

Screening and diagnostic tests

Mikael Hartman, MD PhD
Gustaf Edgren, PhD

Outline

- Introduction to screening
- Assessing the performance of screening tests
- Screening programs and their evaluation
- Summary

Screening

- Screening is the strategy with which tests are performed in selected populations in order to find disease in otherwise asymptomatic individuals
- Importantly, unlike most other areas of medicine, screening is performed in HEALTHY people with NO sign of disease
 - This poses unique challenges!

Rationale for screening

- Intuitively, screening programs are implemented to:
 - Detect disease earlier, before spread has occurred,
 - and before symptoms have begun
 - Ultimately, the goal is to improve prognosis,
 - and to prevent disease spread between people (infectious disease)

When is screening suitable?

- According to WHO guidelines from 1968:
1. The condition should be an important health problem
 2. There should be a treatment for the condition
 3. Facilities for diagnosis and treatment should be available
 4. There should be a latent stage of the disease
 5. There should be a test or examination for the condition
 6. The test should be acceptable to the population
 7. The natural history of the disease should be adequately understood
 8. There should be an agreed policy on who to treat
 9. The total cost of finding a case should be economically balanced in relation to medical expenditure as a whole
 10. Case-finding should be a continuous process, not just a "once and for all" project

When is screening **really** suitable?

1. The disease should have serious consequences
2. The disease should be treatable and early treatment should improve prognosis
3. There should exist a simple, harmless and valid screening test
4. The prevalence of preclinical, asymptomatic disease should be sufficiently high in the screened population

Prevalence vs. incidence screen

Prevalence screen

- First screening round
- Detects cases prevalent at first screening
- Varying length of pre-clinical phase / stage

Incidence screen

- Subsequent screening rounds
- Detects cases with onset after most recent screening
- Less variation in length of pre-clinical phase

Interval cases

- Cases detected in between screening rounds, presumably due to symptoms
- Often worse prognosis



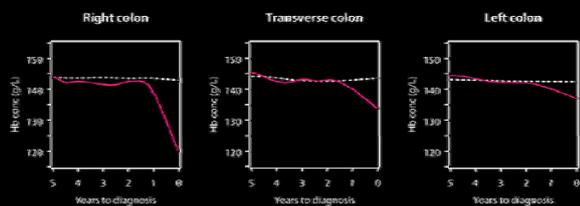
Screening tests

- A screening test is a test by which we want to be able to sort asymptomatic individuals into two groups:
 - Likely to have disease
 - Unlikely to have disease
- This division is usually quite unnatural, whereby most screening tests represent a dichotomy of a continuous measure which is usually:
 - A compromise between *too many false positives* and *too many false negatives*

Screening tests, cont.

- The ideal screening test would classify all persons with the disease as having the disease and all persons without the disease as not having the disease
- The ability of a test to correctly classify diseased and non-diseased is called the **test validity**
- The test's ability to perform exactly the same repeatedly is called the **test reliability**
- Which is more important?

Example – Hemoglobin and colon cancer

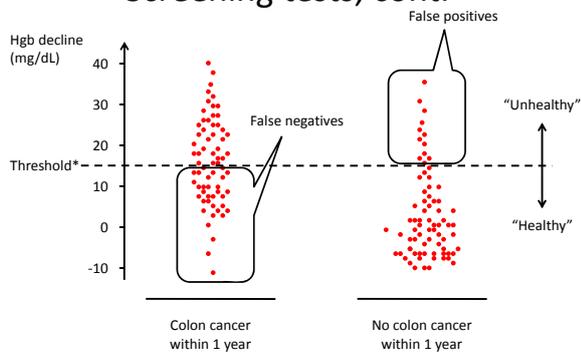


This is great, lets start a screening program!

Hemoglobin screening program

- Lets say we start measuring the hemoglobin concentration of all Singaporeans once every year to screen for colon cancer
- For every individual, we then compare this years result with the previous average to detect declining hemoglobin concentrations
- The question is, where do we draw the line between healthy and unhealthy?

Screening tests, cont.



*Criterion of positivity

The screening 2-by-2

		Gold standard / "truth"		
		+	-	
Screening test	+	True positives (a)	False positives (b)	a+b
	-	False negatives (c)	True negatives (d)	c+d
				a+b+c+d = N

The screening 2-by-2

		Colon cancer within 1 year		
		+	-	
Hemoglobin screening test	+	37	12(00)	
	-	33	58(00)	

Sensitivity

- The test sensitivity is a measure of the test's ability to correctly classify those **with** the disease:

$$Sensitivity = \frac{a}{a+c}$$

i.e. the proportion of those with the disease that are correctly classified as having the disease

Specificity

- The test specificity is a measure of the test's ability to correctly classify those **without** the disease:

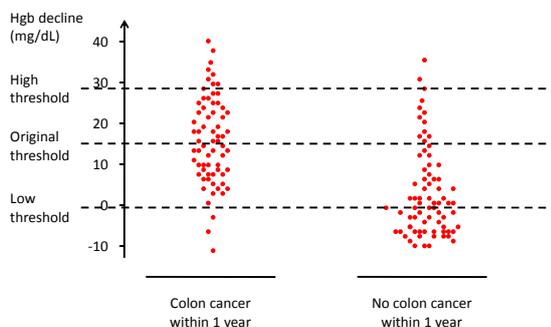
$$\text{Specificity} = \frac{d}{b+d}$$

i.e. the proportion of those without the disease that are correctly classified as not having the disease

Hemoglobin screening

		Colon cancer within 1 yr		
		+	-	
Hgb decline > 15mg/dL	+	37	1200	<ul style="list-style-type: none"> Sensitivity = $37 / (37 + 33) = 53\%$ Specificity = $5800 / (5800 + 1200) = 83\%$ Looks good, or?
	-	33	5800	

Is there room for improvement?



Optimization of screening tests

- There is usually a considerable overlap between test results of healthy and unhealthy
- The choice of test threshold is therefore a compromise between **false positives** and **false negatives**
 - If we increase the threshold, the specificity improves, but this happens at the cost of the sensitivity
 - Similarly, if we decrease the threshold, the sensitivity improves, but at the cost of the specificity

Measures of test feasibility

- As a complement to the sensitivity and specificity, we can also calculate:
 - the Positive predictive value (PPV), which is the proportion of test positives who really DO have the disease, $PPV = \frac{a}{a+b}$
 - and the Negative predictive value (NPV), which is the proportion of test negatives who really DON'T have the disease $NPV = \frac{d}{c+d}$
- Importantly, these measures are functions of the test sensitivity and specificity, as well as the disease prevalence

Hemoglobin screening, cont.

		Colon cancer within 1 yr		
		+	-	
Hgb decline > 1.5mg/dL	+	37	1200	<ul style="list-style-type: none"> • $PPV = 37 / (37 + 1200) = 3\%$ • Thus, only 3% of those that test positively really have the disease • 97% of test positives will undergo an unnecessary colonoscopy
	-	33	5800	

Hemoglobin screening, cont.

		Colon cancer within 1 yr		<ul style="list-style-type: none"> • $NPV = 5800 / (33 + 5800) = 99.4\%$ • Thus, 99.4% of those that test negatively won't have the disease • Alternatively, 0.6% of those that test negatively, actually had the disease
		+	-	
Hgb decline > 15mg/dL	+	37	1200	
	-	33	5800	

Screening programs

- A screening program is the systematic application of a screening test to a population
- Screening programs are typically applied to whole populations (i.e. nationally or regionally)
- Examples of screening programs are:
 - Mammography screening for breast cancer
 - Pap smear (or HPV) screening for cervical cancer
 - PSA screening for prostate cancer

Evaluation of screening programs, cont

- Evaluation of screening programs should consider:
 - Actual gains in morbidity and mortality
 - The cost in relation to gains (i.e., cost efficiency)
 - The over diagnosis and subsequent over treatment
 - Screening test performance and other proxies (*although this is of lesser importance than the actual gains*)

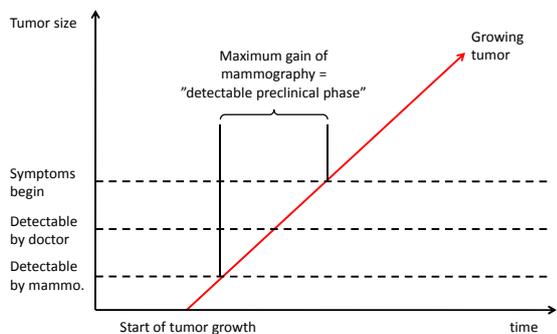
Evaluation of screening programs

- The ultimate goal of all screening programs is to reduce the morbidity (or mortality) associated with a given disease
- Proper evaluation of a screening program must thus address the morbidity or mortality gains associated with the screening program
- Thus, mere assessment of the screening performance (sensitivity, specificity, etc) is not sufficient to properly evaluate screening

Bias in screening program evaluation

- There are three main types of bias in screening program evaluation:
 - Lead-time bias
 - Length-biased sampling (length time bias)
 - Volunteer bias

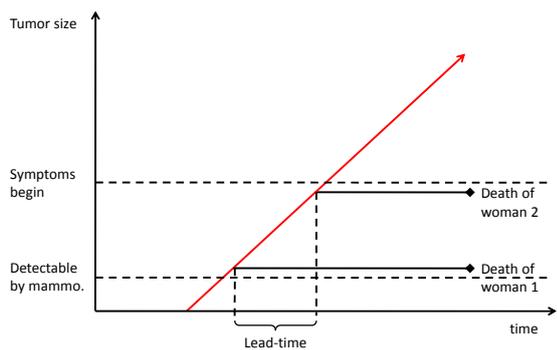
Example – breast cancer screening



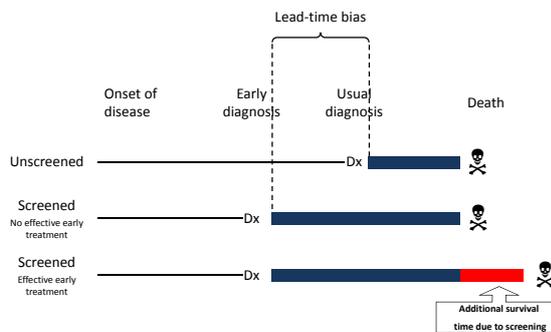
Lead-time bias

- In screening, the goal (and hopefully the result) is to detect disease earlier in the disease progression
- Therefore, in an observational study assessing screening, cases detected through screening will appear to have a superior survival than cases detected clinically

Lead-time bias, cont.



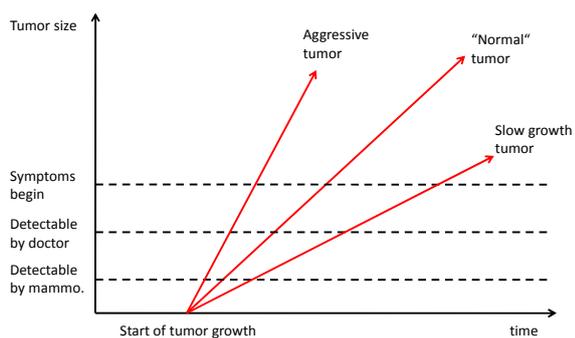
Lead-time bias, alternative explanation



Length-biased sampling

- Since all screening programs only screen participants at certain intervals (often several years), the probability of picking up pre-clinical disease depends on the aggressiveness of the disease (e.g. interval cases)
- Cases detected through screening are therefore often less aggressive and will have a better prognosis

Length-biased sampling



Volunteer bias

- Often, persons that choose to participate in screening are more health conscious than the average person
- Persons that volunteer for screening may therefore have both a better, and sometimes poorer prognosis
- The magnitude and direction of this bias may be difficult to assess

Ideal screening test

- High positive predictive value
- High sensitivity and, more importantly, high specificity
- Simple and safe to execute
- Low cost (e.g. mammograms vs. breast MRI)
- Acceptable to the screened population
- There exists efficient treatment

So, how should we evaluate screening?

- Evaluation of screening must account for lead-time bias, length-biased sampling and volunteer bias, all of which are difficult to handle
- The only generally accepted mode of screening evaluation is therefore RCTs, where healthy individuals are randomized to either screening or no screening

Summary screening

- Screening is unique in medicine in that tests are performed in asymptomatic persons
- The ultimate goal of screening programs is to lessen the burden of disease by:
 - Earlier diagnosis
 - Diagnosis before start of symptoms
 - Prevent spread (both locally, as in disease progression, and between subjects)

Summary screening, cont.

- Not all diseases are suitable for screening:
 1. The disease should have serious consequences
 2. The disease should be treatable and early treatment should improve prognosis
 3. There should exist a simple, harmless and valid screening test
 4. The prevalence of preclinical, asymptomatic disease should be sufficiently high in the screened population

Summary screening, cont.

- The principal measures of the performance of a screening test are:
 - Sensitivity = the proportion of those with the disease that are correctly classified as having the disease
 - Specificity = the proportion of those without the disease that are correctly classified as not having the disease
- Due to unique types of bias, screening programs are best evaluated with RCTs
