

Virtual Mentor

Ethics Journal of the American Medical Association
January 2006, Volume 8, Number 1:34-37.

Clinical Pearl

What Makes a Screening Exam “Good”?

by Cheryl Herman, MD

Screening tests are used to determine whether an asymptomatic individual has an undetected disease or condition. Screening is currently used in many contexts, including blood pressure monitoring for identifying hypertension, prostate-specific antigen measurement for signs of prostate cancer, colonoscopy for detection of colorectal carcinoma, and mammography for evidence of breast cancer. Unfortunately some screening tests lack credible scientific bases, and the risks and benefits of testing are frequently misrepresented to the patient. Many of the tests are marketed directly to the patient [1], so it is important for people to know what makes a screening exam “good.” How do we know that a screening study accurately determines the likelihood that a patient does or does not have the disease in question?

The 2 major objectives of a good screening program are: (1) detection of disease at a stage when treatment can be more effective than it would be after the patient develops signs and symptoms, and (2) identification of risk factors that increase the likelihood of developing the disease and use of this knowledge to prevent or lessen the disease by modifying the risk factors [2]. To fulfill these objectives, a screening test and the disease it screens for must meet the following criteria.

The disease in question should:

- constitute a significant public health problem, meaning that it is a common condition with significant morbidity and mortality.
- have a readily available treatment with a potential for cure that increases with early detection.

The test for the disease must:

- be capable of detecting a high proportion of disease in its preclinical state.
- be safe to administer.
- be reasonable in cost.
- lead to demonstrated improved health outcomes.
- be widely available, as must the interventions that follow a positive result [1].

These criteria bear a closer look.

The Screened-for Disease or Condition

The preclinical phase of a disease starts with the onset of the disease process and lasts until signs and symptoms appear, which is when the clinical phase begins. The

detectable preclinical phase is the interval during which the disease is detectable by screening, but the patient is still asymptomatic. During this period, there is a critical point at which intervention is more effective than if started after the clinical phase begins.

The disease being screened for must be serious enough to warrant testing asymptomatic people. The disease should be one that, if not found in its detectable preclinical phase before the critical point, will become life-threatening or cause significant morbidity. If the critical point occurs soon after the start of the detectable preclinical phase, screening may be too late to be helpful.

Pseudodisease is a condition detected by screening that does not require treatment because it will not adversely affect the patient's life. Type I pseudodisease refers to conditions that might not progress to symptomatic disease and may even regress. A commonly used example of Type 1 pseudodisease is ductal carcinoma in situ of the breast, which may remain in an intraductal state and not progress to invasive carcinoma and may even regress to atypical ductal hyperplasia (ADH). Type II pseudodisease is an indolent, slowly progressive disease found in conditions with a long detectable preclinical phase. Often, this type of pseudodisease cannot be diagnosed until after the patient has died from other causes, when autopsy results reveal histologic evidence of, for example, prostate, breast, or lung cancer that was previously unknown. If pseudodisease conditions such as these are treated, the patient may be considered "cured" because he or she died from a cause other than cancer. But designating such outcomes as "cures" is erroneous because the cancer—even if *untreated*—would not have killed the patient before the time that he or she actually died of other causes.

To justify their cost, screening tests must be able to detect a high number of cases of preclinical disease in the screened population. If prevalence of the condition or disease is low, screening will not identify many cases, rendering the test less cost-effective. In addition to cost considerations, some tests are not without risks of their own (eg, radiation) or discomfort. To justify administering these tests to the population, the potential harm to the patient if the disease is not diagnosed must outweigh the distress or pain of the test.

The Screening Test

In an effective screening program, the test must be inexpensive and easy to administer, with minimal discomfort and morbidity to the participant. The results must be reproducible, valid, and able to detect the disease before its critical point.

Screening tests must be widely available to the population for which they are intended. They cannot be available only at academic or other large medical centers. The tests must not have associated morbidity or mortality—even minor side effects may offset the benefits of screening. The test must also be reasonably priced, otherwise insurers may not provide coverage, and patients may be unable or unwilling to pay for the tests themselves.

The usefulness of the screening test is evaluated by its *sensitivity* and *specificity*. Sensitivity is the true positive rate; that is, the probability that a patient with a positive test result has the disease. As sensitivity increases, the number of patients with preclinical disease not diagnosed by the test decreases. Specificity is the true negative rate; the probability that a patient with a negative test result does not have the disease. A highly specific test produces a small percentage of erroneously positive results. Sensitivity is usually increased at the expense of specificity when the disease is serious and curable in its preclinical phase. However, high specificity may be desired over sensitivity when the costs or risks of further testing are significant, as they are, for example, with surgical biopsy. Patients must be informed that a negative screening result does not mean disease is not present, but rather the likelihood of disease is low. Since few tests have both high sensitivity and high specificity, multiple tests are often used to aid in detection of disease in the preclinical phase.

Screening test results must be reproducible. There are 4 frequent causes of variability: (1) Patient-related variation seen with cardiac motion or changes in patient size; (2) test-related variation, seen in patient positioning changes or technical factors in film development (such as in mammography); (3) intra-observer variability due to the differences in interpretation of a test at different times by the same clinician; and (4) interobserver variability due to variation of interpretation of a test by 2 or more clinicians. The last 2 often occur in interpretations of radiologic screening exams such as mammography. Interobserver variation may be minimized by use of strict criteria during interpretation.

Evaluation of Screening Tests

Comparing the outcomes of screened and unscreened groups can be challenging due to several biases. Lead-time bias refers to the fact that patients whose diseases are detected by screening before they experience symptoms have a longer survival time from diagnosis to death. But this seemingly increased life span is not due to the screening, it is merely the added time interval between the diagnosis of disease at screening and the time at which it would have been detected had the patient waited until the onset of signs and symptoms. Although overall survival—from onset of disease to death—may be the same for both screened and unscreened patients, the cause-specific survival, which is the time from diagnosis to death, may seem longer for screened patients because of their earlier diagnosis. In such instances, there is no advantage for the patient, and there may even be a disadvantage, since the screened patient has knowledge of the diagnosis for a longer period of time, which may increase emotional or psychological stress.

Not all diseases advance at the same rate. Those diseases with a long preclinical phase have more favorable prognoses, regardless of when they are diagnosed. When patients with these diseases are overrepresented among screen-detected cases, length-time bias occurs. Length-time bias could lead to the mistaken conclusion that screening is valuable, when the differences in mortality rate are actually due to the detection of less rapidly fatal diseases, while diseases that are more rapidly fatal were diagnosed after symptoms began.

Comparison of cause-specific mortality rates (the number of deaths in a population due to a specific cause divided by the total population) for screened patients versus rates for those patients whose diagnosis was made after the onset of signs and symptoms offers the best measure of the effectiveness of a screening program. Lead-time and length-time biases are canceled, and, while it is not possible to attribute all differences in mortality rates to screening programs, it is highly likely that at least some of the difference is due to them.

References

1. American Medical Association Council on Scientific Affairs. *Commercialized Medical Screening (Report A-03)*. Available at: <http://www.ama-assn.org/ama/pub/category/13628.html>. Accessed November 25, 2005.
2. Herman CR, Gill HK, Eng J, Fajardo LL. Screening for preclinical disease: test and disease characteristics. *Am J Roentgenol*. 2002;179:825-831.

Cheryl Herman, MD, is the codirector of breast imaging at Vanderbilt Breast Center and assistant professor at Vanderbilt University. She is also a breast imaging examiner and an American Board of Radiology Mammography Quality Standards Act film reviewer.

The viewpoints expressed on this site are those of the authors and do not necessarily reflect the views and policies of the AMA.

Copyright 2006 American Medical Association. All rights reserved.