the natural history and clinical course of the disease. I shall devote the remainder of my comments to the discussion of yardsticks for determining the validity of these screening maneuvers in a clinical situation.

First, what are we going to measure with these yardsticks? In considering the introduction into the general health-care system of a given screening, diagnostic, therapeutic, or rehabilitative maneuver, the scientific evidence for and against its introduction can be classified into four categories: clinical efficacy, effectiveness (or usefulness), availability, and efficiency.

Clinical efficacy asks: Does the maneuver do more good than harm —in terms of mortality or physical, social, and emotional function— to those who will faithfully comply with all instructions?

Effectiveness (or usefulness) questions whether the maneuver actually does more good than harm to those to whom it is offered. Here we see the fragile link between efficacy and effectiveness: namely, will patients comply with the instructions given to them by clinicians?

Availability considers whether effective maneuvers are being made accessible to everyone in the community who can benefit from them.

Efficiency asks: Are effective maneuvers being made available with optimal use of resources?

It is obvious that efficacy and effectiveness must be established prior to studies of availability and efficiency, because the latter are quite insensitive to whether the clinical maneuver under consideration is

DOES THE PERIODIC HEALTH EXAMINATION (PHE) DETECT DISEASES LIKELY TO HAVE AN IMPORTANT EFFECT UPON HEALTH?

	% of those dying from this cause in whom the diagnosis was made at PHE
Cancer	43%
Coronary heart disease	58%
All diseases causing death	51%

Reproduced by permission from Bombardier, C., McLaran, J., and Sackett, D. L.: Periodic health examinations and multiphasic screening. *Canad. Med. Ass. J. 109*:1123-27, 1973.

helpful or harmful to patients. Indeed, we could probably all cite examples of efficiency-maximizing research which has led to making harmful clinical maneuvers available to everyone in the community who could suffer as a result of their application!

I shall now attempt to translate these somewhat abstract theoretical considerations into a series of criteria which health professionals can use to determine the validity of individual and multiphasic screening programs for the early detection of disease. In practical terms, I believe that the issues can be reduced to a series of six questions:[2, 3]

1) Are screening maneuvers able to detect disease which is likely to have an important impact upon health?

2) Will the treatment of risk factors have a major impact upon the subsequent development of disease?

3) What are the prospects that patients will comply with therapeutic regimens initiated as a result of screening programs?

4) Do existing screening programs really alter the outcomes of the target disease?

5) Are we misled by the traditional methods used in evaluating the clinical effectiveness of early detection programs?

6) Have we considered the entire range of possible effects of screening, labeling of individuals as diseased, and long-term therapy?

Are screening maneuvers able to detect disease which is likely to have an important impact on health? When we consider the first question we are in for the first of what may become a series of surprises. Although there is little information available on this topic, the results of an evaluation of the experience of 10 major industrial periodic

health-examination programs in North America are shown in the accompanying table.[4] This analysis determined the portion of individuals dying from specific disorders who had them diagnosed in a periodic health-examination and screening program. Less than half of the individuals who subsequently died of cancer had this disorder diagnosed at a screening examination, and slightly less than two thirds of the individuals who died of coronary heart disease were identified prior to developing overt symptoms of lethal coronary disease. If we were to assume that the early detection of these disorders could lead to improvements in outcome such findings should still be encouraging. However, the screening examination has a relatively low sensitivity for the detection of major disorders with lethal outcomes.

Will the treatment of risk factors have a major impact upon the subsequent development of disease? The identification and modification of a risk factor carries with it no guarantee that an actual change in risk has occurred. The most perplexing current example of this is the disparity between the marked reduction in the risk of cerebrovascular disease which accompanies reductions in elevated blood pressure and the quite meager reductions in the risk of myocardial infarction and sudden death over the periods during which this risk factor was reduced.[5]

What are the prospects that patients will comply with therapeutic regimens initiated as a result of screening programs? To look at the third question we might begin by asking ourselves how successful we each have been in reducing our own waistlines or our own consumption of cigarettes. A systematic determination of the extent to which patients follow clinical instructions can be quite sobering. Much of our own work at McMaster University Medical Centre is devoted to attempting to understand what determines this compliance.[6] We have been forced to conclude that ambulatory patients are unlikely to take more than 50% of the prescribed medications that they receive from clinicians, and that the amount of knowledge which a patient possesses about his or her illness has almost no relation to that patient's degree of compliance with therapeutic instructions. A number of randomized trials of clinical strategies for the improvement of compliance are now underway in Hamilton, Ontario, and elsewhere. Meanwhile, we have no assurance that efficacious therapy will be followed—particularly among patients who are asymptomatic—and therefore we cannot expect

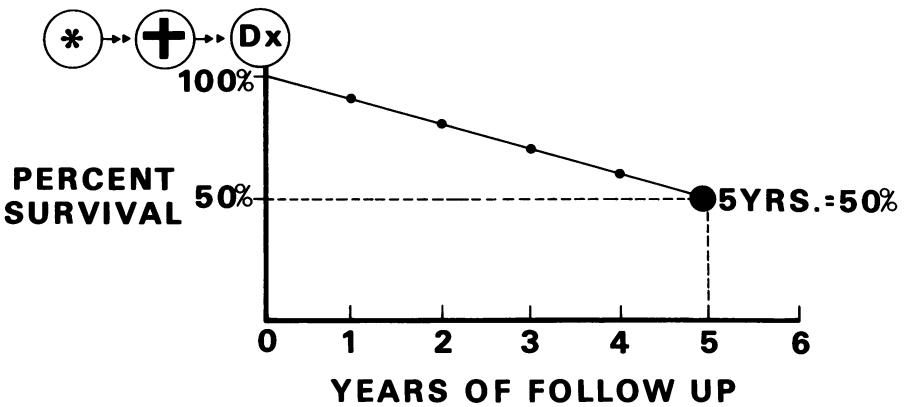

Fig. 1. A typical survival pattern for cancer. * = time of actual onset of disease, + = time of possible early detection, Dx = time of usual diagnosis (onset of symptoms). Reproduced by permission from Bombardier, C., McClaran, J., and Sackett, D. L.: Periodic health examinations from multiphasic screening. *Canad. Med. Ass. J. 109:* 1123-27, 1973, and from Sackett, D. L.: The usefulness of laboratory tests in health-screening programs. *Clin. Chem. 19*:366-72, 1973.

programs of early detection which require high degrees of compliance on the part of patients to achieve the desired benefits.

Do existing screening programs really alter disease outcomes? When we consider the fourth question our attention quickly turns to the important study made by investigators with the Kaiser-Permanente Group.[7] They randomly allocated several thousand participants in the Kaiser Plan into two groups, one of which received intensive encouragement to undergo multiphasic health testing on a regular, recurring basis, while members of the other (control) group used Kaiser Plan services on their own. After seven years of study, these investigators have not found any favorable health result of the multiphasic health-testing among women and most men. Only the group of men between the ages of 45 and 54 at entry showed subsequent—and at times only temporary—differences in disability and absenteeism.[8] Although these differences are statistically significant, I find them clinically unimpressive. In addition, the control group of older men appears to have been less healthy than the experimental group before the beginning of the study; this raises further questions about the validity of the interpretations of the study, particularly as they relate to cost-benefit analyses.[9]

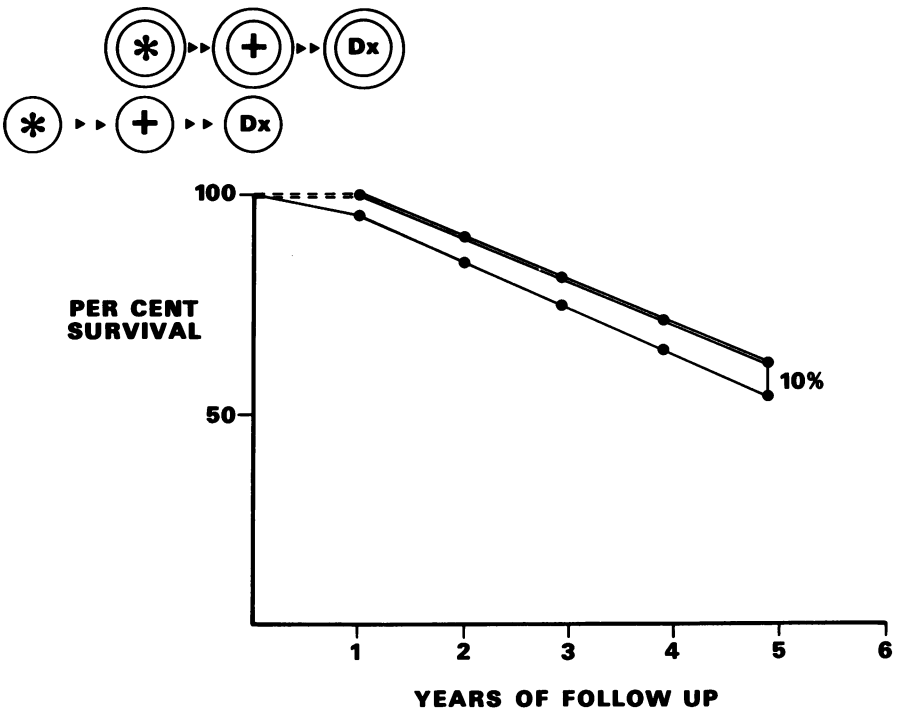

Fig. 2. Comparison of survival patterns for cancer between period of usual diagnosis (single line) and period of early detection (double line), exemplifying a typical error in computation. * = time of actual onset of disease, + = time of possible early detection, Dx = time of usual diagnosis; double circles denote actual time periods involved, single circles denote mistaken assumptions. Reproduced by permission from Bombardier, C., McClaran, J., and Sackett, D. L.: Periodic health examinations and multiphasic screening. *Canad. Med. Ass. J. 109*:1123-27, 1973, and from Sackett, D. L.: The usefulness of laboratory tests in health-screening programs. *Clin. Chem. 19*:366-72, 1973.

Finally, and particularly relevant, is that of 400 deaths occuring in the control group during the experimental period only 60, or 15%, were judged to have been "potentially postponeable" through the optimal application of preventive medical maneuvers.[10]

Are we misled by the traditional methods used in evaluating the clinical effect of screening programs? Much of the foregoing has probably contrasted with many commonly held beliefs about the value of screening. I imagine that every physician can recall at least one patient in whom an early diagnosis was followed by what appeared to be a prolonged survival. However, I submit that the interpretation

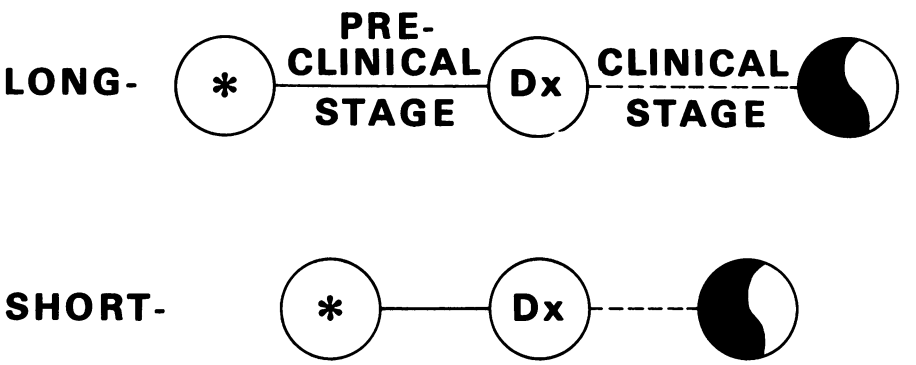

Fig. 3. Relation between duration of preclinical and clinical stages of disease. * = onset of disease, Dx = diagnosis, partially shaded circle = death. Reproduced by permission from Bombardier, C., McClaran, J., and Sackett, D. L.: Periodic health examinations and multiphasic screening. *Canad. Med. Ass. J. 109*:1123-27, 1973, and from Sackett, D. L.: The usefulness of laboratory tests in health-screening programs. *Clin. Chem. 19*:366-72, 1973.

of these patients' subsequent survival was affected by one of several pitfalls in the evaluation of clinical outcomes.

For example, Figure 1 summarizes a typical survival pattern for cancer, with a steady decline in survivors amounting to 50% at five years if we utilize the usual time of diagnosis as the starting point for this measurement. Thus, of a cohort of 45-year-old patients whose cancer was detected by the usual clinical means, we would expect half to be alive at age 50. Let us now assume that early detection techniques could identify this carcinoma an average of one year prior to the usual time of clinical diagnosis: that is, the screening of asymptomatic populations could detect this carcinoma one year prior to the time at which the appearance of symptoms causes the average patient to seek medical care. If we performed the type of survival analysis that frequently appears in clinical journals we would make the mistake shown in Figure 2. Assuming that the therapy for this form of cancer is no more effective when applied earlier (double line) than when it is applied at the time of usual diagnosis (single line), we note that the five-year survival among the early diagnosed group is substantially better than that of the group who were not diagnosed until they developed symptoms. However, this is entirely misleading; all we have done is to shift the starting point for the five-year survival measurement backward one

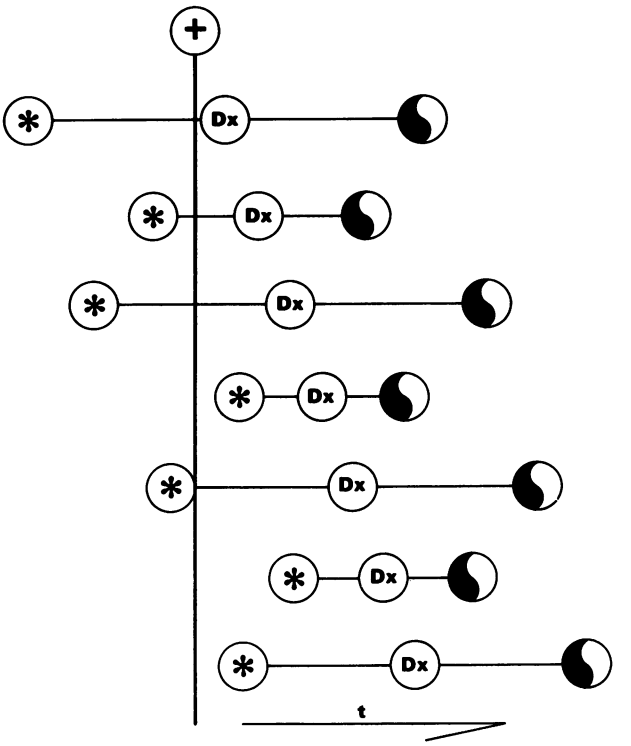

Fig. 4. Expected distribution of long and short preclinical and clinical stages of cancer in a population to be screened. * = onset of disease, Dx = time of usual diagnosis (onset of symptoms), partially shaded circle = death, + and vertical line = time of early detection, t = time elapsed. Reproduced by permission from Sackett, D. L.: The usefulness of laboratory tests in health-screening programs. *Clin. Chem. 19*:366-72, 1973.

year from the usual time of diagnosis to the point at which the early diagnosis could be achieved.[11] The group of 45-year-old patients referrred to earlier would simply have been diagnosed one year earlier, at age 44; the same 50% would be alive at age 50. Thus, we would not have given them an extra year forward of life, but an extra year backward of disease! Selecting an inappropriate starting point for measuring survival, then, is one mistake often made in examining the survival rates of individuals whose disease is diagnosed at an earlier stage. This guarantees an increased survival rate even if the therapy does nothing to control or reverse the course of the disease.

The second common error in analyzing the effectiveness of screen-

ing programs arises out of the relation between the duration of the preclinical (early or asymptomatic) and clinical (late or symptomatic) stages of disease and is illustrated in Figure 3. Studies of cancer of the lung, breast, stomach, and colon indicate that patients with these cancers who have long preclinical stages of disease also tend to have long clinical stages and, conversely, individuals with short preclinical stages tend to have relatively short clinical stages.[11-13] This relation probably characterizes most diseases, but its effect usually has been ignored in analyzing programs of early diagnosis. As Figure 4 illustrates, early diagnosis (the vertical line) will always appear to improve survival because the screening examination will be more likely to pick out those patients whose disease has a long preclinical stage than those where it is short.[14] As a result, when the disease is diagnosed through screening or a periodic health examination, patients can expect longer clinical stages of disease and better short-term survival rates than patients whose diagnosis is made in the usual fashion, even if the therapy instituted as a result of this early diagnosis has no effect whatever.

Have we considered the entire range of the possible effects of screening, labeling of individuals as diseased, and long-term therapy? It has been suggested that, even in the absence of sound evidence that these programs are effective, clinicians simply cannot permit high-risk patients to await the results of proper randomized trials. This is analogous to the individual clinical decisions which we have always made in the case of individual patients with respect to any unproved therapy: the patient simply cannot wait for the treatment to be validated, so we must make decisions on the basis of incomplete evidence.

Instituting the periodic screening of patients as public policy, however, takes on additional dimensions. The individual clinical decision to examine, even if futile, carries with it a relatively low financial cost; as public policy, however, the cost of a massive screening effort becomes sufficiently large that its use may force the reduction, delay, or cessation of other programs of clinical care.

In both the individual and the general case we must also consider the possibility that the intervention, rather than being simply beneficial or useless, may in fact be harmful to health. The magnification of harm through the widespread use of deleterious diagnostic or therapeutic strategies has occasionally had tragic consequences; for example, recall the epidemic of asthma deaths which followed the introduction of

nonprescription bronchodilator aerosols and the risks associated with oral hypoglycemic agents.[15, 16] Further, we have to consider the possibility that labeling patients as diseased may substantially decrease their social, emotional, and occupational ability to function. We have found, for example, some evidence which suggests that the labeling of an individual as hypertensive and the initiation of antihypertensive treatment may cause him to prolong his episodes of absenteeism from work, as compared with other individuals with similar levels of blood pressure who are neither labeled nor treated.[17]

The six clinical questions I have asked were discussed at a meeting of the World Health Organization in 1971 and translated into seven criteria for evaluating screening programs. In their Technical Report No. A24, *Mass Health Examinations as a Public Health Tool*,[18] these criteria are summarized as follows:

1) Screening must lead to an improvement in end-results (defined in terms of mortality; physical, social, and emotional function; pain; and satisfaction) among those in whom early diagnosis is achieved or in the other members of the community.
   a) The therapy for the condition must favorably alter its natural history, not simply by advancing the point in time at which diagnosis occurs, but by improving survival, function, or both. The modification of "risk factors" is not sufficient evidence of effectiveness, nor is the fact that the proposed therapy is "commonly accepted." Claims for therapeutic effectiveness must withstand rigorous methodologic scrutiny, and experimental evidence, such as controlled clinical trials, is a prerequisite. The measurement of survival and other end-results must withstand epidemiologic and biostatistical scrutiny.
   b) Available health services must be sufficient both to ensure diagnostic confirmation among those whose screening is positive and to provide long-term care.
   c) Compliance among asymptomatic patients in whom an early diagnosis has been achieved must be at a level to be effective in altering the natural history of the disease in question.
   d) The long-term beneficial effects, in terms of end-results,

must outweight the long-term detrimental effects of the therapeutic regimen utilized and the "labelling" of an individual as "diseased" or "at high risk."

2) The effectiveness of potential components of multiphasic screening should be demonstrated individually prior to their combination.

3) If the benefits of screening accrue to the community at large rather than, or in addition to, the individual identified (e.g., disease carriers, specific occupations), the community benefit claimed must withstand scientific scrutiny.

   a) The appropriateness of the mix of screening tests to the target population must be considered, acknowledging that differences in the distributions of two diseases may render the combination of their respective screening tests inappropriate.

4) The cost-benefit and cost-effectiveness characteristics of mass screening and long-term therapy must be known. This knowledge is considered essential in developing an appropriate mix of diagnostic and therapeutic services in the face of finite manpower and financial resources. Therefore, a mechanism for the formal periodic weighing of costs against benefits or effectiveness should constitute a basic component of the initial screening activities.

5) The burden of disability for the condition in question (in terms of disease frequency, distribution, severity, and alternative approaches to its detection and control) must warrant action.

6) The cost, sensitivity, specificity, and acceptability of the screening test must be known, and it should lend itself to the utilization patterns of the target population.

7) Ideally, an estimate of the social benefit of preventing, arresting, or curing the condition in question should be known.

These six questions were also applied by task forces in Canada and the United Kingdom, where they have led to conclusions and actions divergent from those prevalent in the United States. With the exception of prescriptive screening among highly selective groups of patients, existing screening and periodic health-examination programs have been found by Canadian and British investigators to have been

conducted either in the absence of, or in direct contradiction to evidence for their clinical effectiveness; they are felt to have very little promise of improving or even maintaining the health of the general population.

These investigations indicate the necessity for further randomized clinical trials of screening, diagnostic, and treatment maneuvers. This need may be somewhat more urgent in Canada, which devotes a higher proportion of its gross national product to health care than does the United States, because it seems likely that no further increments in health expenditures can be anticipated there. On the other hand, excellent trials—such as that performed by Shapiro and his group—are demonstrating that the use of specific screening maneuvers may provide substantial health benefits.[19] We may then witness an irresistible force—a screening program of demonstrated efficacy—meeting an immoveable object: the health budget.

Unless the clinical and scientific community rapidly expands its randomized trials of screening, diagnostic, and treatment maneuvers so that we can free resources currently being spent in worthless clinical procedures to reinvest them in valid clinical innovations, we shall have only ourselves to blame when we are faced with governmental edicts which, for purposes of administrative convenience, arbitrarily restrict health services and the incomes of health professionals.

REFERENCES

1. Garfield, S. R.: Multiphasic health testing and medical care as a right. *New Eng. J. Med. 283*:1087-89, 1970.
2. Bombardier, C., McClaran, J., and Sackett, D. L.: Medical care policy rounds: Periodic health examinations and multiphasic screening. *Canad. Med. Ass. J. 109*:1123-27, 1973.
3. Sackett, D. L.: The usefulness of laboratory tests in health screening programs. *Clin. Chem. 19*:866-72, 1973.
4. Schor, S. S., Clark, T. W., Parkhurst, L. W., Baker, J. P., and Elsom, K. A.: An evaluation of the periodic health examination. *Ann. Intern. Med. 61*:999, 1964.
5. Veterans Administration: Effects of treatment on morbidity in hypertension. I. Results in patients with diastolic blood pressures averaging 115 through 129 mm. Hg. *J.A.M.A. 202*:1028, 1967; Veterans Administration: Effects of treatment on morbidity in hypertension. II. Results in patients with diastolic blood pressures averaging 90 through 114 mm. Hg. *J.A.M.A. 213*: 1143, 1970.
6. Sackett, D. L. and Haynes, R. B., editors: *A Workshop/Symposium on Compliance with Therapeutic Regimens*. McMaster University, Hamilton, Ontario, Canada, May 1974.
7. Cutler, J. L., Ramcharan, S., Feldman, R., Siegelaub, A. B., Campbell, B. A., Friedman, G. D., Dales, L. G., and Collen, M. F.: Multiphasic checkup evaluation study. 1. Methods and population. *Prev. Med. 2*:197-206, 1973.

8. Ramcharan, S., Cutler, J. L., Feldman, R., Sigelab, A. B., Campbell, B., Friedman, G. D., Dales, L. G., and Collen, M. F.: Multiphasic checkup evaluation study. 2. Disability and chronic disease after seven years of multiphasic health checkups. *Prev. Med. 2*:207-20, 1973.
9. Collen, M. F., Dales, L. G., Friedman, G. D., Flagle, C. D., Feldman, R., and Siegelaub, A. B.: Multiphasic checkup evaluation study. 4. Preliminary cost benefit analysis for middle-aged men. *Prev. Med. 2*:236-46, 1973.
10. Dales, L. G., Friedman, G. D., Ramcharan, S., Siegelaub, A. B., Campbell, B. A., Feldman, R., and Collen, M. F.: Multiphasic checkup evaluation study. 3. Outpatient clinic utilization, hospitalization, and mortality experience after seven years. *Prev. Med. 2*:221-35, 1973.
11. Hutchison, G. B. and Shapiro, S.: Lead time gained by diagnostic screening for breast cancer. *J. Nat. Cancer Inst. 41:* 665, 1968.
12. Sutherland, R.: *Cancer: The Significance of Delay.* London, Butterworth, 1960.
13. Feinstein, A. R.: Symptoms as index of biologic behavior and prognosis in human cancer. *Nature 109*:241, 1966.
14. Feinleib, M. and Zelen, M.: Some pitfalls in the evaluation of screening programs. *Arch. Environ. Health* (Chicago) *19*:412, 1969.
15. Inman, W. H. W. and Adelstein, A. M.: Rise and fall of asthma mortality in England and Wales in relation to use of pressurized aerosols. *Lancet 2*:279-85, 1969.
16. University group diabetes program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult onset diabetes. *J.A.M.A.* (Suppl. 2) *19:* 1970.
17. Gibson, E. S., Mishkel, M., Gent., M., Vayda, E., and Sackett, D. L.: Absenteeism from Work Among Hypertensives: The DOFASCO-McMaster Heart Study Group. *Amer. Heart Assoc. Conf. on Cardiovascular Disease Epidemiology.* Tampa, Fla., February 28, 1972.
18. World Health Organization: *Mass Health Examination As a Public Health Tool. Technical Report No. A24.* Geneva, WHO, 1971, pp. 50-51.
19. Shapiro, S., Strax, P., and Venet, L.: Periodic breast cancer screening in reducing mortality from breast cancer. *J.A.M.A. 215*:1777, 1971.