

## Screening tests and screening programmes

Asoka Weerakkody<sup>1</sup>

*Sri Lanka Journal of Obstetrics and Gynaecology* 2011; **33**: 194-196

'Screening' is an important aspect of current medical practice and hence every clinician should know its basic features. It is also a very popular topic among the examiners for the MD and the MRCOG. This article is not meant to be an exhaustive guide to screening (for which the reader is referred to standard texts), but as 'revision notes'.

### *What does a screening program do?*

It diagnoses a condition in its pre-clinical stage, so that it could be treated early (Diabetes) and/or its progression halted (CIN).

### *How does it do it?*

First the screening test identifies a group of people at 'high risk' for having the condition, from the total population at risk (e.g. abnormal cervical smear).

To this selected group, a further diagnostic test is applied to determine which ones actually have the condition (e.g. colposcopy and biopsy). Hence, a screening program consists of these two tests *which work in tandem*.

*The screening test* by necessity has to be simple, cheap, easy to carry out, non-invasive and 'low-tech' (because it is applied to a large population). It does not give an answer to the question, 'do I have it?', but only tells you that 'you are at high risk of having it'.

*The diagnostic test* is 'definitive' in that it does answer the above question. It is usually more complicated, expensive, 'high-tech' and often more invasive - hence cannot be applied to large populations. It should also be nearly 100% accurate. (e.g. cervical biopsy, karyotyping of amniotic fluid, GTT).

### Screening test

It follows that a screening test will return a positive result in some people without the disease; it could

also miss some with the disease. A good screening test should have low figures for both. The actual values vary according to the cut-off point used, which is a compromise between sensitivity required and an acceptable level of false positives (see below).

### *How good is a particular screening test?*

Certain parameters have been identified to measure how good a screening test is.

They do so by asking a series of questions. The first of these concerns those who have the disease and those who do not.

*Sensitivity*: How many with the disease tested positive? This is the most fundamental question, as this is the whole point of having the program. So, it should be at least 60% or more.

*Specificity*: How many without the disease tested negative?

*False positive rate (FPR)*: How many without the disease tested positive?

*False negative rate (FNR)*: How many with the disease tested negative?

In practice, sensitivity and FPR are the most useful and therefore the most often quoted.

### Practice point 1

- Find a mathematical relationship between specificity and FPR.
- FNR is not often quoted. Contemplate why.
- What is the practical use of FPR?

Next series of questions relate to those who tested positive or negative.

*Positive predictive value (PPV)*: How many of those tested positive actually have the disease?

*Negative predictive value (NPV)*: How many of those tested negative do not have the disease?

<sup>1</sup> Consultant Obstetrician and Gynaecologist.

Correspondence: Asoka Weerakkody

E-mail: asokaweerakkody@hotmail.com

A worked example would help:

It is known that the prevalence of GDM in a population is 10%. 100 women were screened by using random blood sugar with a cut-off point of 8.2 m mol/lt:

20 tested positive, out of which 8 were confirmed a GDM by GTT.

You should know how to fit this data into a 2x2 contingency table (a favourite in the exam):

*Number needed to test (NNT):*

How many need to undergo diagnostic test to detect one case?

This parameter is often used to describe/measure the cost effectiveness of a program – the smaller the number, the better.

**Table 1. GDM/No GDM vs. Screen positive/Screen negative**

	<i>Scr pos</i>	<i>Scr neg</i>		
GDM	8	2	subtotal GDM	10
No GDM	12	78	subtotal no GDM	90
	subtotal scr pos 20	subtotal scr neg 80	<b>Grand total</b>	100

Table 2 illustrates how to use the above data to obtain the parameters we require:

**Table 2. Parameters and their calculations**

<i>Parameter</i>	<i>Question</i>	<i>Denominator</i>	<i>Numerator</i>	<i>Result (%)</i>	<i>Comment</i>
Sensitivity	How many with GDM tested positive?	10	8	80	The most important measure (asks ‘does it work?’)
Specificity	How many without GDM tested negative?	90	78	87	Needs to be in the eighties
FPR	How many without GDM tested positive?	90	12	13	Limited by the availability of resources
PPV	How many screened positive actually had GDM?	20	8	40	Important in cost-effectiveness terms
NPV	How many screened negative did not have GDM?	78	80	97.5	Note the high value. See below.

(Tip: Get the right denominator first; that is where most people go wrong.)

**Number Needed to Test (NNT):**

How many need to undergo diagnostic test to detect one case?

This parameter is often used to describe/measure the cost effectiveness of a program – the smaller the number, the better.

**Practice point 2**

- a. Find a relationship between PPV and NNT.
- b. Whereas a low PPV might be acceptable (even 10% or less, e.g. amniocentesis for Down's), NPV must be in the nineties for the test to be acceptable. Contemplate why.

**Learning point:**

**Sensitivity** and **FPR** are unique to a screening test and do not change, whatever the population it is applied to (provided the cut-off point remains the same).

**PPV** (and **NPV**) however, do change according to the prevalence of the condition in the population tested; higher the prevalence, the higher the PPV.

**Likelihood Ratio** (LR – sometimes called 'Positive Likelihood Ratio'):

As both the sensitivity and the FPR are important factors, it is convenient to have one measure combining both, in order to assess the overall value of a test. This is what the LR does. It is simply:

$$\frac{\text{Sensitivity}}{\text{FPR}}$$

This allows comparison to be made between two different screening tests for the same condition.

**Practice point 3**

- a. To detect a certain condition, two screening tests have been offered:

Test A:  
sensitivity 90%; specificity 85%

Test B:  
sensitivity 85%; specificity 90%

Which one would you choose?

- b. How does prevalence of a disease affect the LR of a test?

*Further practical problems:*

1. Given that sensitivity and FPR for the same test could vary with different cut-off points, how do we know which is the best cut-off point?
2. How do we know that a particular test is any good at any cut-off point?
3. How do we compare two different tests overall, i.e. over several cut-off points?

The answer to all three questions lies in the 'ROC' curve.

*Receiver operating characteristics (ROC) curve*

- o This is a curve obtained by plotting FPR (X axis) vs. sensitivity (Y axis) at different cut-off points for the same screening test.
- o The line 'y = x' represents a 'useless' test, and is called the 'the line of non-discrimination'.
- o To be useful therefore, any curve must be above this line. The further it is from the line, the better.
- o The point furthest away from the line represents the most optimal cut-off point.
- o If there is more than one test, the test for which the curve is furthest away overall, is the best – in other words, the one which has the greatest the area under the curve ('AUC').
- o For any given test, unless the AUC is at least 0.8, it is not a worthwhile test.

**Screening programmes**

A screening test, however good it is, does not constitute a 'screening programme'.

The condition to be screened must be an important health problem in the locality i.e. serious enough and common enough. For e.g., screening for toxoplasmosis is considered useful in the European continent, but not in the British Isles.

It should cover the majority of the population at risk, especially those at high risk. For e.g., having a cytology screening service confined to residential parts of Colombo, is not going to reduce the incidence of cervical cancer in SL.

There should be adequate facilities and expertise to deal with the screen positives swiftly. The diagnostic tests needed must be affordable and available. Diagnosed cases must be dealt with quickly.

There should be an effective counselling and support service to deal with all screen positives. There should be robust mechanisms for record keeping, data protection, audit and follow-up.

It is now recommended that before a new screening service is introduced to a population, a pilot trial (preferably cluster-randomised) should be run to prove its efficacy and also to address the above points.