
# Meta-regression in Stata: `metareg`

2010 UK Stata Users Group Meeting
10 September, LSHTM, London

Roger Harbord

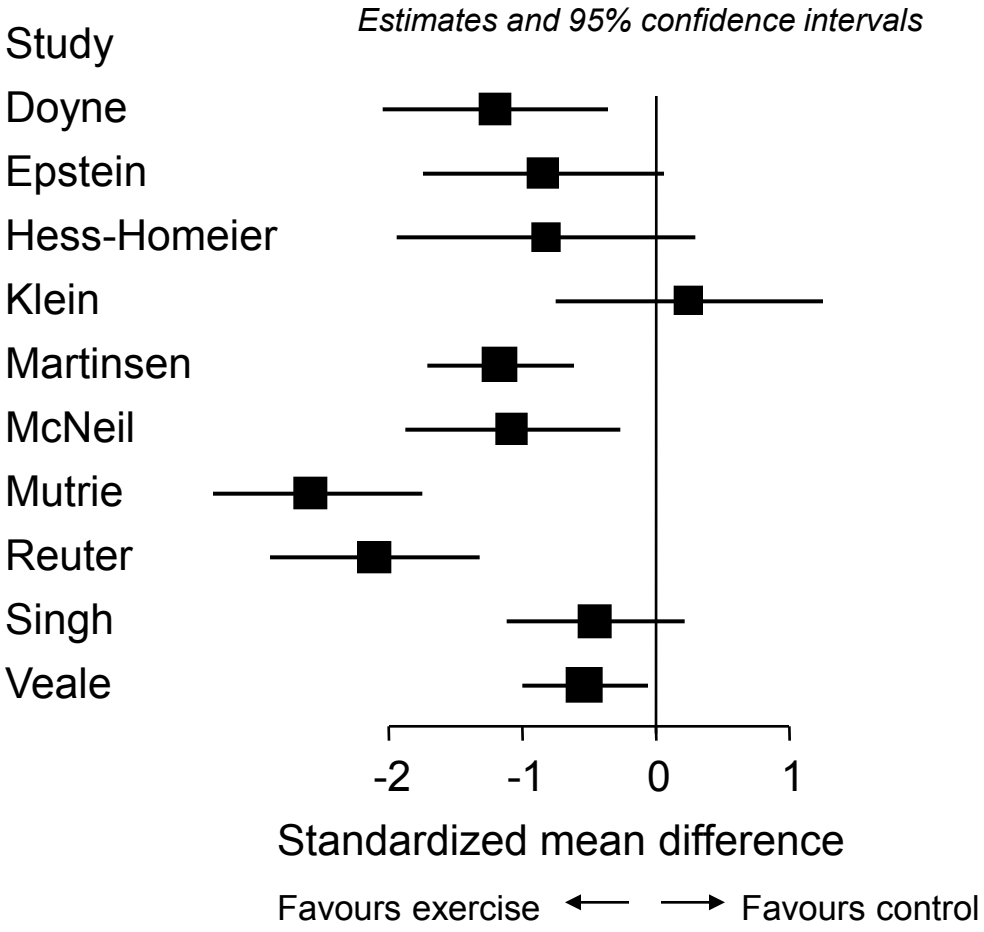
# Acknowledgements


- Stephen Sharp (MRC Epidemiology Unit, Cambridge, UK)

- Julian Higgins
- Simon Thompson

  (MRC Biostatistics Unit, Cambridge, UK)

# Example: exercise for depression



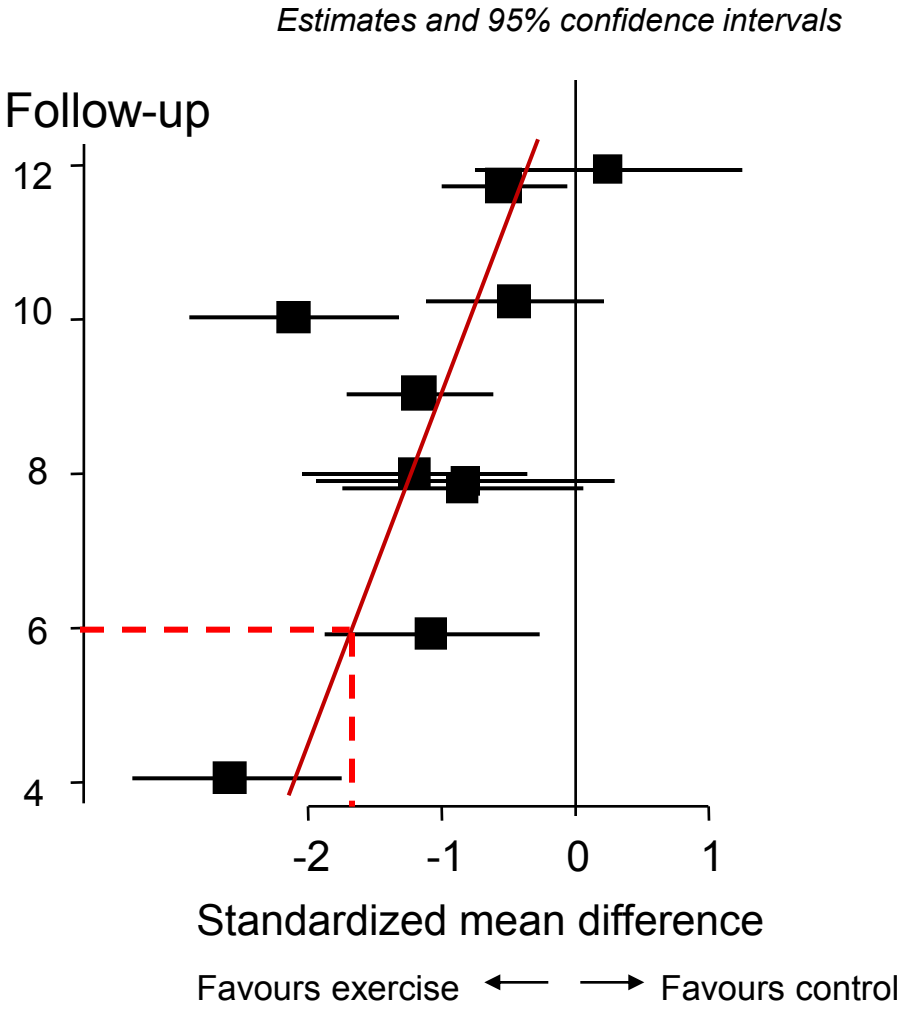

Lawlor DA, Hopker SW.
*BMJ* 2001; 322: 763-7

# Meta-regression to compare subgroups

- Assumes the between-study variance $\tau^2$ is the *same* in all subgroups
  - Sensible when some or all subgroups have few studies
- Estimates the difference in treatment effect between subgroups
- Example: Long duration *vs.* short duration
  Difference in SMD = 0.5 (95%CI: –0.5 to 1.5), $p = 0.32$
  - longer duration trials have a *less negative* SMD
  - i.e. treatment effect is *smaller* in long duration trials
- Weak statistical evidence for this being a true effect
  - but dichotomization reduces statistical power

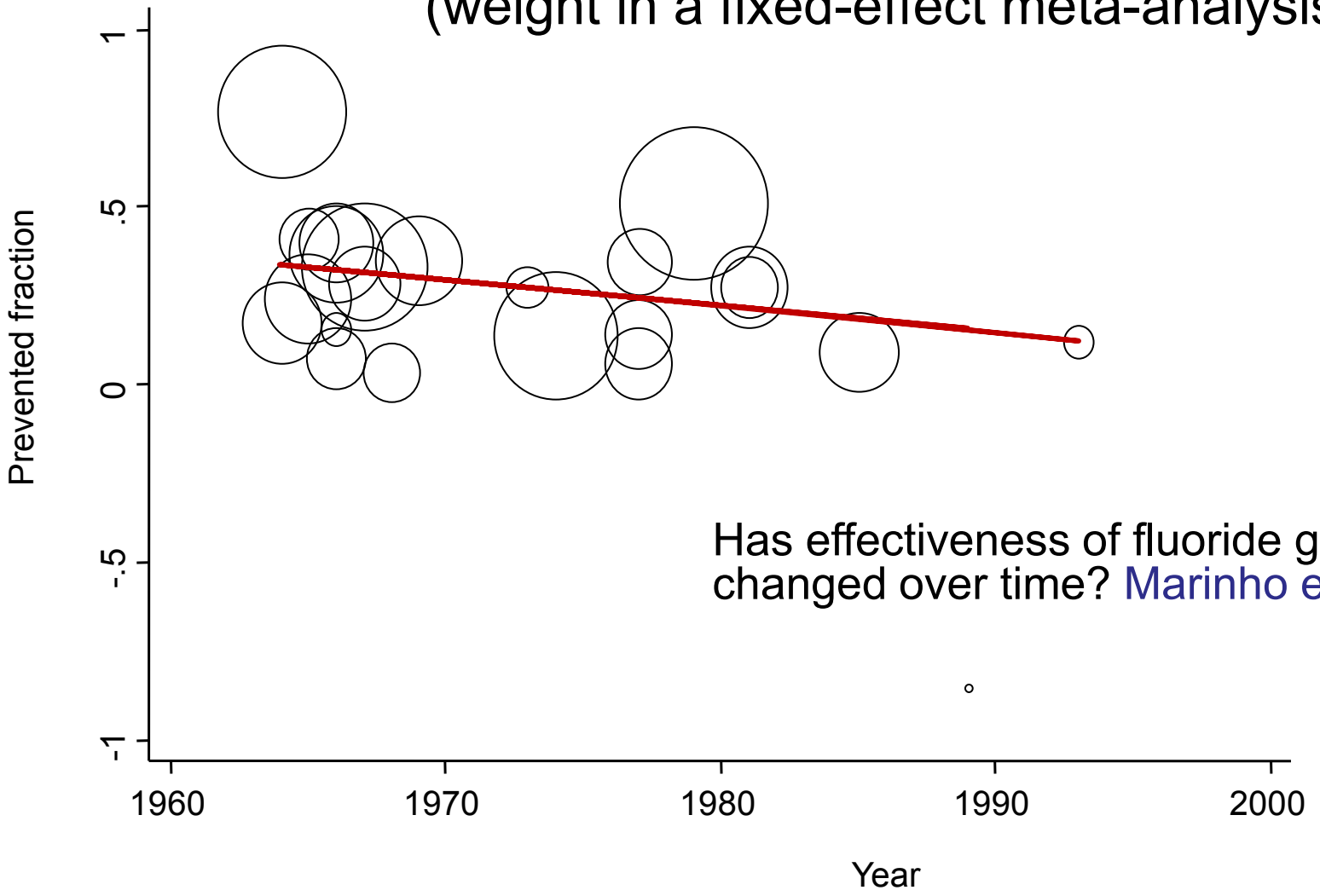
# Meta-regression with a continuous study characteristic

- Predict effect according to length of follow-up
  - SMD decreases by 0.18 (95%CI: 0.02 to 0.34) for each extra week of treatment
  - ($p$ = 0.008)



# ‘Bubble plots’

Circle sizes vary with inverse of within-study variance
(weight in a fixed-effect meta-analysis)




Has effectiveness of fluoride gels changed over time? Marinho et al (2004)

# Problems and pitfalls

- Choice of explanatory variables and spurious findings
- Confounding by study-level characteristics
- Lack of power
- Aggregation bias (ecological bias)

# Further reading

For more on planning and interpretation of meta-regression analyses see:

- Thompson S & Higgins JPT *Statistics in Medicine* 2002
- Higgins & Thompson *Statistics in Medicine* 2004
- Thompson & Higgins *The Lancet* 2005

# Meta-regression in Stata

- **`metareg`** was originally written by Stephen Sharp in 1998 (sbe23 in STB 42)
- Rewritten by me from 2004 onwards
- To install, type **`findit metareg`** in Stata
- For more explanation and discussion, see Harbord & Higgins *Stata Journal* 2008*;* **8**(4):493-519

# My enhancements to `metareg`

- Improved algorithm for the estimation of the between-study variance, $\tau^2$, by residual max. likelihood (REML)
- Modification to the calculation of SEs, p-values, and Cis for coefficients suggested by Knapp and Hartung (2003)
- Various enhancements to the output
- Optional graph of the fitted model with a single covariate
- Option to calculate permutation-based p-values, including an adjustment for multiple testing based on the work of Higgins and Thompson (2004)
- Support for many of Stata's postestimation commands, including `predict`

# Possible future extensions

- I consider `metareg` to be fairly mature, but there are a couple of possible extensions:
- Restricted iterative generalized least squares (RIGLS) is equivalent to REML (Goldstein *Biometrika* 1989) — option for when REML fails?
- Hedges, Tipton & Johnson (*Research Synthesis Methods* 2010) investigate use of cluster-robust variance estimation for dependent study results.

University of BRISTOL

# Meta-analysis of diagnostic test accuracy studies: `metandi` & `midas`

2010 UK Stata Users Group Meeting
10 September, LSHTM, London

Roger Harbord

# Acknowledgements

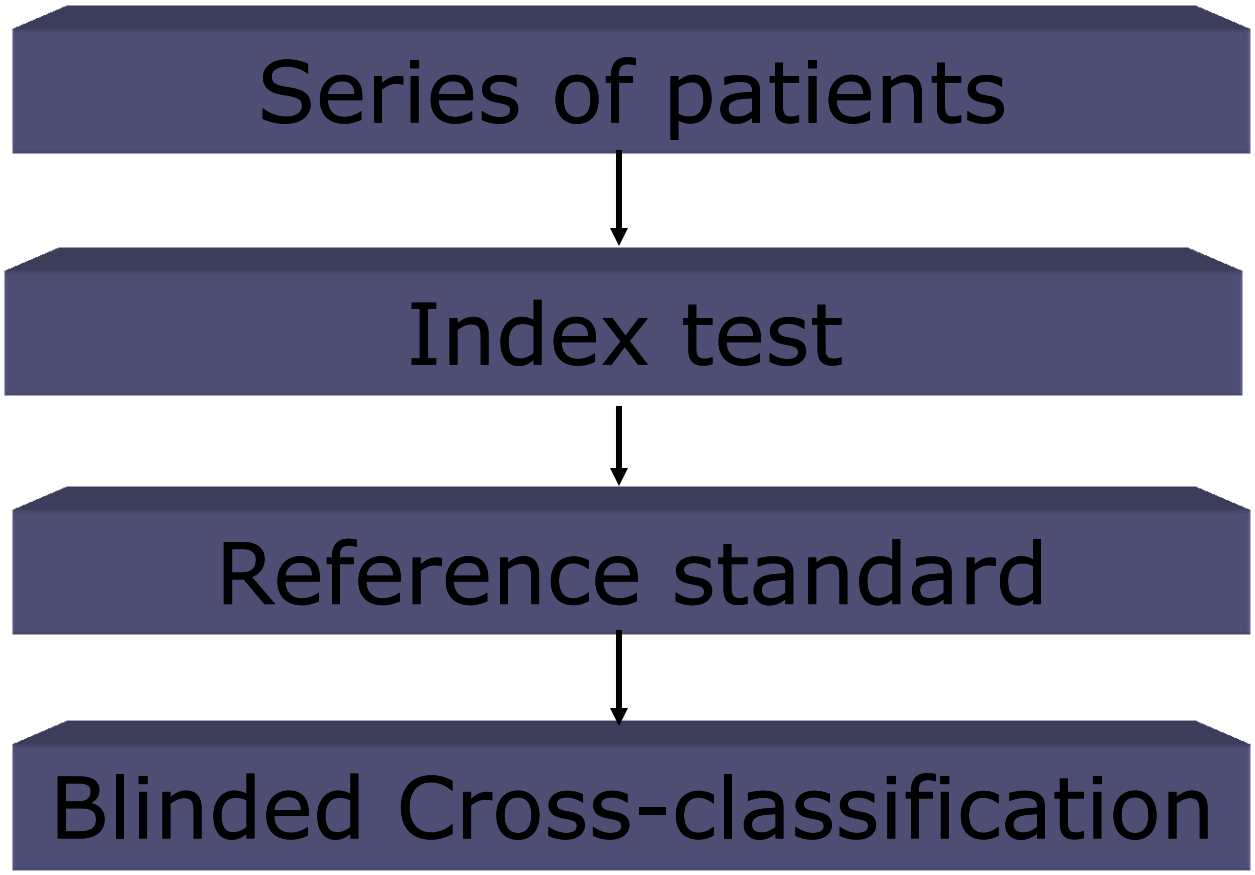

- Penny Whiting
- Jonathan Sterne
- MRC (grant no. G0801405)

- Petra Macaskill
- Constantine Gatsonis
- Jon Deeks
- Joseph Coveney
- Sophia Rabe-Hesketh
- Susan Mallett

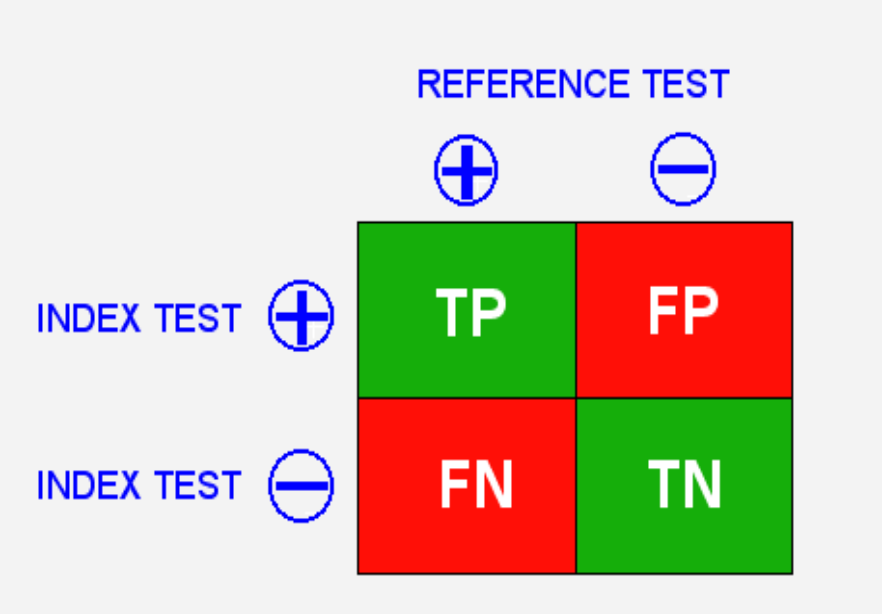
# Outline

- Diagnostic test accuracy (DTA)
  - Study design
  - Measures
- Methods for meta-analysis of DTA
  - Simple methods
  - Hierarchical models
- Software for meta-analysis of DTA
  - Stata: `metandi` and `midas`

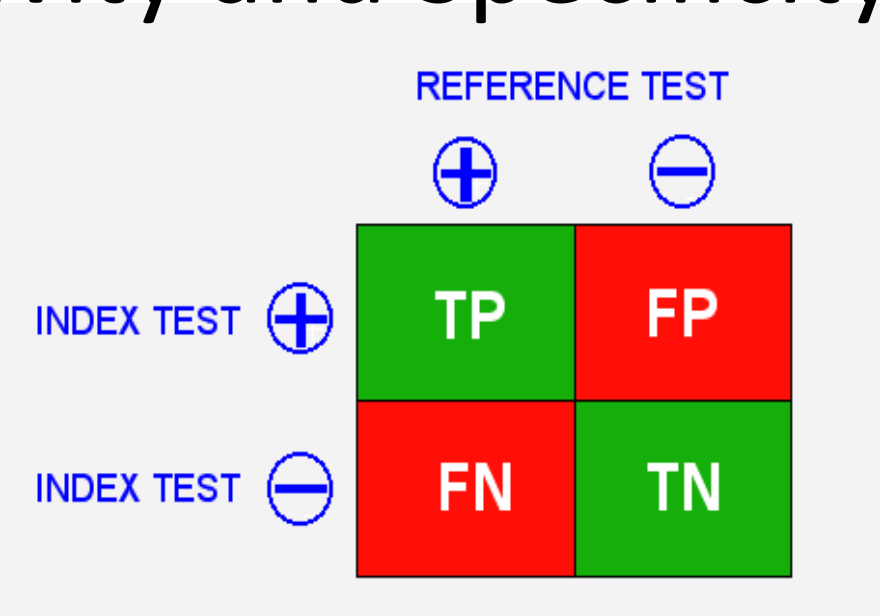
# Basic Test Accuracy Study



# Measures of diagnostic accuracy



# Sensitivity and Specificity



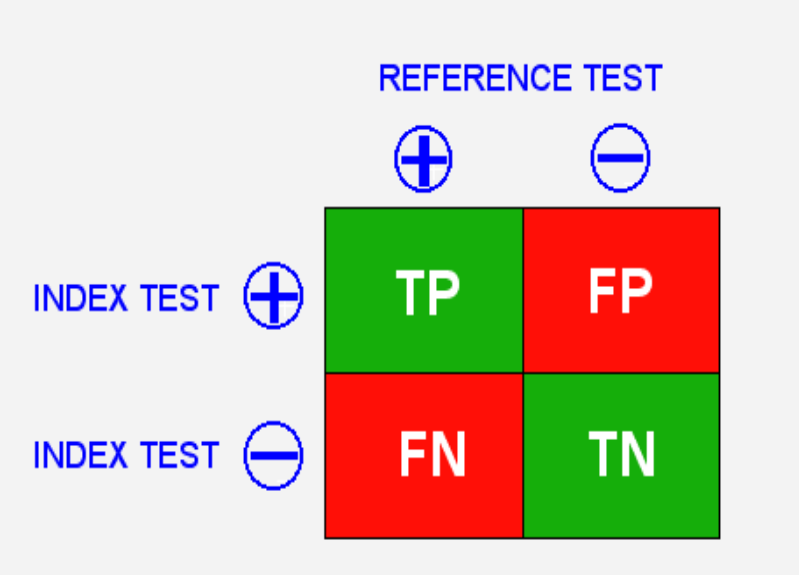

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{a}{a + c}$$

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

# Sensitivity and Specificity

- Computed along the columns of the 2x2 table:
  - The proportion of those with the condition who have positive test results
  - The proportion of those without the condition who have negative test results
- Clinically not directly useable
- In theory not influenced by the prevalence of the target condition – in practice they are
- Will often vary across populations due to differences in patient spectrum
- Values depend on choice of thresholds

# Likelihood ratios



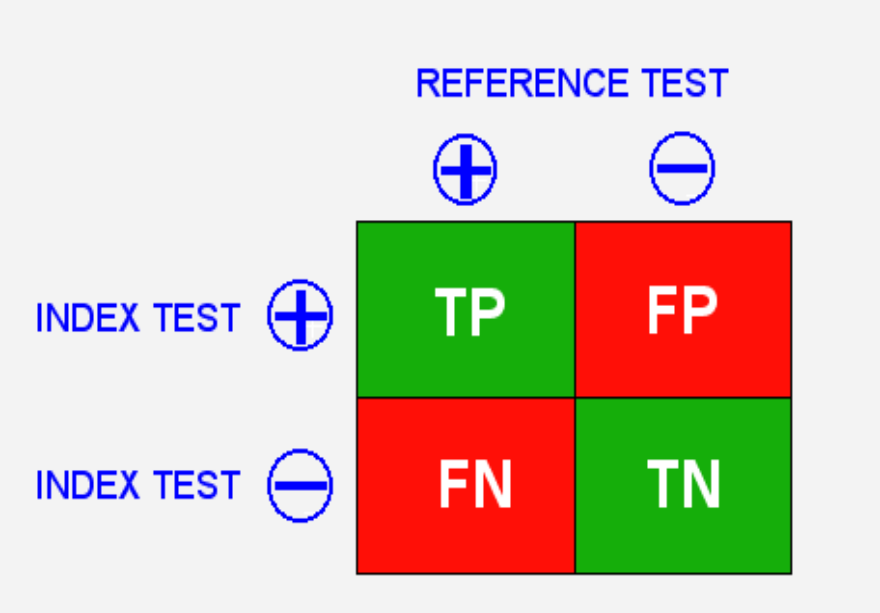

$$\text{LR+:}\quad \frac{\text{True positive rate}}{\text{False positive rate}} = \frac{\text{Sensitivity}}{\text{1-Specificity}}$$

$$\text{LR-:}\quad \frac{\text{False negative rate}}{\text{True negative rate}} = \frac{\text{1- Sensitivity}}{\text{Specificity}}$$

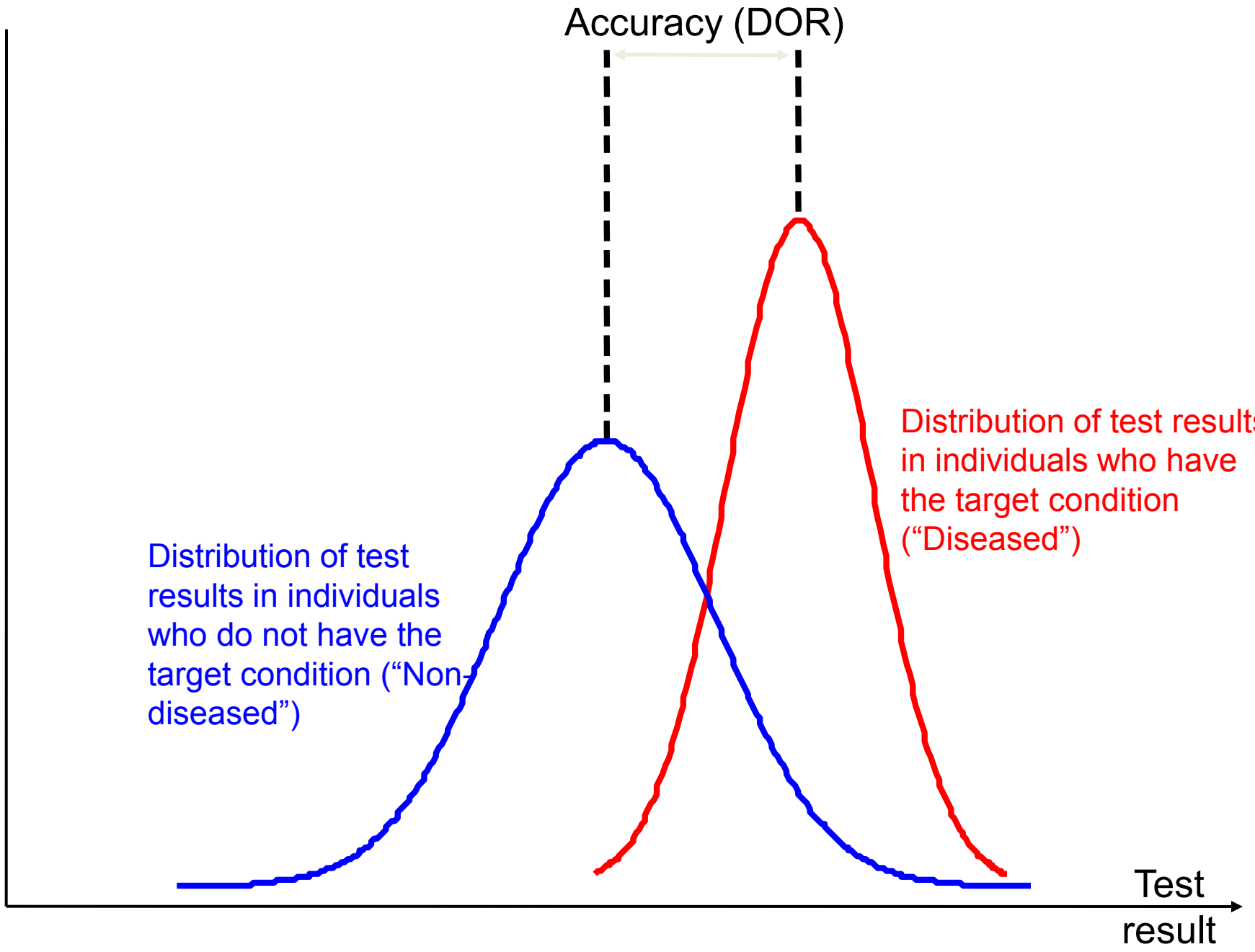
# Likelihood ratios

- Tell us how many times more likely a test result is to be expected in a person with compared to a person without the disease
- Gives direct information on the power of a test to rule in / rule out a condition
- Allows the calculation of post-test probabilities from pre-test probabilities:
  - Post-test Odds= Pre-test Odds x LR

# Diagnostic odds ratio




$$\text{DOR} = \frac{\text{TP}}{\text{FP}} \Big/ \frac{\text{FN}}{\text{TN}} = \frac{\text{sens}}{\text{(1-spec)}} \Big/ \frac{\text{(1-sens)}}{\text{spec}} = \frac{\text{LR +}}{\text{LR -}}$$


Accuracy (DOR)
Distribution of test results in individuals who have the target condition ("Diseased")
Distribution of test results in individuals who do not have the target condition ("Non-diseased")
Test result

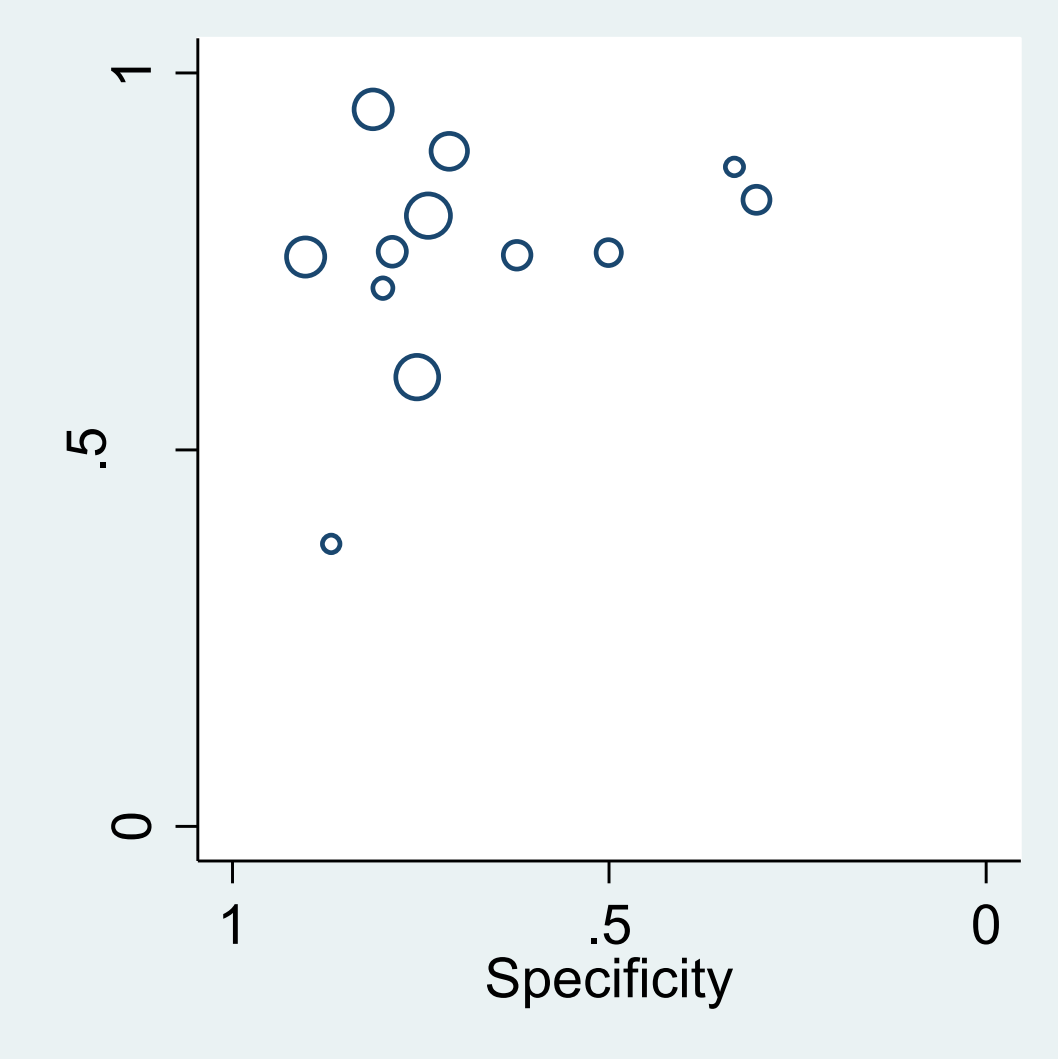
# Diagnostic odds ratio (DOR)

- Overall measure of diagnostic accuracy
- Not useful in clinical practice since it doesn't tell us the implications of positive and negative results

# What's different about meta-analysis of *diagnostic* studies?

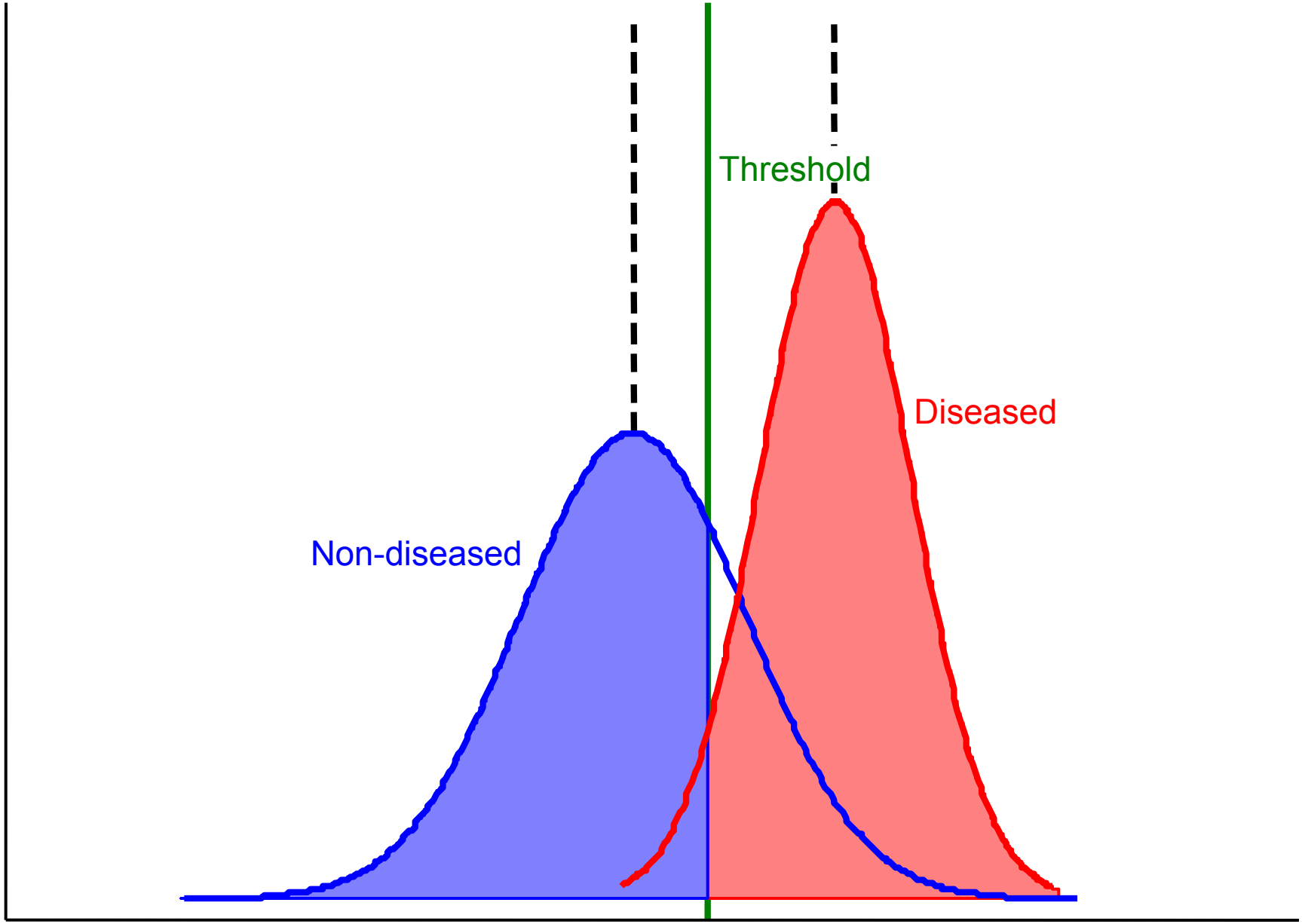
- Observational – not randomized, not balanced
- Odds ratios are often huge (100 or more)
- Studies are often small
- Vary widely in design, reporting and risk of bias
- Substantial statistical heterogeneity is usual
- Need to estimate two parameters: sensitivity *and* specificity
- These are typically negatively correlated

# Example: Alvarado score in suspected acute appendicitis

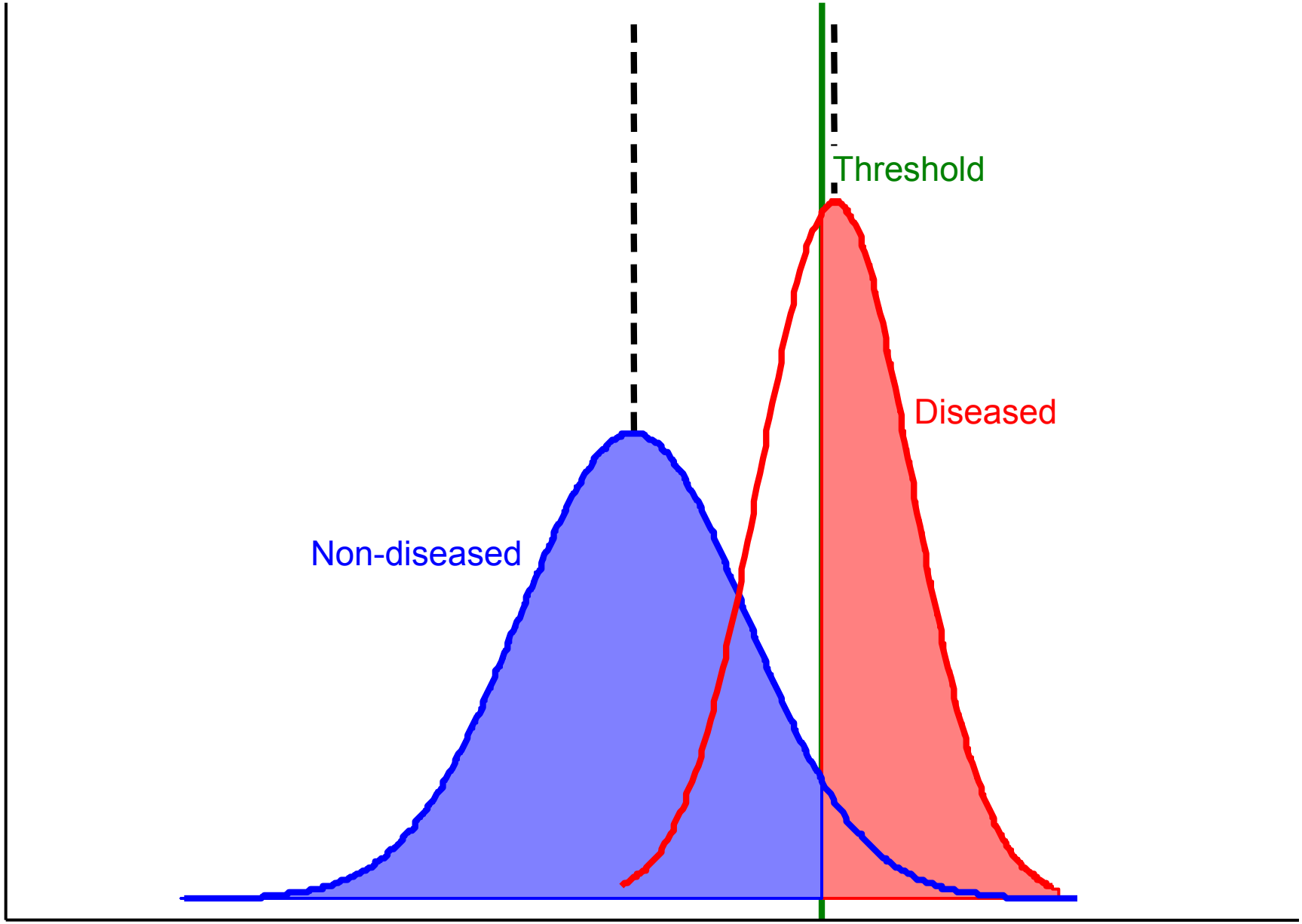


# Variability between study results

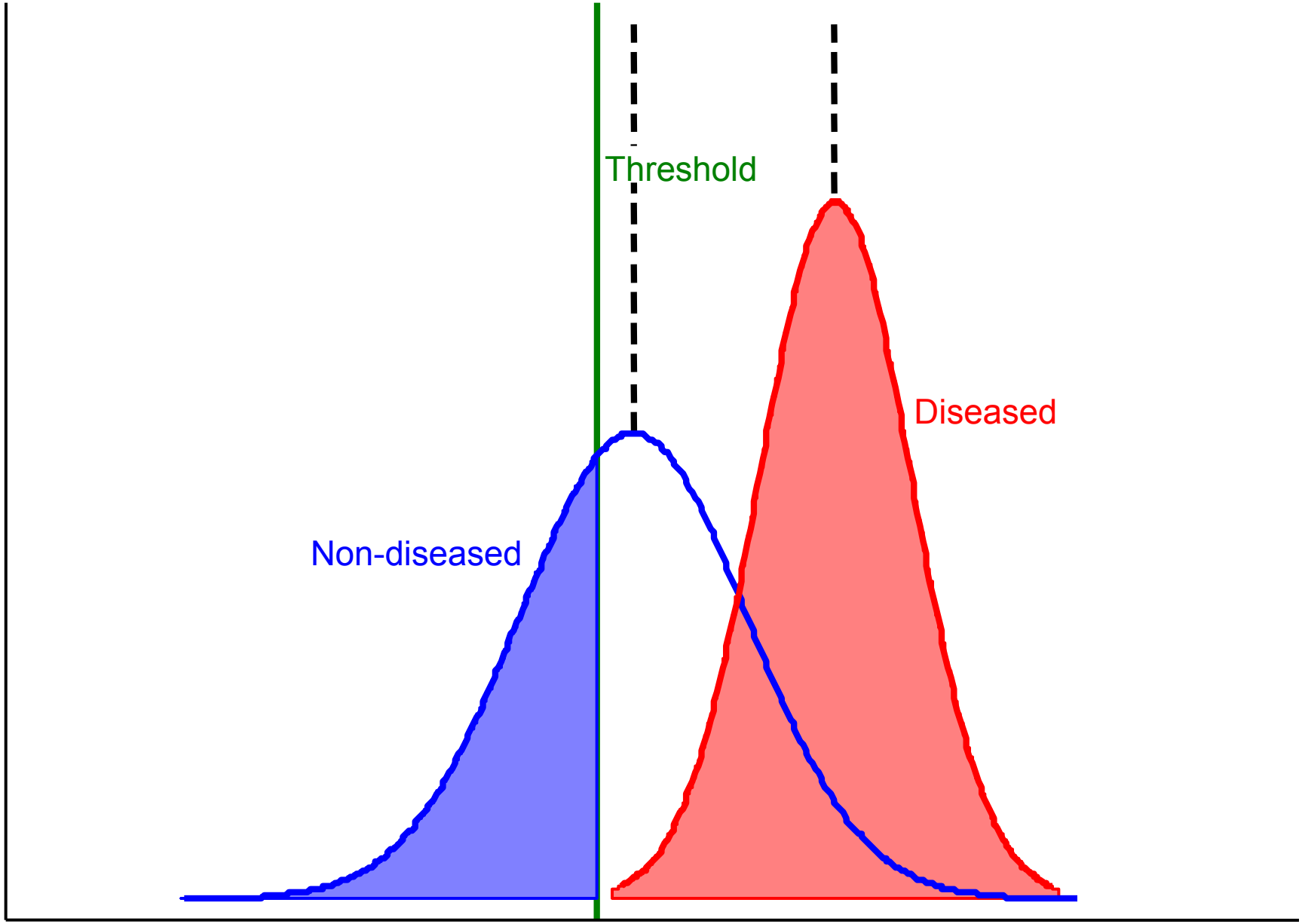
- Chance (sampling variation)
- Threshold effects
- Study quality


Threshold
Diseased
Non-diseased

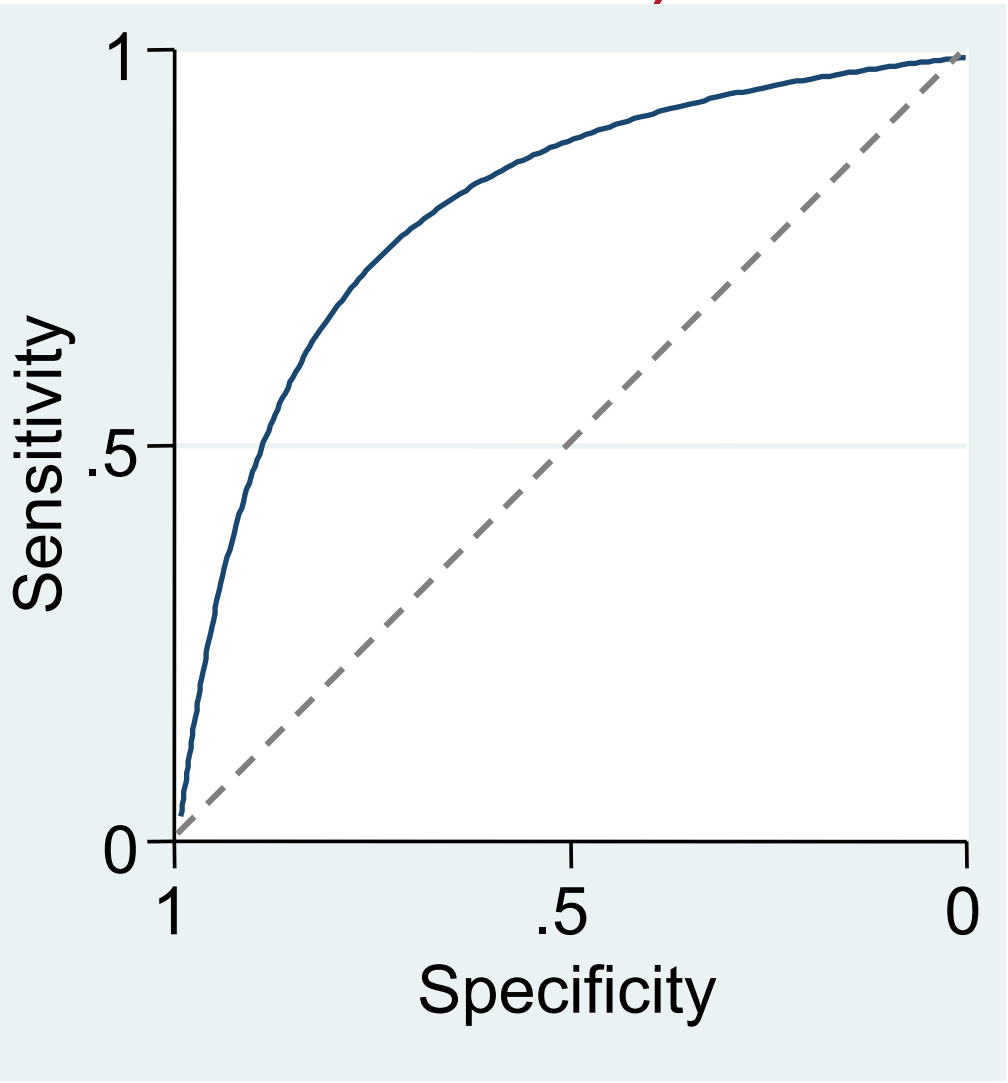
# Increase specificity, decrease sensitivity



# Increase sensitivity, decrease specificity

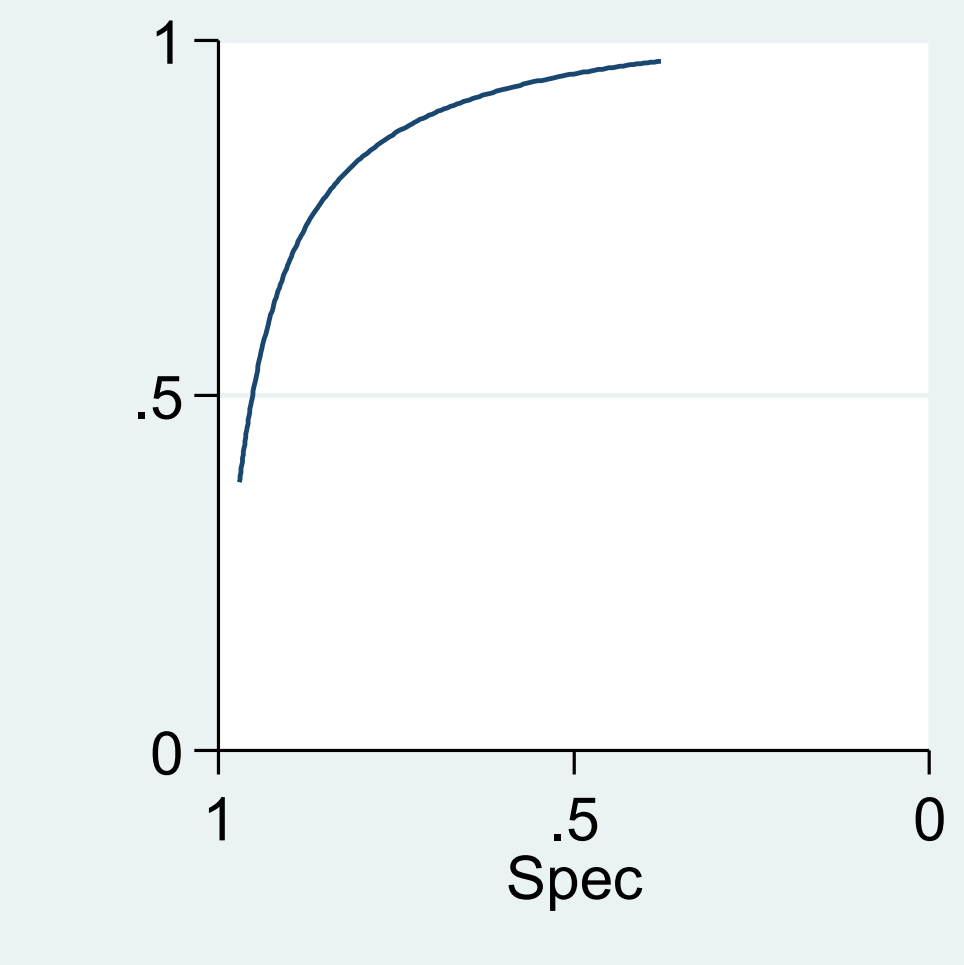


# ROC (receiver-operating characteristic) curve

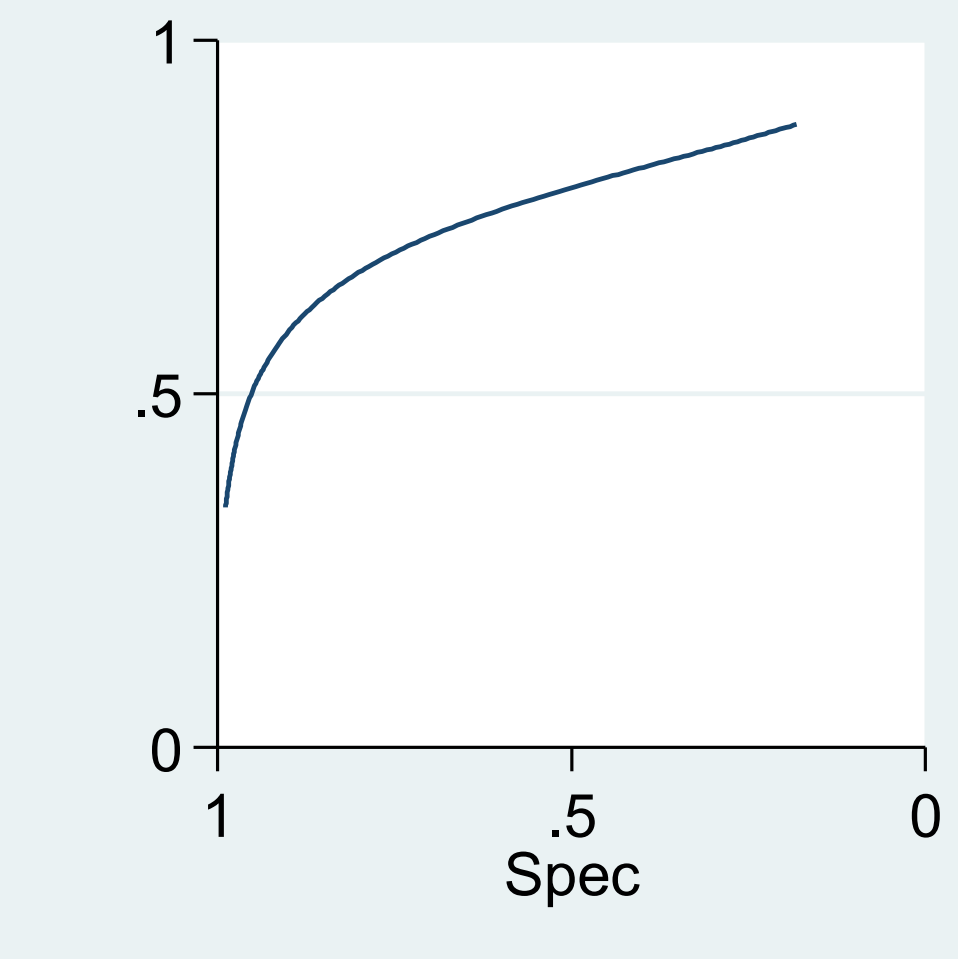



- Display the trade-off between sensitivity and specificity
- The curve must always go from the bottom left to the top right of the graph
- A useless test joins these points with a straight line
- The Area Under the Curve (AUC) is a measure of test accuracy
- Disadvantage: we can't see the threshold corresponding to a particular choice of sensitivity and specificity

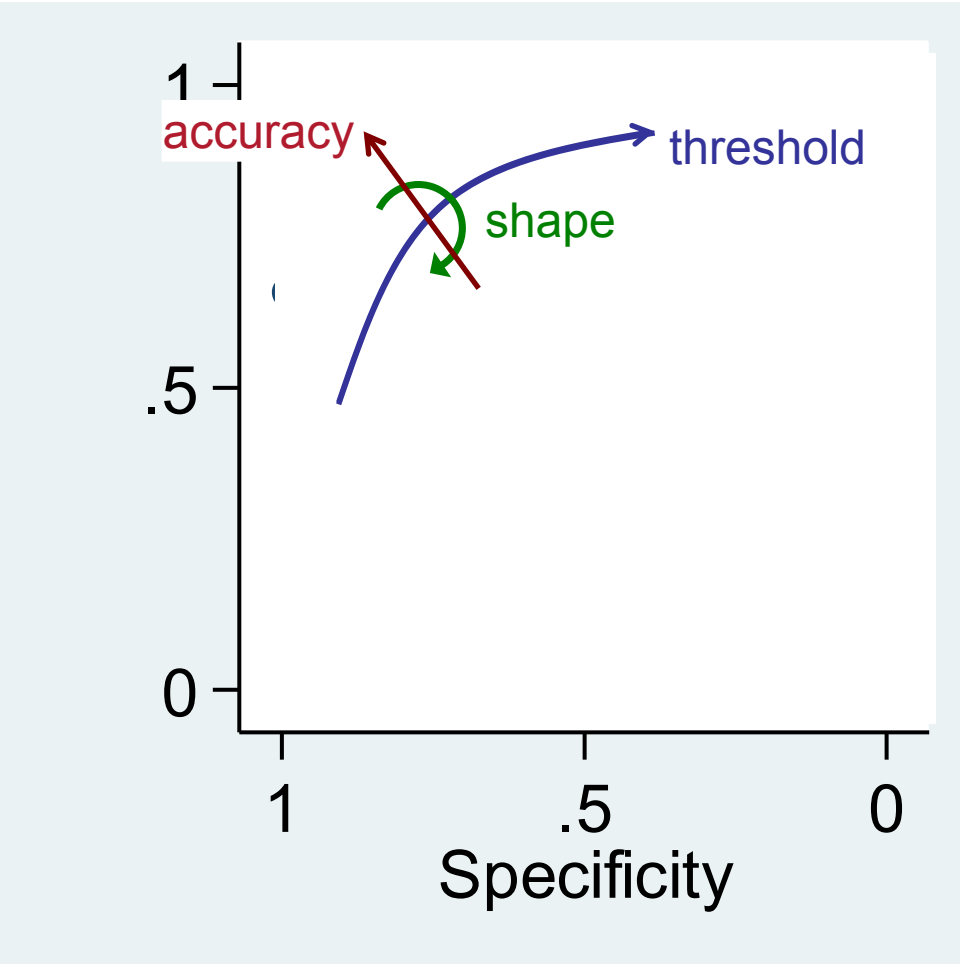
# ROC curve: increase accuracy



# ROC curve: Change shape parameter (different variances)



# ROC plot – single study



# Meta-analysis of Dx accuracy: What methods are available?

All based on 2×2 table from each study

- Easily implemented methods:
  - M-A of (Diagnostic) Odds Ratios (DORs)
  - Separate M-A of Sensitivity & Specificity
  - Moses-Littenberg Summary ROC (SROC) curve
- Hierarchical models:
  - Hierarchical Summary ROC curve (HSROC) model
  - Bivariate random-effects meta-analysis

Gatsonis & Paliwal, *Am. J. Roentgenol.* 2006; **187**:271-288
Leeflang *et al. Annals Int. Med.* 2008; 149:889-897.

# Hierarchical models

- Recommended for formal statistical inference (CIs, *p*-values…)
- 2 forms:
  - Hierarchical Summary ROC (HSROC)
  - Bivariate model

# Binomial *within*-study model

- $TP_i \sim Binomial(N_d, Sens_i)$
  $TN_i \sim Binomial(N_h, Spec_i)$

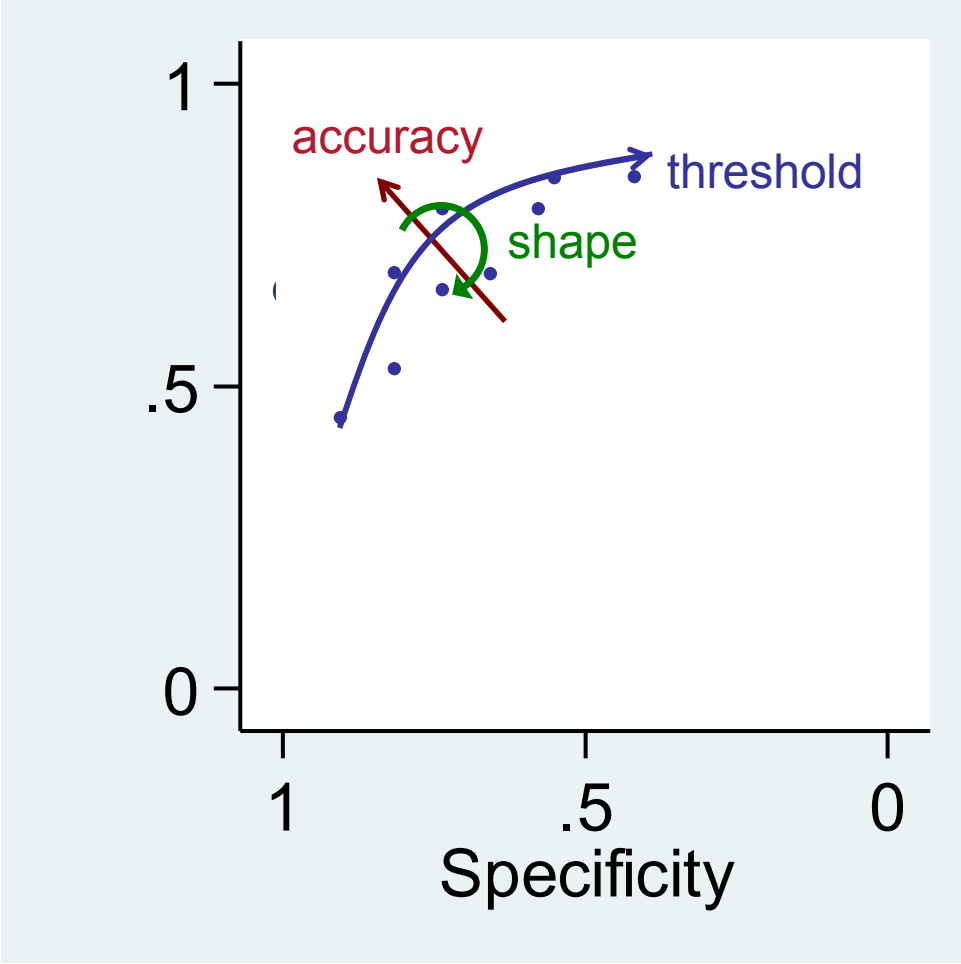
| | Target condition | |
|---|---|---|
| **Test result** | Present | Absent |
| Positive | *TP* | *FP* |
| Negative | *FN* | *TN* |
| Total | $N_d$ | $N_h$ |

- No *within*-study correlation as 'diseased' and 'healthy' are different groups of participants
- Better to use *generalized* mixed models than to linearise using empirical logit transforms

Chu & Cole *J Clin Epidemiol* 2006; **59**: 1331-2
Hamza, van Houwelingen & Stijnen *J Clin Epidemiol* 2008: **61**: 41-51.
Hamza, Reitsma & Stijnen *Medical Decision Making* 2008: **28**: 639-649

# HSROC model

## *summary* ROC plot




5 parameters:

Mean + variance of both accuracy & threshold

Shape (scale, asymmetry) parameter

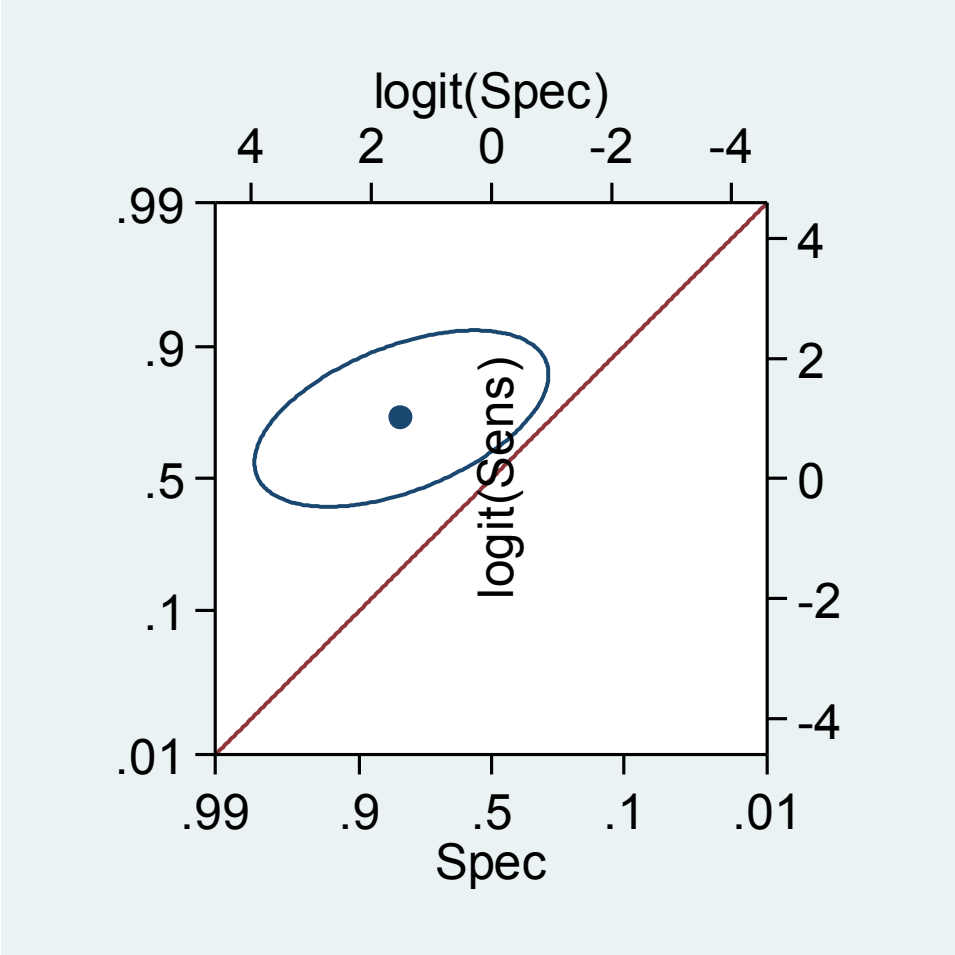
Rutter & Gatsonis. *Statist. Med.* 2001; 20:2865-84

# Bivariate model

- Based on logit-transform of sensitivity and specificity

  - van Houwelingen et al. *Statist. Med.* 2002; 21: 589-624
  - Reitsma et al. *J. Clin. Epidemiol.* 2005; 58: 982-990

# Bivariate model




5 parameters:

Mean & variance of logit(Sens) & logit(Spec)

+ correlation

# Bivariate between-study model

- $(\mu_{i1}, \mu_{i2})$ = (logit $Sens_i$, logit $Spec_i$)
- $\Sigma$ : variance-covariance matrix
- Between-study model is $\mu_i \sim N(\mu, \Sigma)$

$$\Sigma = \text{var}\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix} = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}$$

# How are the hierarchical models related?

- **Without study-level covariates**, the HSROC and bivariate models are precisely the same
  - just different parameterisations

- Harbord, Deeks, Egger, Whiting & Sterne *Biostatistics* 2007; **8**: 239-251
- Arends, Hamza, van Houwelingen, Heijenbrok-Kal, Hunink & Stijnen *Medical Decision Making* 2008; **28**:621-638

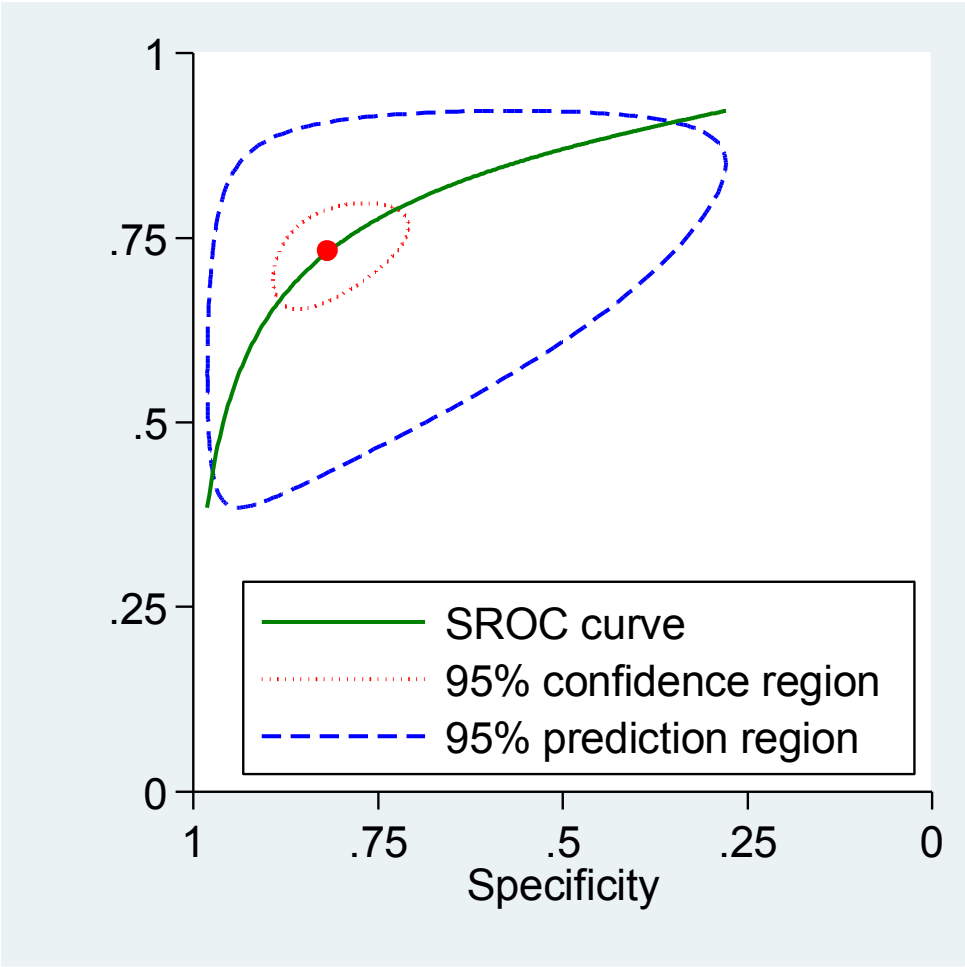
# How are the hierarchical models related?

- Can allow a **covariate** to affect:
  - *bivariate model*: sensitivity or specificity
  - *HSROC model*: threshold or accuracy
- If a covariate affects both, models are equivalent
- Only the HSROC model can also allow a covariate to affect shape

# Comparison

| | Bivariate | HSROC |
|---|---|---|
| Coordinates | Sensitivity & Specificity (logit-transformed) | Accuracy & Precision |
| Usual interpretation | Summary point and region | Summary ROC curve |

# Model outputs




Can get same point, line and regions from either HSROC or bivariate parameter estimates

# Interpretation

- Summary point is meaningful if:
  - Threshold is similar in all studies
  - Between-study heterogeneity is modest
- Summary curve is meaningful if threshold chosen at random
  (Arends & Stijnen 2008; conference presentation)
- I recommend prediction regions to indicate heterogeneity visually

# Software for hierarchical methods

- Need software for mixed models – preferably *generalized* mixed models
- HSROC with covariates is a generalized *non*linear mixed model
  - WinBUGs or SAS NLMIXED
- Bivariate model is a GLMM
  - Above plus `xtmelogit` & `gllamm` in Stata, `lmer` or `INLA` in R
- Packages or sample code available for most or all of these

# metandi

- Harbord & Whiting *Stata Journal* 2009; **9**: 211-229
  (type `findit metandi` within Stata)
- Fits bivariate model using `xtmelogit` or `gllamm`
- `xtmelogit`
  - Requires Stata 10 or above
  - Fast
- `gllamm`
  - Slower
  - Sometimes more reliable?

# Features of `metandi`

- Command has a simple syntax
- Package includes (several) help-files
- Easy to generate a default plot
- Plot can be customized using `metandiplot`
- Post-estimation facilities for model-checking and identification of influential studies

- Does not handle covariates (yet)

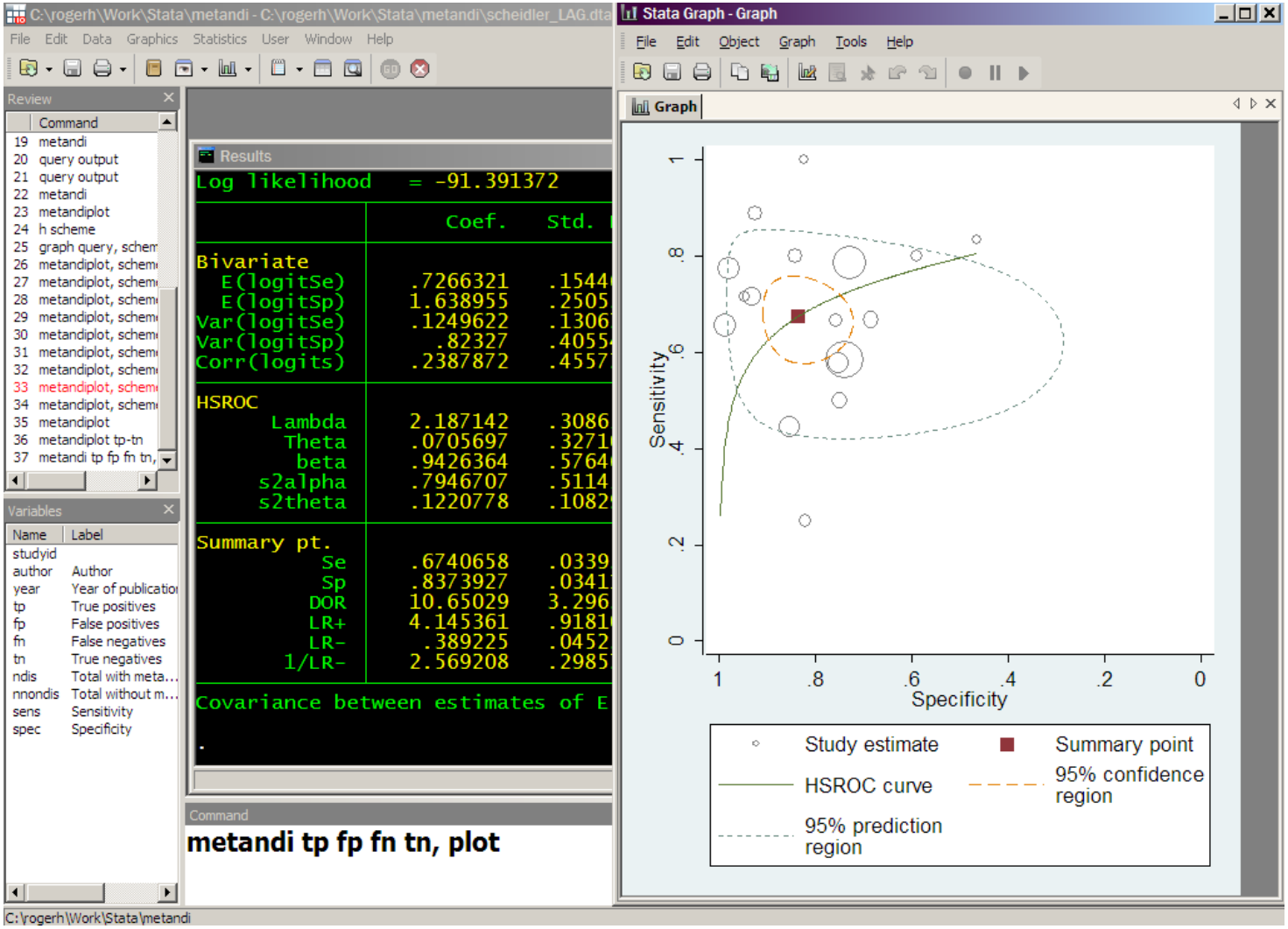
# midas

- By Ben Dwamena, University of Michigan
- Available from SSC
- *SJ*-style paper recently added to package
- Fits bivariate model using `xtmelogit` or `gllamm`
- Additionally provides many other analyses and plots: Model checking plots, study quality tables & charts, Forest plots of Sens & Spec, Deeks' test for study-size effects, Galbraith plot, Fagan nomogram, probability updating curve, LR scatter plot ...

# Comparison

| `metandi` | `midas` |
| --- | --- |
| • Concentrates on hierarchical model | • Comprehensive range of facilities for diagnostic test meta-analysis |
| • Text output formatted as standard for Stata estimation command | • Text output formatting somewhat ad-hoc |
| • e-class, supports post-estimation commands including predict | • r-class |

# Demonstration



# Few studies

What to do when there are $\lesssim 4$ studies?

- Not enough to fit hierarchical models with all 5 parameters by maximum likelihood

# Few studies: approaches

- Just plot individual study estimates
- M-A of ORs
- Separate univariate M-A of Sens & Spec
- Fit 4-parameter model by assuming equal between-study variances, giving symmetric SROC curve
- Use Bayesian approach with (weakly) informative priors for variance parameters

# Future extensions to `metandi`

- Univariate M-A and symmetric model
  (implemented in unreleased version)
- Allow one (or more?) covariates
- Interface to R package `INLA` (Integrated Nested Laplace Approximations) to allow Bayesian approach?

  Paul M, Riebler A, Bachmann LM, Rue H, Held L. Statistics in Medicine 2010; 29(12):1325-1339