Screening for Cancer: 
Evaluating the Evidence

THOMAS J. GATES, M.D., Lancaster General Hospital, Lancaster, Pennsylvania

Many patients expect to undergo screening tests for cancer. In evaluating screening procedures, physicians must take into account the known effects of lead time, length and screening biases, all of which can result in an overestimation of the benefits of screening. The gold standard by which a screening test is evaluated remains the perspective, randomized controlled trial, demonstrating reduced morbidity and mortality. The magnitude of benefit from screening is best expressed in terms of the number of patients needed to screen. This value ranges from approximately 500 to 1,100 for proven screening interventions. These concepts are illustrated by controversies in current screening recommendations for cancers of the cervix, lung, colon, breast and prostate, which together account for more than 50 percent of cancer deaths in the United States. (Am Fam Physician 2001;63:513-22.)

Patients frequently consult family physicians with the expectation of undergoing screening tests for cancer. However, controversy remains as to which screening procedures are justified and how proposed screening procedures should be evaluated. Failure to understand basic concepts can lead to unrealistic expectations for screening programs and sometimes to inefficient use of resources. This article reviews the kind of evidence required to justify screening tests for cancer, with the goal of guiding family physicians through current and future screening controversies.

Key Issues Involving Screening

Screening can be defined as the application of diagnostic tests or procedures to asymptomatic people for the purpose of dividing them into two groups: those who have a condition that would benefit from early intervention and those who do not. The importance of screening is deeply embedded in primary care: family physicians believe in the value of detecting disease at an early, asymptomatic stage when it is more likely to be amenable to treatment and cure. However, it is important to recognize that the ultimate purpose of screening is to reduce morbidity and mortality. If improved outcomes cannot be demonstrated, the rationale for screening is lost. Early diagnosis by itself does not justify a screening program. The only justification for a screening program is early diagnosis that leads to a measurable improvement in outcome.

ETHICAL ISSUES

Although screening is unquestionably important, other issues of equal importance include using scarce resources efficiently and rationally, refraining from unproven or ineffective interventions and doing no harm. In fact, sometimes there are good reasons not to screen. Basic to this is the understanding that the proposal to screen an asymptomatic patient involves a fundamental shift in the physician-patient relationship.

In ordinary medical practice, the patient initiates an encounter because of a troubling symptom. The physician pledges to help but can make no guarantee and is not responsible if the symptom turns out to represent something beyond the ability of current medical practice to cure. By contrast, a screening test is usually initiated by the physician (or indirectly, by professional or advocacy groups) and, in this situation, there is an "implied promise" not just that the screening procedure might be beneficial, but that it is in fact beneficial, that it will do more good than harm.2,3

See editorial on page 440.
These considerations add an often neglected ethical dimension to screening.\(^2\)\(^4\)\(^5\)

**THE IMPACT OF FALSE-POSITIVE AND FALSE-NEGATIVE TESTS**

Even if the benefits of a screening test have been proved, the benefits accrue to only a few persons. By contrast, all persons participating in a screening program are at risk for harm.\(^2\) Over and above the cost and discomfort of the actual test, the most important potential harm is the risk of a false-positive result. Because diseases being screened for have a low prevalence, even the best screening test will have a low positive predictive value (often 10 to 20 percent). Thus, most positive results will be false-positive, leading to further work-up and patient anxiety. False-negative results can also be harmful, by providing false reassurance and thereby encouraging patients to neglect important symptoms. Finally, a screening test may correctly diagnose a disease but, if the resulting therapy is ineffective or harmful, the patient has been harmed rather than helped.\(^3\)

In light of these considerations, justification of screening tests requires a more rigorous standard of evidence than is usually applied in ordinary clinical practice. Table 1 defines the characteristics of an ideal screening test.

**Bias in the Evaluation of Screening Tests**

The evaluation of screening tests is complicated by certain biases that occur when a disease is diagnosed by screening in the asymptomatic period. The nature of these biases is such that “early diagnosis will always appear to improve survival, even when the therapy is worthless.”\(^3\)

**FEASIBILITY OF SCREENING**

Figure 1\(^3\) shows how the “critical point” in the natural history of a disease determines the feasibility of screening. The disease progresses from its biologic onset at the cellular level (e.g., a mutation) and through an asymptomatic period when it is theoretically detectable by a screening test, eventually culminating in the onset of symptoms, when the diagnosis can be made by the usual clinical means. The critical point can be defined as that point in the natural history of the disease before which therapy is relatively effective and after which therapy is relatively ineffective. In the case of cancer, the critical point would represent the time at which regional or distant metastasis occurs.

Whether a screening test is effective depends on whether the critical point occurs before, during or after the asymptomatic period. If the critical point occurs early, as in lung cancer, screening will not be effective because the disease will have “escaped from cure” before it was detectable by screening. If the critical point is late, as in endometrial cancer, screening is

---

**Table 1**

**Characteristics of the Ideal Screening Program**

<table>
<thead>
<tr>
<th>Features of the disease</th>
<th>Significant impact on public health</th>
<th>Asymptomatic period during which detection is possible</th>
<th>Outcomes improved by treatment during asymptomatic period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features of the test</td>
<td>Sufficiently sensitive to detect disease during asymptomatic period</td>
<td>Sufficiently specific to minimize false-positive test results</td>
<td>Acceptable to patients</td>
</tr>
<tr>
<td>Features of the screened population</td>
<td>Sufficiently high prevalence of the disease to justify screening</td>
<td>Relevant medical care is accessible</td>
<td>Patients willing to comply with further work-up and treatment</td>
</tr>
</tbody>
</table>

unnecessary because the disease is curable even when it presents with clinical symptoms. Screening will have a potential effect on the natural history of the disease only when the critical point occurs sometime during the asymptomatic period, as in cervical cancer.

SCREENING BIAS

Screening bias (a type of selection bias) occurs because subjects who volunteer for screening (or, in randomized trials, those who comply with screening recommendations) tend to be healthier than those who do not volunteer or do not comply, with lower rates of mortality not just from the disease in question but from all causes. Thus, an observed benefit may be due not to the screening intervention but only to the self-selection of a healthy cohort of volunteers.

LEAD-TIME BIAS

Figure 2 illustrates the important concept of lead-time bias, which is best explained schematically. Lead-time bias occurs when the asymptomatic period in the natural history of the disease is not taken into account. For example, assume a hypothetic disease that is 100 percent fatal, with an average survival rate of three years from the time of clinical presentation and a preceding asymptomatic period of four years. In an unscreened population, the diagnosis will be based on clinical symptoms, and the median time from diagnosis to death will be three years. However, in a screened population, the diagnosis will be made in the asymptomatic period, an average of two years before clinical symptoms occur. The time from diagnosis to death will now be five years (two years in the asymptomatic period plus three years in the clinical period).

FIGURE 2. Schematic representation of lead-time bias. Assuming a fatal disease with an average survival of three years from the time of clinical presentation, the screened population will seem to have better outcomes in terms of median survival. However, these patients are not actually living longer than the unscreened population but are merely finding out about their disease at an earlier point in the natural history.

Natural History of a Disease
When the two populations are compared in terms of median survival (or five-year survival), the screened population will appear to have better outcomes even without therapy. However, as can be seen in Figure 2, the screened patients are not actually living longer but only finding out about their disease at an earlier point in its natural history. Screening has given them not an extra two years “forward” of life but an extra two years “backward” of disease.3

In studies of screening efficacy, the only way to avoid lead-time bias is to compare actual mortality rates in the screened and unscreened populations. Mortality rates are not influenced by the timing of diagnosis, whereas surrogate measures like median survival or five-year survival are sensitive to the elapsed time from diagnosis until death and will therefore be skewed by a screening program.

LENGTH BIAS

Length, or length-time, bias (Figure 3) occurs because of the heterogeneity of disease, which presents across a broad spectrum of biologic activity. In the case of cancer, some patients will have fast-growing, aggressive tumors with short asymptomatic periods and rapid progression from symptoms to death. Other patients will have slower-growing, less aggressive tumors that are less likely to metastasize and, therefore, have a better prognosis. These less aggressive tumors have a longer asymptomatic period and are therefore more likely to be identified in a screening program.

In the hypothetical example portrayed in Figure 3, with a screening interval of one year, aggressive tumors have at most a 50 percent chance of being diagnosed by screening (the other 50 percent progressing from onset through the asymptomatic period to clinical symptoms in the interval between screening examinations). By contrast, less aggressive tumors have a long asymptomatic period and are therefore almost certain to be found on annual screening. When a cohort identified by screening (e.g., mammography) is compared with a cohort identified by clinical presentation (e.g., palpable mass), less aggressive tumors will be overrepresented in the screening cohort, and more aggressive tumors will be overrepresented in the clinical presentation cohort. Even in the absence of therapy, the cohort identified by screening will have a better prognosis. A screening program may appear to improve survival when in fact it has only preferentially selected out the subgroup with the best prognosis.3

Evaluation of Screening Tests

Because of the complex and unpredictable nature of these various biases, the only reliable way to prove the effectiveness of a proposed screening program is to demonstrate...
Because of lead time and length bias, early diagnosis will always appear to improve survival, even when therapy is ineffective.
Current Controversies in Cancer Screening

An appreciation of these concepts can help us understand many of the current controversies in cancer screening. The discussion that follows is not meant to be comprehensive but simply to illustrate the importance of these considerations in evaluating the evidence for screening. The following information highlights controversies involving five types of cancer (Table 4) that together account for more than 50 percent of cancer deaths in the United States.

CERVICAL CANCER

Papanicolaou smear screening for cervical cancer represents the prototype of a successful cancer screening program. Although there has never been a randomized controlled trial to demonstrate its effectiveness,\(^1\) historical data from British Columbia document an 80 percent decrease in mortality caused by cervical cancer between 1955 and 1988.\(^2,3\) Using the absolute risk reduction from this data, we can calculate an NNS of 1,140, meaning that 1,140 women would have to be regularly screened over 10 years to prevent one death from cervical cancer.

Current controversies in screening for cervical cancer include the proper interval for Pap smears\(^4,5\) and the role of new technologies that may increase sensitivity but at a high marginal cost. At present, no reliable data exist with which to resolve these controversies.

LUNG CANCER

By contrast, screening for lung cancer with chest radiography, sputum cytology, or both, is the paradigm of an ineffective screening program. Although most physicians are aware that these tests are no longer indicated for screening, many are not aware of the history of screening for lung cancer.\(^6\) Various proposals to screen smokers for lung cancer were made throughout the 1950s, and by 1959 published reports documented a shift to earlier-stage disease and improved five-year survival rates in patients diagnosed by screening, compared with those diagnosed clinically. Screening smokers by annual chest roentgenograms was subsequently endorsed by the American Cancer Society (ACS), but eventually three large, randomized controlled trials documented no reduction in mortality in the screened population, and the ACS rescinded its recommendation in 1980.

The main lesson in this is that lead-time and length biases are not just theoretical; they confound our ability to evaluate screening programs. Furthermore, in the case of lung cancer, the false promise of early detection by screening detracted from the real solution, which is prevention of illness and death through cessation of tobacco use.

Recently, investigators have shown that with the use of spiral computed tomographic

---

**TABLE 4**

Annual Mortality Associated with Selected Types of Cancer in the United States

<table>
<thead>
<tr>
<th>Type of cancer</th>
<th>Number of deaths per year</th>
<th>Percentage of total cancer deaths per year</th>
<th>Numerical rank as cause of cancer deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>160,100</td>
<td>28.0</td>
<td>1 (overall)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>56,500</td>
<td>10.0</td>
<td>2 (overall)</td>
</tr>
<tr>
<td>Breast</td>
<td>43,900</td>
<td>7.7</td>
<td>2 (in women)</td>
</tr>
<tr>
<td>Prostate</td>
<td>39,200</td>
<td>6.9</td>
<td>2 (in men)</td>
</tr>
<tr>
<td>Cervical</td>
<td>4,900</td>
<td>0.9</td>
<td>9 (in women)</td>
</tr>
</tbody>
</table>

*—Percentage of 565,000 U.S. cancer deaths annually. (The four most common cancers account for >51 percent of total cancer deaths in the United States.)

(CT) scanning in asymptomatic smokers, it is possible to detect small, potentially resectable lung cancers with greater sensitivity and at an earlier stage than with conventional chest roentgenograms. As yet, no data suggest that CT screening decreases mortality, but it appears likely that the debate over screening for lung cancer may soon be reopened.

**COLORECTAL CANCER**

Three randomized controlled trials have documented a reduction in colorectal cancer mortality in populations screened with fecal occult blood testing. The first study\(^1\) used volunteers, and rehydration of slides resulted in a very high colonoscopy rate, with a 33 percent reduction in relative risk of death from colorectal cancer. Two subsequent studies\(^2\) were community-based and did not use rehydration, demonstrating more modest relative reductions in mortality associated with colorectal cancer of 15 and 18 percent. The absolute risk reduction from the latter two studies gives an NNS of 1,000\(^3\) and 588.\(^4\)

Even though the relative reduction in mortality associated with screening is only 15 to 18 percent (compared with 80 percent for Pap smears), the NNS compares favorably with that of cervical cancer screening because deaths from colorectal cancer are much more common than those from cervical cancer.

Controversy remains over the role of sigmoidoscopy in screening for colorectal cancer. Screening is supported by case-control studies\(^5\) but no randomized controlled trials. Two recent trials have demonstrated that in asymptomatic patients found to have advanced proximal neoplasms on colonoscopy, more than one half had no distal abnormalities and, thus, would have been missed on sigmoidoscopic screening.\(^6\) This finding would suggest that colonoscopy would have a significantly higher yield than sigmoidoscopy but, of course, at a higher overall cost.

Another controversy is the high cost of colorectal cancer screening (about $300,000 per death prevented),\(^7\), combined with the fact that none of the three studies was able to demonstrate a reduction in overall mortality but only in disease-specific mortality. This finding raises the possibility that, in return for such a large investment of society’s resources, colorectal cancer screening programs may not actually save lives but only shift patients to competing causes of mortality.

**BREAST CANCER**

Several randomized controlled trials have demonstrated a reduction in breast cancer mortality in cohorts screened by annual or biannual mammography, making mammography one of the best-documented screening procedures. However, considerable controversy remains about the age at which screening should be initiated.

A meta-analysis of eight randomized controlled trials of women 50 to 74 years of age shows a relative risk of breast cancer mortality in the screened group of 0.77.\(^8\) Using this data as well as the background incidence of breast cancer, one can estimate an NNS of 543, which represents the number of women who would have to be enrolled in an ongoing screening program over 10 years to expect to save one life from otherwise-fatal breast cancer. By contrast, the same meta-analysis showed that in women 40 to 49 years of age the relative risk in the screened group was not significant (0.92), giving an NNS of 3,125 in this younger cohort. Calculations based on gains in life expectancy\(^9\) and cost per year of life saved\(^10\) also show that screening in the younger age group is about five times more expensive than screening in the over-50 group and only one fifth as effective.

A 1997 National Institute of Health consensus conference concluded that current data...
did not warrant routine screening mammography in women 40 to 49 years of age and recommended that decisions be individualized on the basis of the patient’s risk factors and preferences. This recommendation resulted in a firestorm of protest, culminating in a U.S. Senate vote (98 to zero) that endorsed universal screening for women 40 to 49 years of age.24 In view of this highly politicized environment, it seems prudent to offer screening mammography to women in this age group, but only after a discussion of the limited benefits and the high cumulative rate of false-positive results, which can approach 50 percent after 10 annual mammograms.25

The status of screening mammography has been further questioned by the recent publication of a new meta-analysis of the same eight trials. These authors argue that six of the eight trials in the original meta-analysis show evidence of inadequate or flawed randomization and allocation. When these six trials are eliminated, the remaining two adequately randomized trials show no effect of screening on breast cancer mortality.26 Although most commentators have not accepted the authors’ conclusion that “screening for breast cancer with mammography is unjustified,”26 it does underscore the fact that the evidence supporting screening mammography remains controversial.

PROSTATE CANCER

Screening for prostate cancer with the prostate-specific antigen (PSA) test is perhaps the most controversial issue in cancer screening. This subject was recently reviewed in American Family Physician.27 The introduction of the PSA test in 1986 was initially followed by a dramatic increase in the incidence of prostate cancer and a fourfold increase in the rate of radical prostatectomy.28 A decline in incidence has been noted during the past few years, presumably because the initial backlog of undiagnosed asymptomatic tumors has been eliminated by widespread PSA screening.

By contrast, mortality rates from prostate cancer have been fairly constant, increasing by 1 percent per year until 1992 and declining by about 1 percent per year starting in 1993.29 There has also been a documented shift to earlier-stage disease and an increase in five-year survival rates.30 Because prostate cancer is often slow-growing with a long asymptomatic period, these results are almost certainly affected by lead-time and length bias, and some have maintained that the current situation is exactly analogous to the situation with screening programs for lung cancer.13,30

In the absence of randomized controlled trials, PSA screening remains an unproven intervention, and it is therefore impossible to calculate measures of screening effectiveness such as NNS. Randomized controlled trials are under way in the United States and Europe, but results will not be available for several years. In the meantime, the American College of Physicians and the American Academy of Family Physicians recommend that men older than 50 years be counseled about the “known risks and unknown benefits” of PSA screening, and that informed consent be obtained from those who wish to proceed with screening.

Final Comment

Because patients often expect to undergo screening tests for cancer, physicians must know how to evaluate the evidence in support of screening and how to convey that evidence to patients in an understandable way. In an era when advocacy groups and subspecialty organizations have taken the lead in
promoting new screening tests, it is especially important for family physicians to understand the various types of bias that can lead to an exaggeration of the efficacy of screening. Physicians must also attend to the unique ethical dimension of screening, which involves either requiring a high standard of evidence of efficacy or, in the absence of such evidence, engaging in a process of informed consent.

Table 5 summarizes the current state of the five most common screening tests for cancer.

### Table 5

<table>
<thead>
<tr>
<th>Test</th>
<th>Strength of recommendation</th>
<th>Quality of evidence</th>
<th>RRR†</th>
<th>NNS‡</th>
<th>Controversies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pap smear for cervical cancer</td>
<td>A</td>
<td>II-2, II-3</td>
<td>&gt;0.80</td>
<td>1,140</td>
<td>Interval, new technologies, when to stop</td>
</tr>
<tr>
<td>Mammography Age &gt;50 years</td>
<td>A</td>
<td>I, II-2</td>
<td>0.23</td>
<td>543</td>
<td>Interval (annual vs. biannual), when to stop</td>
</tr>
<tr>
<td>Age 40 to 49 years</td>
<td>C</td>
<td>I</td>
<td>0.08</td>
<td>3,125</td>
<td>Some evidence of significant reduction with follow-up &gt;10 years; false-positives</td>
</tr>
<tr>
<td>FOBT for colorectal cancer</td>
<td>B</td>
<td>I</td>
<td>0.15-0.20</td>
<td>588-1,000</td>
<td>Interval (annual vs. biannual), compliance, role of sigmoidoscopy, cost/benefit</td>
</tr>
<tr>
<td>PSA for prostate cancer</td>
<td>D</td>
<td>II-2</td>
<td>NA</td>
<td>NA</td>
<td>Unproven; RCTs in progress</td>
</tr>
<tr>
<td>Chest film for lung cancer</td>
<td>D</td>
<td>I, II-1</td>
<td>NA</td>
<td>NA</td>
<td>Recent reports of spiral CT for screening may reopen controversy</td>
</tr>
</tbody>
</table>

Pap = Pap smear; FOBT = fecal occult blood test; PSA = prostate-specific antigen; RCTs = randomized clinical trials; NA = not applicable; CT = computed tomography.

*—Per U.S. Preventive Services Task Force rating (see Table 3).
†—RRR = relative risk reduction: the proportion of deaths from the cancer in question in the control group that could have been prevented by the screening intervention.
‡—NNS = number of patients needed to screen over a 10-year period to prevent one death from the cancer in question. Calculated as the reciprocal of the absolute risk reduction.

Established screening tests for cancer have an NNS ranging from approximately 500 to 1,100. Although these numbers may initially seem discouraging, they indicate that by diligent attention to accepted screening modalities, a family physician can be expected to prevent several cancer deaths over the course of a career. These numbers can also provide a benchmark by which to evaluate proposed new screening programs, as well as providing comparisons to screening tests for diseases other than cancer.9
Cancer Screening

REFERENCES


