Slide 1: Introduction

The screening module is one of a series created through the funding from the Centers for Disease Control and Prevention and the Association for Prevention Teaching and Research.

Slide 2: Acknowledgements

APTR wishes to acknowledge the following individuals that developed this module:

- Anna Zendell, PhD, MSW
  Center for Public Health Continuing Education
  University at Albany School of Public Health
- Joseph Nicholas, MD, MPH
  University of Rochester School of Medicine
- Mary Applegate, MD, MPH
  University at Albany School of Public Health
- Cheryl Reaves, MS, MLS
  Center for Public Health Continuing Education
  University at Albany School of Public Health

This education module is made possible through the Centers for Disease Control and Prevention (CDC) and the Association for Prevention Teaching and Research (APTR) Cooperative Agreement, No. 5U50CD300860. The module represents the opinions of the author(s) and does not necessarily represent the views of the Centers for Disease Control and Prevention or the Association for Prevention Teaching and Research.

Slide 3: Presentation Objectives

The objectives of this presentation are as follows: Define screening and the appropriate conditions for screening; evaluate screening tests in terms of validity, results, and generalizability; evaluate the effectiveness of a screening program; discuss common biases in evaluating screening programs; and discuss common ethical considerations in screening.

Slide 4: Introduction to Screening

We will begin with an overview of screening.
Slide 5: Oprah’s Full Body Scan

Take a few minutes to watch a video clip related to screening. Oprah Winfrey has completed a full body scan. Fully body scans are one of the newer “fads” in healthcare. Some are done in a medical office and others are conducted out of mobile vans. Oprah is discussing the results of her scan on national TV with her doctor. As you watch the clip and complete this module think about implications for patient screening. Are there medical concerns that may result from using this technology? How about any ethical issues? What barriers to accessing full body scans may exist and for what groups of people? In particular, think about what happens after the scan. Every human body has some imperfections. Do these imperfections necessarily require medical intervention, or may uncovering them cause the patient unnecessary physical or psychological harm?

Slide 6: Preventive Medicine & Public Health

Preventive medicine and public health share many common goals. First and foremost, both strive to promote quality of life by preventing disease. This goal is accomplished through health promotion programs and prevention of common diseases, like diabetes, cancer, and injuries. The difference is that preventive medicine works toward these goals at both the individual and population levels, while public health focuses on populations. [The distinguishing factor about preventive medicine that makes it special is its focus on populations!]
Slide 7: Prevention – Brief Overview

Prevention occurs at three levels. The first level is primary prevention. These are preventive measures that are undertaken to prevent the onset of illness and injury. This is done through the elimination of causal risk factors or by increasing resistance to the condition. An example of primary prevention is childhood vaccination against infectious disease. Secondary prevention entails measures that lead to early diagnosis and prompt treatment of illness or injury. Here we try to interrupt the disease process by detecting and treating it before symptoms emerge. For a test to have the “screening” characteristic it must be done in the non-symptomatic phase of the disease. Cancer screening is an example of secondary prevention. Tertiary prevention involves measures aimed at minimizing disability after disease symptoms have appeared. Cardiac rehabilitation following a diagnosis of heart failure is an example of tertiary prevention.

Slide 8: Screening Defined

Screening is defined as the presumptive identification of an unrecognized disease or defect through tests, examinations, or other procedures that can be applied rapidly and easily. Screening tests differentiate apparently healthy persons who may have a disease from those who probably do not.
Module 4: Screening

Slide 9: Importance of Screening

Screening is widely considered the bedrock of secondary prevention. Periodic health screening can lead to early detection and diagnosis of a disease. This early detection then leads to earlier treatment with a goal of decreasing mortality and morbidity related to that disease. In the case of infectious diseases, screening can also break the chain of transmission and prevent development of new cases. Screenings can be cost-effective if the disease is common enough and the test is accurate enough. It’s also cost-effective if affordable treatments that work are accessible to those patients whom test positive. Because we may develop diseases at many points throughout our lifespan, many screenings are only effective if done periodically.

Slide 10: Screening-Diagnosis Connection

The best screening protocols generally incorporate a patient history, physical exam, and laboratory test. Pre-test probability, also known as demographically determined risk, is probably the biggest reason to screen someone. Positive results of the screening test will trigger a diagnostic work-up and preventive or treatment interventions.

Slide 11: Screening versus Diagnostic Tests

Many people use the terms screening and diagnosis interchangeably. They are not the same.
Slide 12: Screening versus Diagnostic Tests

Screening tests are used for a presumptive identification of an unrecognized disease or illness.

Slide 13: Screening versus Diagnostic Tests

Diagnostic tests, on the other hand, are used to determine the presence or absence of a disease when the patient is showing symptoms of the disease OR has been targeted through a positive screening test. In some circumstances, the same test can function as either a screening or diagnostic test.

Slide 14: Characteristics of a Good Screening Test

A good screening test must meet several important criteria. It needs to be simple and quick to administer. It should be inexpensive and safe to use. It also needs to be readily available, along with an accessible plan of treatment in place in case of positive results. A good screen must be acceptable to the population in which it will be used. It must also be well researched and proven to be valid, reliable, and to have good predictive values. We will discuss these criteria in more detail a bit later.

Slide 15: Common Screening Tests

Next, we will discuss some common screening tests.
Slide 16: Common Disease Screenings

Take a few moments to review this list of screenings commonly conducted in the US. Do you know what conditions they screen for?

Slide 17: Common Disease Screenings

Medical practitioners have access to numerous screenings for potentially life-threatening conditions. We will briefly discuss some common ones and then focus on several in more detail. All women age 21 or older should receive Pap smears to test for cervical cancer. Pap smears may be administered even earlier for younger women and girls who are sexually active. These are routine ongoing screens, usually performed every 1-3 years. Pap smears may be done more often if abnormalities are found, or if a family history is reported. A fasting blood sugar test is conducted for anyone at any age at risk for diabetes, for instance pregnant women, people who are overweight, or those with a family history of diabetes. The fecal occult blood test is a common screening for colorectal cancer. We will learn more about this a bit later. Blood pressure is taken at annual exams and during each visit to the primary care provider. If hypertension is found, patients may be advised to come regularly for blood pressure tests. In some cases, people may receive blood pressure cuffs which enable them to check and log their blood pressure daily. Osteoporosis or osteopenia are often diagnosed in adults over age fifty, or in adults of any age with metabolic or eating disorders. Bone densitometry is indicated for women over the age of 65. People with a family history of these conditions or other risk factors may also have a bone density scan. Men over the age of 50 can be screened for prostate cancer, through an annual PSA test (or prostate-specific antigen). The typical tuberculosis screening is a TB skin test, also called a purified protein derivative or PPD test. TB tests are mandated for children and adults attending educational institutions. Anyone working in a healthcare setting or correctional facility is also mandated to be tested annually. Mammography is used to screen for breast cancer. The guidelines for screening are quite controversial. We will discuss these more in a few moments.
Slide 18: Common Wellness Screenings

There are numerous federal, state, and local initiatives to fight obesity in children and adults. Weight is one of the standard screens that occur at each medical visit. In many cases, BMI or body mass index is also calculated. A standard oral examination can identify dental caries (cavities) as well as oral cancers and many other conditions. The NIDA-Modified Alcohol, Smoking, and Substance Involvement Screening Test is a commonly used screening to detect the use and abuse of alcohol, tobacco, and other legal and illegal drugs. The screening test starts out with preliminary questions to assess use and then asks questions to identify abuse and/or dependence. Now, let’s delve into more detail on a few screenings.

Slide 19: Breast Cancer Screening

There has been a great deal of controversy recently on the protocols for breast cancer screening. Standard protocol for decades has been annual mammogram for women age 40 and older. Younger women are also screened annually if there is a family history of breast cancer. In 2009, the United States Preventive Services Task Force reviewed the medical evidence on regular use of mammograms. Issues such as population prevalence, costs, and consequences of false positive results were considered. Based on this review, the recommendations changed in late 2009. Mammograms are no longer recommended for women under 50, unless they and their providers feel it’s necessary. Women 50 and older are now urged to get mammograms every other year. The opposition to the revised mammogram recommendations has been strong. Many healthcare providers’ organizations, particularly radiologists and breast surgeons, have long advocated for yearly screenings. They have worked hard to promote the importance of regular mammograms. For more information on the reasons behind the controversy, please see articles appended in the resources section.
Slide 20: Colon Cancer Screening

We have a number of screening options available for colorectal cancer. The gold standard, or most accurate screening, is the colonoscopy. The colonoscopy allows immediate biopsy and immediate excision of any lesions. However, it’s also the most expensive and is very dependent on patient preparation and practitioner expertise. Sigmoidoscopy examines the most distal part of the colon. While this screening tool allows immediate biopsy and excision, it can only assess part of the colon and misses polyps that are beyond its range. Virtual colonoscopies, commonly called CT colonoscopies and barium enemas both require the same bowel preparation as the colonoscopy; however, there is no ability to do biopsies or excisions during the procedures. These screenings involve significant radiation exposure. Fecal occult blood testing (FOBT) is an easy, inexpensive, and non-invasive screening tool. Fecal testing is, however, less sensitive and specific than the other tests we have mentioned. Occult blood may be due to other health conditions, leading to many false positive results. New fecal DNA tests remain less sensitive than the other tests mentioned, but they have fewer false positive test results than FOBT. Recommendations for age of initial screening and frequency of screening vary by test and family history.

Slide 21: Case Study

You will find a case study on Colorectal Cancer Screening in the small group activities section of this module. This case study, used in classes across the United States, will allow you to practice evaluating diagnostic tests and screening programs. You will also discuss concepts related to preventive medicine. You will be able to apply these concepts at the individual patient and population levels.
Newborn screening is mandatory in all states. It looks for serious metabolic, hormonal, hematologic, and infectious conditions. States vary widely in what diseases they test for and how many. For example, NY tests for over 40 conditions, including HIV. Newborn screening involves a heel prick blood test done 24-48 hours after birth. Most states now use a tandem mass spectrometer to test blood, which tests for many metabolic conditions with one drop of blood. The 24 to 48-hour window of time is important. If testing is done too early, the presence of diseases may not show up in the baby’s blood. Some of the more common screens include: phenylketonuna (PKU), galactosemia, medium-chain acylcarnitine deficiency (MCAD), sickle cell anemia, and HIV. In addition to a state’s mandated tests, family history or ethnicity may indicate a need for additional screens. Non-mandatory screens are not paid for by the state. They may or may not be covered under a family’s personal health insurance.

**Slide 23: Evaluation of Screening Tests**

Now, we will briefly discuss how screening tests are evaluated for use.
Slide 24: Evaluating Tests

Two central concepts in evaluation of tests are validity and reliability. Validity is defined as how well the test result corresponds to the “true” condition of the patient, i.e. whether or not the person has the disease. Reliability, on the other hand, evaluates how consistent or reproducible the test results are over time or under different testing conditions. For this module, we will not discuss validity in detail, as the concepts are more relevant to research studies than clinical tests. To learn more about validity, see the Resources Section.

Slide 25: Characteristics of a Screening Test

Testing both validity and reliability is important. You can have a test that is quite reliable over time and across groups, but that isn’t valid. It misses the mark. Likewise, you can have a test that is valid, but not very reliable for certain groups or in certain testing situations. We want screening tests that are both valid measures of what is being tested and reliable in various groups and at all times.

Slide 26: Reliability

Reliability, commonly referred to as consistency, is the ability of a test to yield the same results with repeated measurements. Put another way, reliability is the degree to which results are free from error across testing occasions. If we have a large random error in a test, we see decreased reliability.
Slide 27: Common Types of Reliability

Intra-subject reliability refers to the consistency of measurement scores taken on the same subject across testing occasions. We are evaluating the degree of change in subject performance on the test from one time to another.

Slide 28: Common Types of Reliability

Intra-rater reliability refers to the consistency of measurements taken by the same tester on two or more testing occasions.

Slide 29: Common Types of Reliability

Inter-rater reliability is considered one of the most important indices of reliability for screening tools. Here we are looking at the consistency of measurement scores taken by two different testers.

Slide 30: Common Types of Reliability

Instrument reliability is another important indicator of reliability. It refers to the internal consistency of the measurement tool itself. All of these forms of reliability serve an important purpose in evaluating screening and diagnostic tests.
We will now talk about sensitivity and specificity. These concepts relate to validity, or how well a test result reflects reality. **Sensitivity** evaluates the ability to find the patients who do have the disease. In other words, sensitivity is the percent of people with the disease who are correctly identified. A sensitive test will reduce or eliminate false negatives. False negatives are test results that suggest absence of the disease when the disease is in fact present. These are also called beta or Type II errors. We want to aim for high sensitivity when the consequences of a missed diagnosis are serious. An example might be screening donated blood for the presence of HIV. An easy way to remember sensitivity is the SNOUT acronym. Sensitive test with Negative results rules OUT disease.

**Slide 32: Specificity**

Specificity evaluates the ability to identify which patients do not have the disease. Specificity seeks to reduce or eliminate false positives. False positives are test results that suggest presence of disease when the disease is not actually present. They are also commonly called alpha or type I errors. We want to aim for high specificity when there are serious adverse consequences for patients incorrectly identified as having the disease. For example we want a high degree of specificity in breast tissue biopsy outcomes to confirm cancer before doing a mastectomy or starting chemotherapy or radiation. If the specificity for a test is low, we need follow-up testing to rule out false positives. An easy acronym to remember specificity is SPIN. Specific test with Positive results rules IN disease. We want screening tests to be both sensitive and specific.
Slide 33: Relationship Between Sensitivity and Specificity

This table shows the relationship between sensitivity and specificity in prostate-specific antigen (PSA) levels. For screening tests whose results fall along a continuum of values, such as the PSA, there is a trade-off between specificity and sensitivity, based on where you set the “cut-off” score. In setting the cut-off point, we need to balance the risks of Type I and Type II errors for a given disease.

Slide 34: The 2x2 Table

The next few slides will show you how to calculate sensitivity and specificity. The gold squares represent the 2x2 table, while the blue boxes are labels. To understand the results from the 2x2 table, we need information from the test being evaluated and from the gold standard (or reference test) results.

Slide 35: Sensitivity

To calculate sensitivity, or the proportion of people with the disease who test positive, fill in the squares of the 2x2 table with the numbers with and without the disease who tested positive or negative. People with the disease who test positive are the true positives (TP). People with the disease who test negative are the false negatives (FN). People without the disease and who test positive are the false positives (FP). People without the disease and testing negative are the true negatives (TN). To calculate the sensitivity, divide the true positives by the total number with the disease, i.e. TP / (TP+FN).
Slide 36: Specificity

To calculate specificity, use the same 2x2 table, and divide the true negatives by the total number without the disease, i.e. TN / (FP+TN). The specificity tells us the likelihood that test will come back negative in someone without the disease.

Slide 37: Predictive Values

Predictive value of a test is a measure of the percentage of times the result (whether positive or negative) is the correct result. The percentage of all positive tests results that are true positives is the positive predictive value. The percentage of all negative test results that are true negatives is the negative predictive value. Predictive values are useful tools in clinical decision-making. They allow us to determine the probability of the patient having a disease based on test results.

Slide 38: Positive Predictive Value (PPV)

PPV is not an inherent characteristic of a screening test. PPV is affected by both the specificity of the test and disease prevalence.
Slide 39: Negative Predictive Value (NPV)

NPV is also not an inherent characteristic of a screening test. NPV is also affected by disease prevalence. The NPV can tell us the probability that the patient is disease-free based on negative results.

Slide 40: Test Characteristics and Population Tested

Sensitivity and specificity are constant for a particular test, but it’s surprising and counter-intuitive how dramatically the PPV and NPV vary depending on the prevalence of the condition in the specific population being tested. In a group with low prevalence, a test will have a low PPV and a high NPV. In a group with high prevalence, a test will have a high PPV and low NPV.

Slide 41: Predictive Value and Prevalence

Here is a visual example of the relationship between predictive values and prevalence.
Now that we have done the basic 2x2 tables, let’s figure out how to calculate PPV.

Let’s try some sample calculations to give you a feel for how this works, using the ELISA test to screen for HIV. The test’s sensitivity and specificity are both over 90%. Let’s start with a population with low prevalence – only 1.5% -- for instance, a population of 1,000 patients at a prenatal clinic. Using the ELISA in this population gives us a positive predictive value of 12% and a negative predictive value of 99.9%. Even among the patients who test positive, the probability that they have HIV is low. Put another way, the probability that these patients got false positive results and are not HIV-infected is quite high.

Now let’s look at a population with a high prevalence of HIV. Here we have a population of 1,000 patients at an STD clinic. The HIV prevalence is 6%. Using the same ELISA test, the positive predictive value is 37%, much higher than the neonatal clinic population. The NPV remains high at 99.6%.
Moving to a setting with a very high HIV prevalence, for example in Zambia, has the opposite effect. The PPV rises to 75%, and the NPV falls slightly to 98.3%. The probability that these test-positive Zambian patients have HIV is quite high.

Sometimes, it is prudent to use a series of tests to screen for a condition. There may be no gold standard screening tool available for a condition. The available tools may be too invasive, costly, or otherwise impractical to use for an initial screening. We may use multiple screening tests simultaneously or sequentially. Let’s take screening for the presence of Down syndrome in a fetus. A common approach is to simultaneously test for three key bio-markers in the mother’s blood. Each of these tests individually has low sensitivity and specificity. Together, however, they provide a viable alternative to amniocentesis. Amniocentesis is a highly sensitive and specific test. However, it is also invasive and has a small but significant chance of causing a miscarriage. It is better suited as a follow-up diagnostic test than a widespread screening test.

Gestational diabetes testing generally employs a two-stage screening protocol. In the first trimester, a risk assessment is done to identify women at risk for the condition. The Oral Glucose Tolerance Test (OGTT) is then done for women identified as at risk. This two-stage protocol minimizes the number of women who need the Oral Glucose Tolerance Test. The OGTT is a time-consuming test. Women must stay at the clinic or lab for several hours. Most women’s risk of diabetes is low early in pregnancy. Women who are obese or have a prior history of gestational diabetes need to be screened earlier in pregnancy using the OGTT.
Slide 48: Multiple Screening Tests

Another use for two-stage screening is designed to maximize predictive values. Let’s take an example of HIV screening in a suburban primary care office. A risk assessment about sexual and drug use history is given to patients. Those with risk factors for HIV infection are then given a blood test. By using the written questionnaire as a preliminary screening test and selecting the higher risk population to receive the blood test, the number of false positive results is reduced, and the PPV of the test is increased. These two examples of sequential screening tests illustrate different benefits of sequential screening. Reducing the need for more invasive screenings and increasing the PPV are just two.

Slide 49: Effectiveness of Screening Programs

The next few slides discuss effectiveness of screening programs.

Slide 50: Screening Effectiveness Evaluation

Sensitivity and specificity alone are never sufficient for a sound decision about whether to use a screening test. We need to weigh other factors in terms of the individual patient, the healthcare system, and for society. We need to analyze whether the benefits outweigh the risks. We saw that amniocentesis carries significant risks and should be used only when other tests indicate a need for more intensive screening. Cancer screening in very old adults with a short life expectancy has lower benefit (and perhaps higher risks) than screening healthy younger people. Likewise, evidence is mounting that mammograms for women in the 40-50 year old age range provide limited benefits because the prevalence of breast cancer in that age group is so low, and they carry potentially unacceptable risks of excess radiation exposure and false positives leading to stress and unnecessary invasive procedures. The level of inconvenience is another factor to consider. Is the screening conducted in a location that the patient can easily get to? If the screening protocol necessitates a follow-up, such as for the PPD or HIV test, will the patient be able to follow through?
We must also look at the overall cost and resource needs to conduct a screening. Will it be covered by insurance? Are healthcare providers available for the screening and any follow-up tests? Finally, does the test fit with patients’ values or cultural norms? If a population firmly believes in a non-medicalized pregnancy experience, pregnant women from this group will be unlikely to agree to an amniocentesis, even if one is indicated. Likewise, some traditional Latinas or Muslim women may not follow through with breast or cervical cancer screenings due to cultural or religious beliefs. These are but a few considerations in designing a screening program.

Slide 51: Study Design

To assess diagnostic or screening tests, studies seek to expose operating properties or characteristics of the test and assess how close a match these properties are to a “gold standard” diagnostic test. Researchers seek to determine the power of the tool to differentiate between those with and those without a target condition.

To study a screening test, we look at the test’s performance in a group of patients NOT known to have the target condition, and compare it to the performance of a “gold standard” test. Gold standard tests are tests that are considered to be the diagnostic standard for the target condition. (If there is no current gold standard test, the researchers may have to follow patients over time to determine who develops symptomatic disease.) The results of both tests are compared. In evaluating these studies, we want to see one of two things. We want to see a great deal of similarity in results between the two tests. Or, we want to see that the test in consideration exceeds the capacity of the gold standard test to discriminate between presence and absence of the condition. For example, we might ask about the test performance of the liquid-based pap for cervical cancer screening. A group of women with no indication of cervical cancer would be offered both the liquid-based pap and standard Pap smear. The rate of positive, false positive, negative and false negative results would be compared for both the liquid and standard Pap smear screens.

Slide 52: US Preventive Services Task Force

The US Preventive Services Task Force (USPSTF) produces thoroughly researched screening recommendations. Let’s spend a few minutes talking about the USPSTF and how it works.
The USPSTF is charged with systematically reviewing the evidence of effectiveness and developing recommendations for clinical preventive services. The recommendations span screening tests, counseling, and preventive medications.

The USPSTF uses a rigorous methodology. First the analytic framework is defined, including the desired outcomes and key questions to be asked. What constitutes relevant evidence is operationally defined. The evidence is then gathered. The overall quality of each of these studies is evaluated individually. The evidence is then synthesized and judged in aggregate. We want to see convincing or adequate evidence at the very least. After this, the Task Force determines the balance of benefits and harms, based on the evidence. From all of this information, a grade is assigned that links the recommendations to judgments about the net benefits. Grades may range from A to D, or an I may be assigned for inadequate evidence.

In evaluating the available evidence, the Task Force asks a number of questions. Do the studies have the appropriate research design to answer key questions? Are the existing studies high quality? Finally, are the study results applicable to the general primary care population and setting?
Slide 56: Critical Appraisal Questions

The number of studies that have been done is evaluated, as well as the size of the studies. The Task Force assesses the consistency of the studies’ results. Finally, any other factors that can help assess the certainty of the evidence are weighed.

Slide 57: Interpret Task Force Grading

This slide shows us how to interpret the Task Force grading system based on the magnitude of net benefit. The task force uses letter grades for their recommendations.

Slide 58: Communicating USPSTF Recommendations

Here we see the letter grades defined, along with suggested clinical practices.
Sometimes, there is only one effective test to screen for a condition. However, often we have choices, such as whether to use a newer test or the existing gold standard. It’s important to assess what are the best tests for the target population. Tests should be valid for your population. A test should have been normed for the population you are working with. Tests should be accessible. This means that tests should be covered under health insurance, and affordable. When we talk about accessibility, we look at other issues, like language barriers and overall literacy. We also look at privacy issues, particularly with tests that elicit sensitive information that could stigmatize a patient if the information leaked out. The healthcare system must be able to support use of a screening test or program. If we screen for colorectal cancer using the FOBT, we need to ensure there is a colonoscopy provider in the area for follow-up testing. If we recommend colonoscopy as the primary screen, we need to have many more colonoscopists available. We also need to ensure that there are sufficient resources for follow-up treatment. We ask where patients will get follow-up care, whether they can get there, and if the care is affordable to them. In sum, there must be a system in place that can deal with positive results.

This diagram illustrates one way to test a screening tool. We start out with patients without any evidence of a target condition. In this case, let’s use the example of cervical cancer. We will compare a relatively new test; the liquid based Pap smear, with our gold standard, the standard Pap smear. We then calculate the specificity and sensitivity of both tests. It’s quite important to start with patients showing no evidence of the target condition. The sensitivity and specificity characteristics of the test could be different in asymptomatic vs. symptomatic patients. Screening tests are meant to be used on asymptomatic patients.
Slide 61: Evaluating Screening Research

The only way to practice evidence-based medicine (EBM) is to read the medical literature. You will find some resources at the end of this module on reliable sources of EBM literature. The first question we always want to ask as we read study results is whether the results are credible. Look for the following information to assess how valid the results were:

- Did the researcher enroll the right group of people? To test a screening test, the right group would comprise individuals without overt symptoms of the target condition.
- Was the study sample large enough to draw conclusions?
- Were the methods employed to test the screen appropriate to the study aims?
- Did all the subjects get both the target AND gold standard test?
- Some diseases necessitate re-screening for inconclusive results or to allow the disease time to be visible to screening tools. For the target condition, was enough time allowed to lapse between re-screenings where appropriate?

Next we evaluate what the researchers say about specificity, sensitivity and predictive value.

- Do the authors convey the importance of classifying patients correctly and the implications of false negatives and false positives?
- Is the predictive value calculated in a population whose disease prevalence is similar to the target patient population?

If the results are valid and sufficiently large and stable, then we look to apply what we’ve learned.

- Who would this screen likely work for? Who might it not work for? Who has it not been “normed” for? For this, we need to know who was represented in this and other studies.
- Also, what would be the net impact of the screening? For example, what are the benefits and any likely side effects or adverse reactions?
- How does the test improve on the current state of preventive medicine? We want to see a test that will enhance accuracy or make screening easier, more accessible, or cheaper. The test should make a positive contribution to medicine and patient care. This may be through more accurate results, reduced cost, diminished pain and intrusiveness, or perhaps less time taken to complete the test.
Module 4: Screening

TRANSCRIPT

Lead time bias: over-estimation of survival rate among screening-detected cases
- When survival is calculated from diagnosis point
- Length bias: over-estimation of survival rate among screening-detected cases
- Due to excess of slowly progressing cases among those identified by screening

When survival is calculated from diagnosis point, length bias is more subtle than lead time bias. Due to excess of slowly progressing cases identified by screening, treatment and test outcomes might be affected. If we look at raw numbers, screening may appear to extend life even though all that has really happened is a potentially terminal diagnosis for longer. If we examine only the raw numbers, screening appears to increase survival time, even though all that has really happened is an earlier diagnosis. This is called lead time bias. Length bias is more subtle than lead time bias. Using uterine cancer again as an example, if we conduct regular screenings for uterine cancer, the chances are excellent that we will pick up many slow-growing cancers. In fact, we will pick up more slower-growing tumors than fast ones. Rapidly growing tumors have shorter asymptomatic phases where they could be found through screening and lower survival rates. Again, if we look at raw numbers, we might incorrectly deduce that people whose cancer is detected by screening live longer, when in fact, the screening has just picked up a disproportionate number of slow-growing cancers.

Slide 62: Screening Outcome Considerations

We need to be cautious about inferring survival rates from screening detected cases alone. There are inherent biases in screen-detected long-term disease outcomes. By screening, we seek to diagnose a disease earlier than it would be found without screening. Without screening, the disease may be discovered later, when symptoms appear. Let’s compare two cases in which a person is diagnosed with uterine cancer. Even if both patients die at the same time, because we diagnosed the cancer early with screening, the survival time after diagnosis is longer with screening. Life has not been extended for the patient who was screened. In fact, the patient who was diagnosed earlier may suffer added emotional effects from living with the knowledge of a potentially terminal diagnosis for longer. If we examine only the raw numbers, screening appears to increase survival time, even though all that has really happened is an earlier diagnosis. This is called lead time bias. Length bias is more subtle than lead time bias. Using uterine cancer again as an example, if we conduct regular screenings for uterine cancer, the chances are excellent that we will pick up many slow-growing cancers. In fact, we will pick up more slower-growing tumors than fast ones. Rapidly growing tumors have shorter asymptomatic phases where they could be found through screening and lower survival rates. Again, if we look at raw numbers, we might incorrectly deduce that people whose cancer is detected by screening live longer, when in fact, the screening has just picked up a disproportionate number of slow-growing cancers.

Slide 63: To Screen or Not to Screen

This table provides criteria for when to screen...
Slide 64: Pseudodisease

Pseudo disease or over-diagnosis is defined as the identification of disease that would be unlikely to impact the patient during their lifetime. This is the major concern with both mammography and prostate screening. We may find histologic evidence of cancer, but we’re often picking up a lot of “cancer” of little clinical importance. Even a gold standard “diagnostic” test isn’t a perfect window on the future. As a result of screening, some individuals will be diagnosed and treated for an illness that would never have become clinically apparent.

Slide 65: Screening and Ethics

Next, we will discuss just a few of the ethical considerations around screening.

Slide 66: Ethical Considerations

We will discuss mandatory screening programs, genetic testing, and health disparities around screening. Finally, we will talk about considerations in creation of a screening program.
There are more mandated screening programs affecting Americans today than we might realize. Every person applying for a marriage license must take a blood test for syphilis. All healthcare workers are tested for tuberculosis. Drug testing is mandated for airline pilots. We discussed newborn screening earlier. Each of these screening protocols has an important role in public health. Yet, there will always be concerns around universal testing. We will get into a few of these.

We need to do a cost-benefit analysis at all levels from the individual all the way to the societal costs and benefits. We must ensure that the benefits outweigh any costs. Let’s take the example of doing drug screens in airline pilots. The benefits are clear. We may avoid airline accidents caused by drug-impaired pilots operating aircrafts. This potentially saves hundreds, even thousands, of lives. The costs of testing include inconvenience of testing or possible flight cancellation if a pilot tests positive. However, the costs are outweighed by the benefits to society.

The benefits to mandated testing are often clear. We can identify serious problems that can harm others. We also identify conditions necessitating immediate treatment.

The costs may not be so clear at times, but every decision has potential consequences. Patients undergoing mandatory screening lose their autonomy. We may not be able to maintain confidentiality in certain situations, such as diagnosis of syphilis or HIV/AIDS testing. If we are testing a low risk population, the positive predictive value is decreased and the consequences of false positives are increased.
Slide 70: Genetics Testing

Genetic testing has many benefits. Chief among them is screening for disease that can affect not only the individual tested, but also other members of his or her family. Individuals can be tested to see if they carry genes for diseases that may be passed on to children. An example would be Fragile X Syndrome, a developmental disability. Unborn children can be screened for diseases, such as Down syndrome. Genetic diseases can be identified in people of all ages before symptoms become evident.

Slide 71: Genetic Testing

Results of genetic testing must be interpreted carefully by a trained medical geneticist or a genetic counselor. A positive result in genetic testing does not necessarily mean that the person will develop the disease. With these results come decisions—often difficult ones. Should someone with a genetic marker for ovarian cancer have their ovaries removed proactively? Likewise, if a fetus tests positive in utero for Down syndrome, will the couple have the baby or abort? While a negative result most often brings peace of mind, a positive result can trigger enormous stress. Take the decision to have pre-emptive surgery, such as ovary removal, to avoid cancer. These are difficult decisions and lead to major life changes. Proper genetic testing must be done through a medical provider. At-home genetic testing kits are generally not effective, except for a very few approved by the FDA. However, some people turn to them due to high costs, lack of insurance coverage for testing, privacy concerns, or aggressive advertising. Genetic testing may bring up confidentiality issues. A breach in confidentiality can lead to risk of job loss, difficulties maintaining health insurance, and difficulty obtaining a life insurance policy. Recent federal legislation forbids use of genetic test results in these settings, but a person might have to defend those rights in court. A final ethical issue with genetic testing involves the ownership of the DNA. When DNA goes to a laboratory for testing, many laboratories will keep the DNA for further testing. They retain proprietary rights to a person’s genetic material. For some patients, this is not a major concern, but for many it will be. It’s important to know the policies of any genetic testing facility and to be sure that patients are fully informed, as well.
Rural areas often lack access to local healthcare. Additionally, many people struggle with transportation to reach healthcare providers. Rural areas tend to have sparse public transportation, if any. Hospital closures in rural areas are on the rise in many states. Hospital-based screenings become more difficult to obtain in these areas. Inner cities have similar challenges in terms of limited resources, though fewer transportation issues. People who lack insurance or are under-insured often have great difficulty receiving health screening. Affordability of screening is a huge issue for the uninsured, as is education on which screening tests are most critical to obtain. Many uninsured individuals do not have primary care doctors, depriving them of crucial advice about screenings. In addition, they receive much of their healthcare in emergency room settings, which are not designed to provide preventive care. Migrant and immigrant populations often face serious barriers to health care, including screening. Migrant farm workers are among the most economically disadvantaged and medically vulnerable groups in the US. They have little or no access to health care or medication. Barriers may include lack of health insurance, language barriers, lack of transportation, fear of deportation, and lost income. Migrant workers also face the very real threat of being fired or not invited back to work because of health issues or time lost due to medical appointments. With such insurmountable barriers, migrant workers are not likely to risk their incomes for health screening. What can minimize some of these barriers? Mobile screening units may improve access to screenings. Many of these mobile units offer free or low-cost screening tests. This may necessitate use of non-gold standard screening tests. Non-gold standard tests may be more affordable and easier to administer in a mobile unit. They may also require fewer follow-up tests or procedures. Additionally, after-hours screening options may be a viable alternative for migrant farm workers. Bi-lingual health educational initiatives may also help improve access to screening.
Slide 73: Disparities

Information about diseases, prevention and screening practices, is disseminated differentially across cultures. For example, not all groups receive culturally sensitive education on screening and preventive health in general. We often see a lack of detailed information provided to non-English speaking patients. Patients from certain cultures may have specific world views of illnesses. For example, in some Asian and Hispanic cultures, a cancer diagnosis may be viewed with a sense of fatalism. There is a growing evidence base on best practices for adapting education, outreach, and screening protocols to be more culturally sensitive. For example, if we look at prostate cancer, African American groups might do better with outreach that details the risk factors for prostate cancer, the benefits of being screened, and ways to get screened. Mexican males, on the other hand, receive the least amount of health information of all ethnic groups researched in the US. An increase in health education outreach to this group is of vital importance. Screening practices often need to be adapted, depending on the target group. For example, an Iraqi woman may have great difficulty accepting screening for breast or uterine cancer from a male doctor, despite having a female nurse in the room. Likewise, having medical students in the room during a Pap smear might lead to future avoidance of these screenings. Services can be made much more sensitive to the patient’s cultural background by a thorough explanation and an examination by a female medical provider.

Slide 74: Developing a Screening Program

As we develop a screening program, we need to address some ethical questions. First, how ethical is it to use a test that may tell people they have a condition when they do not? If we embark on this screening program, will those with false positive test results engage in unnecessary testing or invasive procedures? If they do, who will pay for it? We need to examine the potential adverse effects to unnecessary testing. We also need to consider the potential for emotional distress in patients screened false positive? How ethical is it to use a test that may tell people they do not have a condition when they actually do? What happens to these people? How much later will they be diagnosed, and how will this affect their mortality or morbidity? Again, we need to consider the potential for emotional distress when they are finally diagnosed. Finally, we need to consider the outcomes if we develop a screening program without a system in place to treat those who test positive. What if we embark on an HIV/AIDS screening program in a very remote Kenyan community 20 kilometers from the nearest hospital or...
Strategic targeting for screening

- Groups with higher prevalence – increase PPV
- Provider vs. patient – who is more likely to request?
- Weigh benefits/drawbacks of screening test
- Grow body of evidence-based medicine allows us to:
  - Identify more precise screening protocols
  - Weigh benefits/drawbacks of screening test
  - Strategic screening can be cost-effective

Slide 75: Implications for Practice

Strategic screening protocols can lead to increased precision in screening for disease. Identification of groups with higher prevalence and targeting screenings toward these groups will increase the test’s predictive value and provide earlier intervention opportunities for those identified as having the disease. Strategic targeting also incorporates knowledge of whom to educate on the importance of a particular screening. In some cases, education and outreach to patients is an efficient method. However, in some cases, such as pap smears or mammograms, the doctor often suggests these screens. In this case, the medical provider is an important target for education and outreach. In some cultures, all medical decisions are made by elders in the family, in which case their education is crucial. With shrinking resources and an aging population, evidence based medicine is a valuable resource to identify more accurate screening tools and protocols. Critical review of the literature provides a mechanism for weighing the benefits and costs of particular screens for your patient population. As we have discussed, the gold standard may not be best for your community or your practice. Screening can be a cost-effective method of earlier disease identification and improved treatment outcomes. It is important to keep up with accumulating evidence to determine the best screening practices for your practice and community.

Slide 76: Summary

There are several key points to take away from this module. Screening is the bedrock of secondary prevention. Screening and diagnosis are not the same. Screening is conducted on asymptomatic patients, while diagnosis is conducted on patients showing some signs or symptoms of the target disease. Sensitivity and specificity are characteristics of a screening test that determine a test’s validity. Predictive values are affected by the sensitivity and specificity of the screening test and by the prevalence of the target condition in the population. Targeting screening to high-risk populations increases the positive predictive value. Decisions on screening protocols must weigh the acceptability and applicability to the practitioner, population, and individual patient.
Module 4: Screening

TRANSCRIPT

Collaborating Institutions

- Department of Public Health
  Brody School of Medicine at East Carolina University
- Department of Community & Family Medicine
  Duke University School of Medicine

Advisory Committee

Mike Barry, CAE  
Lorrie Beanigh, MD  
Nancy Bennett, MD, MS  
Ruth Gaare Bernheim, JD, MPH  
Amber Berrian, MPH  
James Cauley, MPH, PA-C  
Jack Dillenberg, DDS, MPH  
Kristine Gollbus, RN, DrPH  
Asim Jari, MD, MPH, FACP

Denise Koo, MD, MPH  
Suzanne Lazorick, MD, MPH  
Rika Maeshiro, MD, MPH  
Steve McCurdy, MD, MPH  
Susan M. Meyer, PhD  
Sallie Rixey, MD, MEd  
Nawraz Shawir, MBBS

APTR

- Sharon Hull, MD, MPH  
  President
- Allison L. Lewis  
  Executive Director
- O. Kent Nordvig, MEd  
  Project Representative