Diagnostic accuracy in screening trials

Diagnostic screening trials

Screening trials refer to situation with

- Low prevalence of disease
- Subjects are not suspicious to have disease (healthy subjects), and if patients have disease it is mostly at low stage.
- Diagnostic accuracy criterion (goldstandard) is often not available for test negatives (verification bias).
- Screening trials are performed as cohort studies (inclusion criterion: test requested, true disease status is yielded by reference method afterwards, in difference to case-control study: true disease status as obtained by reference method is inclusion criterion)

Measures of diagnostic accuracy

Diagnostic accuracy of a diagnostic test can be obtained from following 2x2 table (gray cells) (D – disease, T – result of test) and is expressed by a pair of measures sensitivity (sens) and specificity (spec) or positive predictive value (PPV) and negative predictive value (NPV). As we will see later, predictive values are appropriate measures for reporting diagnostic accuracy for diagnostic screening trial, whereby PPV is of main interest; however it is a must to report **the pair** of measures.

	Disease	
Test result	D_+ Patients with disease	D_- Patients without disease
T+ Patients with positive test results	TP True positive	FP False positive
T- Patients with negative test results	FN False negative	TN True negative

Diagnostic accuracy measures are yielded by following formulas[1].

Sensitivity	Probability of a true pos. test result given disease	TP/(TP+FN)
Specificity	Probability of a true neg. test result given non-disease	TN/(TN+FP)
PPV	Probability of a true pos. test result given pos. test result	TP/(TP+FP)
NPV	Probability of a true neg. test result given neg. test result	TN/(TN+FN)

In addition, knowledge of prevalence (probability of disease) is of interest.

Prevalence	Probability of disease	(TP+FN)/N

[1] Please note, that measures of diagnostic accuracy should be reported together with their confidence intervals. In order to achieve simplicity of this report, they are not presented here. You can use following tool for calculation: http://www.acomed-statistik.de/sensitivity_specificity_PPV_NPV_diagnostic_test_2x2_table.xls

Interpretation of measures of diagnostic accuracy

Sens+spec:

- Measures to describe test performance from view of diagnostic companies etc.
- Can be derived from case control as well as cohort studies.
- Do not depend on prevalence but on spectrum of diseases (this is the reason why case control studies often overestimate diagnostic accuracy, because severe diseases as well as "clean" healthy are overrepresented in this design).
- Are calculated within **columns** of 2x2 table.

PPV and NPV:

- Measures to describe test performance from point of view of physician as well as patient (PPV: If there is a positive test result, what is the probability that the patient has the disease?) Because this is the clinical application situation, predictive values are of major importance when diagnostic accuracy of screening test is reported.
- Depend on prevalence
- Can only be achieved within cohort studies
- Are calculated within **rows** of 2x2 table

Fictive Example

A diagnostic company reports a screening trial investigating a new screening test for cancer with 8000 patients whereby 50 patients have disease. 35 of those where detected by the screening test (sensitivity: 70%). Specificity is 90%. In addition, a high negative predictive value of 99.8% is reported as a highlight of the test. The company concludes that the test evaluation was successful, and that the test can be introduced as a screening test.

However, is this the correct conclusion? Let's fill the 2x2-table (presented together with measures of diagnostic accuracy) by given numbers, first:

	Disease		
Test result	$D_+ = 50$	$D_- = ?$	PPV, NPV
T+	TP=35	FP=?	PPV=?
T-	FN=15	TN=?	NPV=?
Sens, spec	Sens=70.0%	Spec=90.0%	N=8000

Then we can calculate:

D- =	$N - D_+ =$	8000 – 50 =	7950
TN =	spec · TN =	0.9 · TN =	7155
FP =	D_- - TN =	7950 – 7155 =	795
PPV =	**TP/(TP+FP) =**	**35 / (35 + 795) =**	**0.042 → 4.2%**
NPV =	TN/(TN+FN) =	7155/(15+7155) =	0.9979 → 99.8%
Prevalence =	$D_+/N =$	50/8000 =	0.00635 → 0.62%

Following 2x2 table results:

	Disease		
Test result	$D_+ = 50$	$D_- = 7950$	PPV, NPV
T+	TP=35	FP = 795	**PPV=4.2%**
T-	FN=15	TN = 7155	NPV=99.8%
Sens, spec	Sens=70.0%	Spec=90.0%	N=8000

Conclusions:

A PPV of 4.2% means, that among 100 patients with a positive test result, only 4 have the disease. This is a worse performance, and suitability of the diagnostic test as a screening test is in question. However, the test might be applicable in other situations (e.g. high risk groups, population with elevated prevalence).

Reporting only sens and spec hides true performance of the test. In addition, NPV (here: 99.8%) near 100%-prevalence (here 99.38%) does not indicate a good diagnostic performance of a diagnostic test in screening situation.

A diagnostic test with a false positive rate (=100%-specificity, here: 10.0%) considerably larger than prevalence (here: 0.62%) is not suitable as a screening test.