Screening for Disease: Making Evidence-Based Choices

Margaret M. Fields, MSN, ACNP, AOCNP, and Eric Chevlen, MD

Screening for illness should be an evidence-based activity. Screening tests are useful only if they reduce mortality or morbidity. Therefore, healthcare professionals must know how to evaluate research about screening tests to be sure that, in fact, the tests actually accomplish their goals. Tests that generate many false-positive results may cause harm from anxiety and unnecessary procedures. Tests that generate many false-negative results may worsen outcomes by leading to delayed diagnosis and treatment. Characteristics that make a disease amenable to screening include a significant negative impact on health, an identifiable asymptomatic period, and improved outcomes with early intervention. A useful screening test must have sensitivity and specificity for the disease being screened. It also must be cost effective and acceptable to patients. Sensitivity, specificity, and disease prevalence all interact to determine a test’s positive predictive value—the likelihood that a positive test result indicates that the disease is present. Several types of test bias can undermine the validity of a screening trial. Screening bias occurs when the sample of patients used in a trial to evaluate a screening test is not representative of the patient population to be screened. Another bias results from the fact that indolent disease is more likely to be detected in a screening program than aggressive disease. The apparent improved outcome that results is called length bias. Finally, lead-time bias occurs when survival of a screened population is measured from the date of screening, whereas survival of an unscreened population is measured from detection of symptomatic disease. In screening for illnesses, the goal must not be merely to do something. It must be to do something useful.

Caveat emptor: Let the buyer beware. Healthcare dollars are limited. Evidence-based practice mandates that decisions be based on science, not hunches. Healthcare professionals cannot assume that all screening tests are useful; all tests must be critically evaluated for use in general and high-risk populations. Healthcare professionals should critically evaluate current and newly developed screening tests for general and high-risk populations. As new developments are made, more and more nurses are being asked to provide and interpret statistical information to patients, especially in the statistic-filled areas of genetics and cancer screening. Therefore, nurses must develop an understanding of the terminology and statistical relationships among variables to be able to communicate accurately and effectively (Hanoch & Pachur, 2004).

Factors Affecting Screening Efficacy

The risk versus benefit ratio must be considered in every medical intervention provided or endorsed. Few people undergoing a screening test actually have the disease; all, however, experience the cost and discomfort of the test. Two additional hazards of screening must be recognized. False-positive results may cause unnecessary anxiety, expense, and even a risk of hazardous intervention in unaffected individuals. False-negative results may speciously reassure and delay diagnosis of people who, in fact, have a disease (Gates, 2001). Breast self-examination may be an

At a Glance

- Characteristics that make a disease amenable to screening include a significant negative impact on health, an identifiable asymptomatic period, and improved outcomes with early intervention.
- The positive predictive value of a screening test is determined by the test’s sensitivity and specificity and the disease’s prevalence in a population.
- A useful screening test must reduce mortality or morbidity and have high sensitivity and specificity. It also must be cost effective and acceptable to patients.
example of the problem of false-positive results causing unnecessary cost and anxiety. For many years, public health advocates have argued that breast self-examination should be promoted because it must be beneficial and surely does no harm. A recent randomized trial showed no survival benefit from breast self-examination. Furthermore, screening generated needless costs and invasive procedures; 50% more biopsies were performed in the self-examination group compared with the control group (Thomas et al., 2002). However, no conclusions can be drawn from one study.

Disease Characteristics to Consider

Healthcare professionals must consider a disease itself when deciding whether to screen for its presence. Diseases that best lend themselves to screening have significant negative effects on public health in terms of morbidity and mortality, and have an identifiable asymptomatic period, and demonstrate improved outcomes with early intervention (Gates, 2001). Much of the problem in screening for cancer lies in the fact that cancer is not one disease, and even a single type of malignancy may express itself differently in different individuals.

Cancer may develop in three basic patterns. The first is the early and rapid dissemination of disease to metastatic sites. Screening for such cancer has little value unless it is a disease that is unusually responsive to systemic therapy. That is a rarity, so screening for this pattern of cancer has little impact on survival. The potential for cure is lost by the time the cancer can be detected. An example is small cell lung cancer, which often is diagnosed because of symptoms of metastatic disease. Many studies have attempted to find an acceptable, cost-effective, and efficacious method of screening for lung cancer. Modalities that have been tested include sputum cytology, chest x-ray, and, most recently, spiral computed tomography (CT) scanning (Humphrey, Teutsch, & Johnson, 2004). Spiral CT can detect lung cancer earlier than other modalities. However, it has not yet shown any benefit on survival. Additionally, spiral CT as a screening tool is prohibitively expensive. The U.S. Preventive Services Task Force (2004) does not recommend lung cancer screening, even for high-risk populations, until more evidence is available of its benefit.

The second pattern of cancer development is slow, local progression without early metastasis. Such cancer may be identified and treated but often to scant advantage. Limited potential exists for metastatic spread. Very often, the treatment carries greater morbidity than the natural progression of the disease. Well-differentiated prostate cancer in a man with limited life expectancy because of age or comorbidities may not require intervention and treatment and, therefore, should not be a target of screening.

The final pattern of cancer development is an early asymptomatic period during which cancer is detectable and curable, followed by incurable metastasis. Clearly, early diagnosis and intervention may lead to better outcomes. An example is cervical cancer (Meyers, 1995). The decrease in mortality rates in the past 40 years, since the introduction and widespread use of the Pap test, is clear evidence of an effective screening modality. Cervical cancer is the second most common cancer diagnosed among women worldwide, with about 80% of cases occurring in third-world countries where screening is unavailable (American Cancer Society, 2004). From 1995–2001, the National Breast and Cervical Cancer Early Detection Program screened more than 750,000 women and detected more than 6,300 high-grade lesions. The study uncovered 465 cases of invasive cancer, with 51% of them being stage I disease. The results confirm the ability of the Pap test to identify preinvasive disease, thereby preventing cancer and reducing mortality (Benard et al., 2004).

Even with a disease whose pattern of progression lends itself to screening, the prevalence of the disease must be sufficient in a definable population to justify screening the population. Both criteria are important. If a disease is not prevalent in a population, or if the window of opportunity between when the disease becomes detectable and when it becomes incurable is too small, then the benefits of screening are limited.

Sensitivity and Specificity of Screening Tests

An ideal screening test must be sensitive enough to detect a disease during the asymptomatic period. That is, the true-positive rate must be high (or, stated differently, the false-negative rate must be low). If a test has high sensitivity and the result is negative, the likelihood of disease is low. For example, in a test with 90% sensitivity, the percentage of diseased people who test positive is 90%. The best use of the most sensitive tests is to rule out suspected disease. A specific test, in contrast, is one with a high rate of true negatives (or, stated differently, the false-positive rate must be low). In a test with 80% specificity, the percentage of disease-free people who test negative is 80% (Wallach, 2000).

The sensitivity and specificity of a test are independent of the prevalence of a disease in a population. On the other hand, the positive predictive value (the probability of a person with a positive result really having the disease) of a test is a function of not only the test’s sensitivity and specificity but also the prevalence of the disease in the population.

For example, consider a hypothetically very prevalent disease, present in 90% of a population. For every 1,000 people, 900 will be affected and 100 unaffected. Using a test with 90% sensitivity and 80% specificity, of the 900 people with the disease, 810 will have positive results (true positives) and 90 will have negative results (false negatives); of the 100 unaffected people, 20 will have positive results (false positives) and 80 negative results (true negatives) (see Table 1). The positive predictive value reveals how reliable a positive result is in the population. It is mathematically defined as the number of true positives divided by the sum of true positives plus false positives.

Positive Predictive Value

In the example of the hypothetically very prevalent disease, the positive predictive value is the number of true positives (810) divided by the true positives added to the false positives (810 plus 20). In the first case, the positive predictive value is 98%. Now consider the same test with a sensitivity of 90% and a specificity of 80%, but this time applied to a population of low disease prevalence. In the second example, 10% of the
population has the condition being screened. Of 1,000 people, 100 will be affected and 900 will be unaffected. Of the 100 people affected, 90 will have positive results (true positives) and 10 will have negative results (false negatives). Of the 900 people unaffected, 720 will have negative results (true negatives) and 180 will have positive results (false positives) (see Table 1). The positive predictive value of the same test in a population with a high prevalence of the condition being tested (Wallach, 2000).

A measurement of the effectiveness of cancer screening tests is the number needed to screen (NNS), which is the number of individuals needed to be screened to prevent one premature death. In currently recommended cancer screening tests, the typical NNS is 500–1,100. For mammography in women older than 50, the NNS is 543; for Pap tests, it is 1,140; and for fecal occult blood testing for colorectal cancer, it is 600–1,000 (Gates, 2001). Even the NNS benchmark may be no more than a theoretical construct in some situations. A screening program with a low NNS may not be worthwhile if appropriate medical care is not available and accessible to people identified as diseased by the test. Moreover, identified patients must be willing to undergo follow-up and treatment (Gates).

Finally, acceptability of a test to patients and acceptability of the test’s cost to patients and third-party payers are essential components of a good screening test. Even if a screening test is costly, it may prove to be cost effective in the long term. Identifying and treating preinvasive disease usually is far less expensive than diagnosing and treating invasive malignancy, especially metastatic disease.

All nurses should be aware of currently recommended screening tests. Additionally, oncology nurses should promote the continued screening of long-term survivors for second malignancies. Mahon, Williams, and Spies (2000) found that flexible sigmoidoscopy and colonoscopy screenings were performed in only 20%–28% of long-term survivors of breast, ovarian, and endometrial cancer. Yet such patients actually are at higher risk for second cancers than people who never have had cancer. The data should be a wake-up call for nurses to implement new clinical practice protocols to improve cancer screening among cancer survivors.

### Feasibility and Bias

Essential to the determination of efficacy of any screening test is the consideration of feasibility and bias. The types of bias to be considered are screening bias, lead-time bias, and length bias.

#### Feasibility

The feasibility of a screening test is determined by the timing of the critical point of the disease in relation to the asymptomatic period. The critical point is the time when early treatment no longer is more beneficial than delayed treatment. In cancer screening, the critical point usually is the time when occult metastases occur. Prior to the critical point, treatment for the disease is relatively effective; after the critical point, treatment is relatively ineffective. If the critical point occurs very early in the asymptomatic phase of the disease, screening is not helpful because no survival advantage is attained. An example is aggressive lung cancer, which often is incurable when identified. If the critical point occurs late in the asymptomatic phase of the disease or even after the development of symptoms, screening may be unnecessary because the benefit of delayed treatment is as great as that of early treatment. Screening is beneficial only for diseases for which the critical point occurs during the asymptomatic period.

### Table 1. Relationship of Prevalence to Positive Predictive Value of a Screening Test With a Given Sensitivity and Specificity

<table>
<thead>
<tr>
<th>TEST RESULT</th>
<th>DISEASE PRESENT</th>
<th>DISEASE ABSENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>810 (true positives)</td>
<td>20 (false positives)</td>
</tr>
<tr>
<td>Negative</td>
<td>90 (false negatives)</td>
<td>80 (true negatives)</td>
</tr>
<tr>
<td>Total</td>
<td>900</td>
<td>100</td>
</tr>
</tbody>
</table>

Note. In this example, the positive predictive value (probability that those with positive test results actually have the disease) is 98% (810 divided by [810 plus 20]), indicating near certainty that a positive test result indicates actual presence of disease. In contrast, the negative predictive value (percentage of those with negative results who really do not have the disease) is 47% (80 divided by [90 plus 80]). Thus, a negative test indicates that a patient still is more likely than not (53%) to have the disease.


### Table 2. Relationship of Prevalence to Positive Predictive Value of a Screening Test With a Given Sensitivity and Specificity

<table>
<thead>
<tr>
<th>TEST RESULT</th>
<th>DISEASE PRESENT</th>
<th>DISEASE ABSENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>90 (true positives)</td>
<td>180 (false positives)</td>
</tr>
<tr>
<td>Negative</td>
<td>10 (false negatives)</td>
<td>720 (true negatives)</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>900</td>
</tr>
</tbody>
</table>

Note. In this example, the positive predictive value (probability that those with positive test results actually have the disease) is only 33% (90 divided by [90 plus 180]). In contrast, the negative predictive value (percentage of those with negative test results who really do not have the disease) is 99% (720 divided by [10 plus 720]). Thus, a negative test indicates 99% probability of no disease.

Screening Bias

Bias is the distortion of findings and may occur in several ways. The appearance of screening effectiveness may be generated to some extent by self-selection bias. If people who have good prognoses are more likely to have a screening test than people with poor prognoses, the screening test will seem to improve outcomes, even if the test and associated early treatment are no better than delayed treatment. Individuals who are health conscious often are most likely to be screened and to comply with follow-up. Additionally, they tend to have fewer comorbidities. Therefore, apparent improved outcomes may not necessarily reflect the efficacy of screening and early treatment, but rather a healthier subset of the population.

Lead-Time Bias

Lead-time bias occurs when the asymptomatic phase is not considered as part of the whole course of the disease. Consider, as an example, a disease that is diagnosed upon symptomatic clinical presentation and has an average survival of two years after such diagnosis. Assume that the disease has an asymptomatic period of two years, during which diagnosis is possible with a screening test. If that same disease were discovered by a screening test performed in the middle of the asymptomatic period, patients would seem to benefit from screening. Even if no benefit of early treatment exists, the survival of the screened group would appear to be a year longer than the survival of patients diagnosed when symptomatic. In fact, however, the screening test merely identifies patients earlier in the unaltered course of illness. When viewed from the same starting point (presentation of clinical symptoms), the average survival time is unchanged. The screened patients do not live longer; rather, their survival time was measured from an earlier starting point in the unaltered course of the disease.

Length Bias

Length bias occurs because of the heterogeneity of the natural history of diseases. An indolent case, for example, a disease that has a longer asymptomatic period than does an aggressive case; therefore, it is susceptible to detection by screening for a longer time period. Less aggressive tumors are slower to metastasize and have more favorable prognoses. In contrast, highly aggressive malignancies have shortened asymptomatic periods and rapid progression. In screening programs, less aggressive malignancies with longer asymptomatic periods are more likely to be detected. Aggressive malignancies are more likely than indolent malignancies to progress through the asymptomatic period to clinical symptoms during the interval between screening tests. Thus, malignancies diagnosed upon clinical presentation often are more aggressive with worse prognoses, and the malignancies identified during screening are more likely to be less aggressive tumors with better prognoses (Gates, 2001).

The extreme point of length bias is overdiagnosis. That is, an indolent cancer may be identified and cured when, in fact, no cure was needed. If a patient’s life expectancy is shorter than the time needed for the disease to progress to the point of needing treatment, then treating the target disease is unnecessary. Such is the case in many of the prostate cancers diagnosed in recent years via screening; often the disease never would have posed a problem in the natural history of its course (Kramer, 2003, p. S72).

Advocating screening of unproven value is scientifically unsound and potentially harmful. Similarly, for healthcare professionals to merely encourage patients to decide for themselves about screening tests is abjuring their duty. Just as new drugs must undergo rigorous testing prior to approval, so healthcare professionals must exercise the same scientific rigor when deciding whether to advocate screening tests (Law, 2004).

Author Contact: Margaret M. Fields, MSN, ACNP, AOCNP, can be reached at pegfields@houston.rr.com, with copy to editor at CJONEditor@ons.org.

References