II-3

31. Newman, G. F., Rerych, M. T., Upton, D. C., and Jones, R. H. Comparison of electrocardiographic and left ventricular function changes during exercise. *Circulation* 62: 1204, 1980.
32. Pauker, S. G., and Kassirer, J. P. Clinical application of decision analysis: A detailed illustration. *Semin. Nucl. Med.* 302: 1109, 1980.
33. Pauker, S. G., and Kassirer, J. P. The threshold approach to clinical decision making. *N. Engl. J. Med.* 302: 1109, 1980.
34. Pozen, M. W., D'Agostino, R., Mitchell, J., et al. The usefulness of a predictive instrument to reduce inappropriate admissions to a coronary care unit. *Ann. Intern. Med.* 92: 238, 1980.
35. Pozen, M. W., D'Agostino, R. B., Selker, H. P., et al. A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease. *N. Engl. J. Med.* 310: 1273, 1984.
36. Ransohoff, D. F., and Feinstein, A. R. Is decision analysis useful in clinical medicine? *Yale J. Biol. Med.* 29: 165, 1976.
37. Schwartz, W. B., Wolfe, H. J., and Pauker, S. G. Pathology and probabilities: A new approach to interpreting and reporting biopsies. *N. Engl. J. Med.* 305: 917, 1981.
38. Simel, D. L., Feussner, J. R., Delong, E. R., and Matchar, D.B. Intermediate, indeterminate, and uninterpretable diagnostic test results. *Med. Decis. Making.* 7: 107, 1987.
39. Singer, J. Value of clinical signs in diagnosis of deep vein thrombosis. *Lancet* 1: 1186, 1980.
40. Smith, A. F. Diagnostic value of serum-creatine-kinase in a coronary care unit. *Lancet* 2: 178, 1967.
41. Spodick, D. H. On experts and expertise: The effect of variability in observer performance. *Am. J. Cardiol.* 36: 592, 1975.
42. Weinstein, M. and Fineberg, H. *Clinical Decision Analysis.* Toronto: Saunders, 1980.

*The table for your cirrhotic with the positive α-fetoprotein test should look like this:

		Hepatoma: Yes	Hepatoma: No		
AFP test result	Positive	245 (a)	75 (b)	320 (a+b)	⊕ PV = 77%
	Negative	5 (c)	675 (d)	680 (c+d)	⊖ PV = 99%
		250 (a+c)	750 (b+d)	1000 (a+b+c+d)	

Your patient's likelihood of hepatoma is now 77%

Sackett DL. Clinical Epidemiology

Early Diagnosis

Up to now, this book has focused on the diagnosis of symptomatic disease. However, this section on diagnosis would be incomplete without a final chapter on the early diagnosis of presymptomatic disease. This cornerstone of clinical preventive medicine is often misunderstood and, although its application often can be of enormous benefit to patients, it is frequently oversold. However, the application of a clinical epidemiologic point of view to the concepts and evidence concerning early diagnosis can dispel much of the current confusion and cut through the rhetoric. Once again, this "basic science for clinical medicine" can help us become more effective clinicians. And because early diagnosis entails major issues in therapy as well as diagnosis, its discussion provides a useful bridge between the first two sections of this book.

The Strategies and Objectives of Early Diagnosis

Early diagnosis is sought through two different strategies. First, members of the general public can be invited to undergo tests of various sorts in order to separate them into those with higher and lower probabilities of disease (the former group are then urged to seek medical attention for definitive diagnosis). This form of early diagnosis is called *screening*, and is epitomized by the "Healthorama" booth at the county fair, where passersby are exhorted to submit their arms for blood pressure measurement, their breath for respiratory function, their blood for lipids and their urine for glucose. The combination of several disparate tests in this fashion constitutes *multiphasic screening*.

Other forms of screening are both compulsory and aimed at specific individuals, as when immigrants submit to testing for tuberculosis or airline pilots to periodic electrocardiograms. Many physicians carry out a particular form of screening in which patients are invited to make appointments for a series of screening tests plus a general history and physical examination. These *periodic health examinations* may include extensive, expensive maneuvers such as sigmoidoscopy and stress electrocardiography, especially when performed on senior executives.

The second strategy of early diagnosis relies on the fact that almost everyone sees a physician once in awhile (for example, 75% of Canadians see a physician at least once a year, and 90% see one at least once every 5 years). As a result, clinicians need not call for volunteers but can simply seek early diagnosis when patients come to them for unrelated, intercurrent illnesses. This strategy is called *case finding* and it is exemplified by testing for hyperlipidemia in a 50-year-old who complains of sinusitis, checking the vision and hearing of a septuagenerian who presents with an upper-respiratory infection, and measuring the blood pressure of every adult patient who walks into the office for any reason.

It should be obvious by now that the targets for early diagnosis are not just disease entities such as tuberculosis and cancer. These targets include the predictors or "risk factors" for disease; thus, screening, case finding, and especially periodic health examinations often include questions about alcohol and tobacco use and measurements of blood lipids, glucose, height and weight, and the like.

Although the focus of this chapter is early diagnosis for the sake of the patient, it should be noted that screening and case finding often are executed with other objectives. For example, considerable screening is carried out in order to protect economic wagers; this is what the "life insurance physical" is all about. When you buy life insurance you are betting the company that you are going to die and it is betting you that you won't (at least not until you have paid up your policy). The company may insist that you undergo a physical examination in order to protect its side of the bet; the examination focuses on predictors of early demise (hypertension, diabetes, obesity and the like) and the company reserves the right, based on your results, to refuse to bet with you. Note that the insurance company's objective here is not to maintain or improve your health; it simply wants to win more bets than it loses.

Another reason for screening also places a low priority on health benefits to those screened, and this objective applies when prospective immigrants or teachers are screened for tuberculosis, and crane operators for uncontrolled hypertension. The goal here is the protection of others from tubercle bacilli or falling objects, and this objective is served by denying immigration or employment; treatment of positive screenees is a secondary priority.

A third reason to carry out periodic health examinations, or even case finding, is to establish "baselines" of what patients' electrocardiograms, blood chemistries, and so on are like when they are well; indeed, some clinicians advocate that patients ought to carry copies of these in their wallets and purses. Then, when something does go wrong, the clinician in attendance will have these earlier findings for comparison. Rigorous evaluation of the effects of such baseline information on the outcomes of subsequent illnesses is rare and, as we will show you shortly, continuing skepticism is in order about their worth.

At any rate, this chapter is not directed toward the use of screening, periodic health examinations, or case finding for benefit of life insurance companies, employers, or clinicians. Rather, the focus here is the use of these strategies to benefit patients, especially through the early diagnosis of asymptomatic disease.

Case finding covers a much larger proportion of the population than screening, since screening is carried out only sporadically and rarely captures more than 5 or 10% of eligible citizens. Moreover, because case finding is executed at the same site as definitive diagnosis and therapy, the problem of linking those who "screen positive" to a source of care is obviated. Case finding is better than periodic health examinations as well. Because in many regions the great majority of individuals visit a physician at least once a year, coverage is much more complete, and the busy clinicians can include a bit of "preventive medicine" in every visit rather than trying to squeeze appointments for periodic health examinations into an already overcrowded appointment schedule. Finally, decisions about what to include in case finding have tended to be hardheaded and related to payoffs for patients. Decisions about what to include in periodic health exams, on the other hand, have tended to be influenced more by what it is technically possible to do *to* patients than by what it is clinically possible to do *for* them. It is for these reasons that the Canadian Task Force on the Periodic Health Examination recommended the abandonment of the Periodic Health Examination and its replacement by case finding [4].

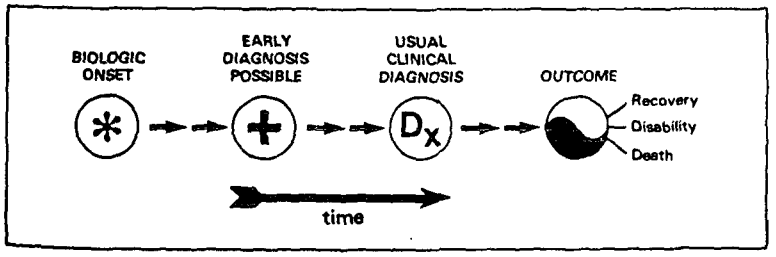
Early Diagnosis and the Natural History of Disease

Regardless of the strategy selected, the objective of early diagnosis is constant: the early detection of presymptomatic disease. This notion of early diagnosis presupposes an orderly biology or "natural history" of disease, divided into the four stages depicted in Figure 5-1.

1. *Biologic onset.* The disease begins with that initial interaction between man, causal factors, and the rest of the environment labelled "biologic onset." We cannot detect the presence of disease at this point in its natural history, but it is there. For some diseases biologic onset occurs at conception, and in many others it probably precedes the later stages by decades.
2. *Early diagnosis possible.* With the passage of time, and although the affected individual remains free of any symptoms, the mechanisms of disease produce structural or functional changes such that, if we applied the correct test, we could achieve the early diagnosis of the disease. At this point, early diagnosis becomes possible by means of screening, case finding, or a periodic health examination.
3. *Usual clinical diagnosis.* In the absence of intervention or spontaneous disappearance, the disease progresses to the point where symptoms appear and the affected individual becomes ill and seeks clinical help. This is the point of "usual clinical diagnosis."
4. *Outcome.* Finally, the disease runs its course and arrives at its "outcome" of recovery, permanent disability, or death.

It is this orderly progression from biologic onset, to the point where early diagnosis is possible, to the time of usual diagnosis, and ultimately to its out-

Figure 5-1. The natural history of disease.



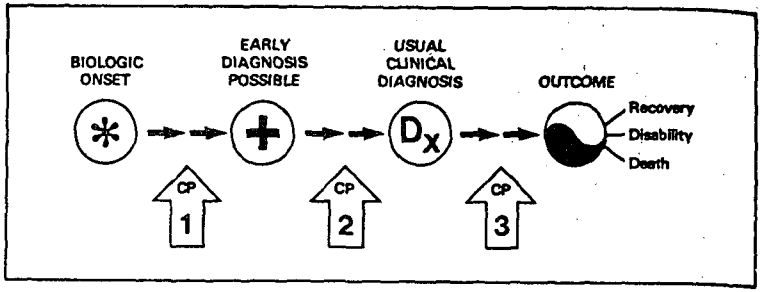


Figure 5-2. Critical points in the natural history of disease. CP = critical point.

come that renders a disease vulnerable to assault through screening, case finding, and the periodic health examination. This orderly progression is not enough, however, because another assumption underlies attempts at early diagnosis. This element was described by Hutchison in 1960 [14] and consists of a "critical point" in the natural history of a disease, *before which therapy is either more effective or easier to apply than afterward.* Now, a disease may have several critical points (arguably pulmonary tuberculosis) or may have none (arguably several cancers), and the location of these critical points along its natural history is crucial to the value of early diagnosis, as shown in Figure 5-2.

If a disease's only critical point were in position 1, between biologic onset and the time that early diagnosis became possible, we can see how screening or case finding would be too late to be of help; the critical point is already passed by the time that detection becomes possible. Similarly, if a disease's only critical point were at position 3, between the time of usual clinical diagnosis and disease outcome, early detection is a waste of time. In this case it would be less trouble all around to wait until symptomatic patients sought clinical help.

It is only when a disease possesses a critical point at position 2, between the time that early diagnosis becomes possible and the time of usual clinical diagnosis, that screening and case finding hold any promise of improving the outcomes of those who have the target disorder.

Admissible Evidence

How do we tell whether a disease has a critical point at position 2 and its detection is worth our critical effort? Unfortunately, the only way to tell for sure is to track down a properly executed randomized trial in which individuals were randomly allocated (by a system analogous to tossing a coin) to receive or not receive the screening or case-finding maneuver. The best standard therapy would have been provided to the experimental patients detected early and to any other patients (experimental or control) detected at the usual time of diagnosis. All patients in both groups would then have been followed up to see whether they succumbed to the target disease.

For example, a group of clinicians and methodologists led by Sam Shapiro [19] randomly allocated over 60,000 New York women (who were enrolled in the Health Insurance Plan [H.I.P.] of Greater New York) to receive or not receive invitations for annual mammography (two views) plus clinical breast examination. These annual examinations were offered for 4 years, and their yield is summarized in Table 5-1.

The potential benefits of mammography, clinical breast examination, and their combination looked promising in terms of the relatively low percent of cases with axillary node metastases at operation. The proof of the pudding, however, was whether deaths from breast cancer were reduced through early diagnosis, and this is shown in Table 5-2 [18, 19].

No benefit could be confirmed among women under age 50, but striking reductions in breast cancer mortality were observed at age 50 and beyond (the mortality from other causes of death was identical, confirming that randomization had produced comparable groups of experimental and control women). This landmark randomized trial (confirmed by additional subsequent trials) demonstrated that a critical point does, in fact, exist in the natural history of breast cancer and that it is located between the point where early diagnosis is possible and the time of usual clinical diagnosis.

Two randomized trials have tested whether multiphasic approaches to early diagnosis are beneficial, and both concluded that they are not. In the first trial [13], Walter Holland, Michael D'Souza, and Anthony Swan at St. Thomas's Hospital Medical School randomly allocated almost 7000 patients from two London general practices to receive or not receive periodic health examinations (that included histories, physical examinations, chest x rays and ECGs, pulmonary function tests, and tests on their blood and feces). Five years later these investigators could demonstrate no beneficial effects on mortality, self-perceived health, risk factors, or absenteeism. All they found was an increase in subsequent visits to the doctor in the screened group.

A short time later, and on the other side of the world, Timothy Durbridge and his colleagues conducted the previously noted randomized trial of hospital admission screening [6]. They randomly allocated about 1500 admissions to an Australian hospital into three groups. One group underwent about 40 tests on admission to the hospital, and the results were given to their clinicians. The second group underwent all the same tests, but the results were not reported. The third group were not screened on admission at all. These investigators were not able to demonstrate any benefit of admission screening in terms of mortality, morbidity, time to diagnosis, levels of nursing care required during their hospitalization, or in the satisfaction of either the patients or their clinicians. Once again, however, they did find that screening led to higher total costs for medical care. So it appears that although automating the clinical laboratory and cranking out batteries of test results may make the laboratory more efficient, it makes looking after patients less efficient.

The several breast cancer screening trials represent the best current evidence that any early diagnostic maneuvers are of any help to cancer patients. Another

Table 5-1. Breast cancers diagnosed early in the H.I.P. study

Mode of early diagnosis	Age at diagnosis								Percentage with positive axillary nodes
	40–49		50–59		60+		Total		
Only by mammography	6	(19%)	27	(42%)	11	(31%)	44	(33%)	16%
Only by clinical exam	19	(62%)	26	(40%)	14	(38%)	59	(45%)	19%
Detected by both modes	6	(19%)	12	(18%)	11	(31%)	29	(22%)	41%
	31	(100%)	65	(100%)	36	(100%)	132	(100%)	

Data modified from S. Shapiro. Evidence of screening for breast cancer from a randomized trial. *Cancer* (Suppl.) 39:2772, 1977.

Table 5-2. Some results of the H.I.P. randomized trial of early diagnosis in breast cancer

	Deaths per 10,000 women per year				
	From breast cancer			From all other causes	From cardiovascular disease
	40–49	50–59	60–69		
Control women	2.4	5.0	5.0	54	25
Experimental women	2.5	2.3	3.4	54	24

Data modified from S. Shapiro. Personal communication, 1976; and Evidence of screening for breast cancer from a randomized trial. *Cancer* (Suppl.) 39:2772, 1977.

trial [1] tested whether 6-monthly chest x rays would benefit lung cancer victims and documented no decrease in mortality in the x-rayed group; by the time a lung cancer is visible on a chest x ray, any critical point in its natural history has already been passed. Subsequent randomized trials tested whether adding sputum cytology to periodic chest x rays would be beneficial, and they were negative as well [7]. Finally, several trials are underway which will determine whether periodic stool testing for occult blood can reduce mortality from colorectal cancer [5, 10]. By these means, crucial knowledge about the clinical benefits of early detection strategies is coming to hand.

Inadmissible Evidence

Isn't this insistence on randomized trials overly pedantic and does it not represent another example of the classic failure of academics to realize what is going on in the front lines of clinical medicine? After all, all clinicians who have had any real experience *know* that early diagnosis works, and each of us accumulates several patients who have done extremely well after having their cancer diagnosed in an early, presymptomatic stage. Common sense tells us that early diagnosis works.

Unfortunately, making clinical judgments about the value of early diagnosis is one of the tasks in which our common sense leads us astray. It goes back to the faulty synthesis of correct observations into incorrect inferences that we discussed back in the chapter on the clinical examination (Chap. 2). In fact, our clinical observations *are* correct: patients whose cancers are diagnosed early do have better 5-year survivals than other patients diagnosed in later, symptomatic stages. It is our inference that these observations prove the value of early diagnosis that is faulty; in fact, *early diagnosis will always appear to improve survival, even when therapy is worthless!*

There are three reasons for this paradox. First, the patients and other people who volunteer for screening and periodic health examinations appear to be healthier before they start. This can be demonstrated by going back to the H.I.P

trial we discussed a bit earlier. A reexamination of the right-hand columns of Table 5-2 confirms the effectiveness of the randomization: the overall experimental and control groups were identical for causes of death for which they were not screened, and deaths from other causes and from cardiovascular disease were identical.

Now, only about two-thirds of experimental women responded to their invitation to attend for mammography and clinical breast examination, and the other third refused. This provided an opportunity to compare those who volunteered with those who refused for causes of death for which they were never screened. The results are shown in Table 5-3 [18, 19].

In Table 5.3 we see that the subset of experimental women who volunteered for breast cancer screening experienced only about *half* the mortality from other causes exhibited by those women who refused; the volunteers fared far better for diseases for which they were never screened and never (on this basis) treated! Volunteers for screening are generally a strange and healthy lot, and we cannot generalize from them to our other patients.

The next reason that early diagnosis will always appear to improve survival, even when therapy is worthless, arises from the way that we measure survival after cancer is detected. This is usually done as shown in Figure 5-3.

Here we see 5 years of follow-up of a group of cancer patients identified at the time of usual clinical diagnosis, 10% of whom die each year, yielding a 5-year survival of 50%. If we diagnosed a group of such patients at age 45, half of them would be alive at age 50.

Now, suppose that we developed a screening test that could detect this cancer one year earlier, before such patients developed symptoms. Suppose further that the treatment of this cancer was no more effective when applied early than when applied at the time of usual clinical diagnosis. If we carry out the standard (but incorrect) comparison of our old and new results, we are likely to make the mistake shown in Figure 5-4.

The old results are expressed in a solid line as before, with zero-time the time of "usual clinical diagnosis." When the new results are charted, however, a zero-time corresponding to the 1-year previous "early diagnosis" is used and

Table 5-3. More results of the H.I.P. trial

		Deaths per 10,000 women per year	
		From all causes	From cardiovascular causes
Control women		54	25
Experimental women	Volunteered	42	17
	Refused	77	38

Data modified from S. Shapiro. Personal communication, 1976; and Evidence of screening for breast cancer from a randomized trial. *Cancer* (Suppl.) 39:2772, 1977.

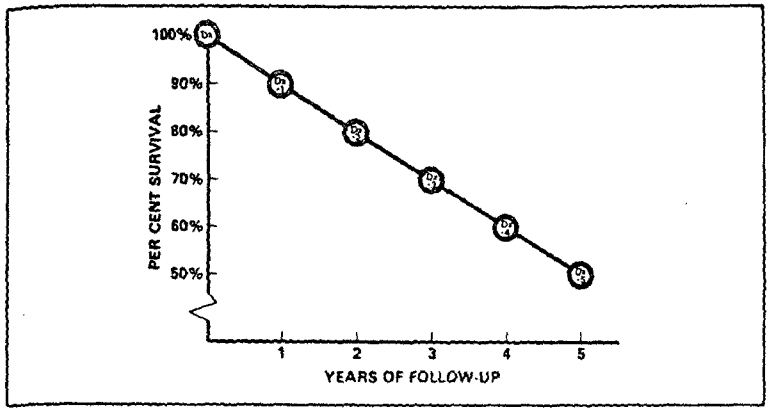


Figure 5-3. Survival following the detection of cancer at the time of usual clinical diagnosis.

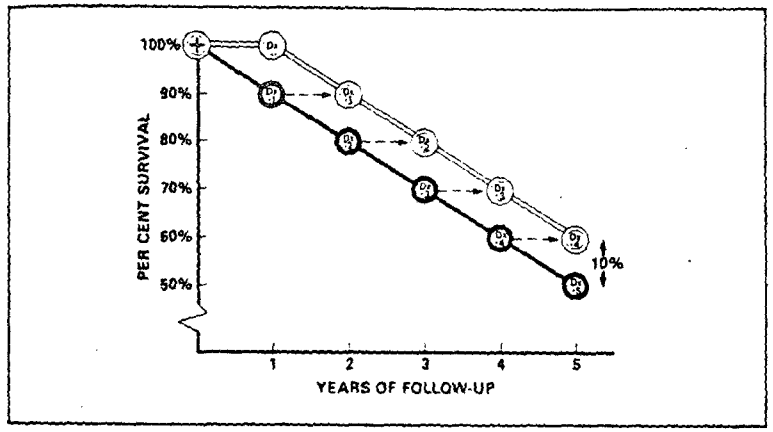


Figure 5-4. Failure to correct for "zero-time shift" in assessing the value of early diagnosis.

the resulting survival curve, illustrated by the open line, is shifted 1 year to the right. As a result of this "zero-time shift," even when therapy is worthless a 10% improvement in 5-year survival is guaranteed.

This apparent improvement in 5-year survival is fictitious, however, for all we have done is to shift the starting point for the five-year survival measurement 1 year backward, from the time of usual clinical diagnosis—the "Dx" sign—to the "+" sign of early diagnosis. Our group of 45-year-old cancer patients are simply diagnosed 1 year earlier, at age 44. Only half of them will be alive at age 50, as before, and we have given them not an extra year forward of life but an

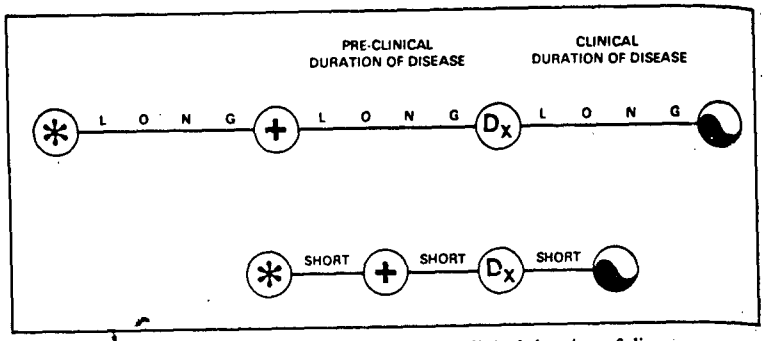


Figure 5-5. The relation between the preclinical and clinical duration of disease.

extra year backward of disease!* This is a key issue, and we will come back to it again later.

The final reason why early diagnosis will appear to improve survival, even when therapy is worthless, has to do with fast and slow growing tumors and is illustrated in Figure 5-5.

Occasionally, after we have diagnosed a cancer of the lung, stomach, or colon we go back to earlier radiographs and discover, to our dismay and chagrin, that the cancer was already there on the previous study. Usually it was too small or indistinct to have been diagnosable on the earlier film, but its location and configuration are enough to convince us that it was there. Studies of such cases and their subsequent clinical courses have taught us that some cancers grow very quickly and others, even of the same organ, grow much more slowly [9, 22]. These fast or slow growth rates tend to operate throughout the natural history of a given patient's disease, as shown in Figure 5-5. Patients with long preclinical durations of disease tend to have long clinical durations of disease, and those with short preclinical durations tend to have short and rapidly fatal clinical durations.

Although this relationship between the preclinical and clinical durations of disease may characterize most diseases, its effect on the assessment of early diagnosis was ignored until it was pointed out by Manning Feinleib and Marvin Zelen in 1969 [8]. This effect is illustrated in Figure 5-6.

Once again, patients detected through early diagnosis (depicted by the vertical staff topped by the "+" sign) will have longer survival than those detected at the time of usual clinical diagnosis, even when therapy is worthless. This is because the slow growing tumors are detectable longer than the fast growing ones and will, therefore, be preferentially identified by any early diagnosis strategy. The fast growing tumors, with their shorter survival, will be left for routine diagnosis, and the comparison distorted.† Note here that even correcting for

*This "zero-time shift" is also commonly referred to as "lead-time bias."

†The preferential detection of slow growing tumors is often called "length-time bias."

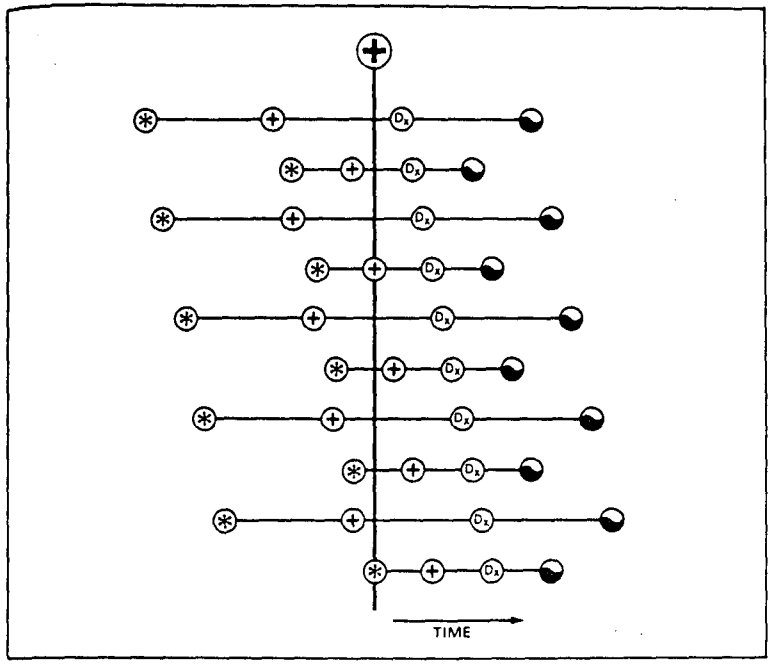


Figure 5-6. Preclinical duration of disease and the likelihood of early diagnosis.

"zero-time shift" will fail to overcome this pitfall in assessing the value of early diagnosis.

In summary, there are three reasons why common clinical inferences about the value of early diagnosis are inadmissible: the volunteer effect, the zero-time shift problem (lead-time bias), and the preferential detection of slowly progressive disease (length-time bias). As a result of these three factors, early diagnosis will always appear to improve survival, even when therapy is worthless. Accordingly, we must insist on evidence from proper randomized trials of early diagnostic maneuvers before we act.

The Hazards of Inappropriate Early Diagnosis

Some readers may still regard our insistence on evidence from randomized trials as excessively rigorous. After all, it will take several years to carry out the necessary randomized trials, and in the meanwhile countless patients will lose their chance for early diagnosis.

We believe that a firm stand requiring hard evidence from randomized trials is correct, both for the reasons already given and for two more: the implied pledge to the patient who undergoes early diagnosis, and the potential for doing harm to the early diagnosed patient.

Consider first our implied (and frequently stated) pledge to symptomatic patients when they seek us out and ask us for help. Although we promise to do our best for them, we do *not* guarantee them better health (in fact, clinicians who do provide such guarantees are often labeled quacks and charlatans). We diagnose and treat these symptomatic patients as best we can, realizing that standard therapies for their conditions may never have been validated and may not even work, and that even the validated ones will not work for everyone. Their symptoms and their call for help force us to act, often on the basis of incomplete evidence about the value of what we are about to do.

Now consider our implied (but almost never stated) pledge to the symptom-free citizens whom we solicit for screening or periodic health examination. The tables are reversed: we are seeking them out. They feel fine, but we are asking them to come for tests and possibly extensive, prolonged therapy. Surely in this case our implied promise to them is better health; not that their subsequent treatment *may* work but that it *does* work; not that we will simply do our best but that we will make them healthier. When we impose ourselves on the public in this fashion, we require very firm evidence that our early diagnosis and subsequent therapy will do more good than harm.

This last comment introduces the second reason for demanding hard evidence for the benefit of early diagnosis: the potential for doing harm to the early diagnosed patient. Advocates of screening and periodic health examinations sometimes act as if there were only two possible outcomes of these undertakings: on the one hand, positive benefit and, on the other, no effect. What is often ignored is the third possible outcome: harm.

This harm can be of two sorts. First, the diagnosis can be wrong. The risk of making a false-positive diagnosis becomes huge when we are looking for rare disorders in symptomless patients. An example of this is shown in Table 5-4.

Although the sensitivity (PiD rate, TP rate) and specificity (NiH rate, TN rate) of the B27 test for ankylosing spondylitis are excellent (90% and 95%, respectively), this disorder is so rare that, of every 100 individuals with a negative family history but a positive B27 test, 98 do *not* have ankylosing spondylitis (in fact, this is probably a low estimate of the extent of misdiagnosis; the pretest probability of this disorder is probably closer to .05%, which means that 99 of 100 individuals with positive B27 tests are misdiagnosed) [11, 21]. We are left with lots of individuals with positive tests but no symptoms, and most of them do not have the target disorder. The best way to resolve this sort of dilemma is to stay out of it in the first place.

Second, the treatment initiated as a result of early diagnosis may, in fact, do more harm than good. A catastrophic example is provided by a widespread treatment for hyperlipidemia. Most periodic health examinations in the 1960s and 1970s included measurements of blood lipids, and both screening and case finding for hyperlipidemia were very popular, especially in the United States. When elevated serum cholesterol was found on these examinations, dietary advice was usually offered and, if the elevation persisted, drug therapy with clofibrate often was initiated to bring the cholesterol level down. As a result, it was not unusual

Table 5-4. The B27 test in the early diagnosis of ankylosing spondylitis

B27 test	Ankylosing spondylitis: Yes	Ankylosing spondylitis: No	
Positive	90 (a)	510 (b)	600 (a+b)
Negative	10 (c)	9390 (d)	9400 (c+d)
	100 (a+c)	9900 (b+d)	10,000 (a+b+c+d)

Sensitivity = $\frac{a}{a+c} = \frac{90}{100} = 90\%$

Specificity = $\frac{d}{b+d} = \frac{9390}{9900} = 95\%$

Positive predictive value = Posttest likelihood or posterior probability of disease = $\frac{a}{a+b} = \frac{90}{600} = 15\%$

Negative predictive value = Posttest likelihood or posterior probability of *no* disease = $\frac{d}{c+d} = \frac{9390}{9400} = 99.9\%$

Prevalence = Pretest likelihood = *Prior probability of disease* = $\frac{a+c}{a+b+c+d} = \frac{100}{10,000} = 1\%$

to find ostensibly healthy, asymptomatic, middle-aged men taking this drug as a result of having undergone a "cholesterol count."

It had been known for some time that clofibrate could lower serum cholesterol levels, and so it was assumed that lowering this "risk factor" with this drug had to be beneficial. However, it was only after this lipid screening and risk factor modification had been going on for several years that a proper randomized trial of clofibrate was carried out in healthy middle-aged men with hypercholesterolemia.* The results were shocking: mortality in clofibrate-treated men was 17% *higher* than among men on placebos, and this excess mortality continued for four years following withdrawal of the drug [23].

Extrapolation of these results to the United States where most of this screening and treatment took place provided a somber lesson. It was estimated that the detection of hyperlipidemia and its treatment with clofibrate led to the deaths of over 5000 asymptomatic Americans [16]. Unless we are willing to permit a future recurrence of a catastrophe of this magnitude, we must insist on sound evidence from randomized trials that early diagnosis and specific forms of therapy do more good than harm *before* we solicit our patients and the public to submit to them.

The third sort of harm that can come from early diagnosis is the damage done when we tell someone who feels well that they are sick. This phenomenon, which we shall call "labeling," is well known in social psychology but only now becoming recognized by most clinical disciplines.

Consider what happens when we tell a symptomless steelworker for the first time that he has hypertension [12, 20]. His absenteeism from work due to illness, which was previously indistinguishable from that of his normotensive coworkers, doubles in the next year and remains high. He is not sick more often, but when he develops minor illnesses he stays away roughly twice as long, treating himself as if he were quite fragile (or, in social psychological terms, he "adopts the sick role"). His psychological well-being declines and he is less satisfied with his work and his marriage. His income suffers, and he perceives himself as not advancing as rapidly as his workmates.

There are a couple of additional surprises in our follow-up of such patients. We cannot pass off their subsequent behavior as simply representing the side effects of antihypertensive drugs, for these same changes occur in men who are only labeled but never treated. Moreover, teaching these men about their disease and its treatment (which, as you will learn later in this book, does nothing for their compliance with their antihypertensive drug regimens) tends to accentuate the effects of labeling.

Other investigators have, with some exceptions, confirmed these observations in hypertensives elsewhere [15], and even more dramatic effects of labeling are seen with other diagnoses. As you may recall from our earlier chapter on the clinical examination, normal children who were misdiagnosed as having organic heart disease show as much deterioration in physical and social function as children who really do have damaged hearts. More recently, a randomized trial of the Denver Developmental Screening Test revealed that children who screened positive and received counselling and referral to all sorts of health professionals and agencies fared no better in their future academic, cognitive, or developmental performance than children whose positive screening results were never revealed [3]. Moreover, the labeled children's parents were far more likely to worry about their school performance, and teachers tended to report more behavioral problems among them.

*This trial was a forerunner of the later, positive trials of other approaches to lowering serum cholesterol.

Labeling happens all the time. We accept it as inevitable when it results from routine diagnosis among symptomatic patients; after all, they come to us asking for it. When labeling occurs through early diagnosis, however, we must consider it in its own right. The screenee has no symptoms and is not asking to be labeled. We must decide whether we really ought to take away someone's health through early diagnosis by labeling him or her with a disease that requires intervention.

This labeling of ostensibly healthy individuals as diseased happens every time we achieve an early diagnosis, regardless of whether the subsequent therapy is effective, worthless, or harmful. When therapy is effective, both we and the patient are usually willing to live with the label, for the ultimate outcome is worth it. If, however, the therapy is worthless or no more effective when applied early than when applied later among symptomatic patients, we have needlessly taken away "healthy" time from our patients and have merely made them sick longer. Moreover, when therapy is harmful we have both robbed our patients of healthy months or years, and have needlessly exposed them to damaging interventions.

Our final set of comments about the hazards of inappropriate early diagnosis steps away from our usual focus on individual patients and looks at populations, professional credibility, and the search for useful new knowledge [17]. The decision to case-find in an individual patient, even when the associated treatment is unproven, costs relatively little and limits the human risk of side effects and toxicity should the remedy later be shown to be useless or harmful. However, this opportunity for waste and harm is progressively magnified when an untested procedure is applied throughout a practice (as universal case finding or as part of a periodic health examination scheme) or in the town at large (as a screening program).

In addition, the widespread implementation of untested methods of early diagnosis renders their subsequent rigorous evaluation much more difficult and less decisive; indeed it may even become impossible to correct the original error. Moreover, research into alternative strategies is discouraged in the interim.

How to Decide When to Seek an Early Diagnosis

This chapter could close with a list of early diagnoses worth pursuing. Instead, we will end it with some guides that clinicians can use to decide when, and on which patients, they should attempt the early diagnosis of presymptomatic disease. We have adopted the strategy of showing you how to decide for yourself, rather than simply telling you which screening tests to do, for two reasons. First, any specific advice could be out of date by the time this book is published.

Table 5-5. How to decide when to seek an early diagnosis

1. Does early diagnosis really lead to improved clinical outcomes (in terms of survival, function, and quality of life)?
2. Can you manage the additional clinical time required to confirm the diagnosis and provide long-term care for those who screen positive?
3. Will the patients in whom an early diagnosis is achieved comply with your subsequent recommendations and treatment regimens?
4. Has the effectiveness of individual components of a periodic health examination or multiphasic screening program been demonstrated prior to their combination?
5. Does the burden of disability from the target disease warrant action?
6. Are the cost, accuracy, and acceptability of the screening test adequate for your purpose?

Important randomized trials are still underway to determine the value of testing stools for occult blood, and of treating certain forms of elevated blood pressure. Their results will render obsolete any set of recommendations we might make.

The second reason returns to the basic objective of this book: helping clinicians make better clinical decisions by showing them how to critically appraise clinical articles. We want to show you how to decide whether to apply a particular maneuver for early diagnosis. To tell you *which* tests to use would defeat the purpose of the book.

When deciding whether to carry out a given test for the early diagnosis of presymptomatic disease, you should ask yourself the questions* that are listed in Table 5-5.

DOES EARLY DIAGNOSIS REALLY LEAD TO IMPROVED CLINICAL OUTCOMES (IN TERMS OF SURVIVAL, FUNCTION, AND QUALITY OF LIFE)?

As you have learned earlier, the therapy for the condition must favorably alter its natural history, not simply by advancing the time at which diagnosis occurs, but by improving survival, function, quality of life or all three. The modification of "risk factors" is not sufficient evidence of effectiveness, nor is the fact that the proposed therapy is widely accepted. Claims for therapeutic benefit must withstand close scrutiny and experimental evidence from randomized trials is a prerequisite.†

You need to be sure that, on average, the long-term beneficial effects of therapy outweigh the long-term detrimental effects of the treatment regimen and labeling of patients as diseased.

*A related set of questions has been developed by David Cadman, Larry Chambers, and William Feldman for use in deciding whether to launch a community-wide screening program [2].
†Detailed guides for scrutinizing the results of randomized trials are presented in Chapter 7.

CAN YOU MANAGE THE ADDITIONAL CLINICAL TIME REQUIRED TO CONFIRM THE DIAGNOSIS AND PROVIDE LONG-TERM CARE FOR THOSE WHO SCREEN POSITIVE?

Increased demands on your time start, not end, with early diagnosis and you need to be sure that you have enough of it. Large numbers of labeled but untreated hypertensives attest to the size of this problem.

WILL THE PATIENTS IN WHOM AN EARLY DIAGNOSIS IS ACHIEVED COMPLY WITH YOUR SUBSEQUENT RECOMMENDATIONS AND TREATMENT REGIMENS?

If patients will not take their medicine, all the foregoing screening and diagnosis, however elegantly they were conceived and executed, are nullified. Once again, all that you have left is a labeled patient.

This question may best be answered on an individual patient-by-patient basis, and we know many clinicians who say they would withhold early diagnostic tests from patients with past histories of noncompliance.*

HAVE THE EFFECTIVENESS OF INDIVIDUAL COMPONENTS OF A PERIODIC HEALTH EXAMINATION OR MULTIPHASIC SCREENING PROGRAM BEEN DEMONSTRATED PRIOR TO THEIR COMBINATION?

The appropriateness of a mix of tests must consider whether differences in the distributions of two diseases render the combination of their respective screening tests nonsensical. It was this consideration that led the Canadian Task Force on the Periodic Health Examination to propose quite different "health protection packages" for patients of different age, sex, and social status [4].

DOES THE BURDEN OF DISABILITY FROM THE TARGET DISEASE WARRANT ACTION?

The disease you are searching for should be either so common or so awful as to warrant all the work and expense of detecting it in its presymptomatic state.

ARE THE COST, ACCURACY AND ACCEPTABILITY OF THE SCREENING TEST ADEQUATE FOR YOUR PURPOSE?

After acquiring one or more "belts" in the previous chapter, this question may appear easy to answer. There is an additional twist, however, because in early diagnosis these tests are applied to asymptomatic people with generally very low pretest likelihoods (or prior probabilities) of disease. Accordingly, you may have to juggle the cutoff for a positive test or, better still, make full use of the likelihood ratios for different levels of the test result if you are to avoid both missing cases who could benefit, on the one hand, and avoiding unnecessary labeling, on the other.

*We have devoted Chapter 8 to some practical clinical aspects of compliance.

References

1. Brett, G. Z. The value of lung cancer detection by six-monthly chest radiographs. *Thorax* 23: 414, 1968.
2. Cadman, D., Chambers, L. W., Feldman, W., and Sackett, D. L. Assessing the effectiveness of community screening programs. *J.A.M.A.* 251: 1580, 1984.
3. Cadman, D., Chambers, L. W., Walter, S. D., et al. Evaluation of public health preschool child developmental screening: The process and outcomes of a community program. *Am. J. Pub. Health* 77: 45, 1987.
4. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can. Med. Assoc. J.* 121: 1193, 1979.
5. Chamberlain, J. Is screening for colorectal cancer worthwhile? *Br. J. Cancer* 62: 1, 1990.
6. Durbridge, T. C., Edwards, F., Edwards, R. G., and Atkinson, M. An evaluation of multiphasic screening on admission to hospital. Précis of a report to the National Health and Medical Research Council. *Med. J. Aust.* 1: 703, 1976.
7. Eddy, D. M. Screening for lung cancer. *Ann. Intern. Med.* 111: 232, 1989.
8. Feinleib, M., and Zelen, M. Some pitfalls in the evaluation of screening programs. *Arch. Environ. Health* 19: 412, 1969.
9. Feinstein, A. R., Schimpf, C. B., and Hull, E. W. A reappraisal of the staging and therapy for patients with cancer of the rectum. *Arch. Intern. Med.* 135: 1441, 1975.
10. Gilbertson, V. A., Church, T. R., Grewe, F. J., et al. The design of a study to assess occult-blood screening for colon cancer. *J. Chron. Dis.* 33: 107, 1980.
11. Hawkins, B. R., Dawkins, R. L., Christiansen, F. T., and Zilko, P. J. Use of the B27 test in the diagnosis of anklyosing spondylitis: A statistical evaluation. *Arthritis and Rheum.* 24: 743, 1981.
12. Haynes, R. B., Sackett, D. L., Taylor, D. W., et al. Increased absenteeism from work after detection and labeling of hypertensive patients. *N. Engl. J. Med.* 299: 741, 1978.
13. Holland, W. W., D'Souza, M. D., and Swan, A. V. Is mass screening justified? *T. Soc. Geneesk* 56: 22, 1978.
14. Hutchison, G. B. Evaluation of preventive services. *J. Chron. Dis.* 11: 497, 1960.
15. Macdonald, L. A., Sackett, D. L., Haynes, R. B., and Taylor, D. W. Labelling in hypertension: A review of the behavioral and psychological consequences. *J. Chron. Dis.* 37: 933, 1984.
16. Round the World: Clofibrate. Correspondents Report. *Lancet* 1: 771, 1981.
17. Sackett, D. L., and Holland, W. W. Controversy in the detection of disease. *Lancet* 1: 357, 1975.
18. Shapiro, S. Personal communication, 1976.
19. Shapiro, S. Evidence of screening for breast cancer from a randomized trial. *Cancer* (Suppl.) 39: 2772, 1977.
20. Taylor, D. W., Haynes, R. B., Sackett, D. L., and Gibson, E. S. Long-term follow-up of absenteeism among working men following the detection and treatment of their hypertension. *Clin. Invest. Med.* 4: 173, 1981.
21. van der Linden, S. M., Valkenburg, H. A., de Jongh, B. M., and Cats, A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. *Arth. Rheum.* 27: 241, 1984.
22. Weiss, W., Boucot, K. R., and Cooper, D. A. Growth rate in the detection and prognosis of bronchogenic carcinoma. *J.A.M.A.* 198: 1246, 1966.
23. World Health Organization. Trial on primary prevention of ischaemic heart disease using clofibrate to lower serum cholesterol: Mortality followup. Report of the Committee of Principal Investigators. *Lancet* 2: 379, 1980.

II

Management