Introduction to
Meta-analysis of Accuracy Data

Hans Reitsma MD, PhD
Dept. of Clinical Epidemiology, Biostatistics & Bioinformatics
Academic Medical Center - Amsterdam
Continental European Support Unit – Cochrane Collaboration

Copenhagen
February, 2009

Outline

- Nature of the outcome measures
- Descriptive analysis: plots and figures
- Statistical models
- REVMAN

Diagnostic Accuracy Data

- Agreement between results of the index test and reference standard
- Many measures of agreement
- Focus on pairs of sensitivity & specificity

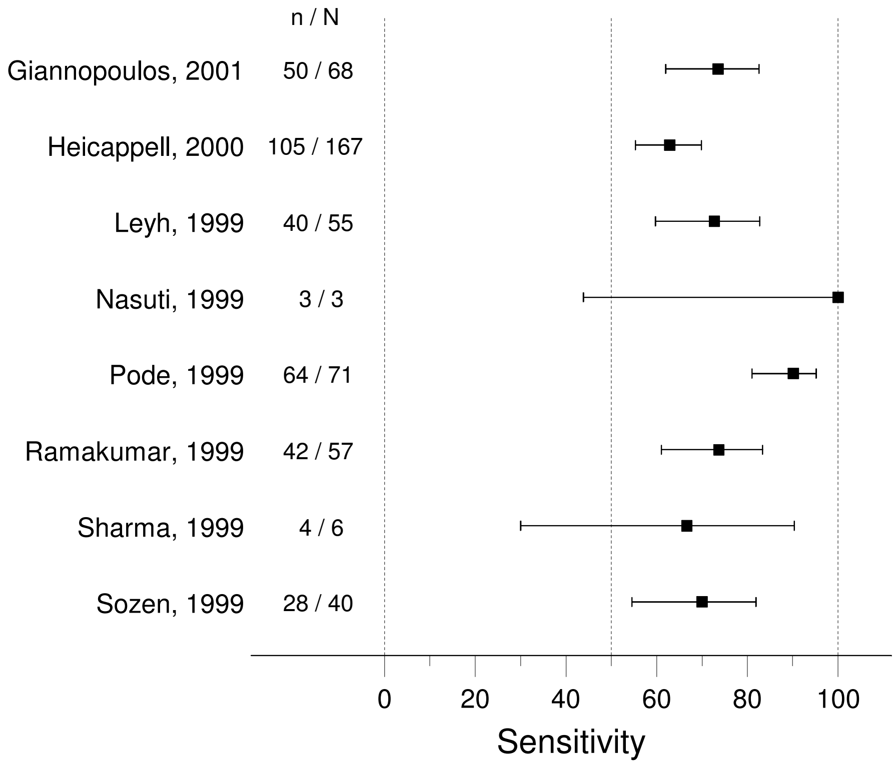
Clinical Example

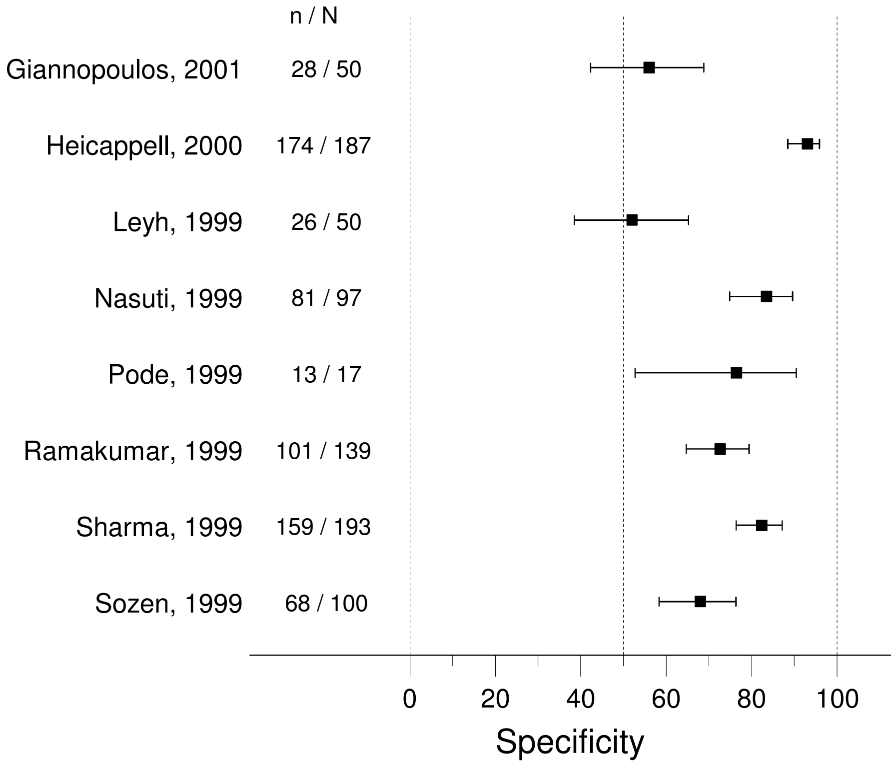
- Tumor markers for the detection of bladder cancer
- Measurement in urine rather than invasive cystoscopy
- Several markers: focus on bladder tumor antigen (BTA stat)
- N=8 studies

Descriptive Analysis

- Forest plots
  - point estimate with 95% CI
  - paired: sensitivity and specificity side-by side

Forest Plot

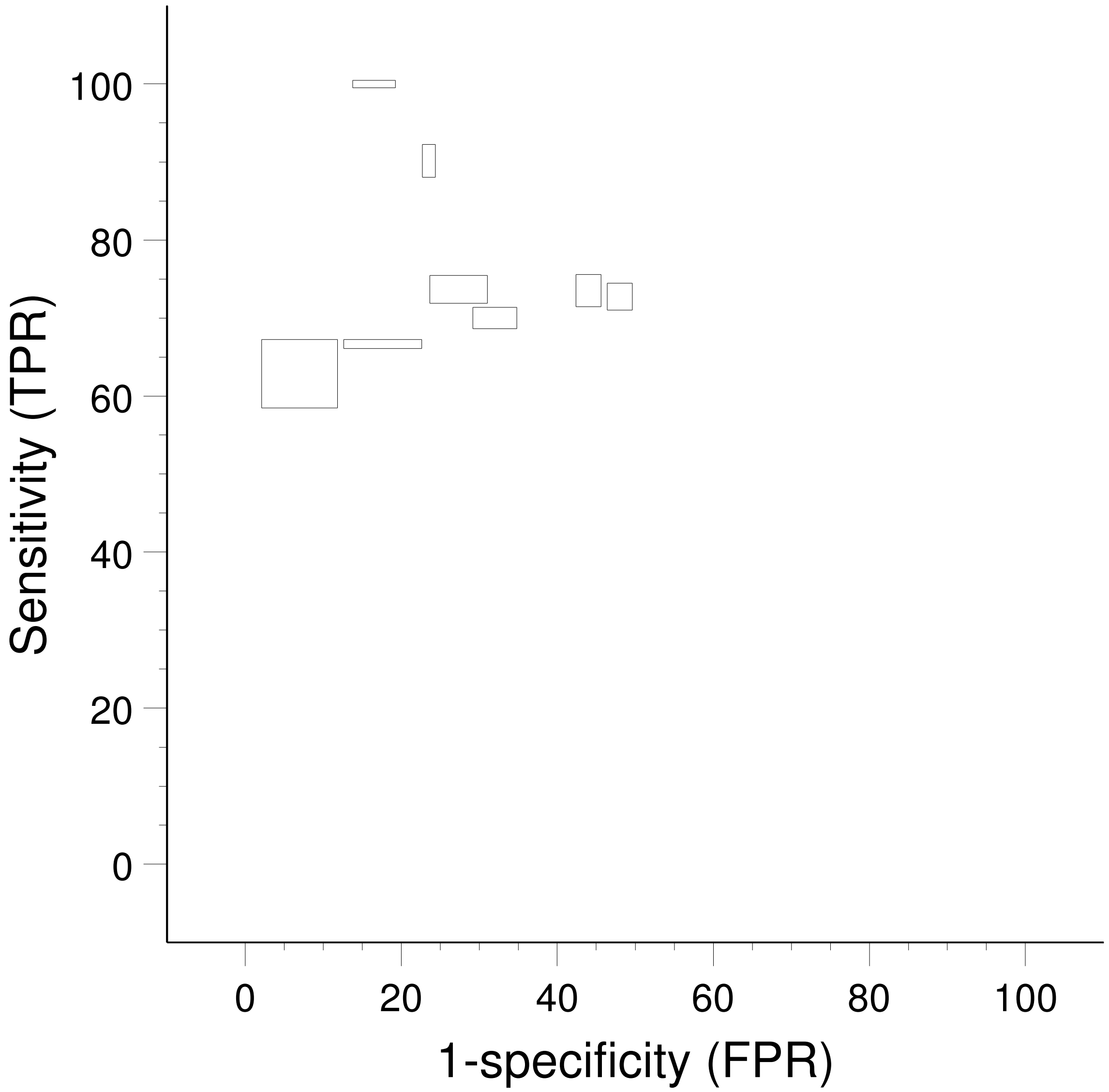

n / N
Giannopoulos, 2001
50 / 68
Heicappell, 2000
105 / 167
Leyh, 1999
40 / 55
Nasuti, 1999
3 / 3
Pode, 1999
64 / 71
Ramakumar, 1999
42 / 57
Sharma, 1999
4 / 6
Sozen, 1999
28 / 40
0
20
40
60
80
100
Sensitivity


n / N
Giannopoulos, 2001
28 / 50
Heicappell, 2000
174 / 187
Leyh, 1999
26 / 50
Nasuti, 1999
81 / 97
Pode, 1999
13 / 17
Ramakumar, 1999
101 / 139
Sharma, 1999
159 / 193
Sozen, 1999
68 / 100
0
20
40
60
80
100
Specificity

Descriptive Analysis

- Forest plots
  - point estimate with 95% CI
  - paired: sensitivity and specificity side-by side
- ROC plot
  - pairs of sensitivity & specificity in ROC space
  - bubble plot to show differences in precision

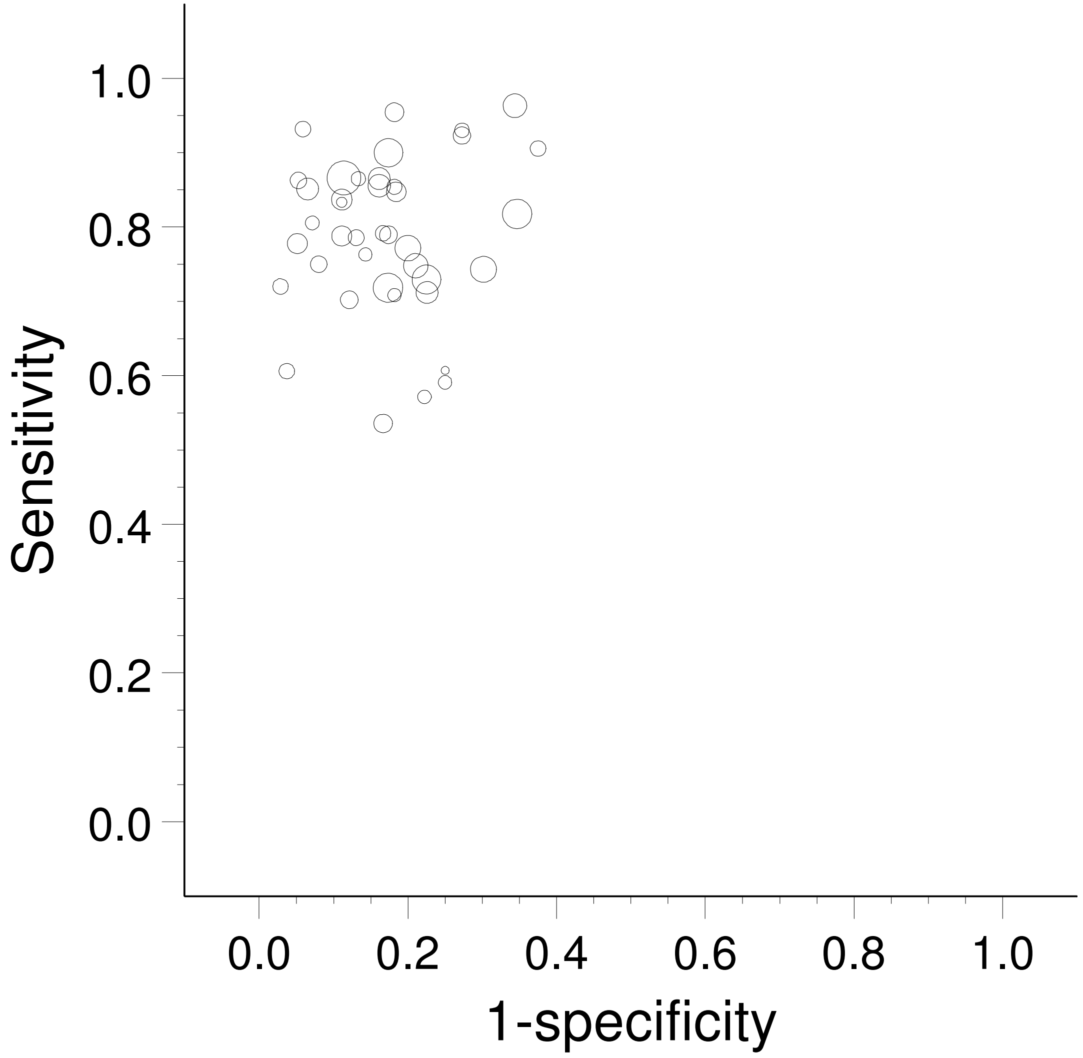
Plot in ROC Space



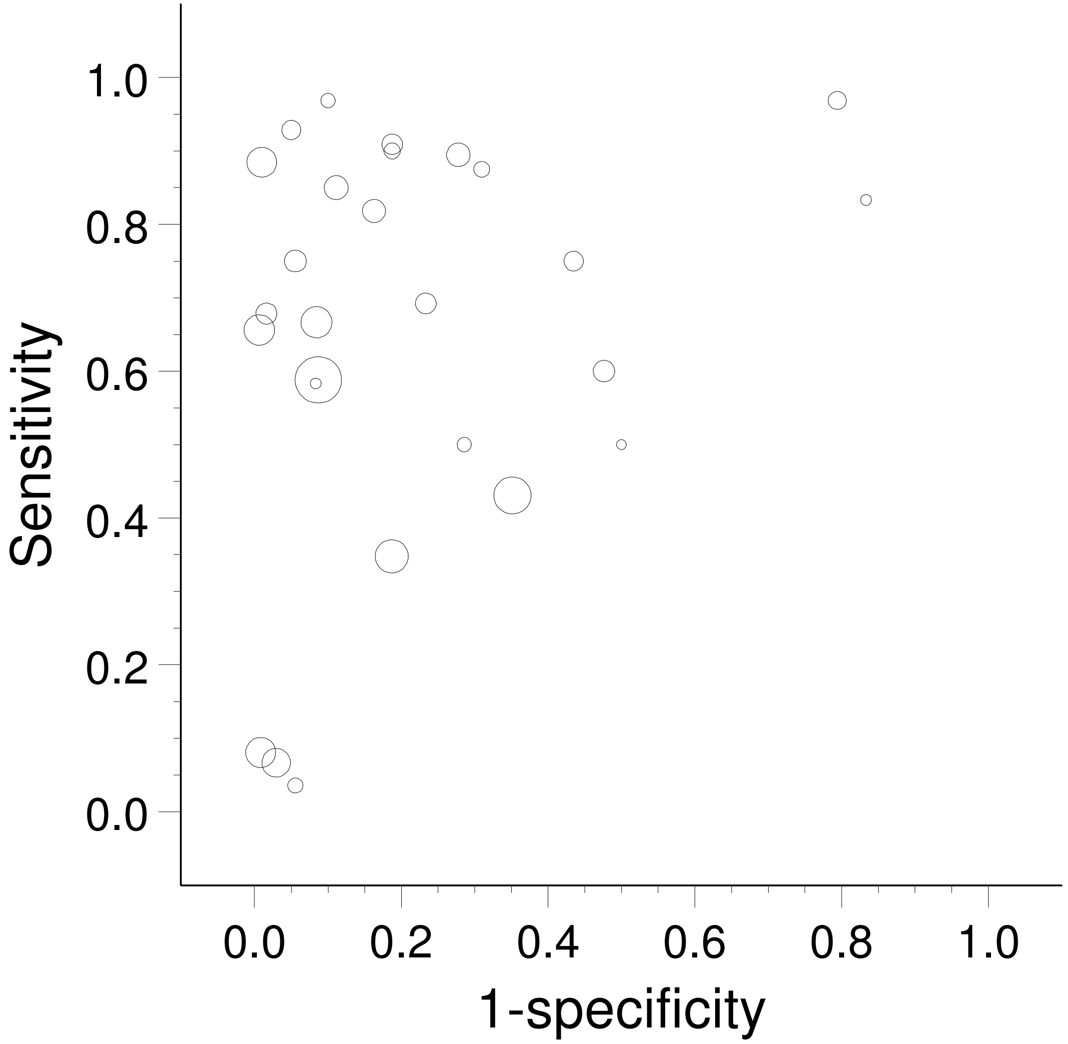
Challenge

- Understanding sources of variation, as results often vary between studies
- Providing informative summary measures of the data
- Drawing robust conclusions with respect to the research question

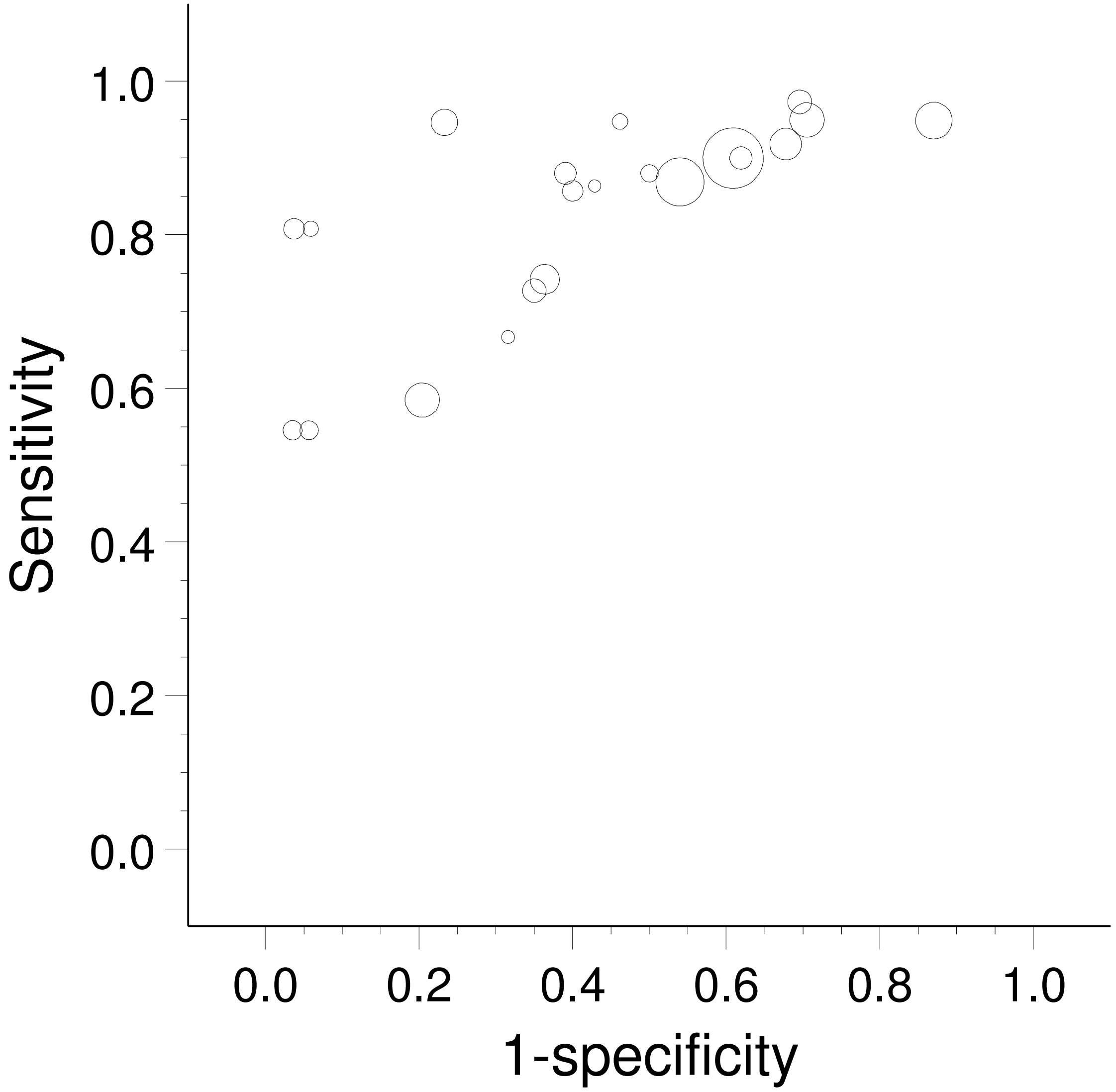
Echocardiography in Coronary Heart Disease



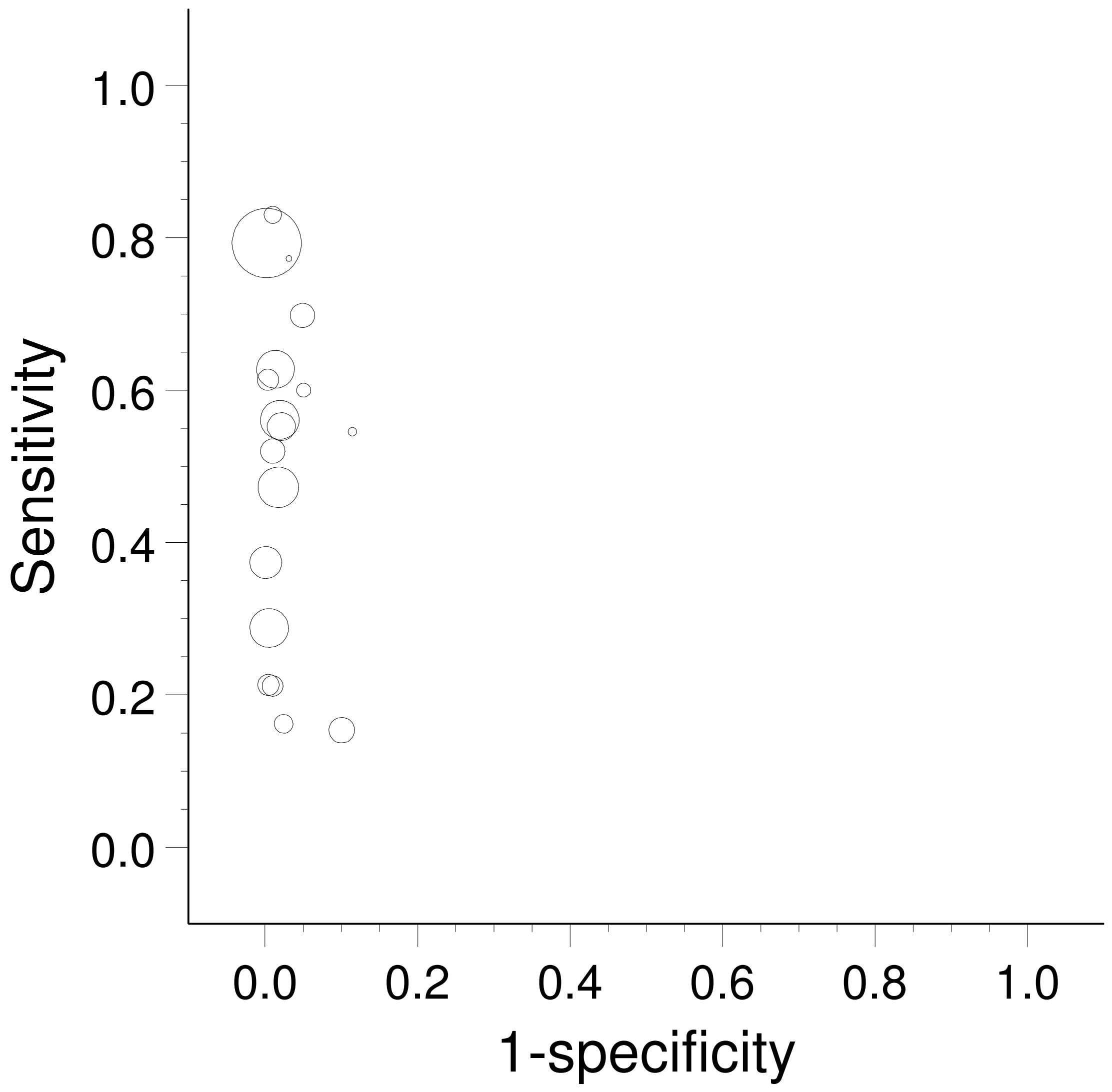
GLAL in Gram Negative Sepsis



F/T PSA in the Detection of Prostate cancer



Dip-stick Testing for Urinary Tract Infection



Sources of Variation

- Why do results differ between studies?

Sources of Variation

I. Chance variation

II. Differences in threshold

III. Bias

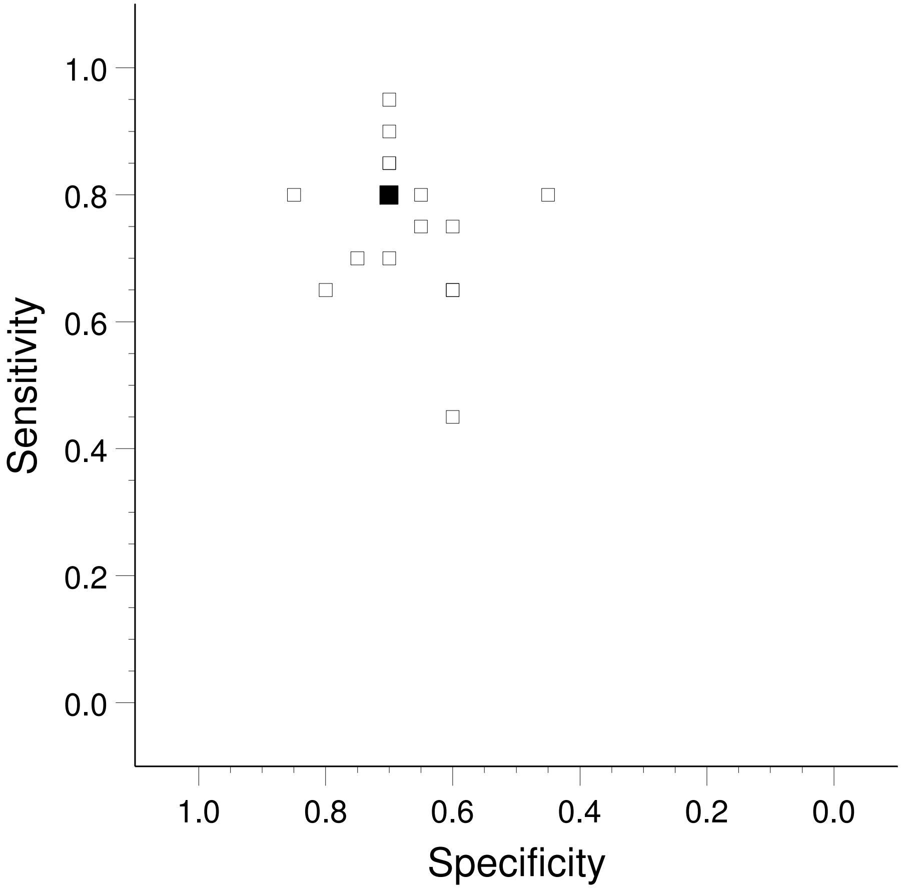
IV. Subgroups

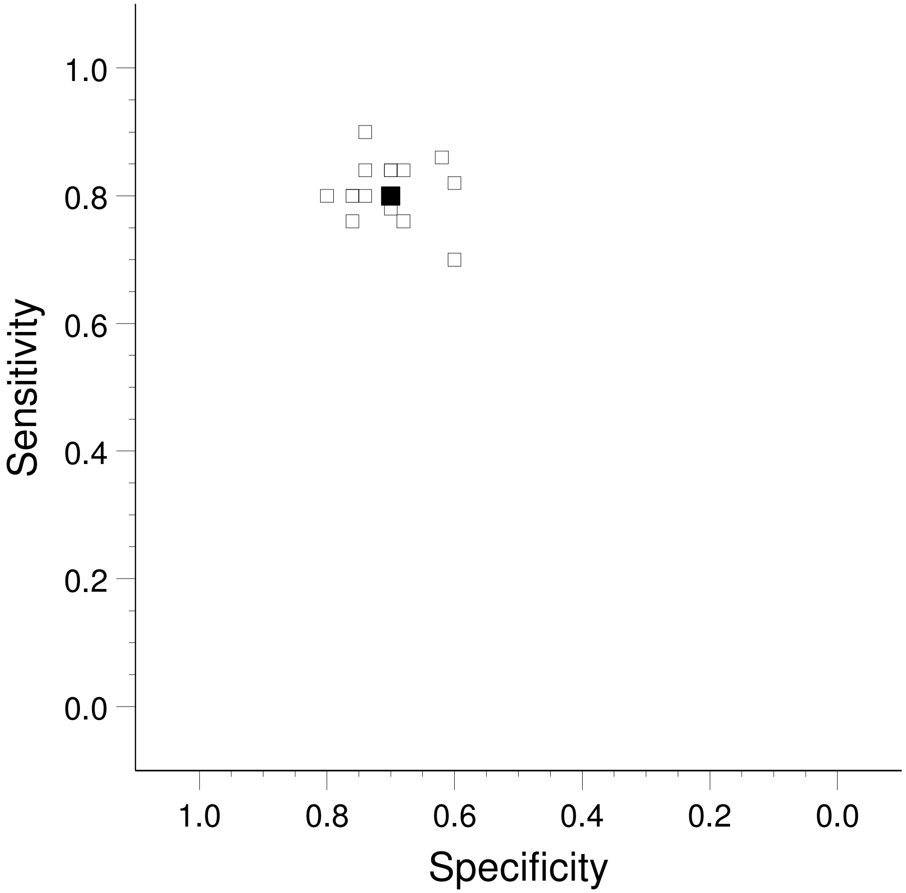
V. Unexplained variation

Sources of Variation: Chance

Chance variability:
sample size=40

Chance variability:
sample size=100

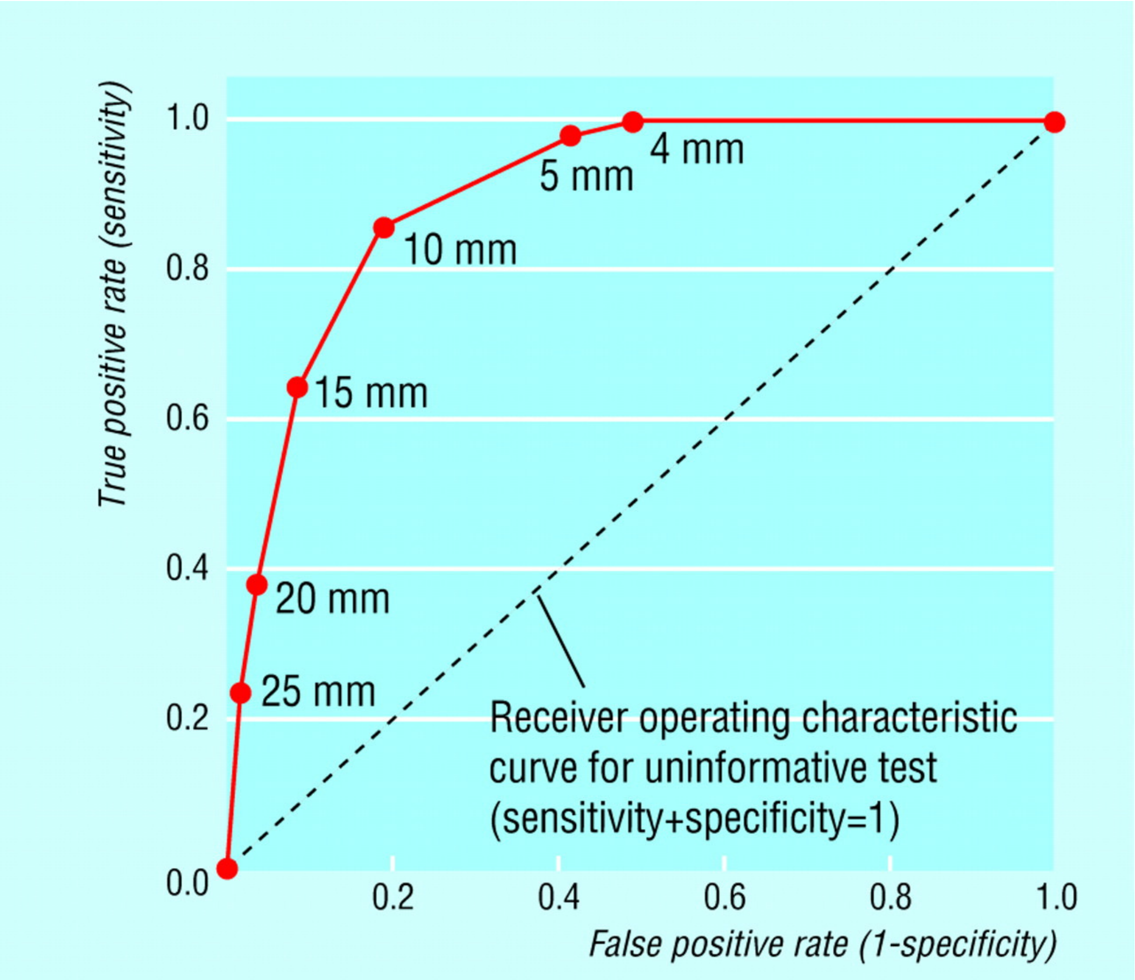




Variation due to Threshold Differences

- Explicit differences
  - studies have used different cut-off values to define positive test results

Receiver characteristic operating (*ROC*) curve



Deeks, J. J BMJ 2001;323:157-162

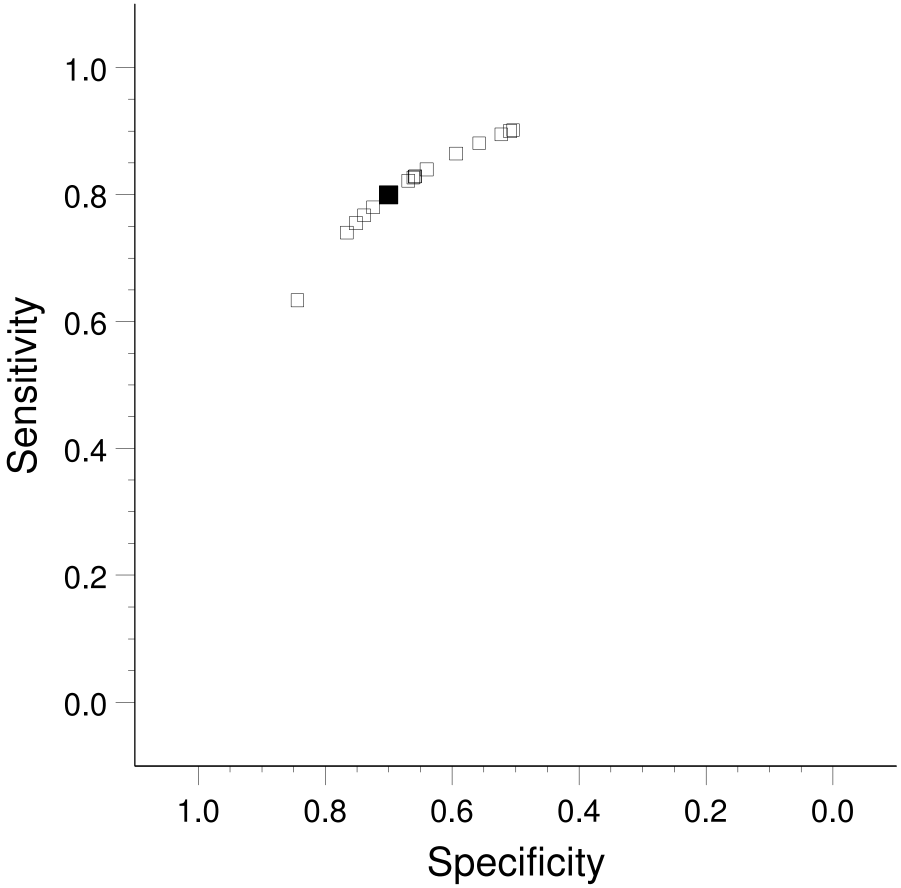
The ROC curve represents the relationship between sensitivity and specificity of the test at various thresholds

Variation due to Threshold Differences

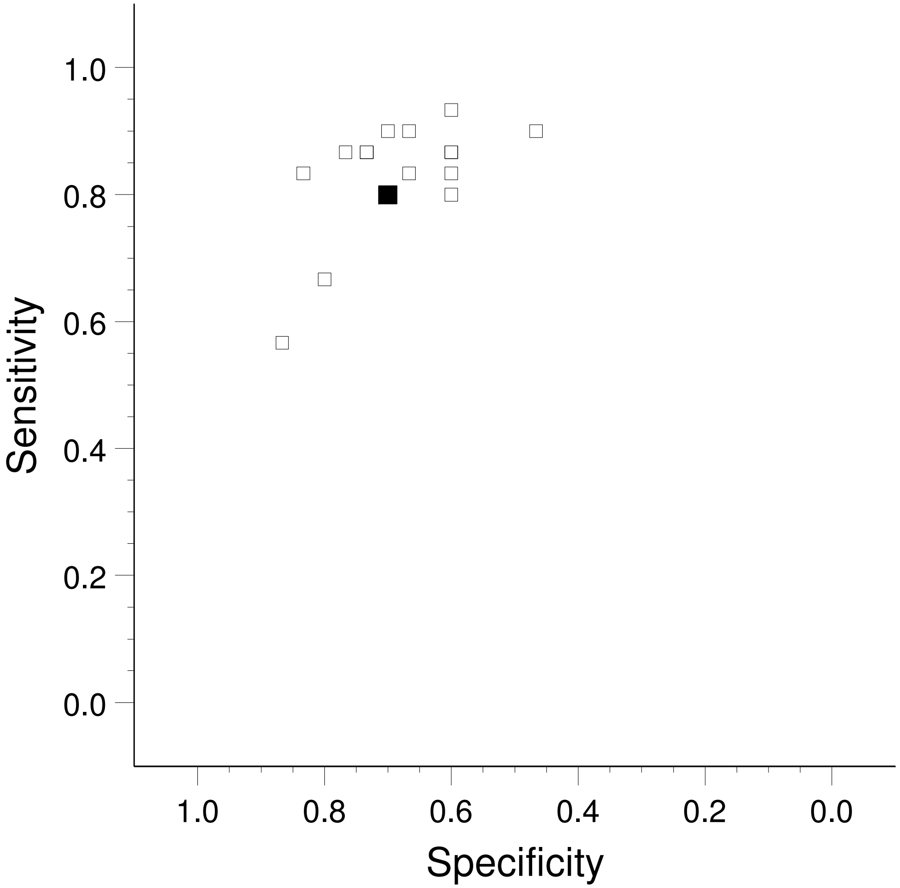
- Explicit threshold differences
  - studies have used different cut-off values to define positive test results
- Implicit threshold differences
  - differences in observers
  - differences in equipment
- Consequence: negative correlation arises between sensitivity and specificity

Sources of Variation: Threshold

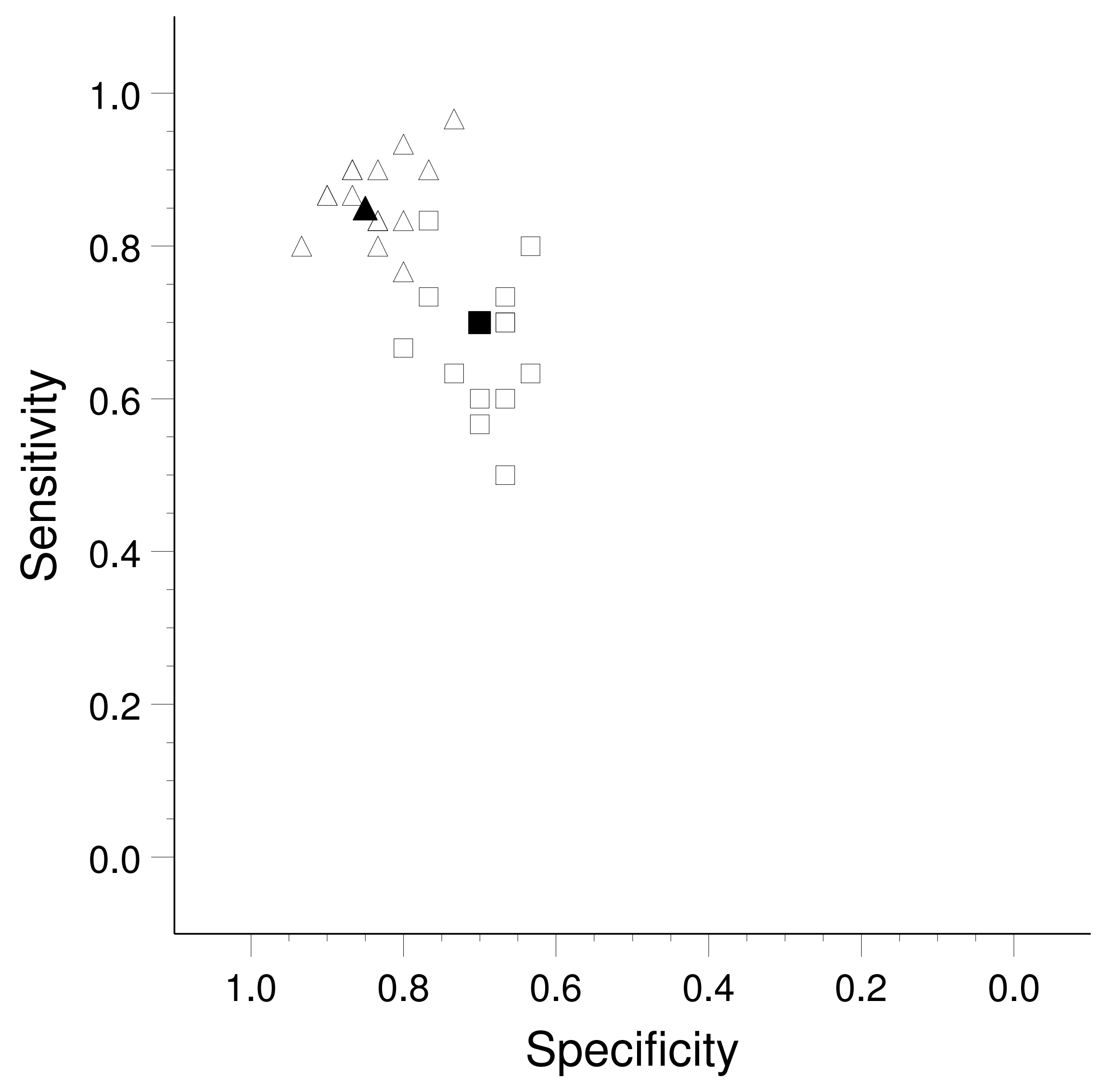
Threshold:
- perfect negative correlation
- no chance variability



Threshold:
- perfect negative correlation
- + chance variability: ss=60



Sources of Variation: Bias & Subgroup



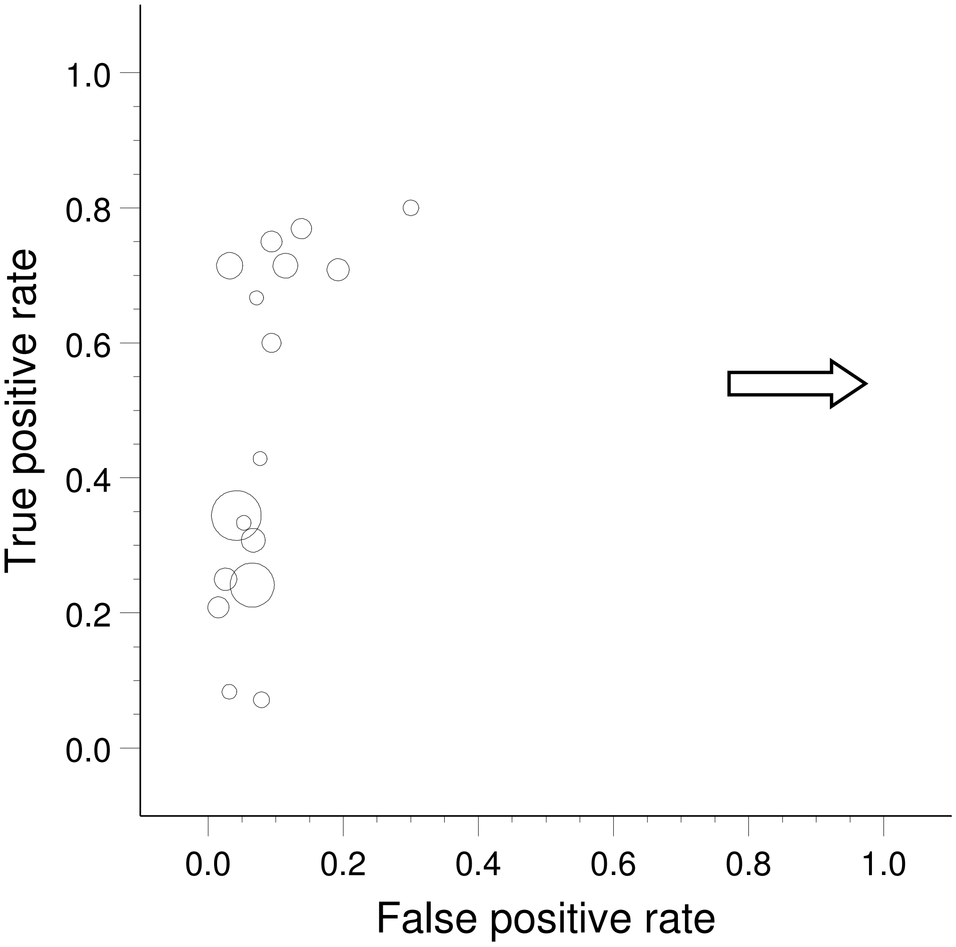
Bias & Subgroup:
- sens & spec higher
- ss=60
- no threshold

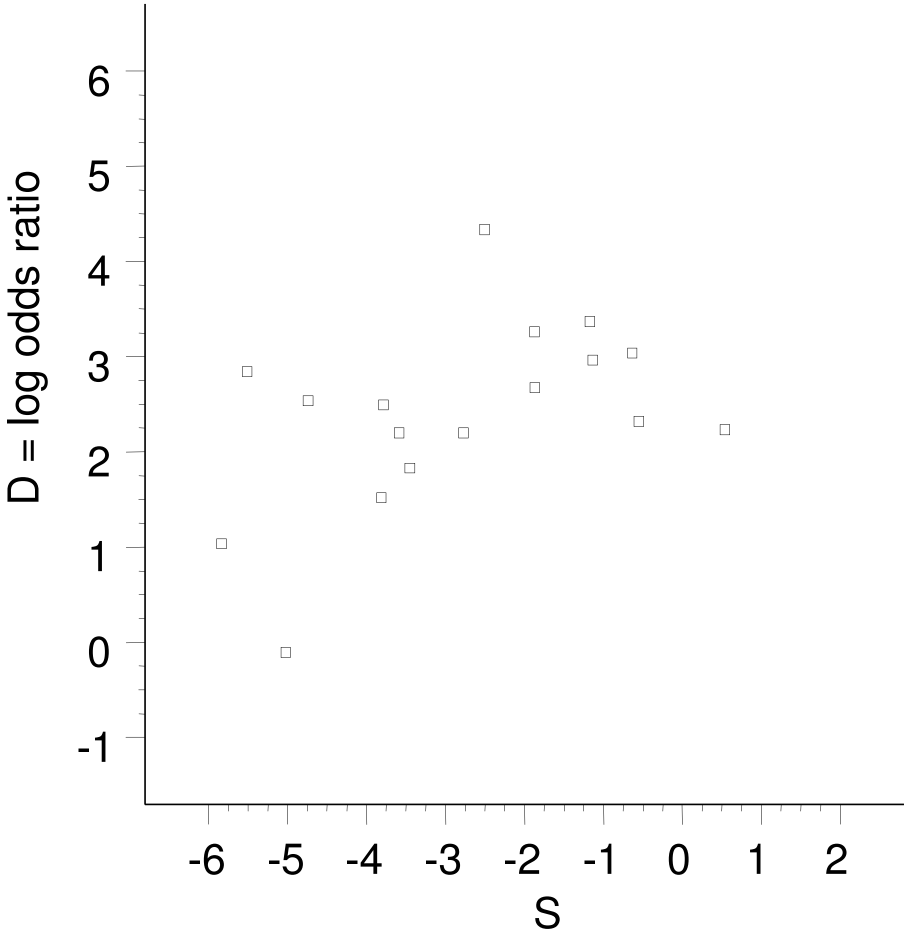
Overview of Statistical Approaches

- Summary ROC model / Moses-Littenberg (ML)
  - Traditional approach, straightforward

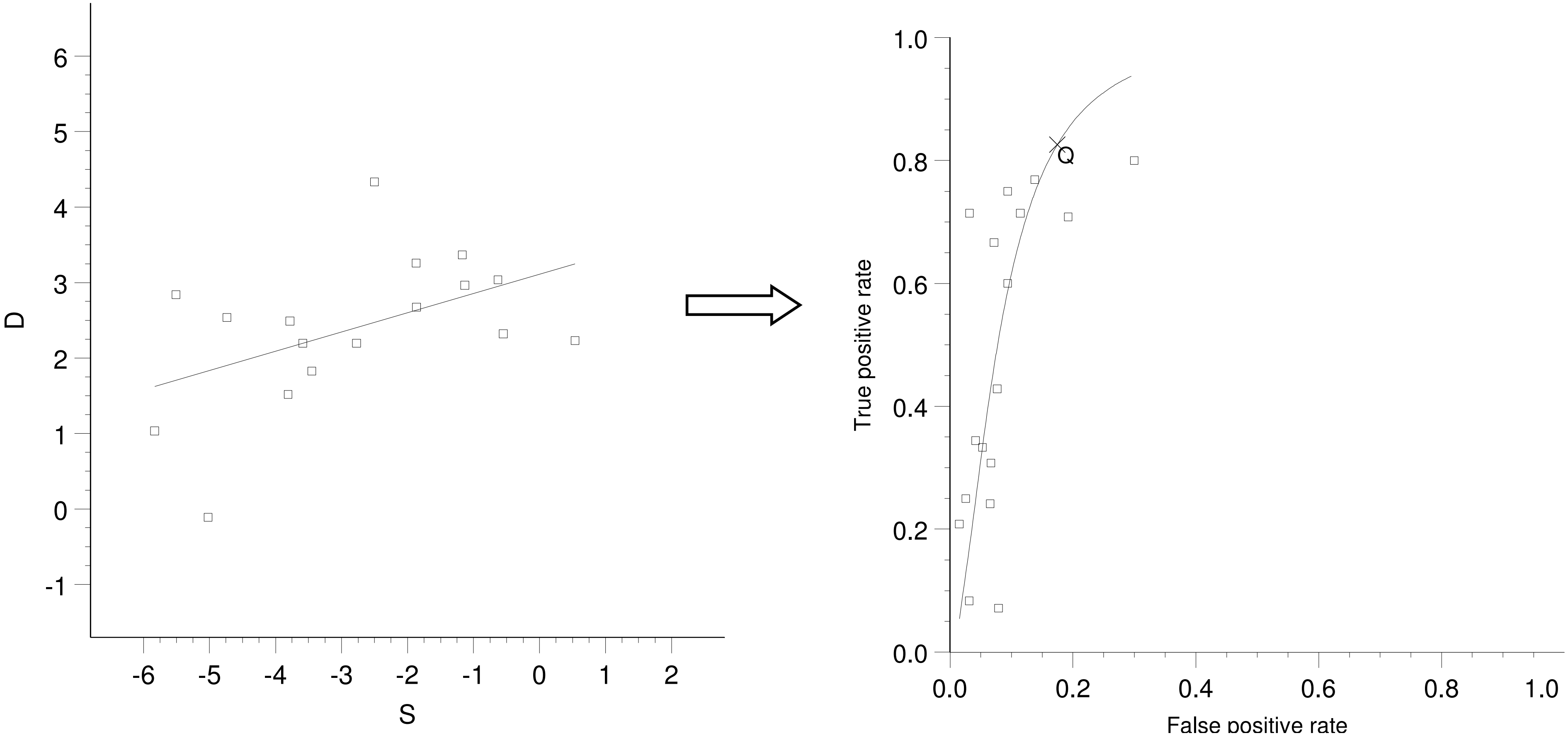
- More complex models
  - Bivariate random approach
  - Hierarchical summary ROC approach

ML approach: Finding Smooth Curve in ROC





Linear Regression & Back Transformation



Drawbacks Moses-Littenberg Approach

- Validity of significance tests
  - Sampling variability in individual studies not properly taken into account
  - P-values and confidence intervals erroneous
- Summary points
  - Average sensitivity/specificity cannot be obtained
  - Sensitivity for a given specificity can be estimated

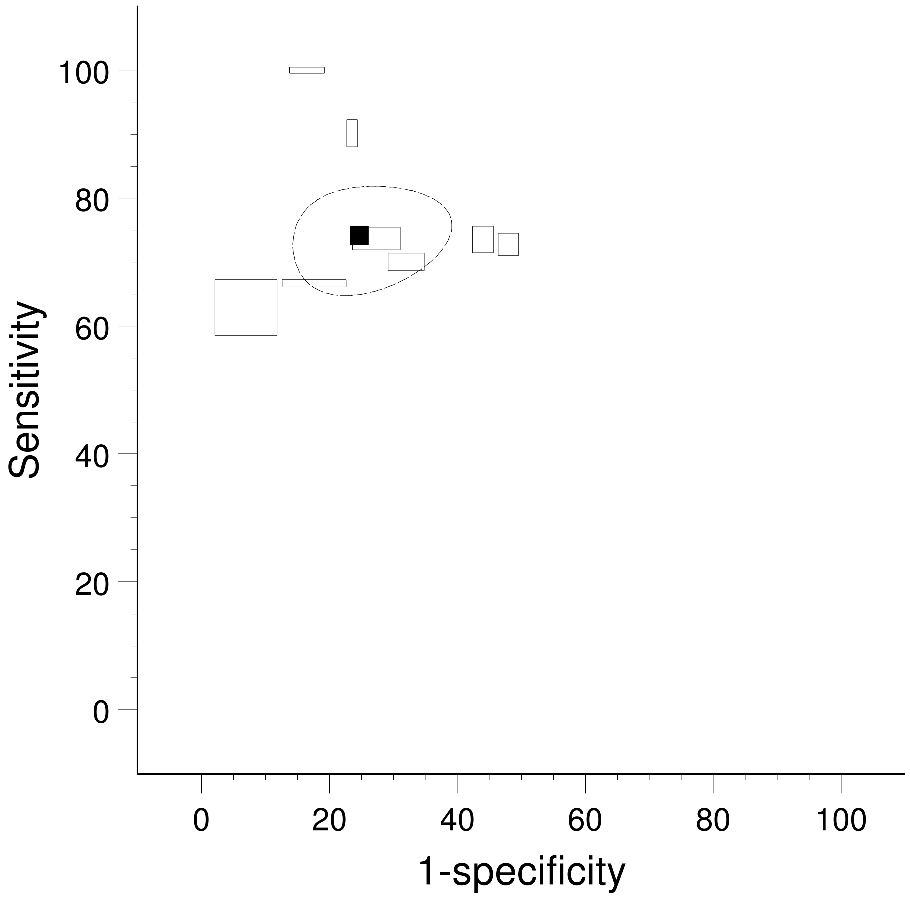
Advanced Models – HSROC and Bivariate methods

- Hierarchical / multi-level random effects
  - allows for both within and between study variability
- Binomial distribution
  - correctly models sampling uncertainty in both sensitivity and specificity
  - no zero cell adjustments needed
- Regression models
  - flexible in examining sources of heterogeneity

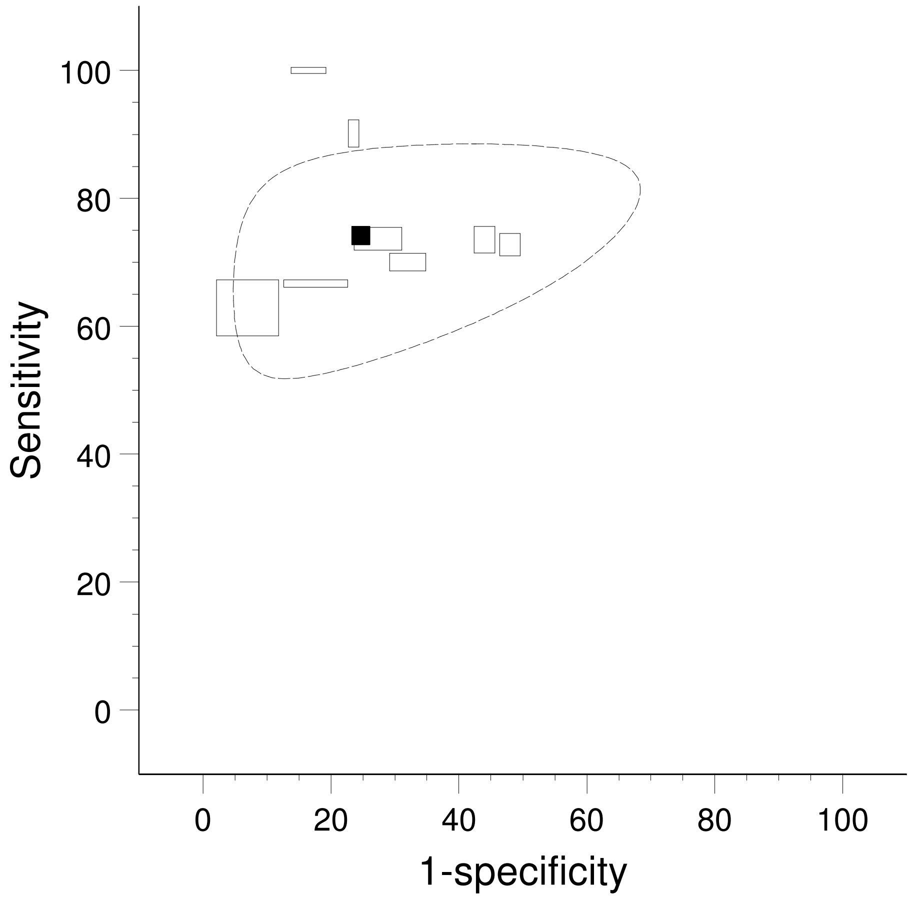
Presentation of Results

Summary Values with Ellipses

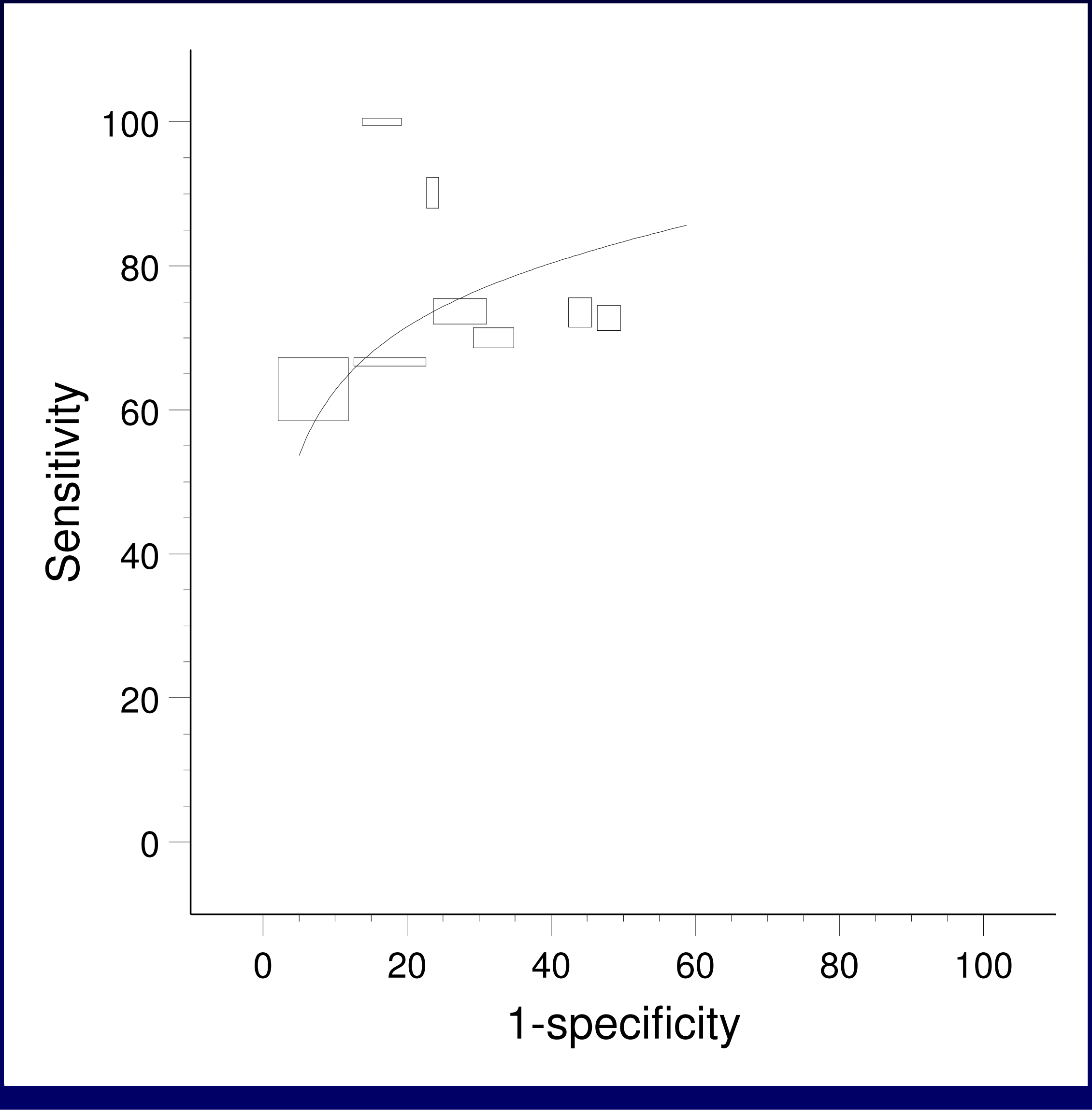
Ellipse around mean value



Prediction ellipse



Curves, Summary Points, Ellipses



Bad News

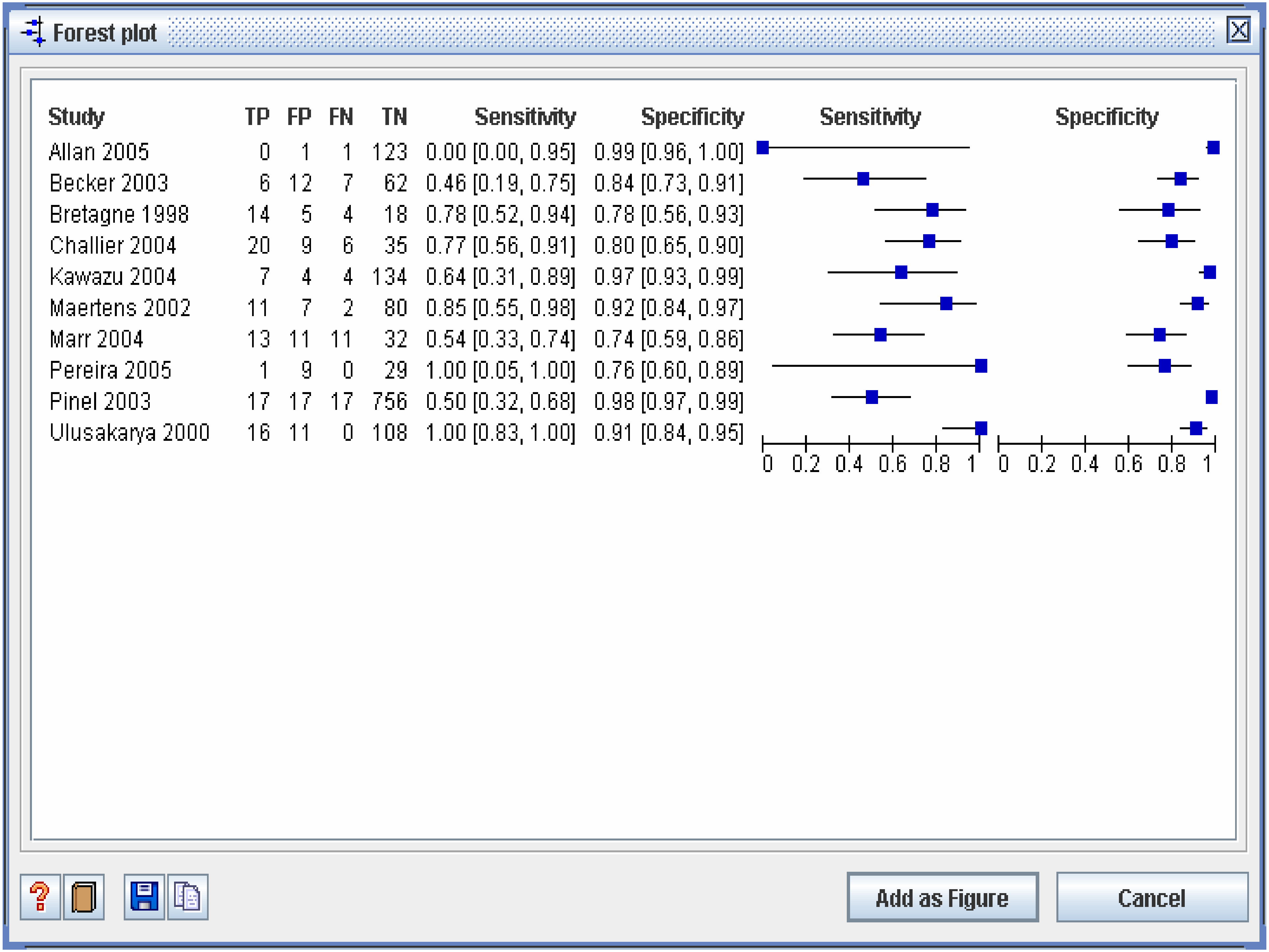
- Straightforward and most-frequently used method (Moses-Littenberg model) is statistically flawed
- Advanced models needed to make inferences (e.g. P-values) and to calculate appropriate confidence intervals
- Fitting and checking advanced models require statistical expertise
- Advanced methods not available in RevMan 5

Good News

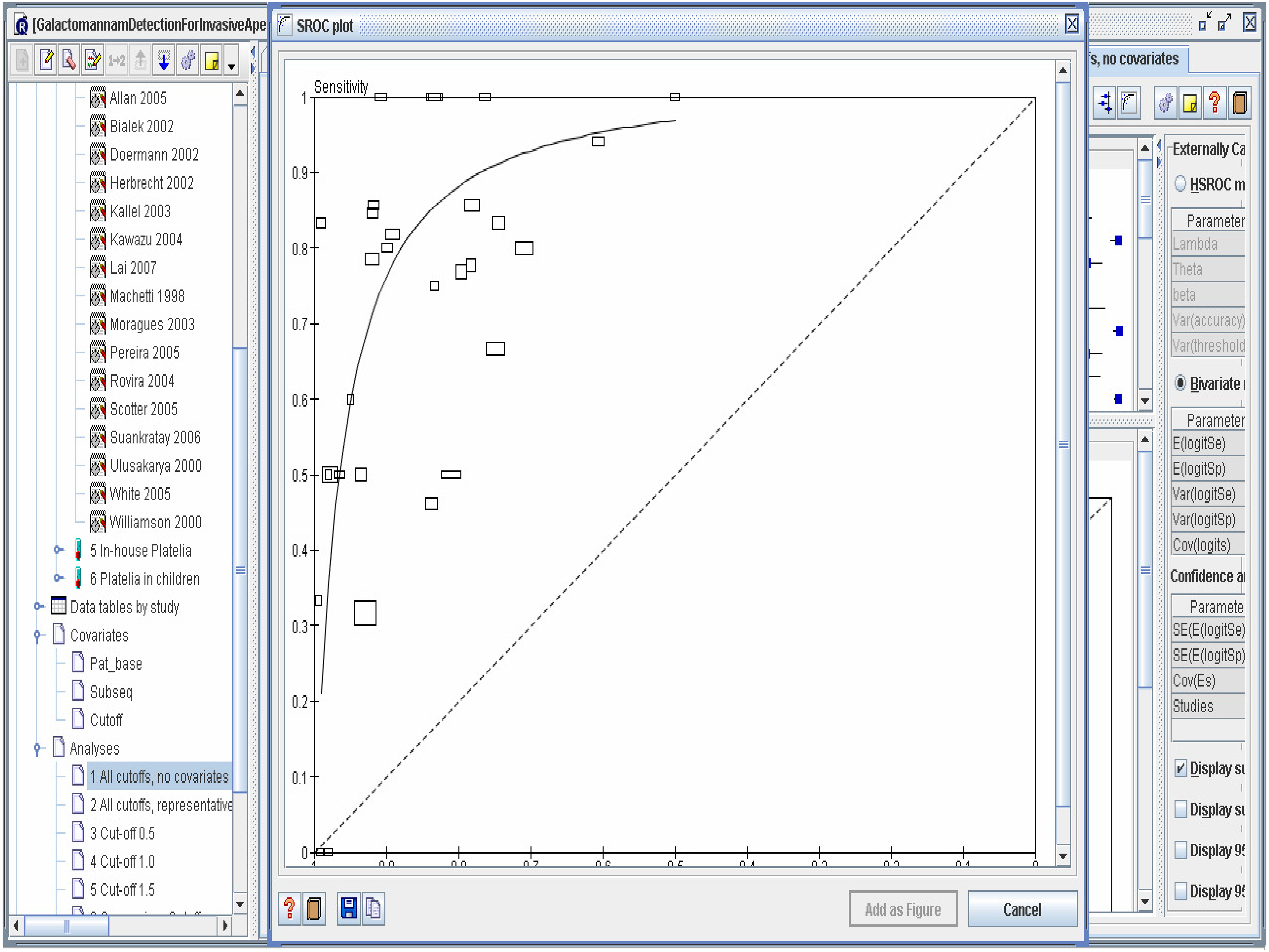
- Syntax to run more complex models in SAS, STATA, WINBUGS, S-PLUS, and R are available
- Results from these packages can be entered into RevMan 5 to make graphs

RevMan 5

- Perform descriptive analyses
- Estimates from hierarchical SROC or bivariate model can be imported into REVMAN to:
  - plot fitted SROC curve
  - display summary points
  - draw confidence or prediction ellipses



Study	TP	FP	FN	TN	Sensitivity	Specificity
Allan 2005	0	1	1	123	0.00 [0.00, 0.95]	0.99 [0.96, 1.00]
Becker 2003	6	12	7	62	0.46 [0.19, 0.75]	0.84 [0.73, 0.91]
Bretagne 1998	14	5	4	18	0.78 [0.52, 0.94]	0.78 [0.56, 0.93]
Challier 2004	20	9	6	35	0.77 [0.56, 0.91]	0.80 [0.65, 0.90]
Kawazu 2004	7	4	4	134	0.64 [0.31, 0.89]	0.97 [0.93, 0.99]
Maertens 2002	11	7	2	80	0.85 [0.55, 0.98]	0.92 [0.84, 0.97]
Marr 2004	13	11	11	32	0.54 [0.33, 0.74]	0.74 [0.59, 0.86]
Pereira 2005	1	9	0	29	1.00 [0.05, 1.00]	0.76 [0.60, 0.89]
Pinel 2003	17	17	17	756	0.50 [0.32, 0.68]	0.98 [0.97, 0.99]
Ulusakarya 2000	16	11	0	108	1.00 [0.83, 1.00]	0.91 [0.84, 0.95]



○ HSROC m

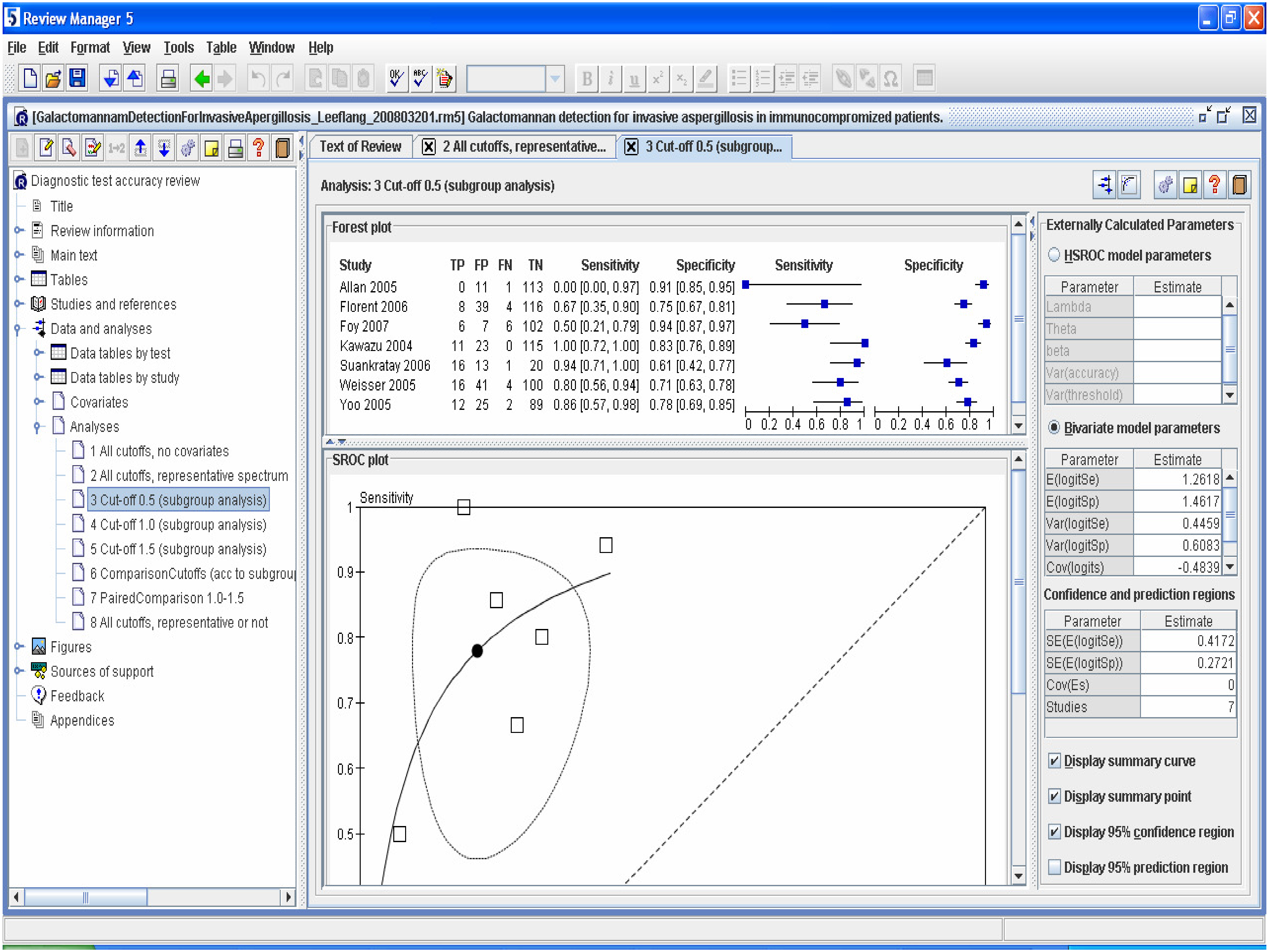
● Bivariate

☑ Display su

☐ Display su

☐ Display 95

☐ Display 95



Analysis: 3 Cut-off 0.5 (subgroup analysis)

Forest plot

Study	TP	FP	FN	TN	Sensitivity	Specificity
Allan 2005	0	11	1	113	0.00 [0.00, 0.97]	0.91 [0.85, 0.95]
Florent 2006	8	39	4	116	0.67 [0.35, 0.90]	0.75 [0.67, 0.81]
Foy 2007	6	7	6	102	0.50 [0.21, 0.79]	0.94 [0.87, 0.97]
Kawazu 2004	11	23	0	115	1.00 [0.72, 1.00]	0.83 [0.76, 0.89]
Suankratay 2006	16	13	1	20	0.94 [0.71, 1.00]	0.61 [0.42, 0.77]
Weisser 2005	16	41	4	100	0.80 [0.56, 0.94]	0.71 [0.63, 0.78]
Yoo 2005	12	25	2	89	0.86 [0.57, 0.98]	0.78 [0.69, 0.85]

Externally Calculated Parameters

○ HSROC model parameters

Parameter	Estimate
Lambda	
Theta	
beta	
Var(accuracy)	
Var(threshold)	

● Bivariate model parameters

Parameter	Estimate
E(logitSe)	1.2618
E(logitSp)	1.4617
Var(logitSe)	0.4459
Var(logitSp)	0.6083
Cov(logits)	-0.4839

Confidence and prediction regions

Parameter	Estimate
SE(E(logitSe))	0.4172
SE(E(logitSp))	0.2721
Cov(Es)	0
Studies	7

☑ Display summary curve

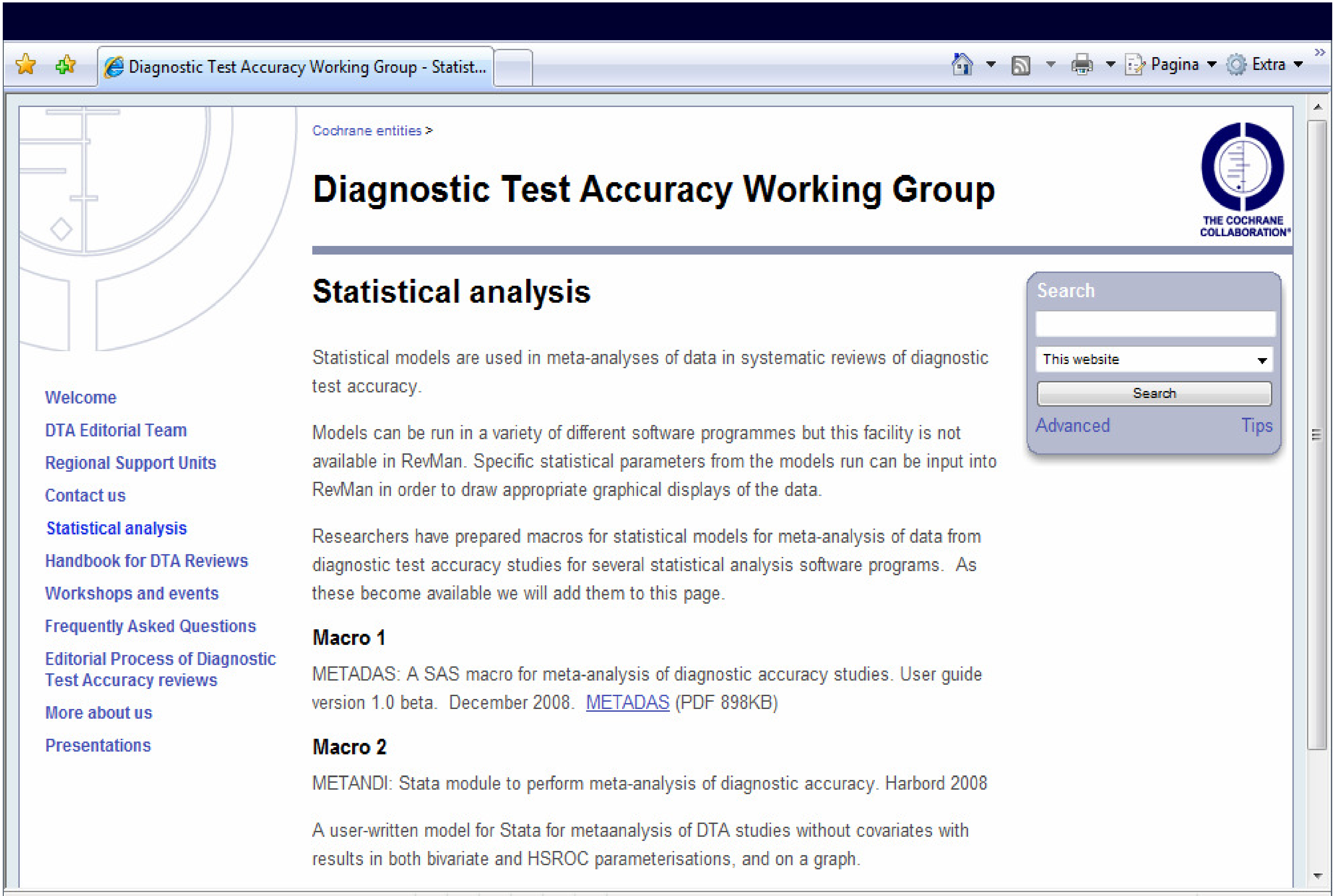
☑ Display summary point

☑ Display 95% confidence region

☐ Display 95% prediction region

Support

- Cochrane support:
  - CESU and UKSU
  - explanatory papers
  - pilot reviews
  - editorial process with specific attention to meta-analysis
  - workshops at Cochrane Colloquia
- Courses
- Overview on website Diagnostic Test Accuracy Working Group (http://srdta.cochrane.org)



Cochrane entities >

Diagnostic Test Accuracy Working Group

- Welcome
- DTA Editorial Team
- Regional Support Units
- Contact us
- Statistical analysis
- Handbook for DTA Reviews
- Workshops and events
- Frequently Asked Questions
- Editorial Process of Diagnostic Test Accuracy reviews
- More about us
- Presentations

Statistical analysis

Statistical models are used in meta-analyses of data in systematic reviews of diagnostic test accuracy.

Models can be run in a variety of different software programmes but this facility is not available in RevMan. Specific statistical parameters from the models run can be input into RevMan in order to draw appropriate graphical displays of the data.

Researchers have prepared macros for statistical models for meta-analysis of data from diagnostic test accuracy studies for several statistical analysis software programs. As these become available we will add them to this page.

Macro 1

METADAS: A SAS macro for meta-analysis of diagnostic accuracy studies. User guide version 1.0 beta. December 2008. METADAS (PDF 898KB)

Macro 2

METANDI: Stata module to perform meta-analysis of diagnostic accuracy. Harbord 2008

A user-written model for Stata for metaanalysis of DTA studies without covariates with results in both bivariate and HSROC parameterisations, and on a graph.

Take Home Messages

- Two potentially correlated outcome measures require more complex statistical models
- Moses-Littenberg model is not appropriate for formal testing
- Bivariate and hierarchical summary ROC model are sound, powerful and flexible models
- These models can not be fitted in RevMan, but results can be incorporated
- Statistical expertise required in review team

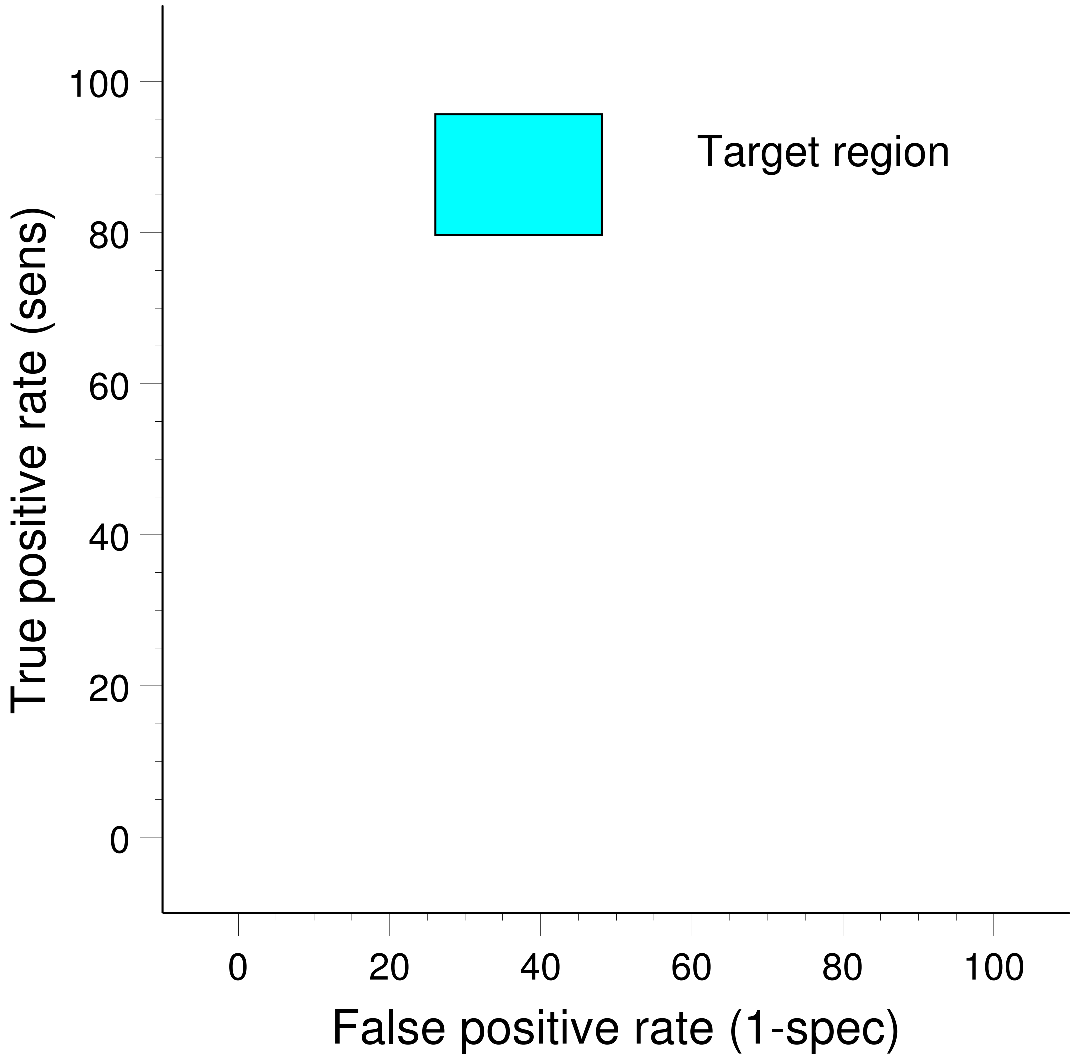
Meta-analysis of Accuracy Studies

- Results often highly heterogeneous
  - differences in design and conduct
  - differences in verification
  - differences in spectrum
  - differences in technology of tests or test execution
  - differences in threshold
  - chance variation

Powerful and Flexible Models

- Examples of multivariate meta-analysis: all advantages apply
- Extension with study-level covariates to explain heterogeneity in results or differences in accuracy between test in accuracy
- Separate effects on sensitivity and specificity
- Testing of joint parameters
- Software: need for non-linear mixed models in SAS, STATA, R, S, WinBugs

Other View


Target region
True positive rate (sens)
100
80
60
40
20
0
0
20
40
60
80
100
False positive rate (1-spec)