Cochrane Diagnostic Test Accuracy Reviews

Meta-analysis of diagnostic accuracy studies

Yemisi Takwoingi
UK Support Unit
Department of Public Health, Epidemiology & Biostatistics
University of Birmingham

Diagnostic Test Accuracy Reviews

- Framing the question
- Identification and selection of studies
- Quality assessment
- Data extraction
- Data analysis
- Interpretation of the results

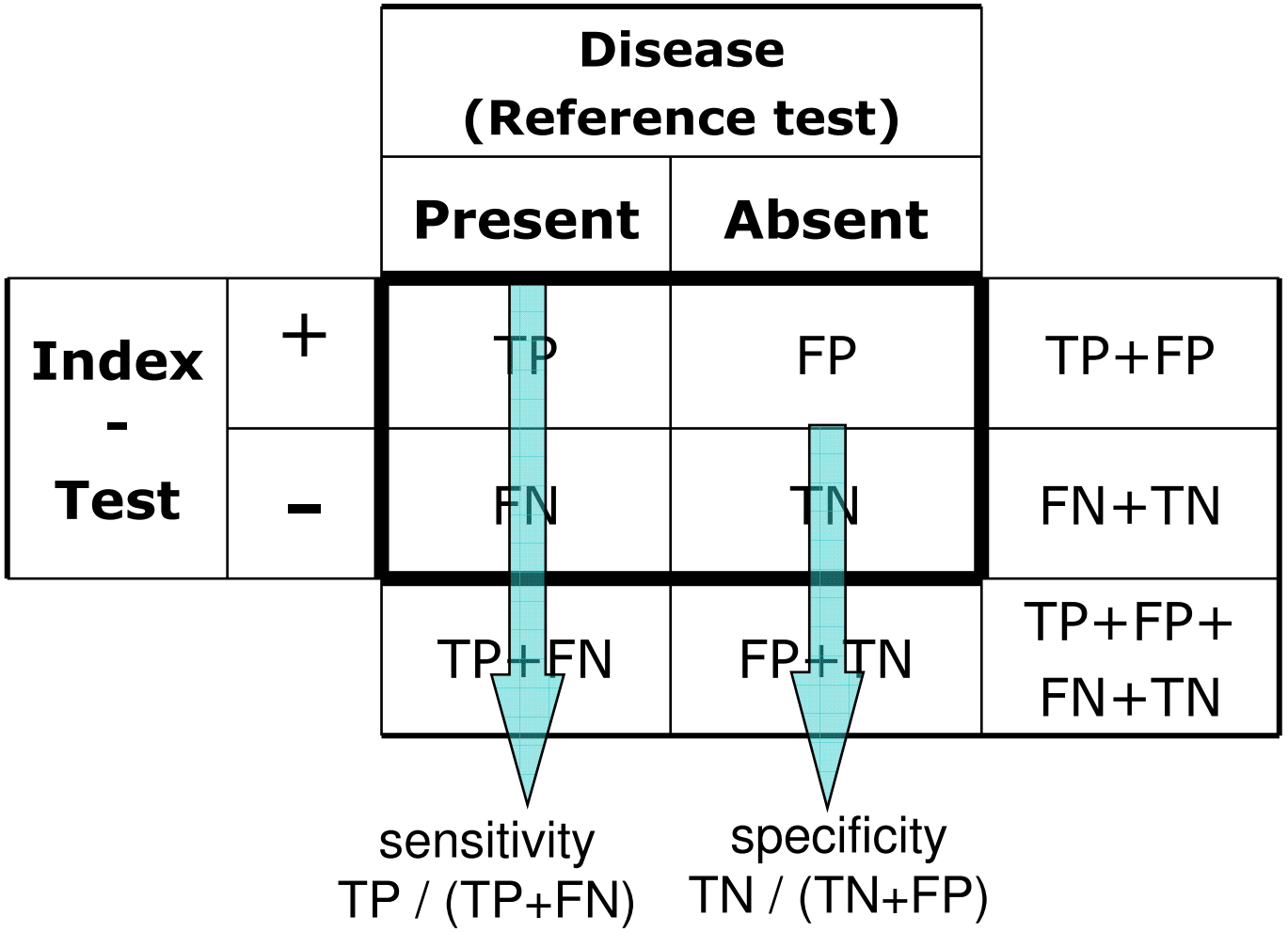
2x2 Table

		Disease (Reference test)		
		Present	**Absent**	
Index - Test	+	TP	FP	TP+FP
	–	FN	TN	FN+TN
		TP+FN	FP+TN	TP+FP+ FN+TN

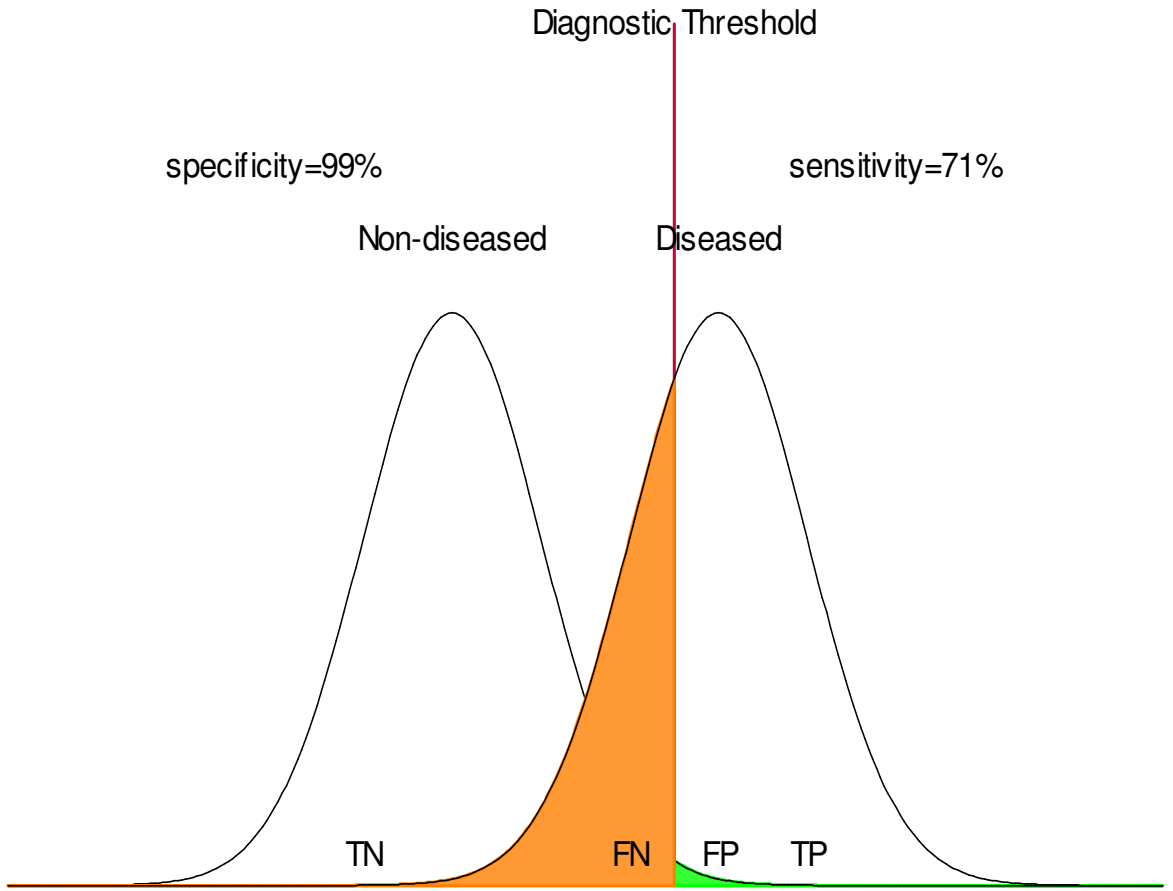
Test accuracy

- Sensitivity
  - describes the proportion of patients with the target condition with index test results above a threshold
- Specificity
  - describes the proportion of patients without the target condition with index test results below a threshold
- Thresholds vary between studies
- Same threshold can imply different sensitivities and specificities in different groups

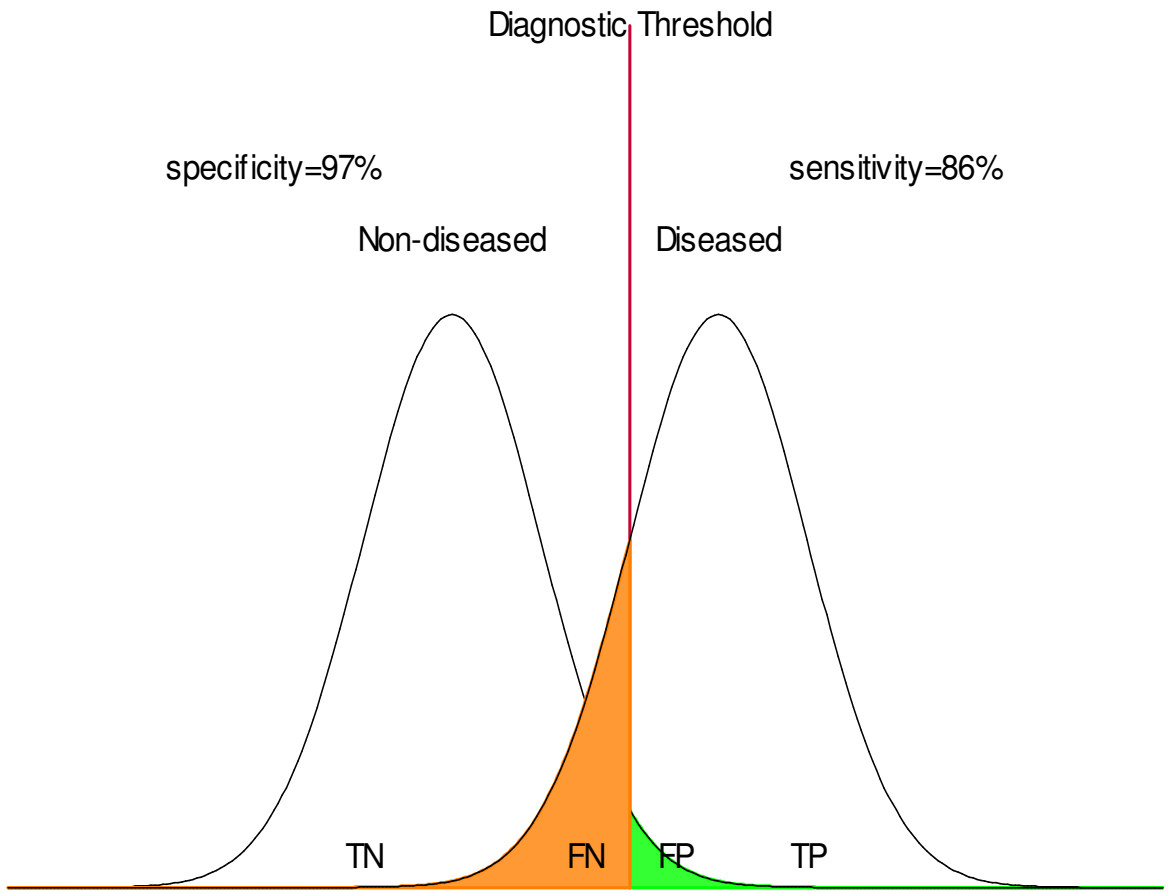
2x2 Table


Disease
(Reference test)
Present
Absent
Index
-
Test
+
–
TP
FP
TP+FP
FN
TN
FN+TN
TP+FN
FP+TN
TP+FP+
FN+TN
sensitivity
TP / (TP+FN)
specificity
TN / (TN+FP)

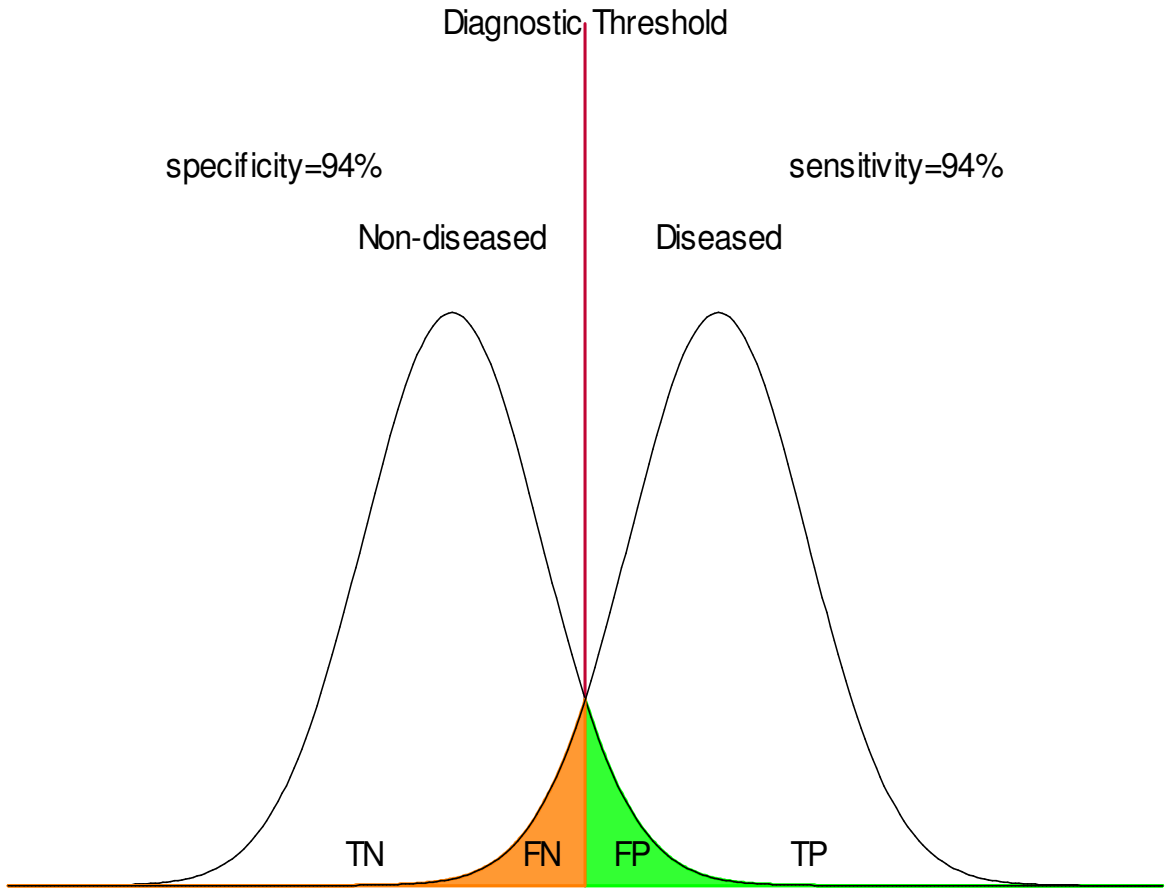
Heterogeneity in threshold



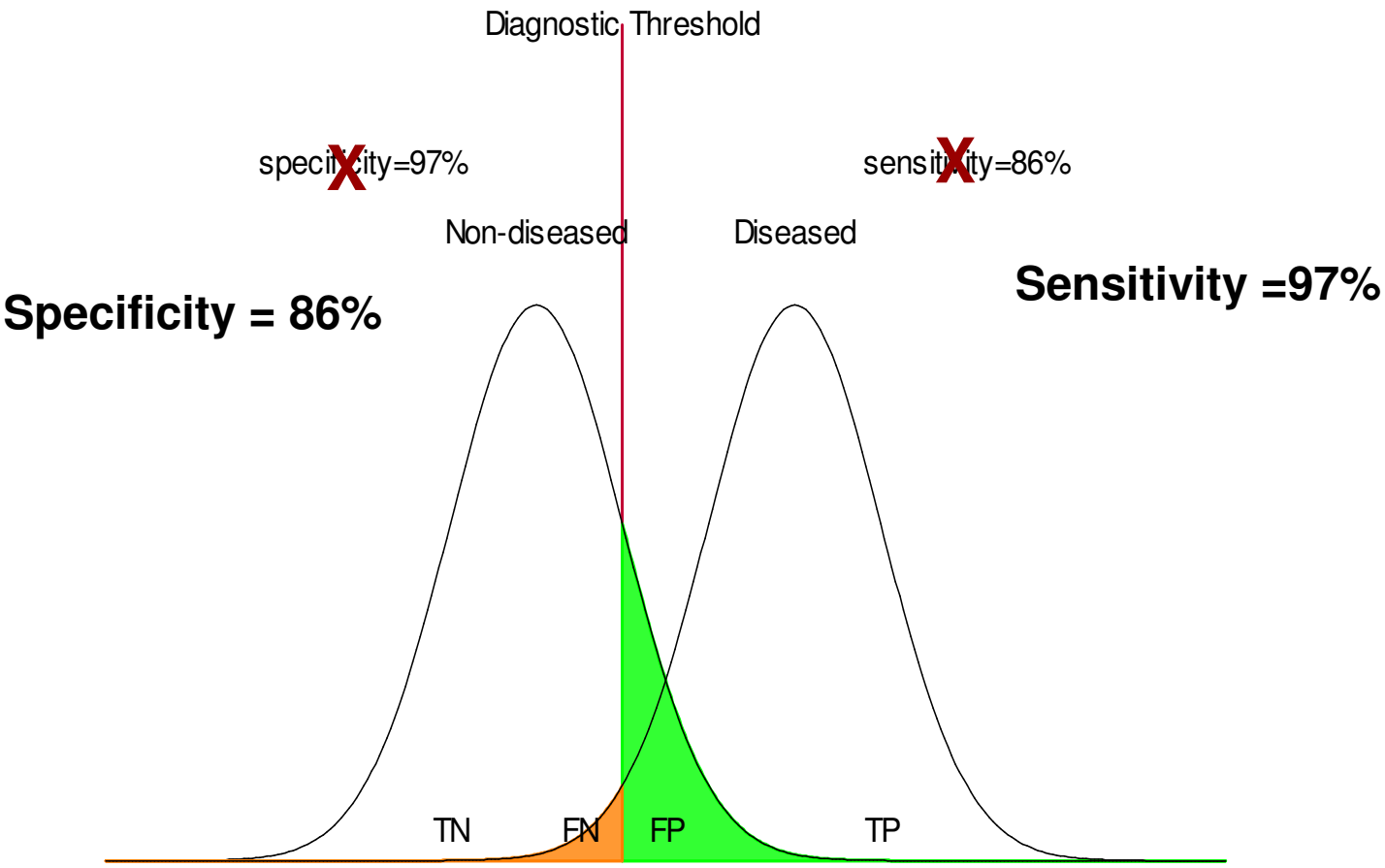
Heterogeneity in threshold



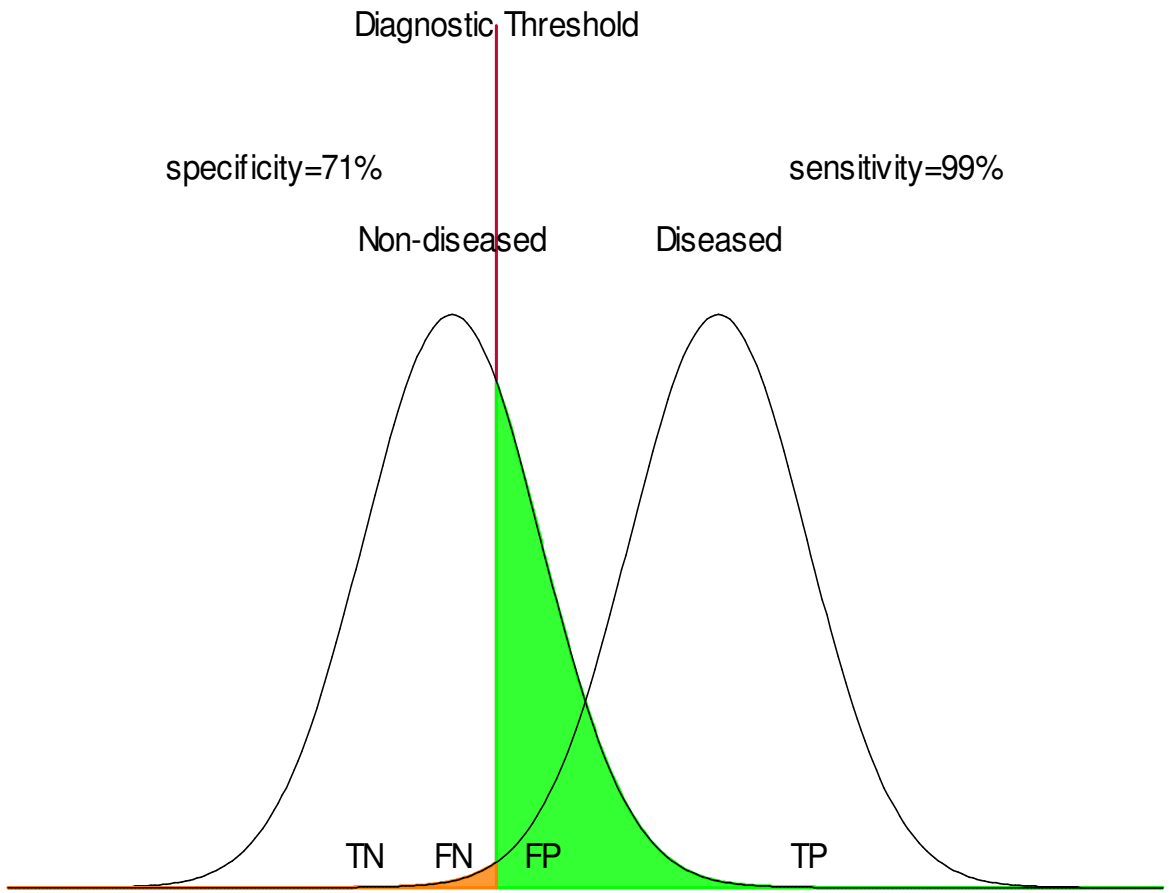
Heterogeneity in threshold



Heterogeneity in threshold

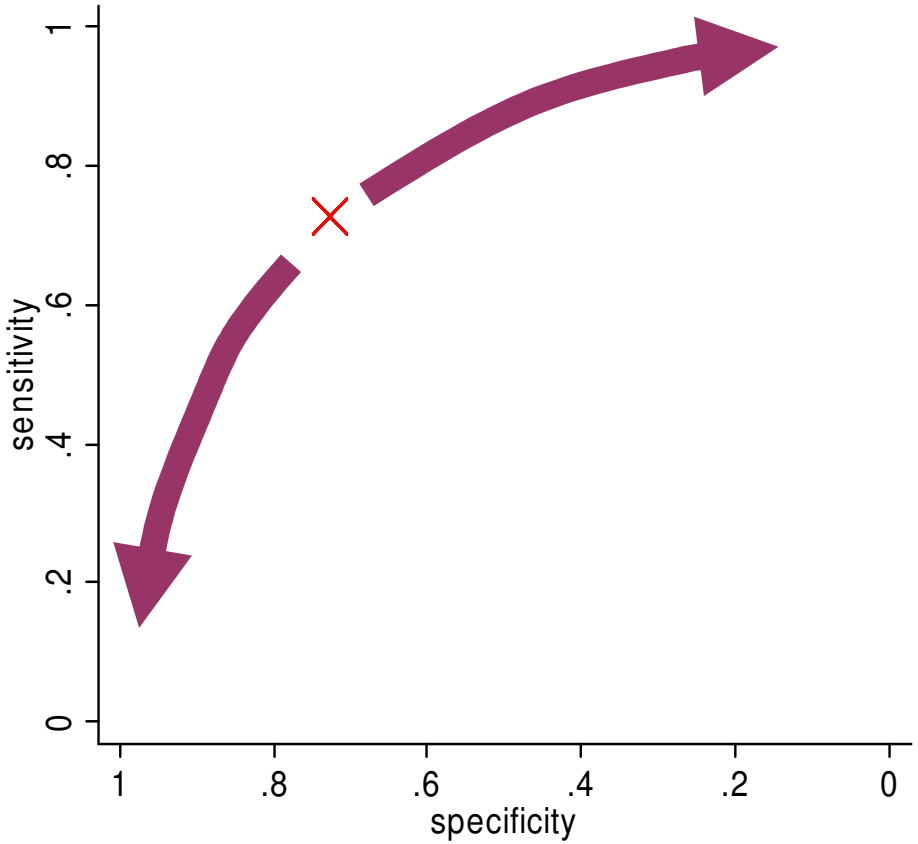


Heterogeneity in threshold



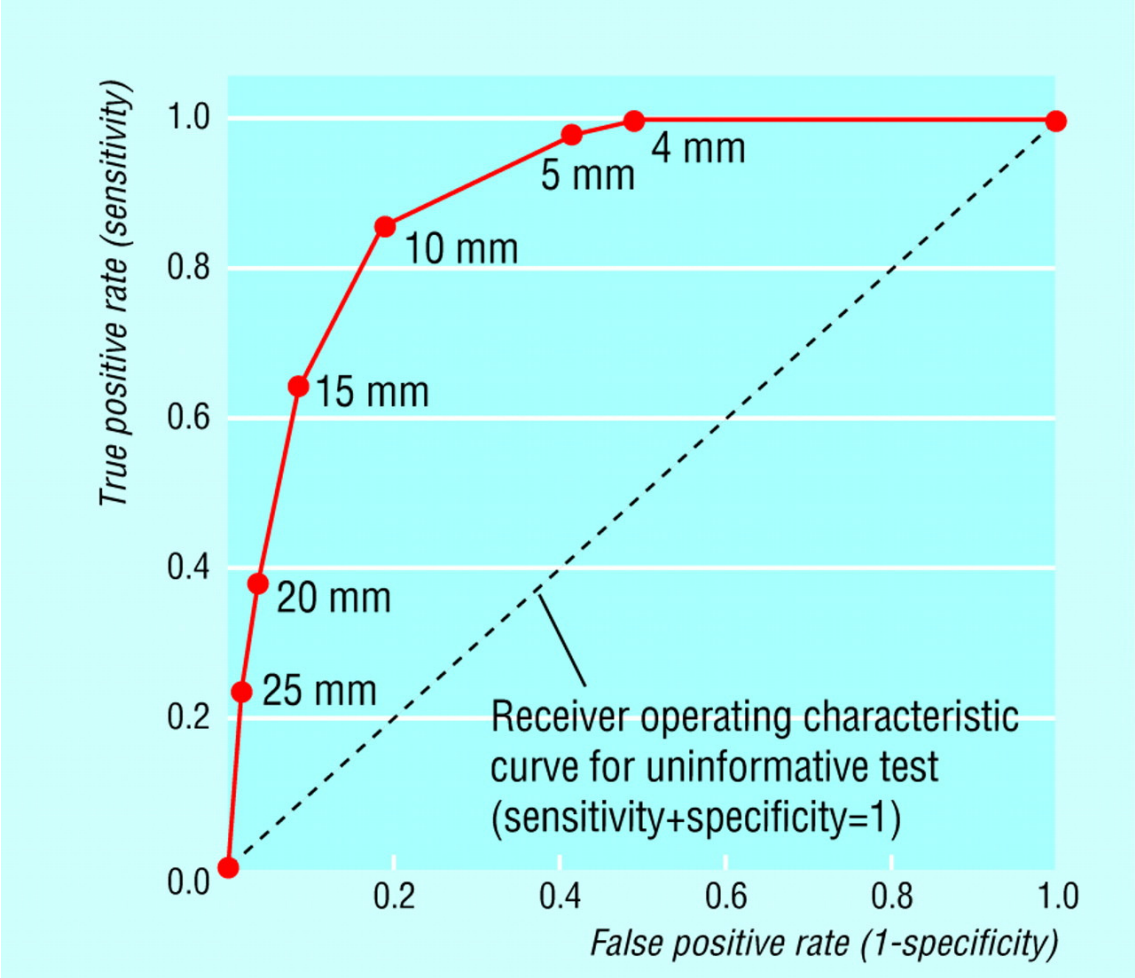
Threshold effects

Decreasing threshold increases sensitivity but decreases specificity



Increasing threshold increases specificity but decreases sensitivity

Receiver characteristic operating (*ROC*) curve



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

The ROC curve represents the relationship between the true positive rate (TPR) and the false positive rate (FPR) of the test at various thresholds used to distinguish disease cases from non-cases.

Diagnostic odds ratios

Ratio of the odds of positivity in the diseased to the odds of positivity in the non-diseased

$$Diagnostic\ OR = \frac{TP \times TN}{FP \times FN}$$

$$DOR = \frac{\left(\frac{sensitivity}{1-sensitivity}\right)}{\left(\frac{1-specificity}{specificity}\right)} = \frac{LR+ve}{LR-ve}$$

Diagnostic odds ratios

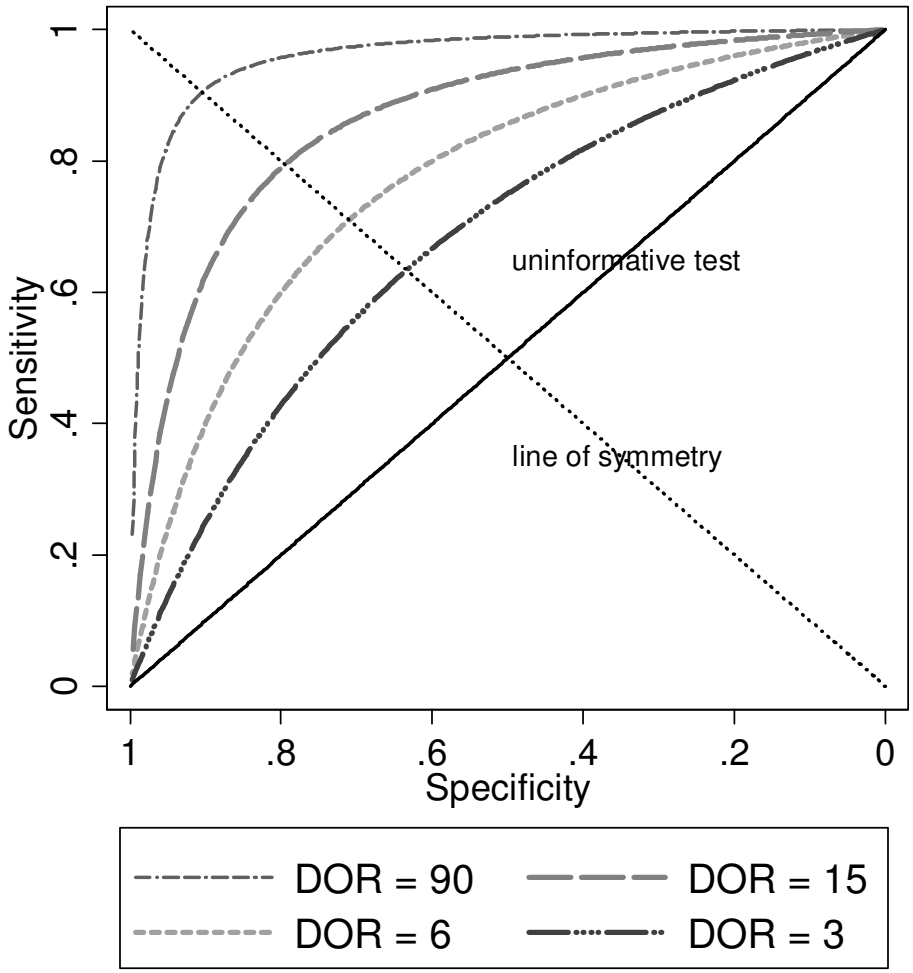
		Cervical Cancer (Biopsy)		
		Present	**Absent**	
HPV Test	+	65	93	158
	-	7	161	198
		72	254	356

$$\text{DOR} = \frac{65 \times 161}{93 \times 7} = 16$$

Diagnostic odds ratios

	Sensitivity						
Specificity	*50%*	*60%*	*70%*	*80%*	*90%*	*95%*	*99%*
50%	1	2	2	4	9	19	99
60%	2	2	4	6	14	29	149
70%	2	4	5	9	21	44	231
80%	4	6	9	16	36	76	396
90%	9	14	21	36	81	171	891
95%	19	29	44	76	171	361	1881
99%	99	149	231	396	891	1881	9801

Symmetrical *ROC* curves and diagnostic odds ratios



As DOR increases, the ROC curve moves closer to its ideal position near the upper-left corner.

ROC curve is asymmetric when test accuracy varies with threshold

The meta-analysis process

1. Calculation of an overall summary (average) of high precision, coherent with all observed data
2. Typically a “weighted average” is used where more informative (larger) studies have more say
3. Assess the degree to which the study results deviate from the overall summary
4. Investigate possible explanations for the deviations

Meta-analysis of studies of diagnostic accuracy

- Pair of related summary statistics for each study
  - Sensitivity and specificity
  - Positive and negative likelihood ratios
- Threshold effects induce correlations between sensitivity and specificity
- Heterogeneity is the norm not the exception
  - Substantial variation in sensitivity and specificity are noted in most reviews

Statistical modelling of ROC curves

- statisticians like straight lines with axes that are independent variables
- first calculate the logits of TPR and FPR
- and then graph the difference against their sum

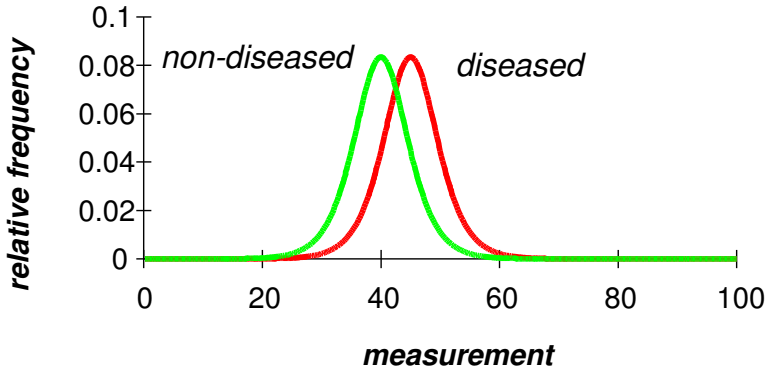
$$\text{logit}(TPR) = \ln\left(\frac{TPR}{1-TPR}\right)$$

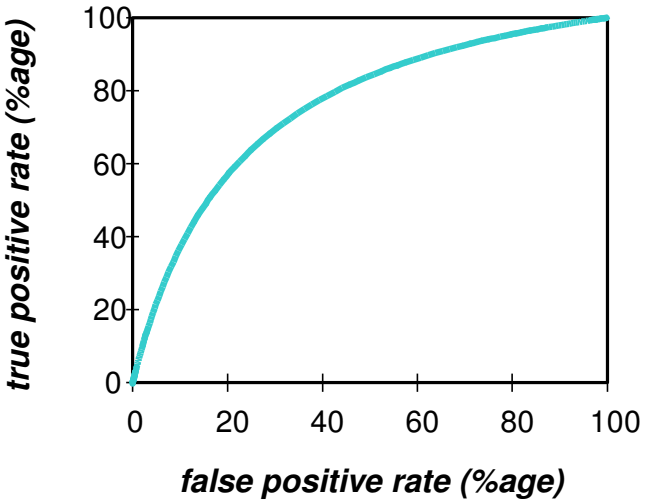
$$\text{logit}(FPR) = \ln\left(\frac{FPR}{1-FPR}\right)$$

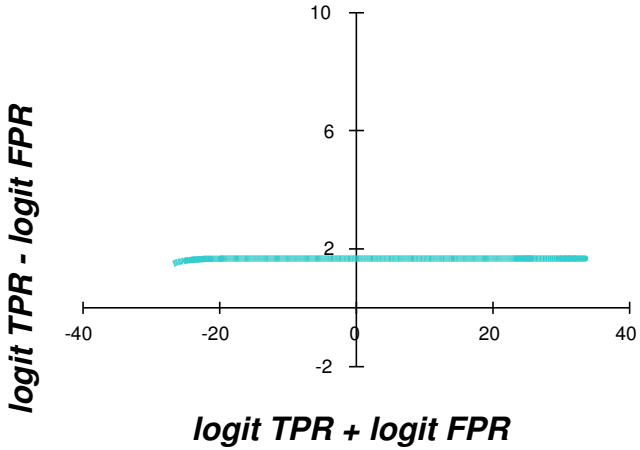
$$S = \text{logit}(TPR) + \text{logit}(FPR)$$

$$D = \text{logit}(TPR) - \text{logit}(FPR)$$

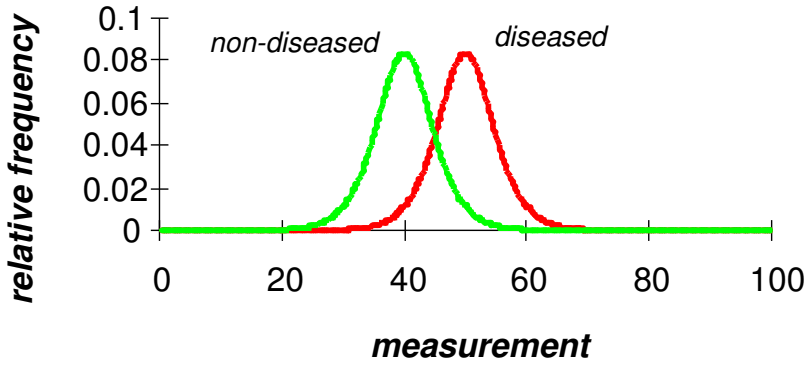
ROC curve and logit difference and sum plot: small difference, same spread

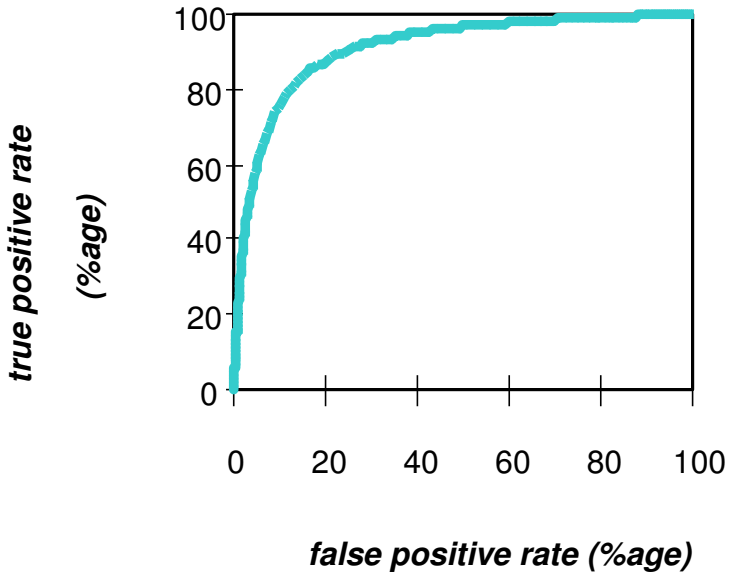


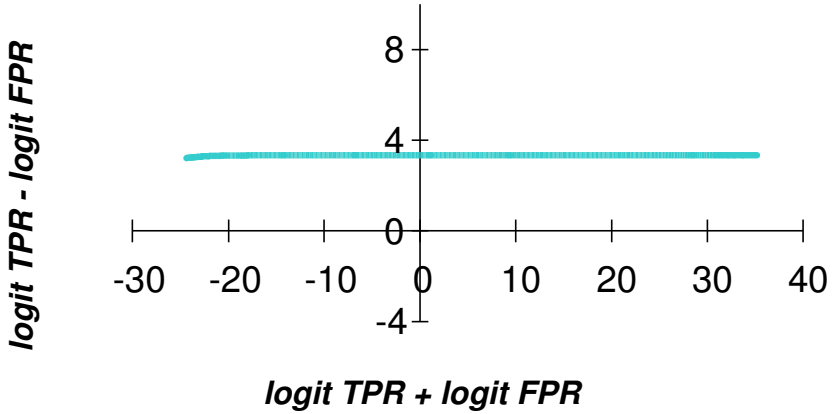




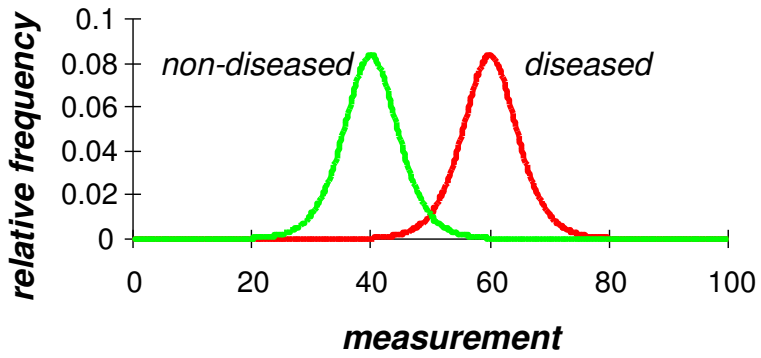
ROC curve and logit difference and sum plot: moderate difference, same spread

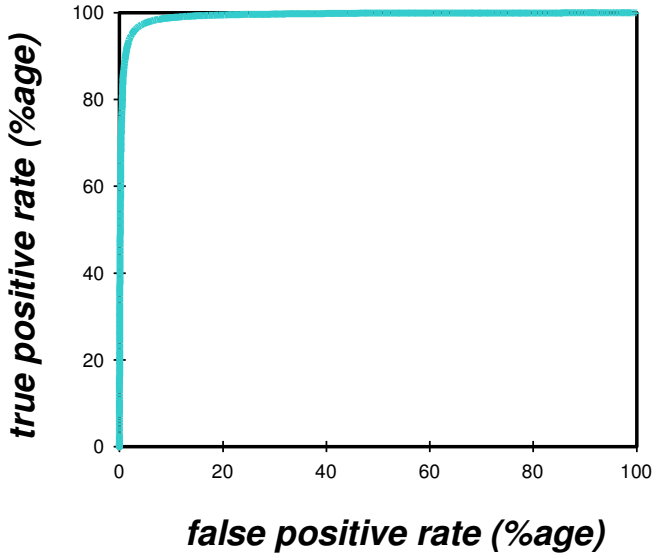


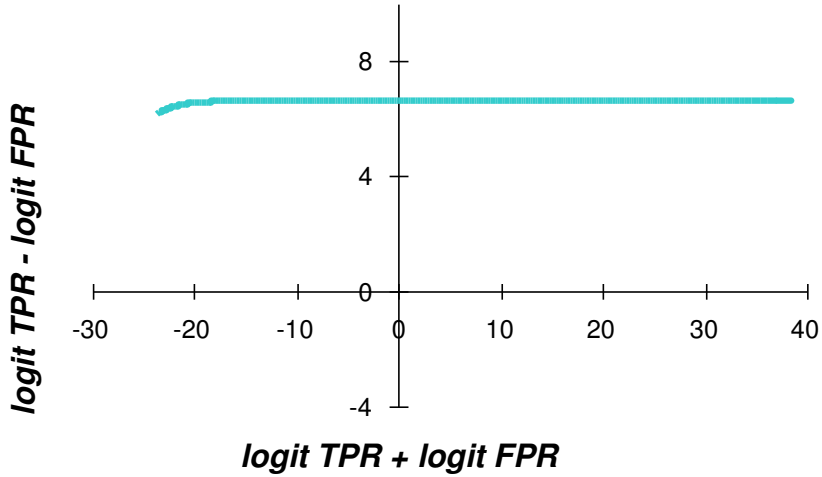




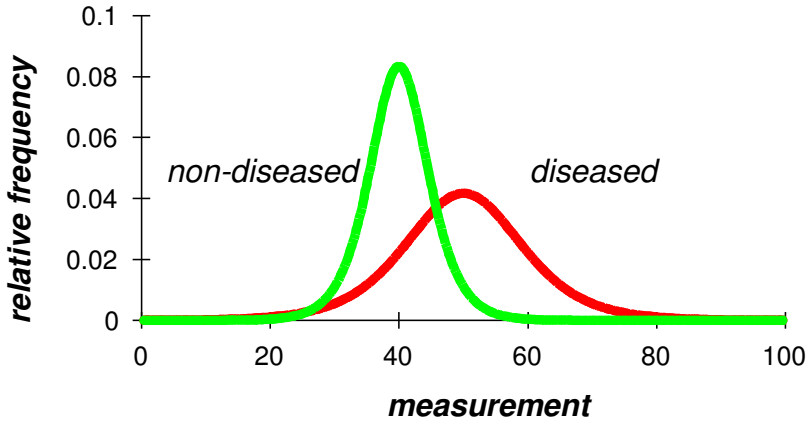
ROC curve and logit difference and sum plot: large difference, same spread

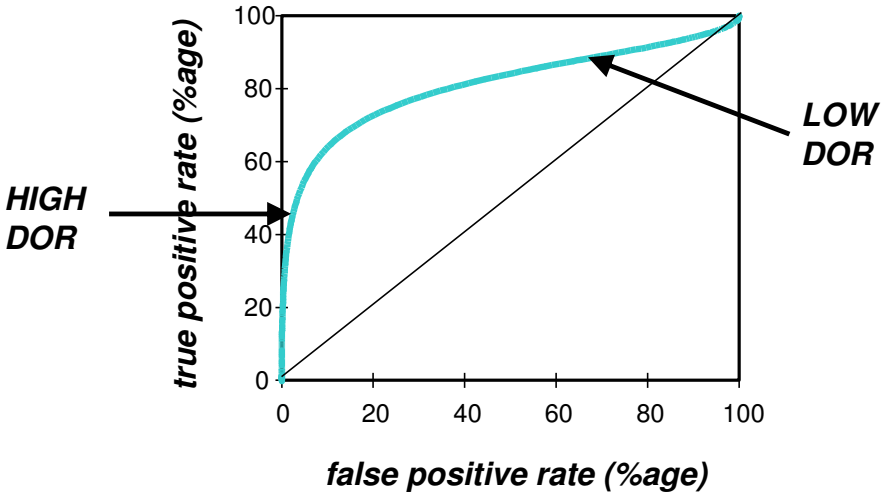


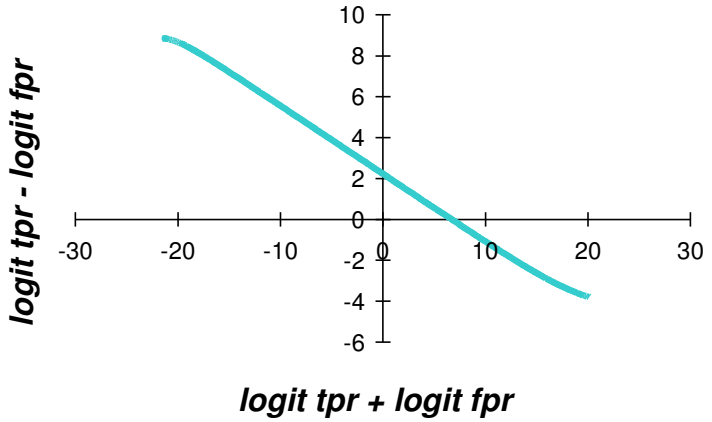




ROC curve and logit difference and sum plot: moderate difference, unequal spread







Moses-Littenberg SROC method

- Regression models can be used to fit the straight lines to model relationship between test accuracy and test threshold

$$D = a + bS$$

  - Outcome variable D is the difference in the logits
  - Explanatory variable S is the sum of the logits
  - Ordinary or weighted regression – weighted by sample size or by inverse variance of the log of the DOR

- What do the axes mean?
  - Difference in logits is the log of the DOR
  - Sum of the logits is a marker of diagnostic threshold

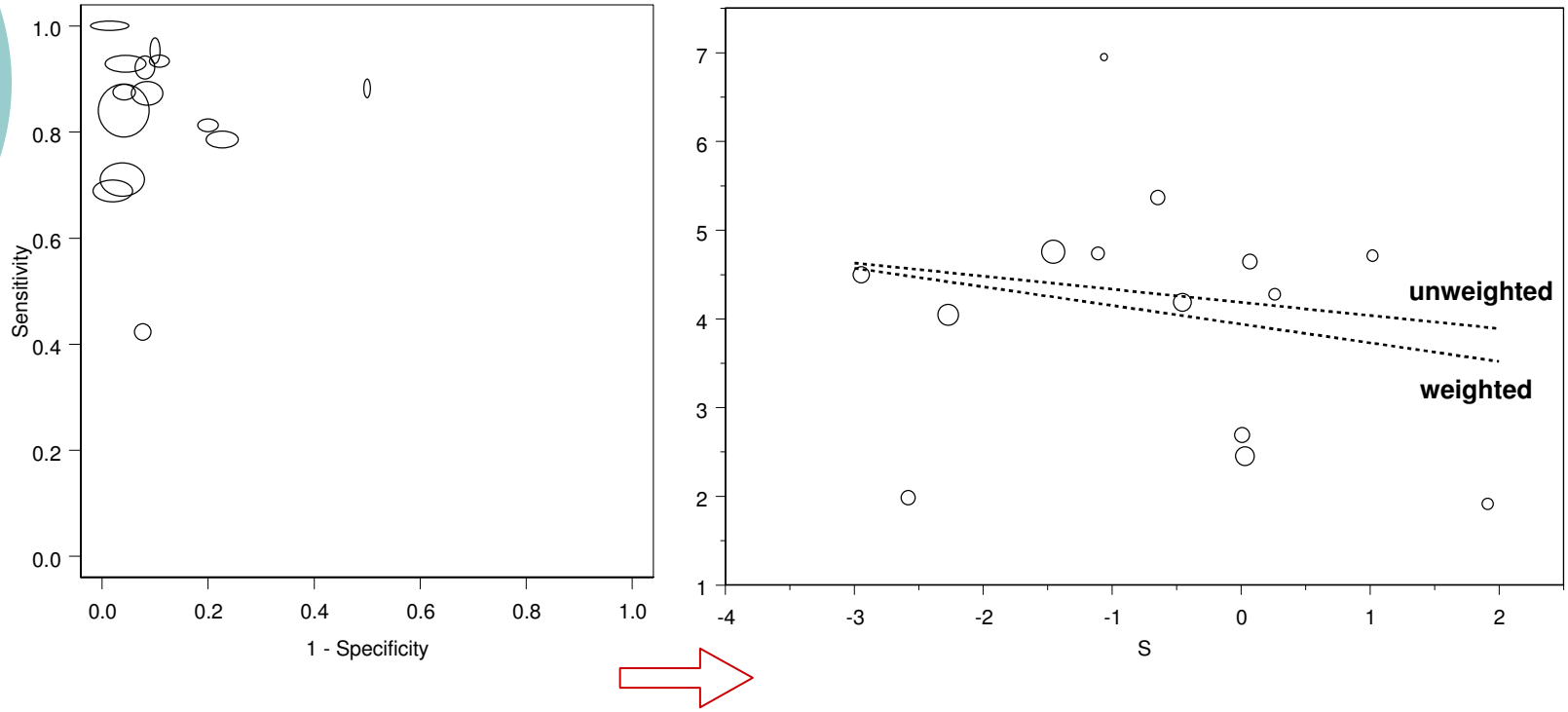
Producing summary ROC curves

- Transform back to the ROC dimensions

$$TPR = \frac{1}{1 + \frac{1}{e^{a/(1-b)}} \times \left(\frac{FPR}{1 - FPR} \right)^{\frac{1+b}{1-b}}}$$

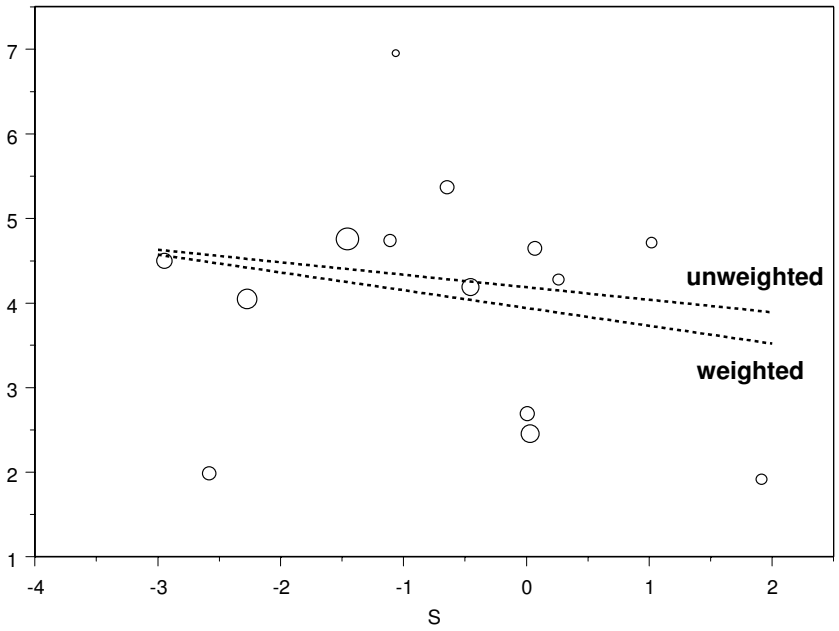
- where ‘a’ is the intercept, ‘b’ is the slope
  - when the ROC curve is symmetrical, b=0 and the equation is simpler

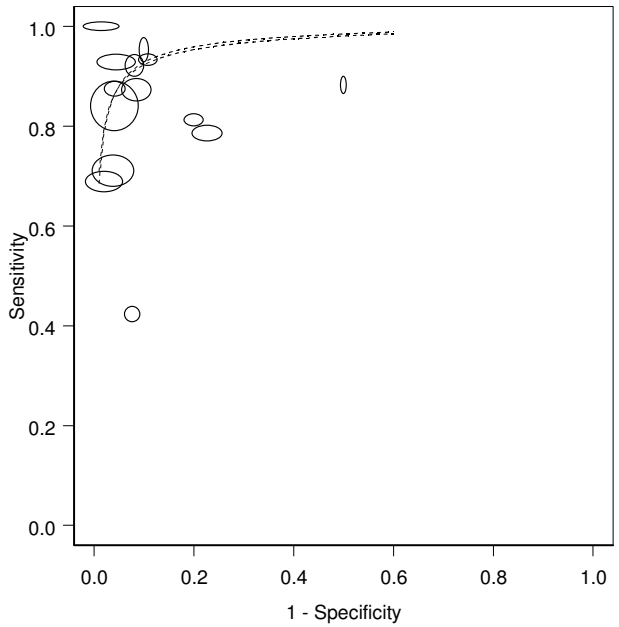
SROC regression: PSV example



Transformation linearizes relationship between accuracy and threshold so that linear regression can be used

PSV example *cont.*





inverse transformation

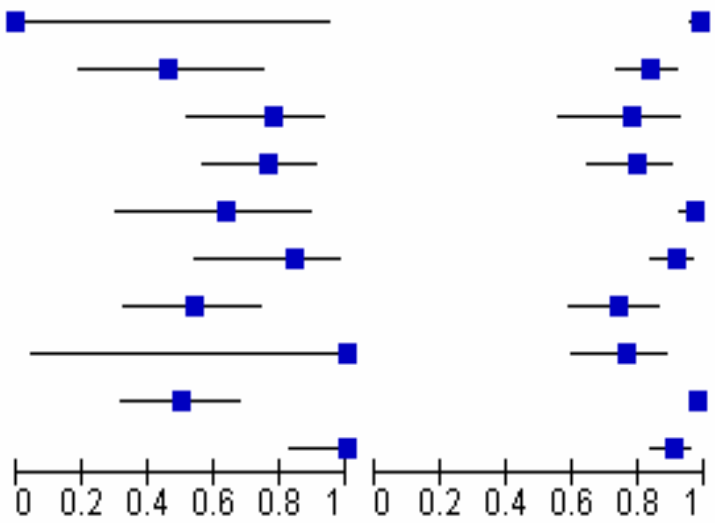
The SROC curve is produced by using the estimates of a and b to compute the expected sensitivity (*tpr)* across a range of values for 1-specificity (*fpr)*

RevMan 5: data and analyses

- Add data by test or study
- Add covariate
  - Study or test level
  - Continuous or categorical
- Add analysis
  - Single test
  - Multiple tests
  - Paired data

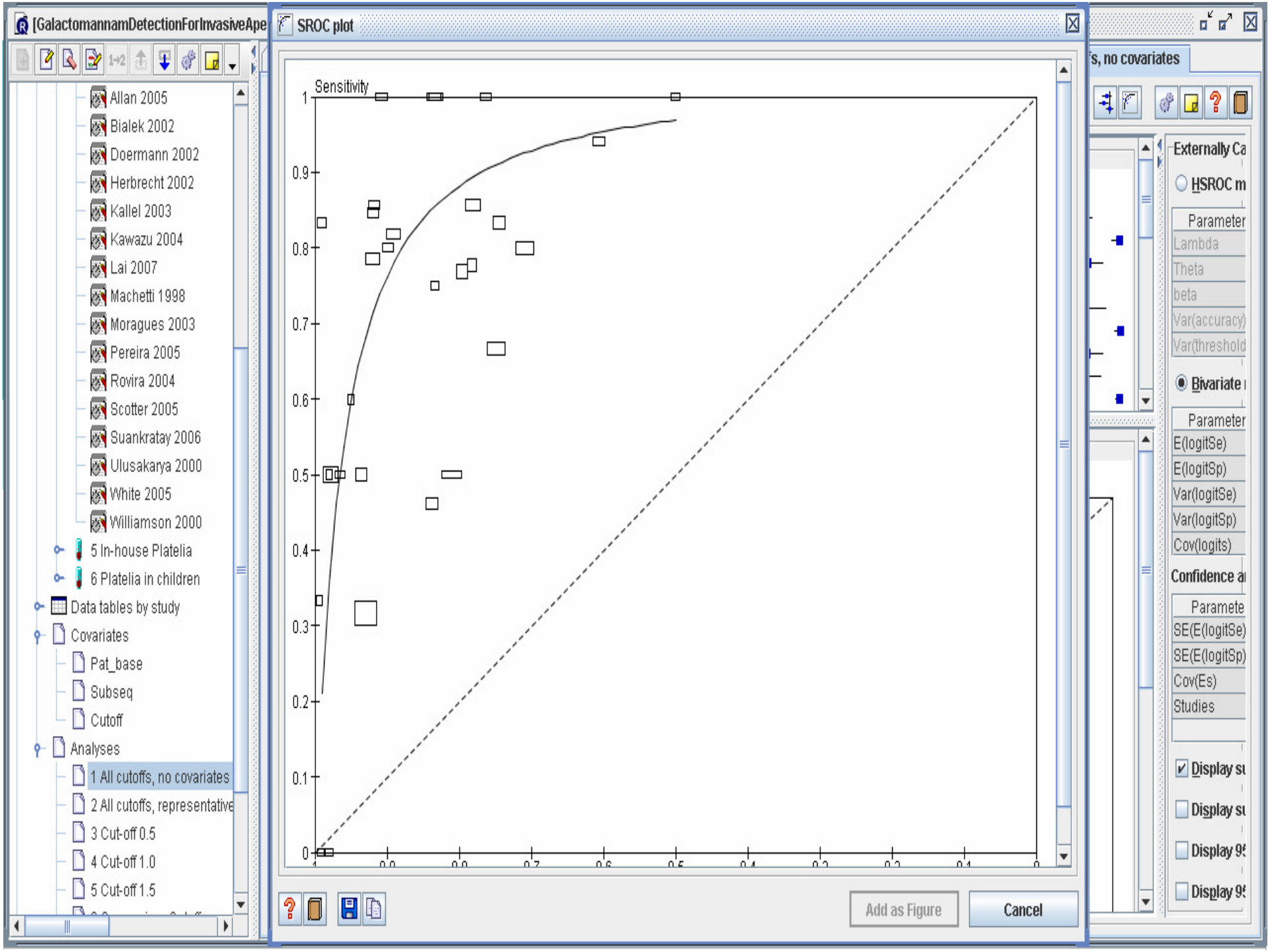
Forest plot

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]



Add as Figure

Cancel


[GalactomannamDetectionForInvasiveApe
Allan 2005
Bialek 2002
Doermann 2002
Herbrecht 2002
Kallel 2003
Kawazu 2004
Lai 2007
Machetti 1998
Moragues 2003
Pereira 2005
Rovira 2004
Scotter 2005
Suankratay 2006
Ulusakarya 2000
White 2005
Williamson 2000
5 In-house Platelia
6 Platelia in children
Data tables by study
Covariates
Pat_base
Subseq
Cutoff
Analyses
1 All cutoffs, no covariates
2 All cutoffs, representative
3 Cut-off 0.5
4 Cut-off 1.0
5 Cut-off 1.5
SROC plot
Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Add as Figure
Cancel
's, no covariates
Externally Ca
HSROC m
Parameter
Lambda
Theta
beta
Var(accuracy)
Var(threshold
Bivariate
Parameter
E(logitSe)
E(logitSp)
Var(logitSe)
Var(logitSp)
Cov(logits)
Confidence a
Paramete
SE(E(logitSe)
SE(E(logitSp)
Cov(Es)
Studies
Display su
Display su
Display 9
Display 9

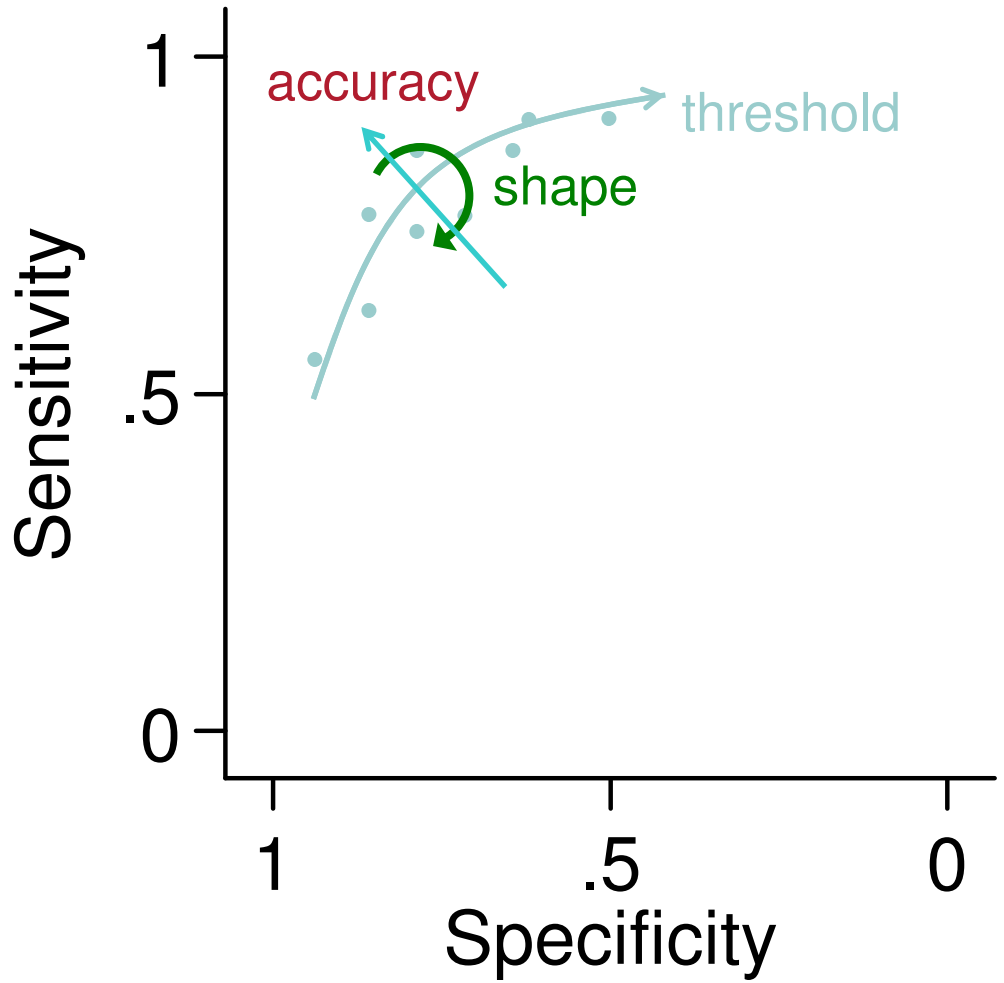
Problems with the Moses-Littenberg SROC method

- Poor estimation
  - Tends to underestimate test accuracy due to zero-cell corrections and bias in weights
- Validity of significance tests
  - Sampling variability in individual studies not properly taken into account
  - P-values and confidence intervals erroneous
- Operating points
  - knowing average sensitivity/specificity is important but cannot be obtained
  - Sensitivity for a given specificity can be estimated

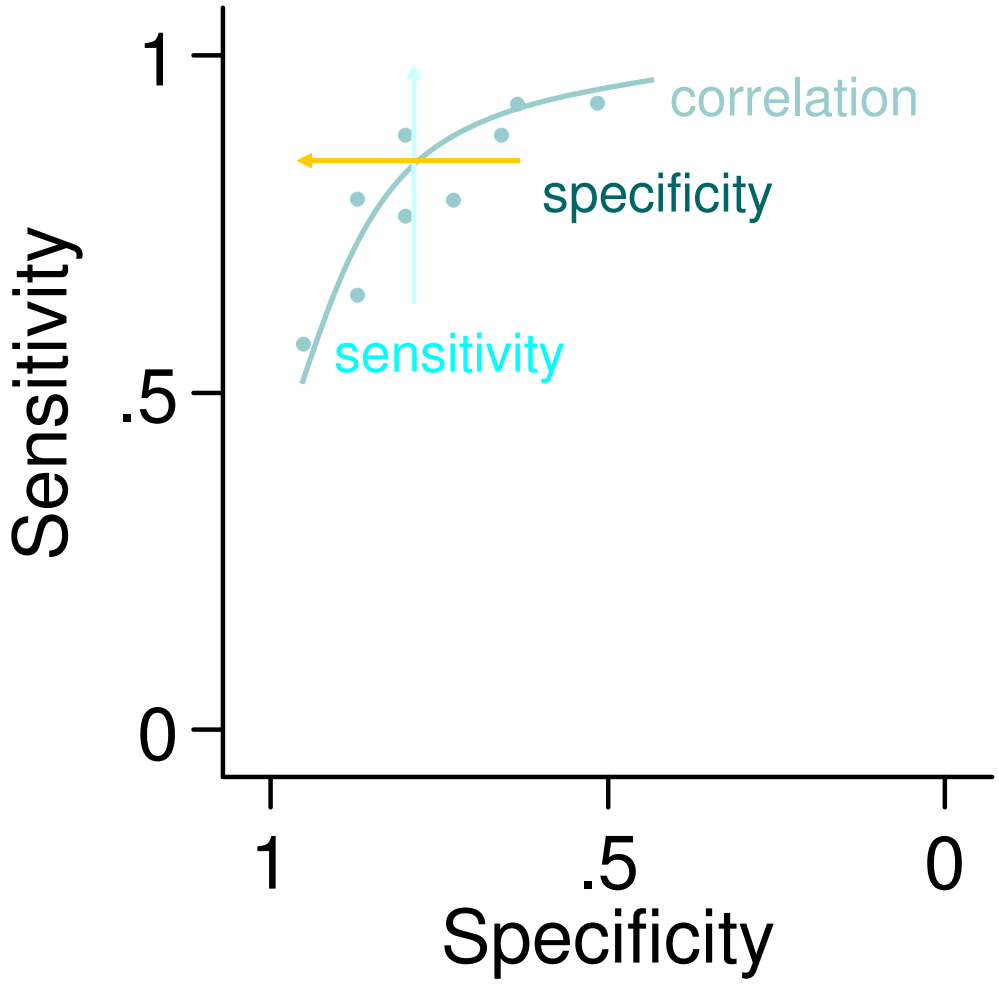
Advanced models – HSROC and Bivariate methods

- Hierarchical / multi-level
  - allows for both within and between study variability, and within study correlations between diseased and non-diseased groups
- Logistic
  - correctly models sampling uncertainty in the true positive proportion and the false positive proportion
  - no zero cell adjustments needed
- Random effects
  - allows for heterogeneity between studies
- Regression models
  - used to investigate sources of heterogeneity

Hierarchical SROC model



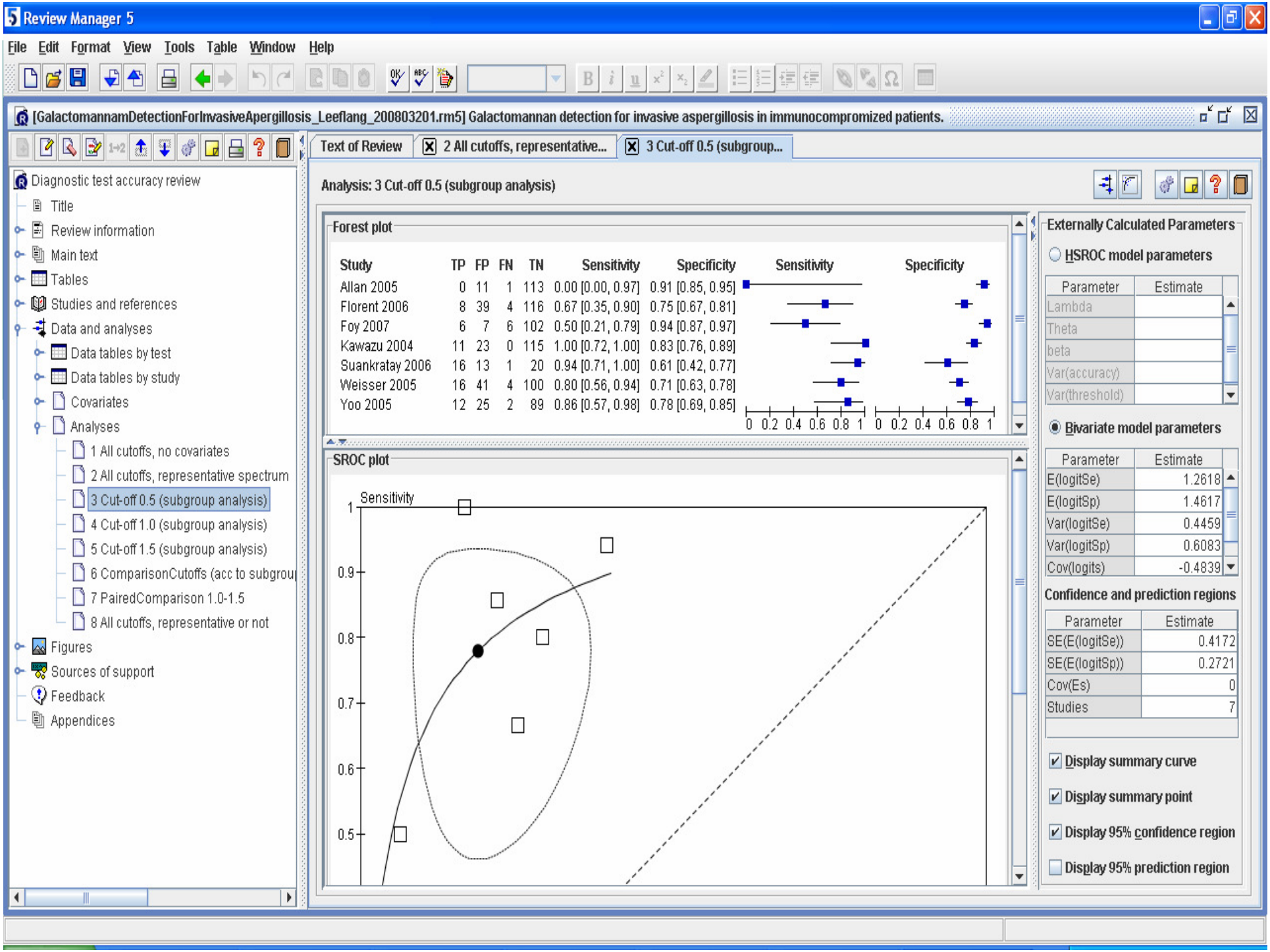
Bivariate model



Summary points or SROC curves?

- Clinical interpretation
  - Need to estimate performance at a threshold, using sensitivity, specificity or/and likelihood ratios

- Single threshold or mixed thresholds?
  - Summary curve describes how test performance varies across thresholds. Studies do not need to report a common threshold to contribute.
  - Summary point must relate to a particular threshold. Only studies reporting a common threshold can be combined.


Review Manager 5
File Edit Format View Tools Table Window Help
[GalactomannamDetectionForInvasiveApergillosis_Leeflang_200803201.rm5] Galactomannan detection for invasive aspergillosis in immunocompromized patients.
Text of Review
2 All cutoffs, representative...
3 Cut-off 0.5 (subgroup...
Diagnostic test accuracy review
Title
Review information
Main text
Tables
Studies and references
Data and analyses
Data tables by test
Data tables by study
Covariates
Analyses
1 All cutoffs, no covariates
2 All cutoffs, representative spectrum
3 Cut-off 0.5 (subgroup analysis)
4 Cut-off 1.0 (subgroup analysis)
5 Cut-off 1.5 (subgroup analysis)
6 ComparisonCutoffs (acc to subgrou
7 PairedComparison 1.0-1.5
8 All cutoffs, representative or not
Figures
Sources of support
Feedback
Appendices
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]
Sensitivity
Specificity
0 0.2 0.4 0.6 0.8 1
0 0.2 0.4 0.6 0.8 1
SROC plot
Sensitivity
1
0.9
0.8
0.7
0.6
0.5
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

Comparative analyses

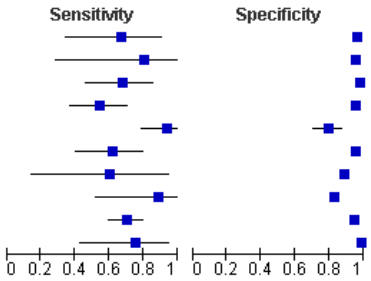
- Indirect comparisons
  - Different tests used in different studies
  - Potentially confounded by other differences between the studies
- Direct comparisons
  - Patients receive both tests or randomized to tests
  - Differences in accuracy more attributable to the tests
  - Few studies may be available and may not be representative

Example of pilot Cochrane Review Down' Syndrome screening review

	Studies	Participants
1st trimester - NT alone	10	79,412
1st trimester - NT and serology	22	222,171
2nd trimester - triple test (serology)	19	72,797

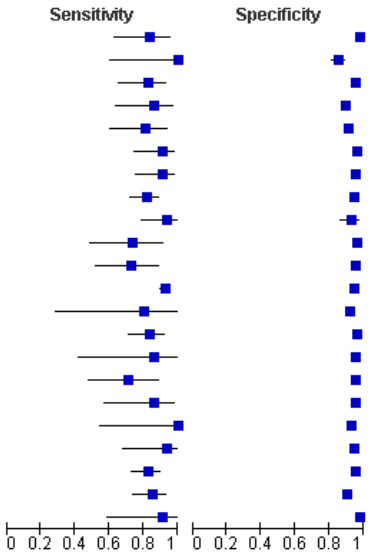
NT alone

Study	TP	FP	FN	TN	Sensitivity	Specificity
Audibert 2001	8	178	4	3940	0.67 [0.35, 0.90]	0.96 [0.95, 0.96]
Bennatar 1999	4	84	1	1567	0.80 [0.28, 0.99]	0.95 [0.94, 0.96]
Borrell 2005	17	89	8	3613	0.68 [0.46, 0.85]	0.98 [0.97, 0.98]
Crossley 2002	20	628	17	11932	0.54 [0.37, 0.71]	0.95 [0.95, 0.95]
Marsk 2006	29	22	2	86	0.94 [0.79, 0.99]	0.80 [0.71, 0.87]
Muller 2003	16	273	10	5184	0.62 [0.41, 0.80]	0.95 [0.94, 0.96]
Niemimaa 2001	3	186	2	1411	0.60 [0.15, 0.95]	0.88 [0.87, 0.90]
Pajkrt 1998	8	249	1	1215	0.89 [0.52, 1.00]	0.83 [0.81, 0.85]
Wald 2003 (NT cohort)	60	2393	25	37505	0.71 [0.60, 0.80]	0.94 [0.94, 0.94]
Wojdemann 2005	9	154	3	8456	0.75 [0.43, 0.95]	0.98 [0.98, 0.98]



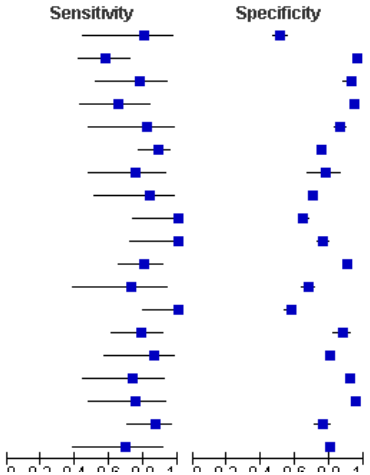
NT with serology

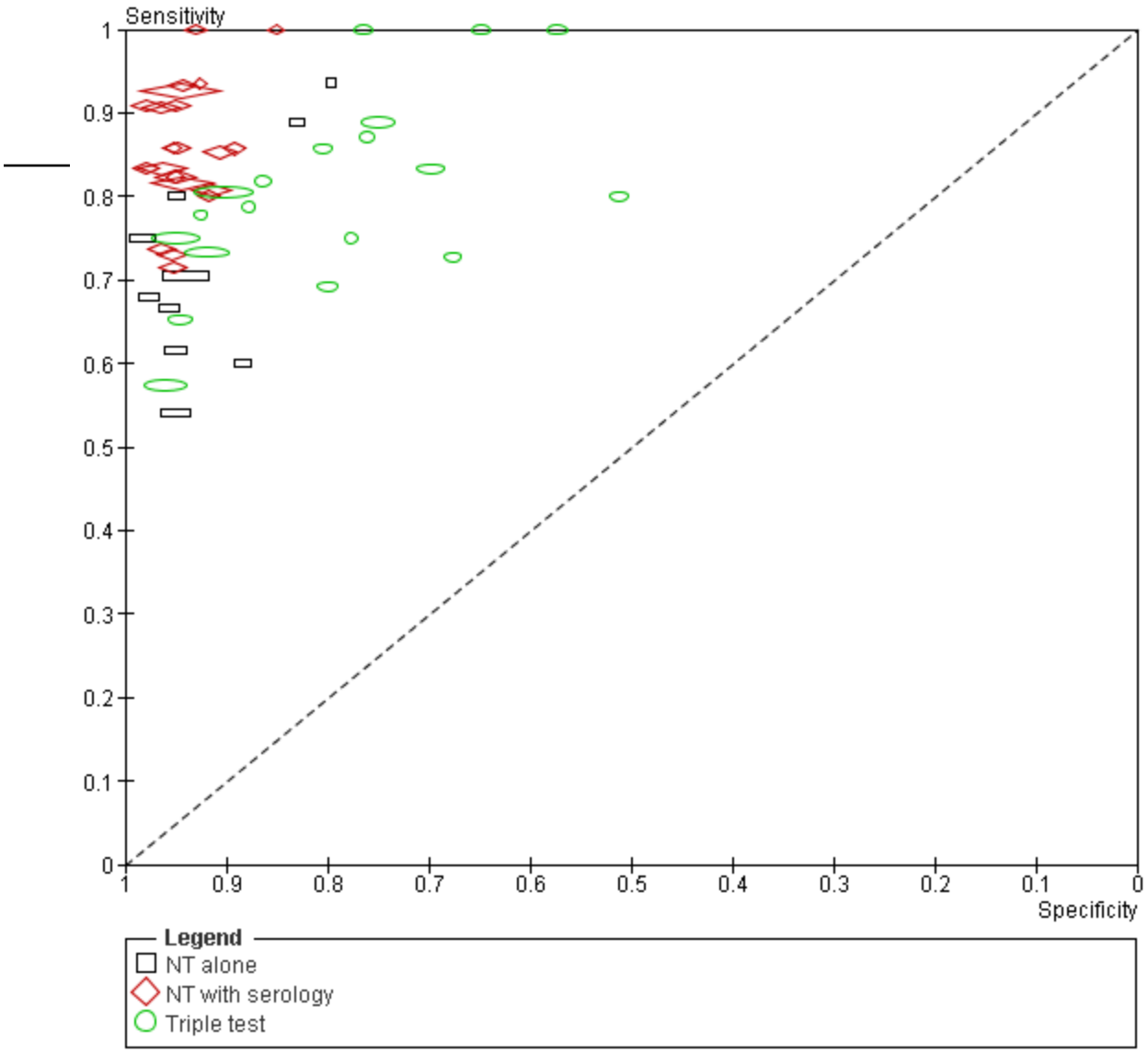
Study	TP	FP	FN	TN	Sensitivity	Specificity
Borrell 2005	20	64	4	2987	0.83 [0.63, 0.95]	0.98 [0.97, 0.98]
Centini 2005	6	60	0	342	1.00 [0.61, 1.00]	0.85 [0.81, 0.88]
Crossley 2002	28	628	6	11932	0.82 [0.65, 0.93]	0.95 [0.95, 0.95]
Go 2005	18	188	3	1550	0.86 [0.64, 0.97]	0.89 [0.88, 0.91]
Gyselaers 2005	21	1130	5	12051	0.81 [0.61, 0.93]	0.91 [0.91, 0.92]
Hadlow 2005	29	374	3	10030	0.91 [0.75, 0.98]	0.96 [0.96, 0.97]
Krantz 2000	30	289	3	5487	0.91 [0.76, 0.98]	0.95 [0.94, 0.96]
Malone 2005	75	2130	17	35903	0.82 [0.72, 0.89]	0.94 [0.94, 0.95]
Marsk 2006	29	8	2	100	0.94 [0.79, 0.99]	0.93 [0.86, 0.97]
Montalvo 2005	14	163	5	4332	0.74 [0.49, 0.91]	0.96 [0.96, 0.97]
Muller 2003	19	256	7	5201	0.73 [0.52, 0.88]	0.95 [0.95, 0.96]
Nicolaides 2005	301	4100	24	70856	0.93 [0.89, 0.95]	0.95 [0.94, 0.95]
Niemimaa 2001	4	132	1	1465	0.80 [0.28, 0.99]	0.92 [0.90, 0.93]
O'Leary 2006	50	827	10	21453	0.83 [0.71, 0.92]	0.96 [0.96, 0.97]
Orlandi 1997	6	35	1	702	0.86 [0.42, 1.00]	0.95 [0.93, 0.97]
Schielen 2006	15	190	6	3822	0.71 [0.48, 0.89]	0.95 [0.95, 0.96]
Schuchter 2002	12	245	2	4543	0.86 [0.57, 0.98]	0.95 [0.94, 0.95]
Scott 2004	5	143	0	1905	1.00 [0.55, 1.00]	0.93 [0.92, 0.94]
Stenhouse 2004	14	283	1	4702	0.93 [0.68, 1.00]	0.94 [0.94, 0.95]
Wald 2003 (serology)	70	51	15	974	0.82 [0.73, 0.90]	0.95 [0.94, 0.96]
Wapner 2003	52	767	9	7388	0.85 [0.74, 0.93]	0.91 [0.90, 0.91]
Wojdemann 2005	10	138	1	6292	0.91 [0.59, 1.00]	0.98 [0.97, 0.98]

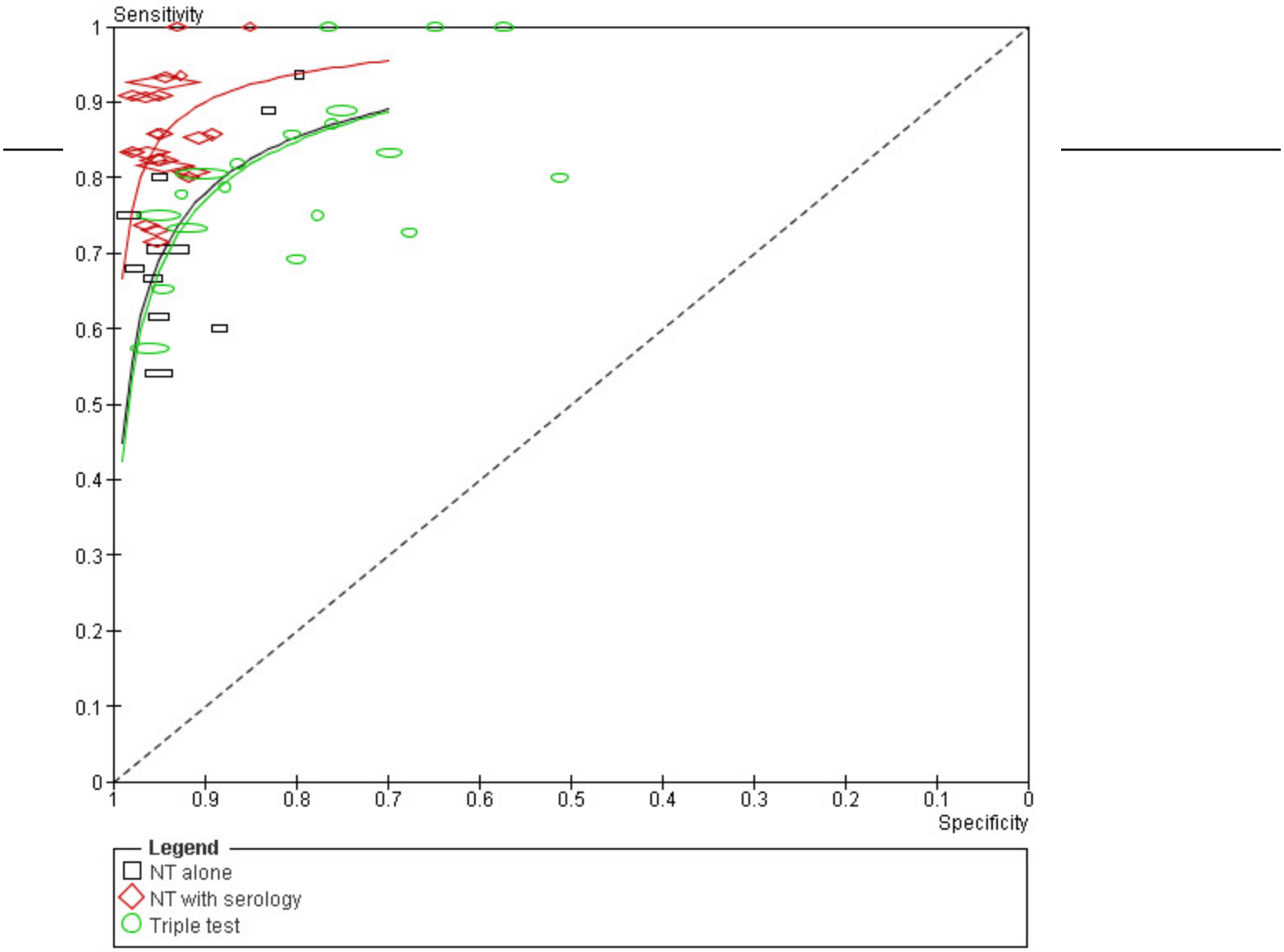


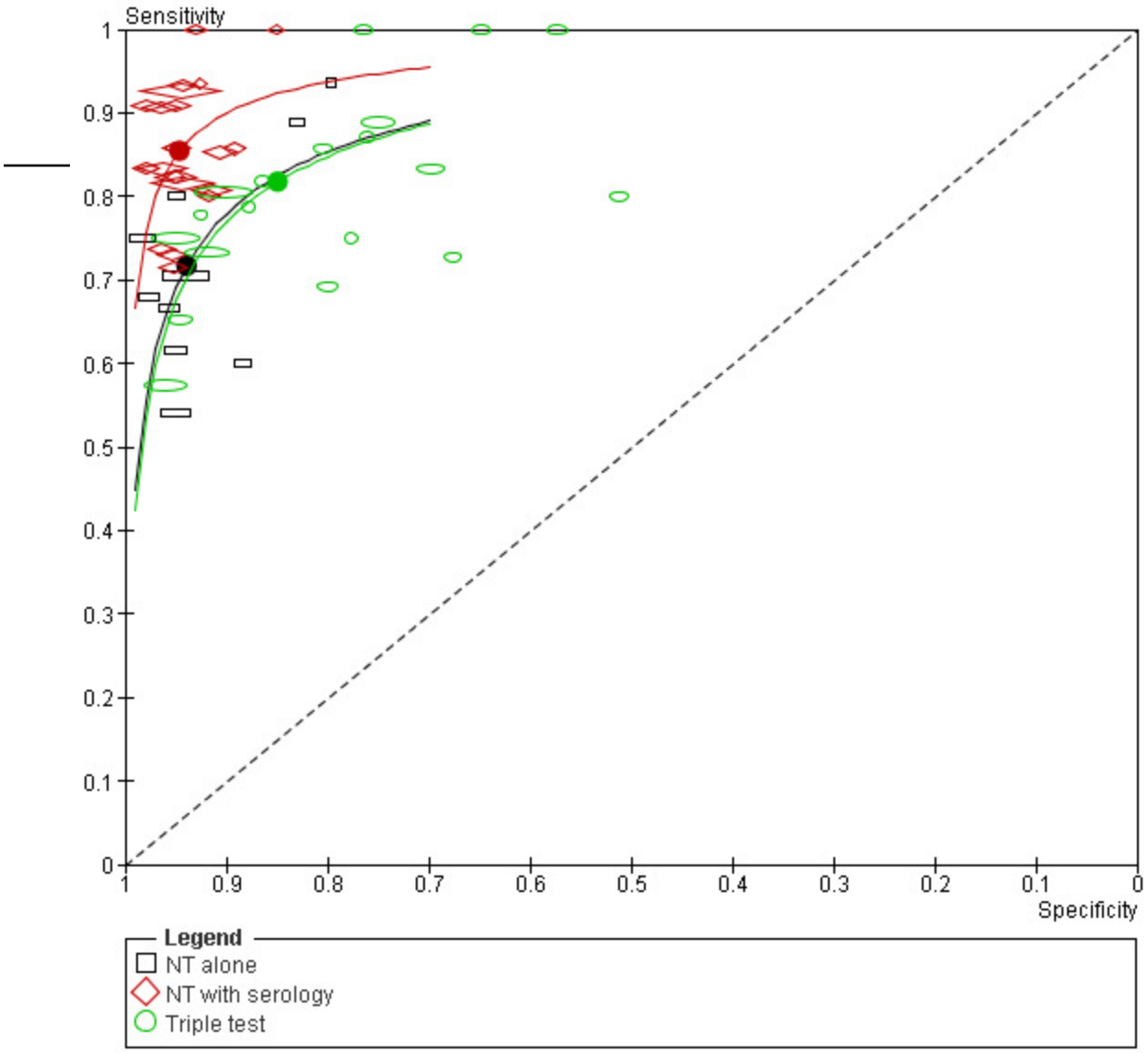
Triple test

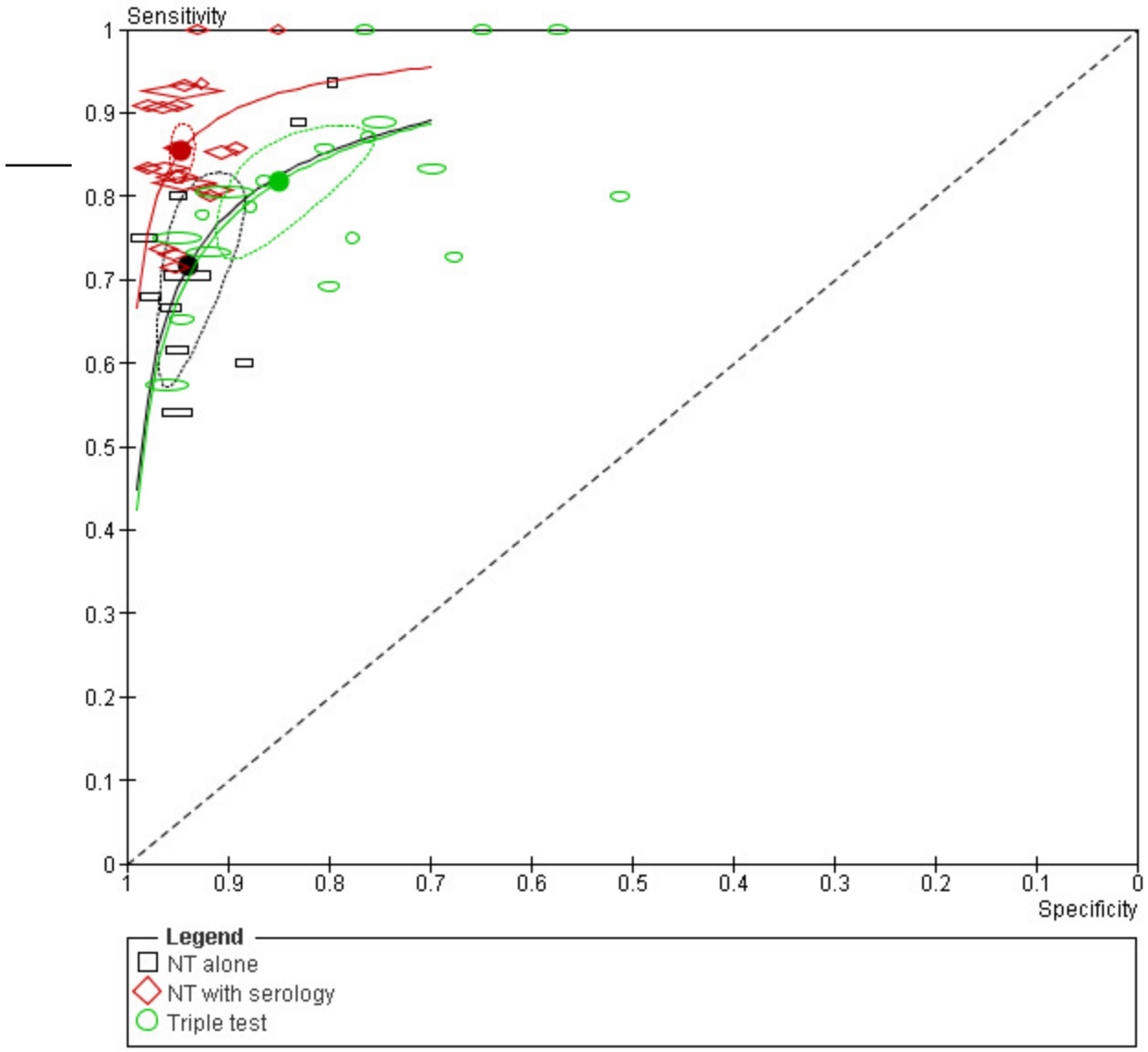
Study	TP	FP	FN	TN	Sensitivity	Specificity
Bartels (I) 1994	8	315	2	330	0.80 [0.44, 0.97]	0.51 [0.47, 0.55]
David 1996	27	372	20	8939	0.57 [0.42, 0.72]	0.96 [0.96, 0.96]
Debieve 2000	14	15	4	185	0.78 [0.52, 0.94]	0.93 [0.88, 0.96]
Extermann 1998	15	137	8	2379	0.65 [0.43, 0.84]	0.95 [0.94, 0.95]
Forest 1995	9	68	2	432	0.82 [0.48, 0.98]	0.86 [0.83, 0.89]
Haddow 1994	48	1321	6	3961	0.89 [0.77, 0.96]	0.75 [0.74, 0.76]
Heyl 1990	12	19	4	66	0.75 [0.48, 0.93]	0.78 [0.67, 0.86]
Huderer-Duric 2000	10	852	2	1969	0.83 [0.52, 0.98]	0.70 [0.68, 0.71]
Kishida 2000	10	368	0	677	1.00 [0.74, 1.00]	0.65 [0.62, 0.68]
Mancini 1991	9	170	0	552	1.00 [0.72, 1.00]	0.76 [0.73, 0.80]
Perona 1997	33	2031	8	18784	0.80 [0.65, 0.91]	0.90 [0.90, 0.91]
Piggott 1994	8	203	3	424	0.73 [0.39, 0.94]	0.68 [0.64, 0.71]
Rosen 2002	13	424	0	569	1.00 [0.79, 1.00]	0.57 [0.54, 0.60]
Sancken 2003	26	23	7	165	0.79 [0.61, 0.91]	0.88 [0.82, 0.92]
Suzimori 1997	12	208	2	856	0.86 [0.57, 0.98]	0.80 [0.78, 0.83]
Verloes 1995	11	841	4	9594	0.73 [0.45, 0.92]	0.92 [0.91, 0.92]
Ward 1999	12	673	4	12922	0.75 [0.48, 0.93]	0.95 [0.95, 0.95]
Wenstrom 1997	27	75	4	238	0.87 [0.70, 0.96]	0.76 [0.71, 0.81]
Wenstrom 1999	9	249	4	994	0.69 [0.39, 0.91]	0.80 [0.78, 0.82]



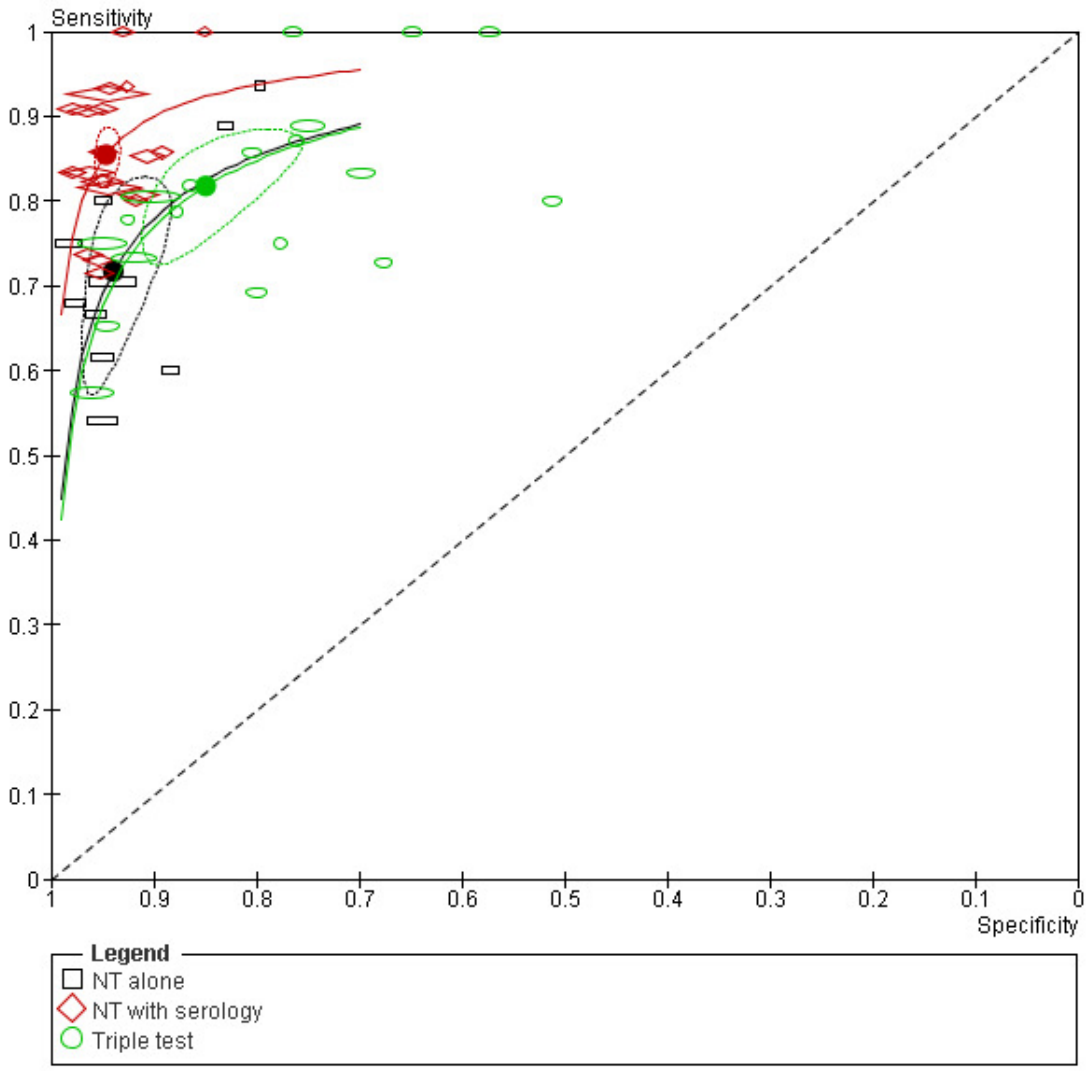

Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Specificity
Legend
NT alone
NT with serology
Triple test


Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Specificity
Legend
NT alone
NT with serology
Triple test


Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Specificity
Legend
NT alone
NT with serology
Triple test


Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Specificity
Legend
NT alone
NT with serology
Triple test

Indirect comparison



NT alone

Sensitivity: 72% (63%-79%)

Specificity: 94% (91% -96%)

DOR: 39 (26-60)

NT with serology

Sensitivity: 86% (82%-90%)

Specificity: 95% (93%-96%)

DOR: 110 (84-143)

RDOR: 2.8 (1.7-4.6), p <0.0001

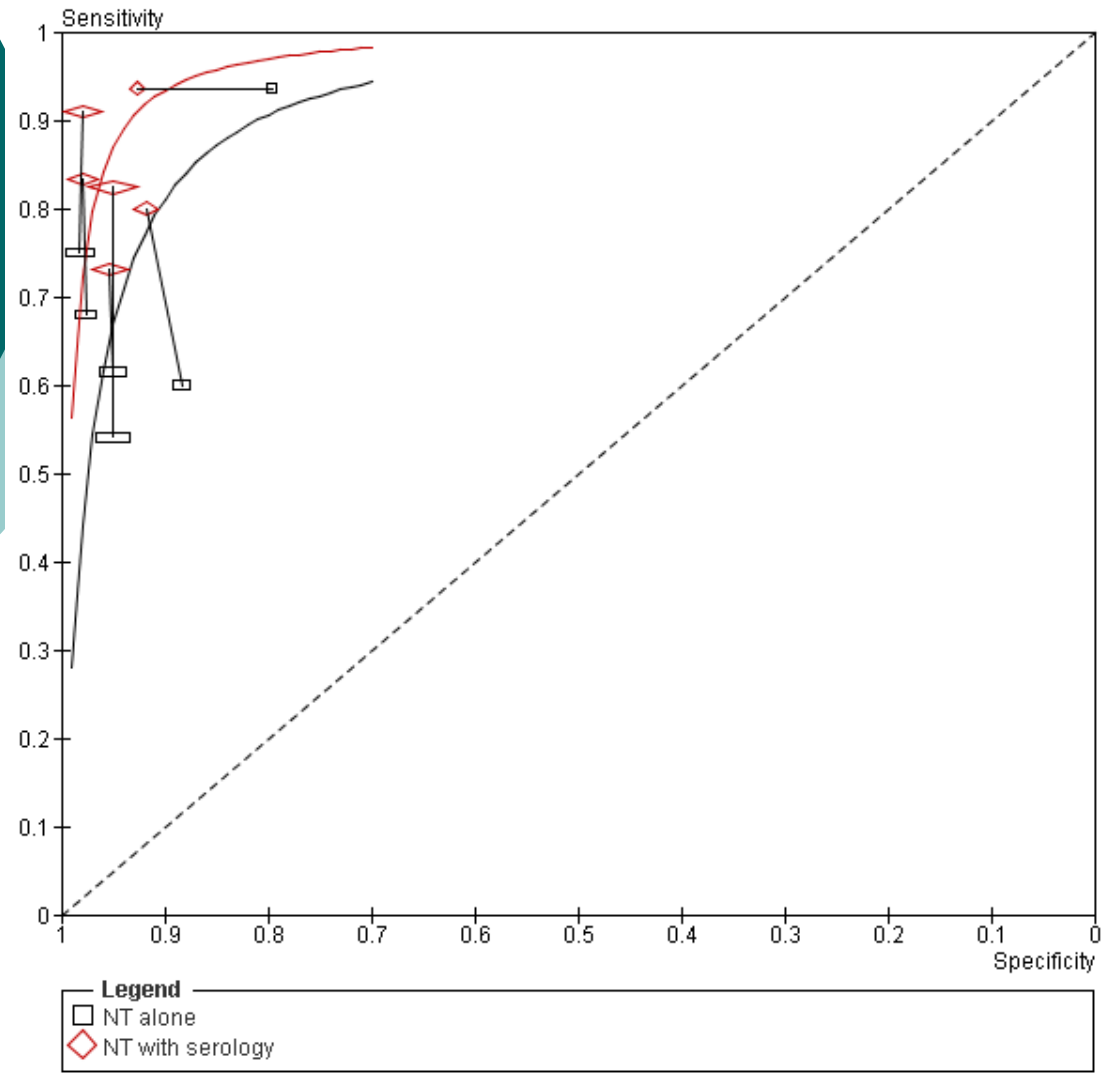
Triple test

Sensitivity: 82% (76%-86%)

Specificity: 83% (77%-87%)

DOR: 21 (15-30)

RDOR: 0.5 (0.3-0.9), p = 0.03



DIRECT COMPARISONS

NT alone

Sensitivity: 71% (59%-82%)

Specificity: 95% (91%-98%)

DOR: 41 (16-67)

NT with serology

Sensitivity: 85% (77%-93%)

Specificity: 96% (93%-98%)

DOR: 123 (40-206)

Triple test

No paired studies available

Summary

- Bivariate nature of the data requires a different approach to traditional meta-analysis
- SROC approach useful for preliminary analyses
- Advanced methods required for making formal inference