Principles of Cancer Screening

Robert Clark, MD
Radiology Service, H. Lee Moffitt Cancer Center & Research Institute

Early detection of cancer, prior to its clinical manifestations, appears to be a worthwhile and desirable goal. Yet, the concept and acceptance of cancer screening remain controversial and often confusing. While the benefits of screening are obvious to those whose screening tests have resulted in successful interventions, attention also needs to be given to the risks, economic costs, and psychological effects of screening procedures. Specific governing principles that define the cancers to be screened, the appropriate screening test, and the measurement of outcomes should be established in order for a screening program to be deemed worthwhile. A beneficial screening strategy detects cancer prior to its systemic spread, alters the natural history of the disease, and defers the time of death.

Introduction

Cancer screening seems intuitively beneficial. The concept of detecting cancer early, when tumor is manageable and has not spread from its primary site, rather than later, when it has metastasized to other vital organs, seems reasonable. Yet, despite being an intuitive or reasonable concept, cancer screening remains controversial and often confusing. Why not screen everyone in order to detect cancer early? Should screening be recommended for the most common or most lethal cancers? Conversely, how can cancer screening be considered useful when most screened individuals never get the disease, yet incur the costs of screening and the anxiety of a possible false-positive test? Why is cancer screening so controversial, so debated, and so rarely recognized as valuable?

Although the answers to these questions often are different for specific cancers, many general principles of cancer screening are common to the various cancers and screening strategies. These principles should be considered ideals; few, if any, cancer screening strategies will fulfill all degrees of proof. However, these principles should be considered when reading the scientific literature and evaluating the rationale for new screening tests or proposed strategies.

Definitions of Terms Relating to Cancer Screening

Clinicians may define terms used in cancer screening in different ways. The following summarizes current usage.

**Cancer Screening** Screening is the application of a test to detect a potential cancer where no signs or symptoms of the cancer are present.[1,2] Testing for cancer screening involves both the traditional tests and the newer tests for risk factors. The traditional test detects cancer before it is clinically apparent, early in its natural history, before it has become systemic, when treatment may be more effective, less expensive, or both. An abnormal screening test in this situation leads to further diagnostic evaluation to determine if cancer is present and leads to subsequent treatment if cancer is detected. Another type of cancer screening, which may become more prevalent in the future, involves screening for risk factors or other markers (eg, genetic or molecular) that designate a high risk for developing cancer. It is not yet clear what an abnormal screening test in this situation means or what recommendations should follow.[3,4] Cancer screening is a secondary form of cancer prevention, as distinguished from primary prevention, a concept which might be avoided cigarette smoking to prevent lung cancer.

The term “cancer screening” is synonymous with other terms, such as detection or early detection,[1] and has been modified by some authors as “mass,” “routine,” “regular,” or “selective” cancer screening. However, these modifiers have no universally accepted definitions. Every cancer screening strategy must identify its target population, the proposed screening test, and the frequency of the screening test. Knowledge of these parameters for any screening strategy makes other modifier terms moot.

**Cancer Screening Test** The cancer screening test is the method used to detect a specific target cancer[1,2] and may consist of a single modality or a combination of tests. Laboratory tests of blood or body fluids, physical examinations, invasive procedures, and imaging tests are examples of screening tests.

**Asymptomatic** The goal of cancer screening is to detect cancer before it is clinically apparent. Therefore, “asymptomatic” is defined in the perspective of the individual who has no known signs or symptoms of cancer prior to the screening test. For example, as a result of a digital rectal examination, a physician detects a prostate nodule on the gland of a man with no previously known signs or symptoms. The patient then has a sign of cancer detected by the physician and the screening test; nevertheless, the man was asymptomatic prior to the screening and the cancer was detected by screening. This example of asymptomatic status is appropriate even if the man had symptoms related to another condition, such as benign prostate hypertrophy. This concept is less clear in the skin cancer screening test involving the inspection of a skin lesion visible to both the individual and the examiner.

**Screened Individual** Screened individuals are often inappropriately referred to as patients. Screening involves the testing of an asymptomatic person. A screened individual does not become a “patient” until the screening test is abnormal. New concerns, anxieties, costs, and discomforts begin when the individual becomes a “patient.”

**Target Population** Certain characteristics identify an individual as a candidate for cancer screening. For example, since prostate cancer is rare in teenage boys, screening is inappropriate in this age group. The target population of a proposed screening strategy defines the characteristics of an individual who would be appropriate to receive the screening test. Typical defining characteristics of a target population include sex, family history, specific known risk factors, geographic region of birth or residence, race or ethnicity, and age.

**Screening Practitioner** The screening practitioner is the health care professional who performs the cancer screening test,[1] including primary care physicians, specialist physicians, nurses, physician assistants, and technicians.

**Diagnosis** Screening is not diagnosis.[1,2,5] The cancer screening test identifies asymptomatic individuals with a high likelihood of having cancer. Screened individuals are divided between two groups: those with normal test results and those with abnormal results. In some individuals with normal results from a screening test (a falsenegative screening test), cancer may be subsequently detected with diagnostic tests such as biopsy. All individuals with abnormal screening test results require some diagnostic evaluation. Some of those with abnormal results and further diagnostic evaluation will not have cancer (a falsepositive screening test). Diagnosis is the clinical problemsolving process applied to symptomatic individuals or asymptomatic individuals with abnormal screening tests.
Symptomatic individuals require diagnostic evaluation to determine the cause of symptoms. A screening test applied to a symptomatic person is not considered a cancer screening event, since diagnostic evaluation is required regardless of the results of the screening test. Moreover, the value of a screening strategy cannot be assessed if symptomatic individuals are included in the target population, since these people may already have advanced disease that needs diagnostic evaluation.

**Screening Strategy or Protocol** A cancer screening strategy or protocol defines the operational parameters of a cancer screening program: who, how, what, where, and when. The screening strategy or protocol defines the population to be screened (the target population) and the screening test to be used, as well as when and how often the screening test should be applied. The screening strategy also may define who should perform the screening test, the conditions under which it should be applied, and the criteria for an abnormal test. A screening strategy or protocol is useful in designing clinical trials and interpreting scientific data about screening. Another function of screening strategies or protocols is to make recommendations to individuals or groups about cancer screening.[1]

The screening protocol design must be clearly understood when interpreting scientific evidence about cancer screening. The protocol of a screening clinical trial often is limited. Therefore, the results obtained by the trial are valid only for the conditions of that protocol. For example, a cancer screening protocol that studies a target population of white European women ages 50 to 69 years, applies a screening test every two years for 10 years, and finds 40% fewer cancer deaths in screened women than in unscreened women suggests strong evidence of the effectiveness of cancer screening. However, it may not be applicable to women older than 69 years of age, to women younger than 50 years, or to Japanese women, and the results of this protocol may not be enough evidence alone to justify recommending cancer screening annually for all black women older than 40 years of age. Alternatively, a cancer screening test applied to men between ages 65 and 75 years showing no benefit to the screened group when compared with unscreened men does not signify a lack of potential benefit to screening men younger than age 65 years. The results simply show that no information about screening younger men is available from the screening test.

The scientific literature focusing on cancer screening is replete with clinical trials that have different screening strategies and protocols for the same target cancer. It is difficult to compare or combine the data from these trials to answer scientific questions that were not posed prior to the design of the clinical trial. However, cancer screening strategies for clinical practice are not limited to those studied by clinical trials; they may be recommended by individual practitioners, professional medical societies, public health agencies, and health maintenance organizations. Their recommendations are based on their best assessment of the available scientific evidence, as well as their best estimate of applicability to individuals or target populations not included in the original scientific protocols, which varies among groups, practitioners, and policymakers. Therefore, it is not surprising that "guidelines" for cancer screening may vary among various organizations and among practitioners.

### Table 1. Outcome Measures in Cancer Screening Programs

**Shortterm Measures** Information on the target population who received screening. The number and proportion of target population who received screening. The number and proportion of individuals in the target population who received screening. The number and proportion of target population who were examined by multiple screens. The number or prevalence of preclinical cancers detected. The proportion of abnormal screened individuals brought to definitive diagnosis or followup. The monetary cost per cancer detected. Sensitivity and specificity of the screening test. The positive and negative predictive values of the screening test.

**Longterm Measures** Information on the stage distribution of detected cancers. The case fatality rate of screened individuals. The site-specific cancer mortality rate of screened target population. The total monetary costs.

---

**Outcomes** The scientific evidence of the value of screening requires that outcomes of a screening protocol be measured. Outcomes are the health and economic results that are related to screening.[1,6,7] Outcomes include the benefits, harms, and costs of screening as well as its incurred diagnostic evaluations. Outcomes are measured by tracking the detailed clinical results of screened individuals (Table 1).

**Effectiveness** The effectiveness of cancer screening is determined by comparing outcomes to ascertain if the benefits outweigh the harms and whether the health outcomes (benefits and harms) justify the costs.[1,6,7] Moreover, the outcomes and effectiveness measures of the screened population must be compared to those of a similar unscreened group. For example, for a screening program to be judged effective, the stage distribution of detected cancers in screened individuals should be lower than cancers detected in unscreened people. Similarly, the casefatality rate and, more important, the sitespecific mortality rate for a screened group should be significantly less than that of an unscreened group.

**Costeffectiveness** Ideally, the cost of the screening program (the total of screening costs, diagnostic evaluations, treatment costs of detected cancers, and value of years of life lost to cancer deaths) should be less than the cost for the unscreened group (the total of diagnostic evaluations, treatment costs of detected cancers, and value of years of life lost to cancer deaths). Relevant costs to be considered in this evaluation are listed in Table 2. However, cost savings that result from screening programs have been difficult to determine.[8] Hence, other cost-associated measures are analyzed, such as cost determination, cost minimization, costeffectiveness, costbenefit, and costutility analyses.[1,9,13]

### Table 2. -- Relevant Costs of a Cancer Screening Program

**Costs of screening tests:**

<table>
<thead>
<tr>
<th>Direct costs or charges</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total monetary costs</td>
<td></td>
</tr>
</tbody>
</table>
indirect costs (time, anxiety)
 Costs incurred by abnormal screening tests:
 direct costs or charges of diagnostic evaluation or biopsy
 indirect costs of complications, morbidity, anxiety, time, loss of work
 Costs related to false-positive screening tests:
 direct costs or charges of diagnostic evaluation or biopsy
 indirect costs of complications, morbidity, anxiety, time, loss of work
 Costs related to false-negative screening tests:
 false sense of security delay in diagnosis due to disregard of clinical symptoms
 direct costs or charges of treatment and rehabilitation
 indirect costs of complications, morbidity, anxiety, time, loss of work
 Costs related to death:
 direct costs or charges related to death
 indirect costs of years of life lost

Prevalence and Incidence
The prevalence rate of cancer denotes the number of cancers that exist in a defined population at a specific time, whereas the incidence rate denotes the number of new cancers that develop in a defined population during a specific period of time.[1,2] Both are commonly expressed as the number of cancers per 100,000 individuals in the defined population. The ideal screening test would detect all the prevalent cases of cancer in the first screen of a previously unscreened population. Subsequent screening examinations would detect incidence cases developing in the population since the prior screen.

The incidence rate for a given cancer is lower than the prevalence rate. Theoretically, in a defined population of individuals who received three cancer screenings at yearly intervals, the first screen would detect all prevalence cases (developing for several years prior to the first screening). The second and third screenings would detect incidence cases, ie, only those cases that developed since the first screening.

Measures of Validity of a Screening Test

Several issues govern the validity of a screening test (Table 3).

<table>
<thead>
<tr>
<th>Results of Screening Test</th>
<th>True Characteristics in the Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Have the disease</td>
</tr>
<tr>
<td>Positive test</td>
<td>80</td>
</tr>
<tr>
<td>Negative Test</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

True-positive tests = 80 Sensitivity = 80/(80 + 20) = 0.80
True-negative tests = 800 Specificity = 800/(80 + 100) = 0.89
False-positive tests = 100 Positive predictive value = 80/(80 + 100) = 0.44
False/negative tests = 20 Negative predictive value = 800/(800 + 20) = 0.98

Positive and Negative Tests. A truepositive screening test is an abnormal test for cancer in an individual who subsequently is found to have cancer within a defined period of time after the test, whereas a truenegative screening test is a normal test for cancer in an individual who subsequently is found not to have cancer within a defined period of time after the test.[14,15] Conversely, a falsepositive screening test is an abnormal test for cancer in an individual who subsequently is found not to have cancer within a defined period of time after the test, and a falsenegative screening test is a normal test for cancer in an individual who subsequently is found to have cancer within a defined period of time after the test.[14,15]

Sensitivity
The sensitivity of a screening test represents its ability to detect those individuals with cancer in the defined population[2,14,15] and is derived from the truepositive ratio, ie, the proportion of positive tests in all individuals with disease. Sensitivity is defined as the number of truepositive (TP) cases divided by the total number of truepositive and falsenegative (FN) cases.

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

Specificity
The specificity of a test represents its ability to identify those free of cancer in the population[2,14,15] and is derived from the truenegative ratio, ie, the proportion of negative tests in all individuals without disease. Specificity is defined as the number of truenegative (TN) cases divided by the total number of true negative and falsepositive (FP) cases.

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

Table 4. Hypothetical Examples of Relationship of Predictive Value of the Screening Test to the Prevalence of Cancer in the Population

Sensitivity of Screening Test = 0.99
Specificity of Screening Test = 0.95
Prevalence of Cancer = 1%

Prevalence of Cancer = 5%
_______________________________________________________________________

Positive Predictive Value

The positive predictive value is the measure of the validity of a positive test, ie, the proportion of positive tests that are true positive cases.

\[
\text{Positive Predictive Value} = \frac{TP}{TP + FN}\]

The predictive value of a test is dependent on the disease prevalence (Table 4). As the prevalence of cancer increases in the population, the positive predictive value of the screening test increases, even though its sensitivity and specificity remain unchanged.[1618] Therefore, for maximum efficiency and cost-effectiveness, screening should be focused on the populations with highest risk (highest prevalence).

Negative Predictive Value

The negative predictive value is the measure of the validity of a negative test, ie, the proportion of negative tests that are true negative cases.

\[
\text{Negative Predictive Value} = \frac{TN}{TN + FN}\]

Governing Principles of Cancer Screening

The principles that should be followed for a cancer screening program to be worthwhile. These principles define characteristics of the disease considered for screening, the screening test, and the outcomes. (Table 5)[1,2,19]

<table>
<thead>
<tr>
<th>Characteristics of the Disease</th>
<th>Characteristics of the Screening Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>High morbidity, mortality, costs</td>
<td>Able to detect disease in preclinical phase</td>
</tr>
<tr>
<td>High prevalence and incidence</td>
<td>Effective (ie, sensitive and specific)</td>
</tr>
<tr>
<td>Known natural history and biology</td>
<td>Safe</td>
</tr>
<tr>
<td>Preclinical phase with high prevalence</td>
<td>Simple, inexpensive</td>
</tr>
<tr>
<td>Effective treatment of early stage disease</td>
<td>Acceptable to individuals</td>
</tr>
</tbody>
</table>

The disease considered for screening should have high prevalence and incidence rates and should have serious clinical consequences measured in mortality, morbidity, and costs. The biology and natural history of the disease should be known. Ideally, the cancer should exist for a long time in a preclinical phase amenable to screening, and this preclinical phase should have a high prevalence rate in the screened population. The disease should have an effective treatment at an early stage, and this treatment should be more effective than treatment at late stage. When a disease has no effective treatment or when treatment in its early stage is no more effective than in its advanced stage, screening is problematic unless counseling is shown to be useful.

An effective screening test should have the ability to detect cancer in its preclinical phase with acceptable sensitivity, specificity, and predictive values. The test should be...
The most important outcome measure of the effectiveness of a screening strategy is the demonstration that the mortality rate from the disease is significantly lower in the total screened population when compared with the cancer mortality rate in an equivalent population of unscreened people, preferably demonstrated by a randomized, controlled, defined population clinical trial.

**Expected Benefits and Potential Harms**

The expected benefits of screening are a lower mortality rate from the target cancer, a reduction in morbidity from the disease, and lower health care costs. Additional benefits may include improved length and quality of life, as well as less pain, anxiety, and disability. Expected benefits of screening are derived from the true positive results of a screening test. While not a benefit that makes a screening program effective, a true negative screening test result provides reassurance that cancer has not developed.

The potential harms of screening are related to the test itself or to its results. Those related to the test are costs, inconvenience, anxiety, and discomfort. Additional potential risks (complications) may be related to invasive screening tests. The potential harms related to the results are associated with false positive and false negative tests. The potential benefits of screening must outweigh the potential risks, since any harm to an asymptomatic person is not to be considered lightly. A falsepositive test result causes anxiety and incurs a diagnostic evaluation, with its attendant costs, potential risks, and side effects. A falsenegative test result can lead to a false sense of security. Subsequent clinical signs or symptoms of cancer may be dismissed because of a prior negative screening test, resulting in further delay in detection.

**Table 6. - Cancer Control Phases: A Systematic Evaluation Process for Proposed Screening Strategies**

<table>
<thead>
<tr>
<th>Basic research and epidemiology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis development</td>
</tr>
<tr>
<td>Methods development</td>
</tr>
<tr>
<td>Controlled intervention trials</td>
</tr>
<tr>
<td>Defined population studies</td>
</tr>
<tr>
<td>Demonstration and implementation projects</td>
</tr>
<tr>
<td>Nationwide dissemination programs</td>
</tr>
</tbody>
</table>

**Evaluation of a Proposed Screening Strategy**

A systematic approach to cancer control research has been developed that provides a framework for the evaluation of a proposed screening strategy (Table 6).[20,21] The evaluation begins with knowledge about the basic biology and epidemiology of the cancer and incorporates information about characteristics of the populations at high risk, cancer prevalence and incidence rates, tumor growth rates, mortality rates, and costs of care and disability. The next step, hypothesis development, synthesizes the available scientific information and proposes possible interventions to be applied to the cancer problem. Cancer screening may not always be the appropriate intervention; primary prevention, if possible, is the optimum prevention. The proposed intervention strategy should be expressed as a testable hypothesis that can be evaluated in an objective, scientific fashion.

Next, methodological research is necessary to characterize the variables to be controlled or monitored in subsequent clinical trials. This phase might include pilot studies to identify target populations or compliance rates of screened individuals, to evaluate application or acceptability of screening tests, or to estimate the efficacy of the screening test. Methods that have been tested adequately and proven may be incorporated into clinical intervention trials. Initial trials may be uncontrolled but, ideally, these interventions should be controlled. Cohort studies or casecontrol trials may be used to estimate benefits from a screening intervention; however, randomized, controlled trials are likely to provide the most convincing results.

Measures of the quantitative impact of a screening intervention using a defined population study would identify not only barriers to widespread adoption of the intervention, but also methods for overcoming these barriers. The defined population must be comprised of a large number of people in order to show a significant intervention benefit. The screening strategy is beneficial if the defined population study demonstrates a significant reduction in diseasespecific mortality rate when compared with the unscreened group.

When screening is shown to be beneficial in defined population studies, demonstration and implementation programs are appropriate. The purpose of these programs is to apply the proven intervention in a community at large with measurement of the public health impact. A surveillance system should be in place to ensure that the application, accuracy, and effectiveness of screening in the community are equal to that demonstrated in clinical trials. Quality control processes may be developed during this phase, as well as assessment of the adequacy of diagnostic evaluation and treatment in the community.

Finally, when demonstration and implementation programs ensure that community dissemination can be achieved, nationwide screening programs and policy recommendations may be developed.

**Potential Harms and Biases in Evaluation**

All the steps in the evaluation process described above are not always adhered to. Pressures to circumvent them include immediate clinical acceptance and dissemination of a new screening test, expense of defined population studies, and preliminary recommendations for screening from professional organizations. However, without the assurance of this process, an incompletely evaluated screening strategy may deliver more harms and costs than benefits. Moreover, without the demonstrated benefit of a screening intervention in a defined population trial, the potential benefit of a screening strategy may be overestimated and invalid.

Almost invariably, individuals with cancer identified by screening will have longer survival times than those diagnosed with usual clinical detection. However, these apparent increased survival times are not always equivalent to reduction in mortality from cancer. Three biases that contribute to this spurious survival increase and potentially mask the lack of screening benefit are lead time bias, length bias, and overdiagnosis bias. Randomized, controlled clinical trials can control for these biases and can identify and quantify more accurately the benefits of a screening strategy.

![Fig 1. - Natural history of a lethal neoplasm is depicted schematically. The time line extends from left to right, depicting onset of cancer, its systemic spread, the subsequent clinical diagnosis, and death. The diamonds represent examples of two possible screening test applications. If a screening test is applied at Screen #1, the time of diagnosis is advanced from the time of usual clinical diagnosis by Lead Time #1. Survival time is apparently increased by this lead time, even though the natural history of the disease is unchanged. If a screening test is applied at Screen #2, the lead time is increased by Lead Time #2. If Screen #2 is applied prior to the systemic spread of the cancer, the screening benefit is lost.](image)
A beneficial screening strategy detects cancer prior to its systemic spread, alters the natural history of the disease, and defers the time of death. This alteration of the natural history of the disease, with prolongation of life, cannot be recognized without a randomized, controlled trial that prevents lead time bias.

References


