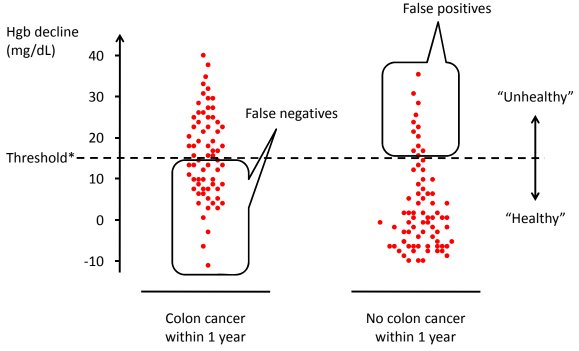
Screening and diagnostic tests

Gustaf Edgren, PhD

Outline

- Repetition of screening
  - Measures
  - Errors associated with screening
  - How to evaluate screening programs
- Advanced measures of diagnostic test performance
- Evaluation of screening <u>programs</u>
- Summary

Screening tests



*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

Sensitivity

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

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

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:

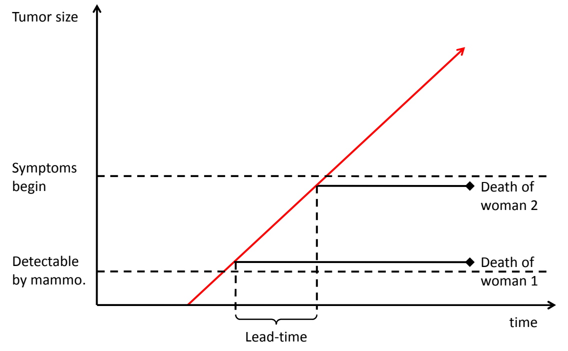
$$\boxed{Specificity = \frac{d}{b+d}}$$

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

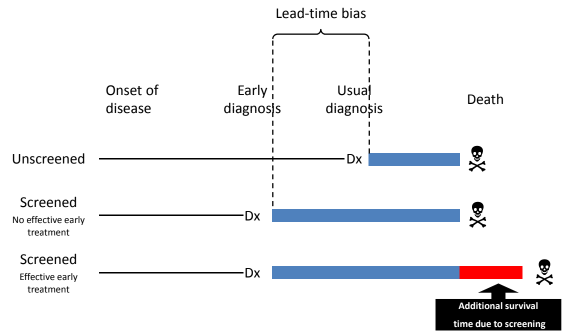
Lead-time bias

- In screening, the goal (and almost always 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.



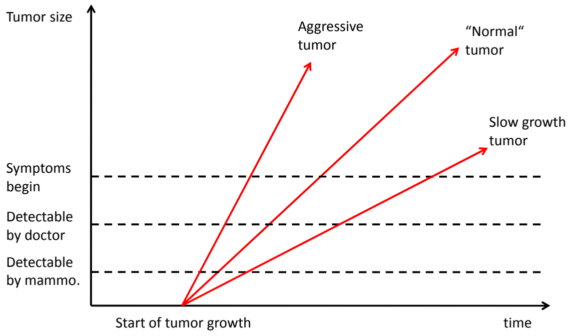
Lead-time bias, alternative explanation



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



Over-diagnosis bias

- In addition to the classic biases (lead-time, length-bias and volunteer bias), observational studies are also susceptible to over-diagnosis bias:
  - For some conditions, the natural course of illness is often difficult to predict
  - For prostate cancer, as an example, there is considerable clinical heterogeneity and a paucity of methods for prognostication

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)
- Due to unique types of bias, screening programs are best evaluated with RCTs

Summary screening, cont.

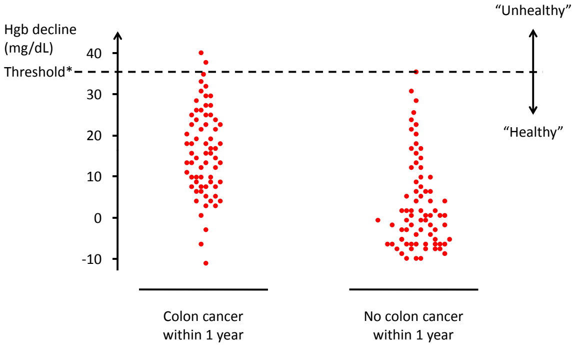
- Not all diseases are suitable for screening:

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

Advanced test performance measures

- In most screening tests, a continuous variable is measured and dichotomized into “sick” / “not sick” using a “criterion of positivity”
- The choice of threshold may seem arbitrary, but it can be optimized depending on the application of the test
- A common tool is the receiver operating characteristics (ROC) curve

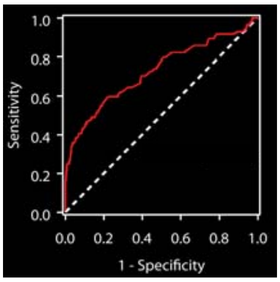
ROC curves



*Criterion of positivity

ROC curves (2)

- So, what can we do with all these sensitivity/specificity values?
- Plot them of course!

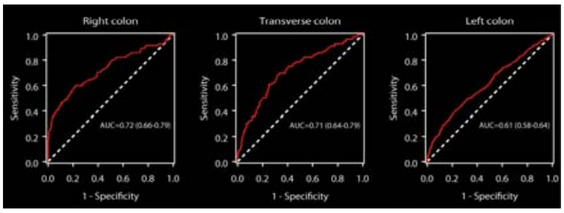


Comparing ROC curves

- The ROC curve is a graphical representation of what possible sensitivity/specificity values can be achieved with a certain test
- For each ROC curve, the area under the curve (AUC) can be estimated
- The AUC value can take any value between 0 and 1 (higher is better) and gives a summary measure of the test performance

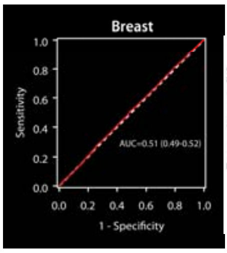
Comparing ROC curves

- The ROC curve is a graphical representation of what possible sensitivity/specificity values can be achieved with a certain test

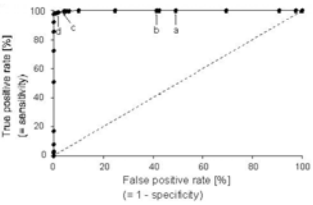


Comparing ROC curves!

- Is hemoglobin changes a good test for breast cancer diagnostics?



HIV PCR test for reference, AUC=0.994



Numbers needed to screen

- A classic measure of the successfulness of a screening program is numbers needed to screen (NNS)
- NNS is an analogue to NNT – Numbers needed to treat
- It is calculated as 1/ARR (=absolute risk reduction)
- NNS tells us how many individuals we need to screen to prevent one death (or whatever the outcome is)

Cost efficiency

- Cost efficiency calculations of screening notoriously difficult as they require factoring in of “costs” on many levels:
  - Cost of the screening program itself
  - Cost of “unnecessary” investigations and treatments*
  - Burden of false alarms
  - Burden of false reassurance
  - Pain and suffering

Screening RCT:s

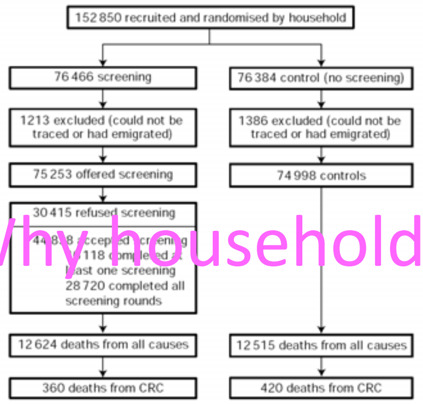
- We’ve already concluded that RCT:s are typically the study design of choice for screening program evaluation
- But, how would one design such a trial?

Pointers for screening RCT:s

- Depending of the expected gain of the screening, randomization can be on an individual level or in natural clusters
- In fact, in some cases, natural clusters can be necessary to promote acceptance
- Design the study for a HARD outcome: i.e. death*, or death from a certain cause
- Recall the ethical dilemmas of offering something only to one group – equipoise!

Screening RCT:s – limitations

- The screening RCT is typically an enormous undertaking, but ensures a high validity
- There are still some caveats, however:
  - If the prevalence of the screened disease is low, the trials have to be VERY large
  - If the disease of interest attracts public attention, your comparison group may be heavily polluted* and dilute the effects of the screening
  - While no threat to the internal validity, volunteer bias may limit the external validity**


152 850 recruited and randomised by household
76 466 screening
76 384 control (no screening)
1213 excluded (could not be traced or had emigrated)
1386 excluded (could not be traced or had emigrated)
75 253 offered screening
74 998 controls
30 415 refused screening
44 838 accepted screening
[illegible] 118 completed at least one screening
28 720 completed all screening rounds
12 624 deaths from all causes
12 515 deaths from all causes
360 deaths from CRC
420 deaths from CRC

Why households?

Hardcastle et al, Lancet, 1996