Meta-analysis of diagnostic studies

Khalid Ashfaq

Technical Analyst, NICE

Outline

- Diagnostic studies – basics
- Formulating diagnostic questions
- Diagnostic outcome measures
- When meta-analysis can be done
- Plots and figures
- Advanced techniques

Diagnostic studies – re-cap

- Effective medical treatment usually depends on accurately diagnosing a patient's condition
  - Tests can take many forms
- The purpose of a diagnostic test is to perform a specific function, in a specific population that is believed to have a specific condition

Uses of diagnostic tests

- Diagnostic tests subserve one of 5 functions
  - Screening test
  - Routine test
  - Test used to establish a diagnosis
  - Staging test
  - Monitoring test

Formulating diagnostic questions

- Similar to the well known PICO model
  - **Population**
  - **Index test**
  - **Reference test**
  - **Outcome**

The 2 x 2 table

- For each study:

	Reference test +ve (disease/condition present)	**Reference test –ve (disease or condition absent)**
Index test +ve	True positive (TP)	False positive (FP)
Index test -ve	False negative (FN)	True negative (TN)

True positives = known patients

False negatives = ‘stealth’ patients

False positives = stigmatised people

True negatives = relieved people

Summary measures

- Paired (and related) measures
  - Sensitivity & specificity
  - PPV & NPV
  - Likelihood ratios (for positive & negative test)
- Single measures
  - Diagnostic odd ratios (DORs)
  - Area under the curve (AUC) of summary Receiver Operating Characteristics (sROC) plots

Summary measures (contd)

	Reference test +ve	**Reference test -ve**	Total
Index test +ve	TP	FP	TP + FP
Index test -ve	FN	TN	FN + TN
Total	TP + FN	FP + TN	TP+FP+FN+TN

PPV = TP/TP + FP

NPV = TN/FN + TN

Sensitivity
TP/TP + FN

Specificity
TN/FP + TN

LR for positive test result = [TP/(TP+FN)]/[FP/(FP+TN)] = sensitivity/(1-specificity)

LR for negative test result = [FN/(FN+TP)]/[TN/(TN+FP)] = (1-sensitivity)/specificity

What do they really mean?

- For example, sensitivity = 95% and Specificity = 99%.

  Sensitivity of 95% means that of those with disease, 95% tested positive and 5% tested negative (i.e. 5% had a false negative result).

  Specificity of 99% means that of those without disease, 99% tested negative and 1% tested positive (i.e. 1% had a false positive result).

- For example, PPV = 93% and NPV = 90%

  PPV of 83% means that 83% of those with a positive test have the disease/condition

  NPV of 90% means that 90% of those with a negative test result do not have disease

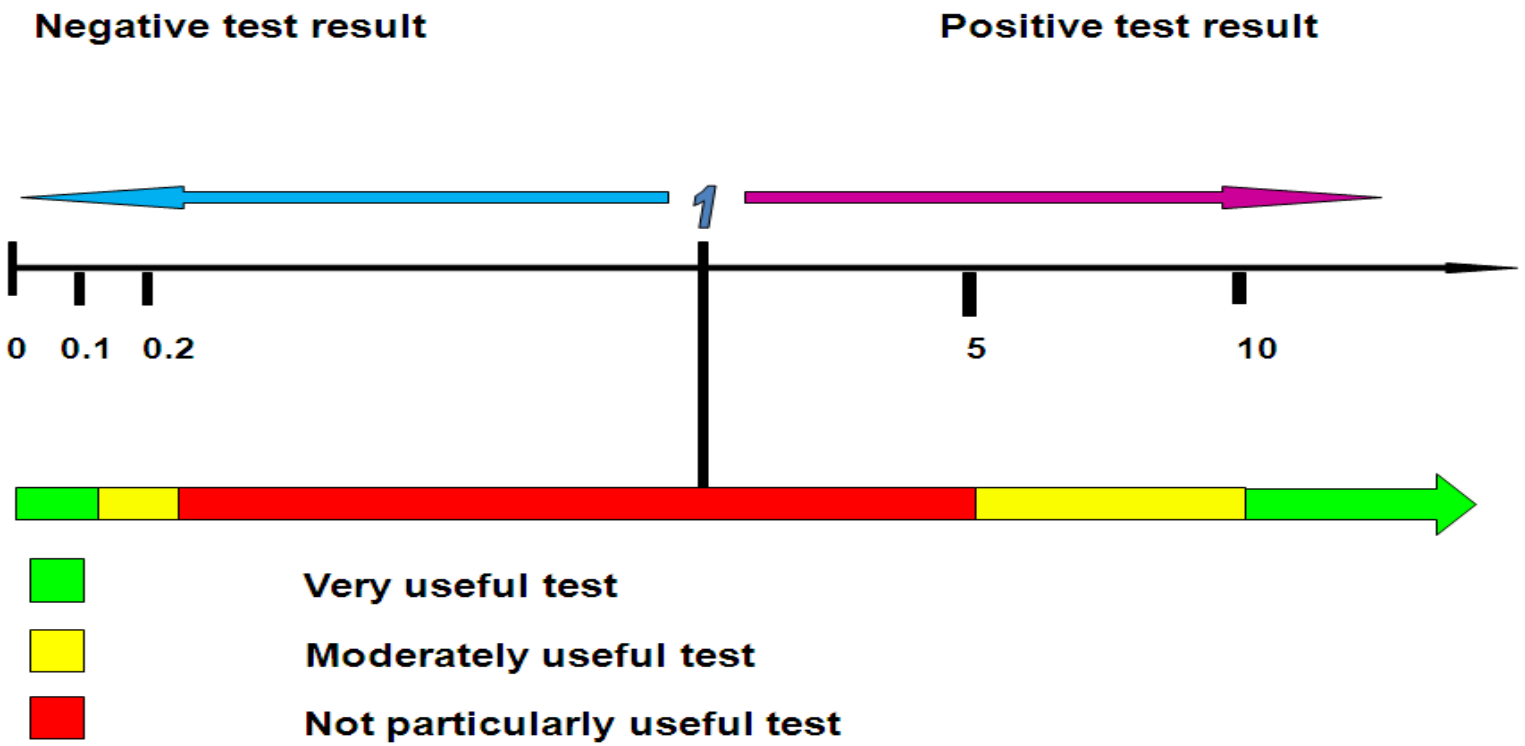
- Note that
  - A sensitive test is a good *'rule out'* test *(Snout)*
    i.e. a negative result of a sensitive test is likely to indicate no disease (low FN rate)
  - A specific test is a good *'rule in'* test *(Spin)*
    i.e. a positive test result of a specific test is likely to indicate disease (low FP rate)
  - predictive values vary highly with disease prevalence therefore should not be pooled in a meta-analysis

Likelihood ratios

- Positive and negative likelihood ratios describe the discriminatory properties of positive and negative tests
- How much more likely particular test results are in patients with disease than in those without disease
  - LR+ >10 and LR- <0.1 conclusive evidence
  - LR+ 5-10 and LR- 0.1-0.2 strong diagnostic evidence
  - LR+ 2-5 and LR- 0.2-0.5 weak diagnostic evidence
  - LR+ 1-2 and LR- 0.5-1 negligible evidence

(Jaeschke 1994)

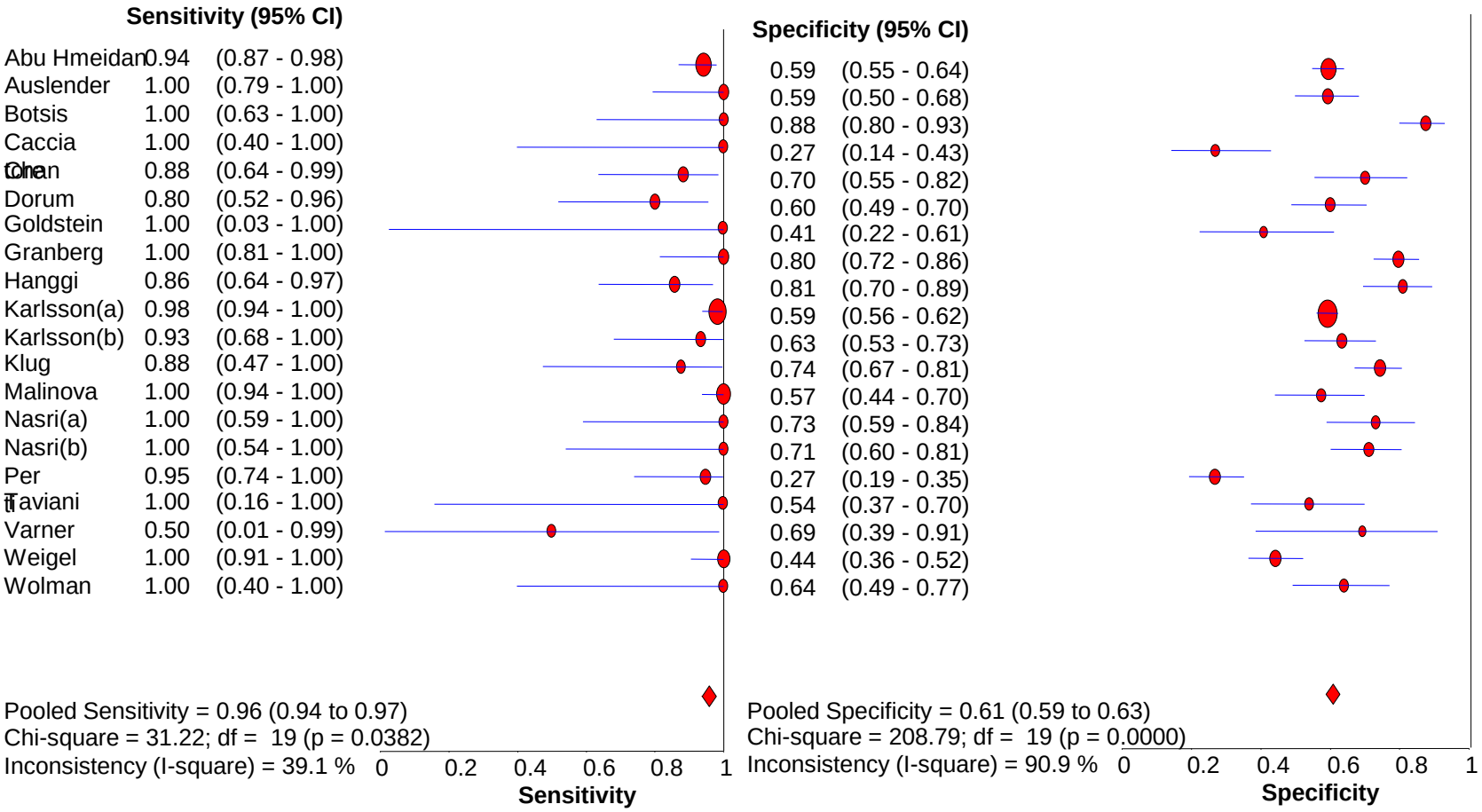
Interpreting Likelihood ratios



Meta-analysis?

- 2 or more clinically similar studies
  - spectrum
  - thresholds
- Fixed effects model vs random effects model
  - FEM - all studies detect the underlying common effect
  - REM - heterogeneity between studies
  - If presence of clinical or statistical heterogeneity - use REM
  - If in doubt - use REM

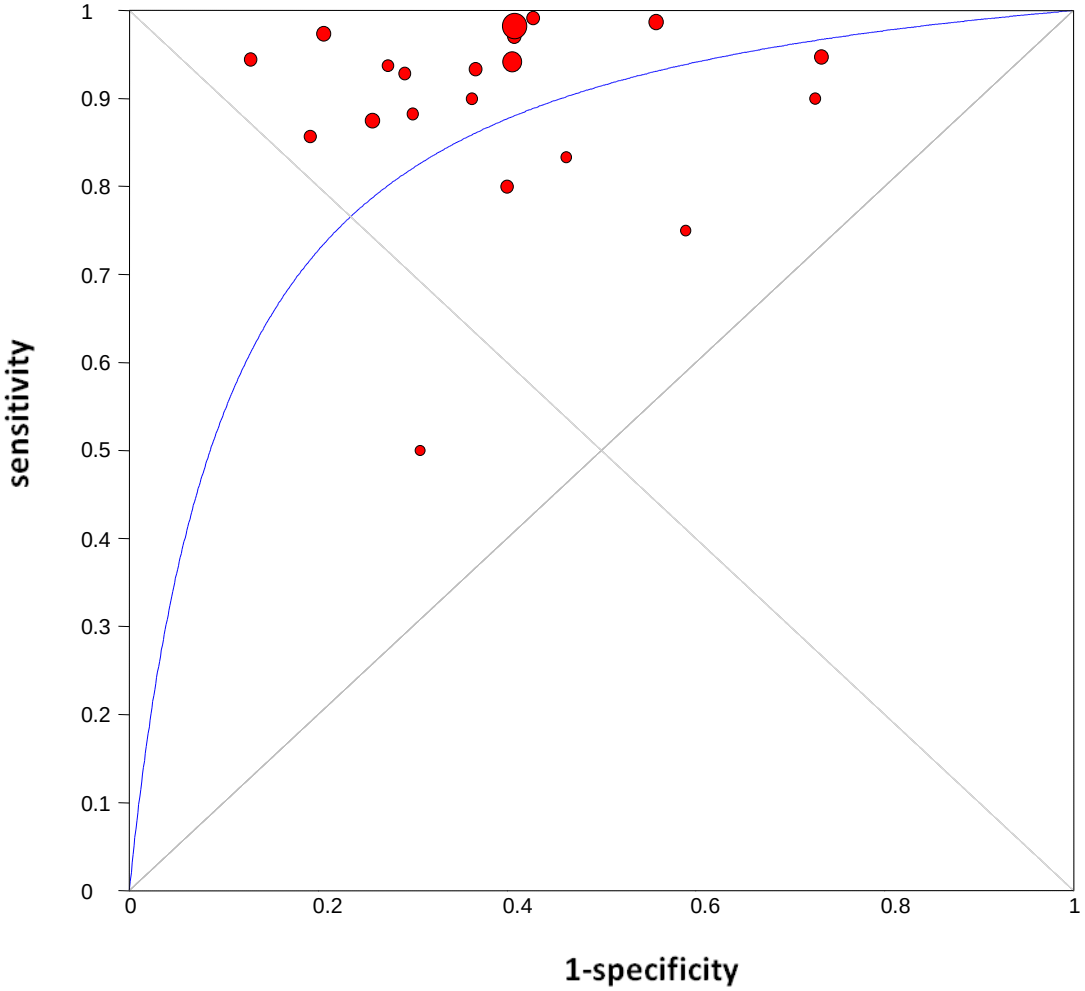
Forest plot


Sensitivity (95% CI)
Abu Hmeidan 0.94 (0.87 - 0.98)
Auslender 1.00 (0.79 - 1.00)
Botsis 1.00 (0.63 - 1.00)
Caccia 1.00 (0.40 - 1.00)
Chan 0.88 (0.64 - 0.99)
Dorum 0.80 (0.52 - 0.96)
Goldstein 1.00 (0.03 - 1.00)
Granberg 1.00 (0.81 - 1.00)
Hanggi 0.86 (0.64 - 0.97)
Karlsson(a) 0.98 (0.94 - 1.00)
Karlsson(b) 0.93 (0.68 - 1.00)
Klug 0.88 (0.47 - 1.00)
Malinova 1.00 (0.94 - 1.00)
Nasri(a) 1.00 (0.59 - 1.00)
Nasri(b) 1.00 (0.54 - 1.00)
Per 0.95 (0.74 - 1.00)
Taviani 1.00 (0.16 - 1.00)
Varner 0.50 (0.01 - 0.99)
Weigel 1.00 (0.91 - 1.00)
Wolman 1.00 (0.40 - 1.00)
Pooled Sensitivity = 0.96 (0.94 to 0.97)
Chi-square = 31.22; df = 19 (p = 0.0382)
Inconsistency (I-square) = 39.1 %
0
0.2
0.4
0.6
0.8
1
Sensitivity
Specificity (95% CI)
0.59 (0.55 - 0.64)
0.59 (0.50 - 0.68)
0.88 (0.80 - 0.93)
0.27 (0.14 - 0.43)
0.70 (0.55 - 0.82)
0.60 (0.49 - 0.70)
0.41 (0.22 - 0.61)
0.80 (0.72 - 0.86)
0.81 (0.70 - 0.89)
0.59 (0.56 - 0.62)
0.63 (0.53 - 0.73)
0.74 (0.67 - 0.81)
0.57 (0.44 - 0.70)
0.73 (0.59 - 0.84)
0.71 (0.60 - 0.81)
0.27 (0.19 - 0.35)
0.54 (0.37 - 0.70)
0.69 (0.39 - 0.91)
0.44 (0.36 - 0.52)
0.64 (0.49 - 0.77)
Pooled Specificity = 0.61 (0.59 to 0.63)
Chi-square = 208.79; df = 19 (p = 0.0000)
Inconsistency (I-square) = 90.9 %
0
0.2
0.4
0.6
0.8
1
Specificity

Heterogeneity

- Heterogeneity is present if:
  - Chi square $p < 0.05$
  - $I^2 > 33\%$
- $I^2 > 66\%$ = significant heterogeneity
- Sources of heterogeneity:
  - Study design and conduct
  - Population spectrum
  - Test technology/execution
  - Chance

ROC Curve



Area Under the Curve (AUC)

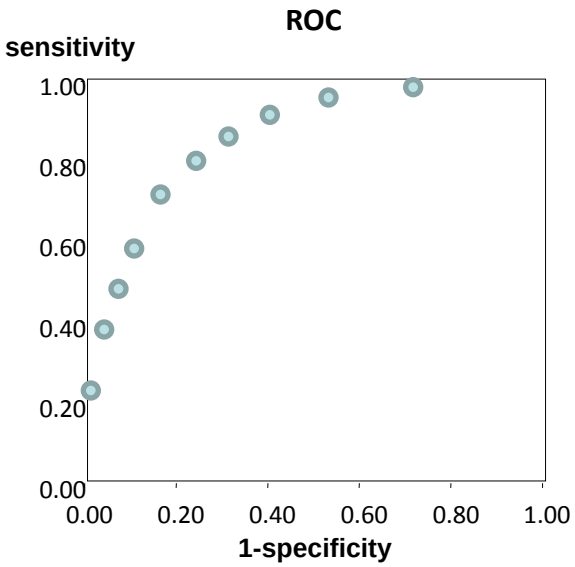
Ranges between 0.5 - 1

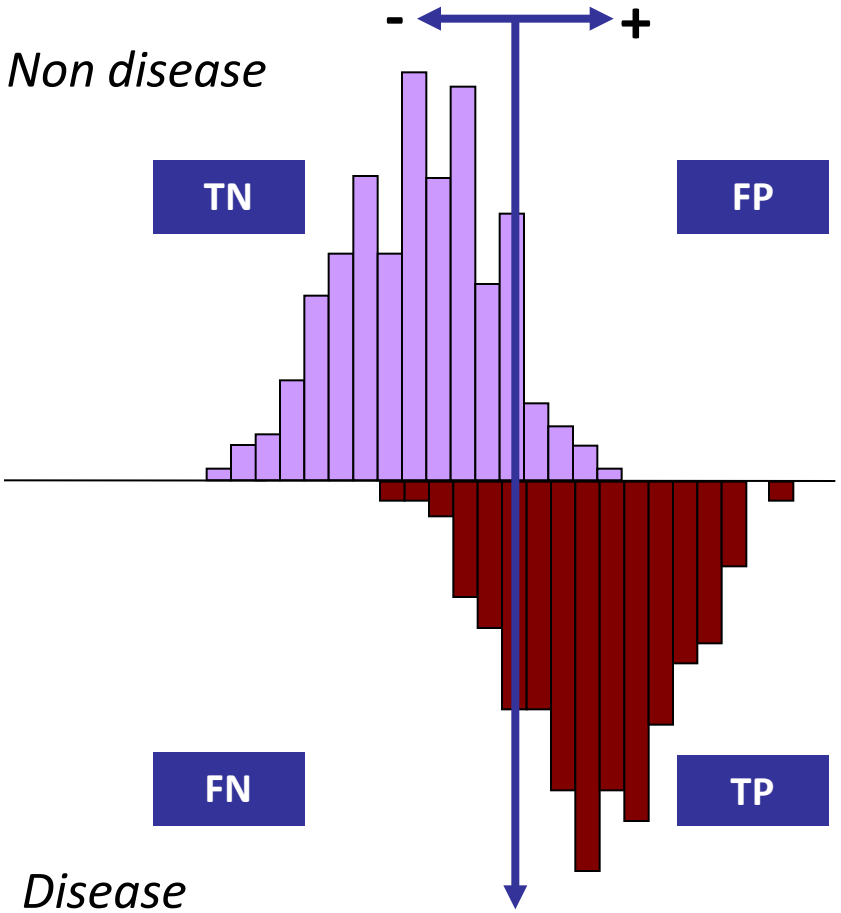
The greater the better

Unuseful test AUC = 0.5

Perfect test AUC = 1

Summary ROC plot





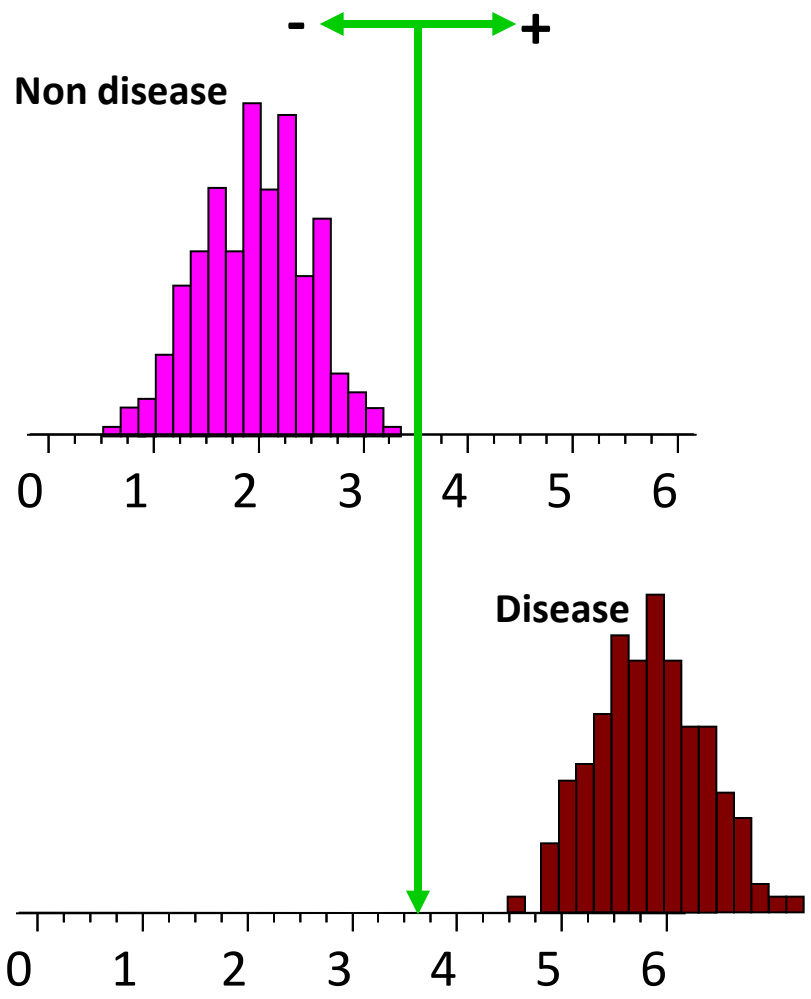
Statistical packages

- Simple to use and freely available
  - Meta-DiSc
  - RevMan
  - Meta-Analyst
- Packages requiring more skills
  - R
  - SAS
  - STATA
  - Win-bugs

Pooling sensitivity and specificity

- Sensitivity and specificity are
  - related to each other
  - affected by spectrum, but not by prevalence
- Pooling should only be done in the absence of the threshold effect
  - ROC plot
  - Tests of heterogeneity
  - Spearman correlation
- Independent pooling of sensitivity and specificity may underestimate accuracy (Shapiro 1995)

Threshold effect: What does it mean?



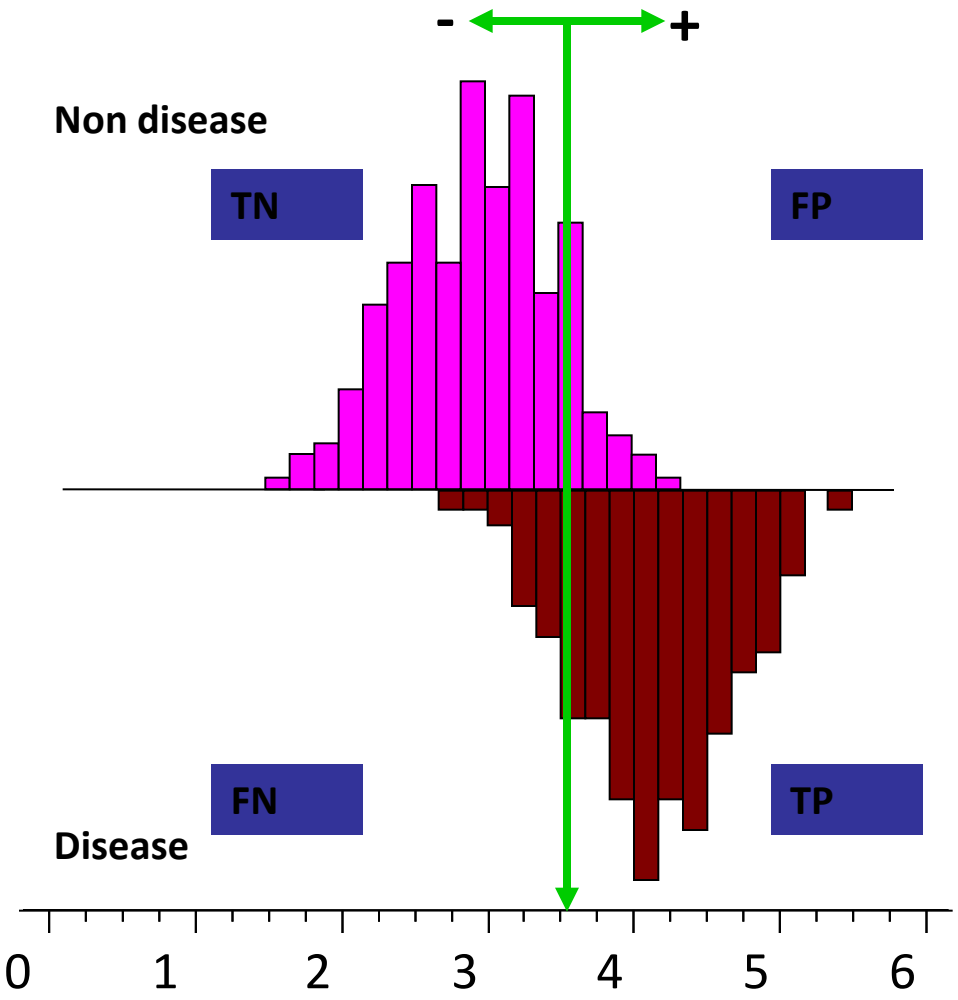
Threshold effect: What does it mean?

sen = TP / (TP + FN)

spec = TN / (TN + FP)

↓ sen = TP / (TP + **FN**)

↑ spec = **TN** / (TN + FP)



Advanced techniques

- Bivariate random effects meta-analysis
- Hierarchical summary ROC curves (HSROC)
  - allow for the relationship between sensitivity and specificity
  - recommended in by Cochrane
  - Can be performed in R, SAS, Stata, Win-bugs
  - Bivariate output from these programs can be input into RevMan to produce summary statistics and figures

In summary

- Carry-out meta-analysis with 2 or more clinically similar diagnostic studies (population and test)
- Decide on summary measures to be pooled – usually sensitivity & specificity, also LRs
- Be aware of sources of heterogeneity – investigate and consider subgroup analysis
- Consider bivariate meta-analysis

If choosing one summary statistic rather than another can even occasionally affect the clinical judgement of physicians reading a published article, then scrupulous attention must be paid to the use of summary statistics in the medical literature

Furrow, Taylor & Arnold 1992

Thanks for listening

Questions/discussion