

# Generalized Evidence Synthesis for Diagnostic Test Data

*Pablo E. Verde*

*Coordination Centre for Clinical Trials  
University of Duesseldorf  
Germany*

*IceBUGS  
11-12.02.06  
Finland*



# Systematic Reviews

Motivations for the analysis and synthesis of published information:

- Collect background information to design a new clinical trial.
- Compare new results with previously published ones.
- Exploratory analysis e.g., identify subgroups of patients who are likely to respond well to a treatment.
- Health Care Evaluation: evaluate critically published information about a new medical technology.

# Running Example



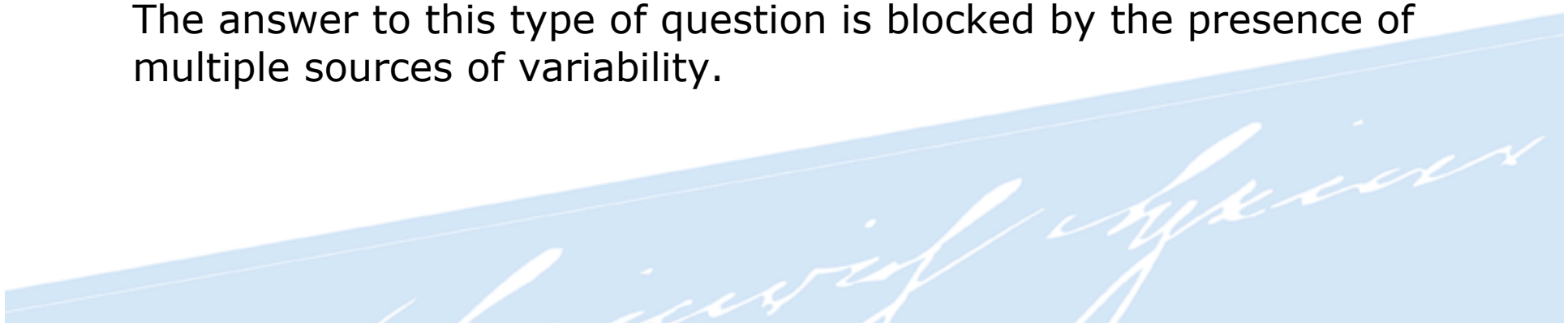
A health technology assessment problem:

Acute appendicitis is one of the most common acute surgical events (e.g. 250,000 per year in USA). Traditional diagnoses reported false positive rates of 20% to 30%.

Computer Tomography scans (CT) have been advocated to be of high potential diagnostic benefit in suspected appendicitis.

Question: Which diagnostic performance has this technology ?

The answer to this type of question is blocked by the presence of multiple sources of variability.



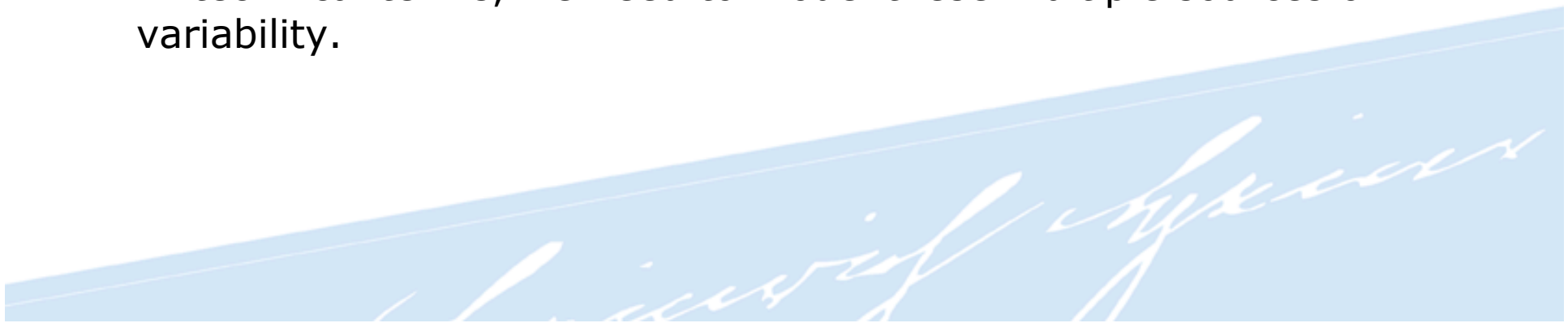
# Generalized Evidence Synthesis



The information we wish to combine could be heterogeneous:

- Published studies coming with different designs
- Different quality
- Different populations
- Indirect comparisons
- Combined database data with published studies results

In technical terms, we need to model these multiple sources of variability.



# Running Example



A systematic review was performed (Ohmann et al. 2005) to evaluate the diagnostic benefit of this technology. They selected 52 studies for analysis:

author	country	n	tp	fn	fp	tn	se	sp
.....								
Applegate2001	USA	96	87	2	4	3	98	43
Brandt2003	USA	179	168	1	3	7	99	70
Cho1999	AU	36	21	0	1	14	100	93
D'Ippolito1998	Brazil	52	40	4	0	8	91	100
Hershko2002	Israel	197	67	5	7	118	93	94
.....								

A classical approach using a random effects model gives:

Sensitivity: 94% [94, 95]

Specificity: 94% [91, 94]

After modeling variability introduced by study quality:

Sensitivity: 94% [78, 99]

Specificity: 94% [79, 99]

## The Diagnostic Test Data

Test results are usually summarized in a 2×2 table giving the number of positive and negative test results for patients with and without disease:

		Reference result		Total
		Present	Absent	
Test Outcome	(+)	a	b	a+b
	(-)	c	d	c+d
Total		a+c	b+d	a+b+c+d

- True positive rate (TPR) or Sensitivity =  $a/(a+c)$
- True negative rate (TNR) or Specificity =  $d/(b+d)$
- The false positive rate: FPR =  $b/(b+d)$
- Diagnostic Odds Ratio DOR =  $(a \times d)/(b \times c)$

# Test Data Generation Process

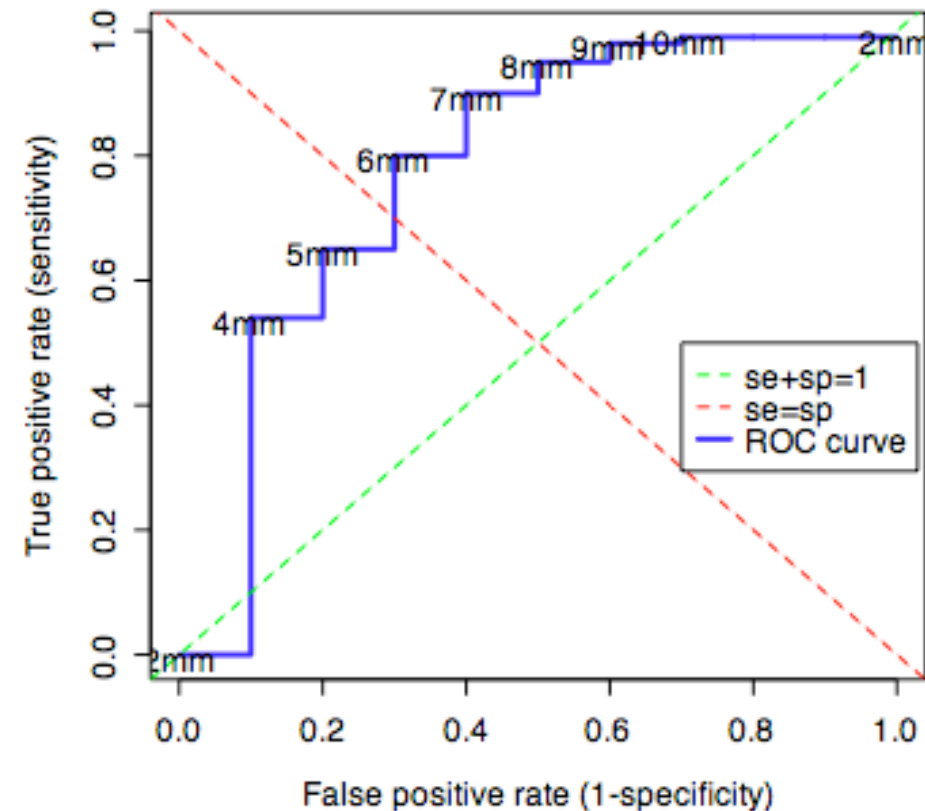
Let  $y_j$  be a test measurement of patient  $j$  and "k" a threshold value.

Then the test outcome is

$$T_j = \begin{cases} \text{test positive if } y_i \geq k \\ \text{test negative otherwise.} \end{cases}$$

The test threshold is chosen as a trade off between TPR and FPR.

The construction of diagnostic tests introduces in a correlation between TPR and FPR.



# Smooth fitting of the SROC curve

The idea of SROC curve is to represent the relationship between TPR and FPR across studies, assuming that they may use different “*threshold values*” (don’t take this literally!).

The SROC is a meta-regression model (Moses et al. 1993).

- 1) For each study convert the estimates of TPR and FPR to their logistic transform:

$$U = \text{logit}(\text{TPR}), \quad V = \text{logit}(\text{FPR})$$

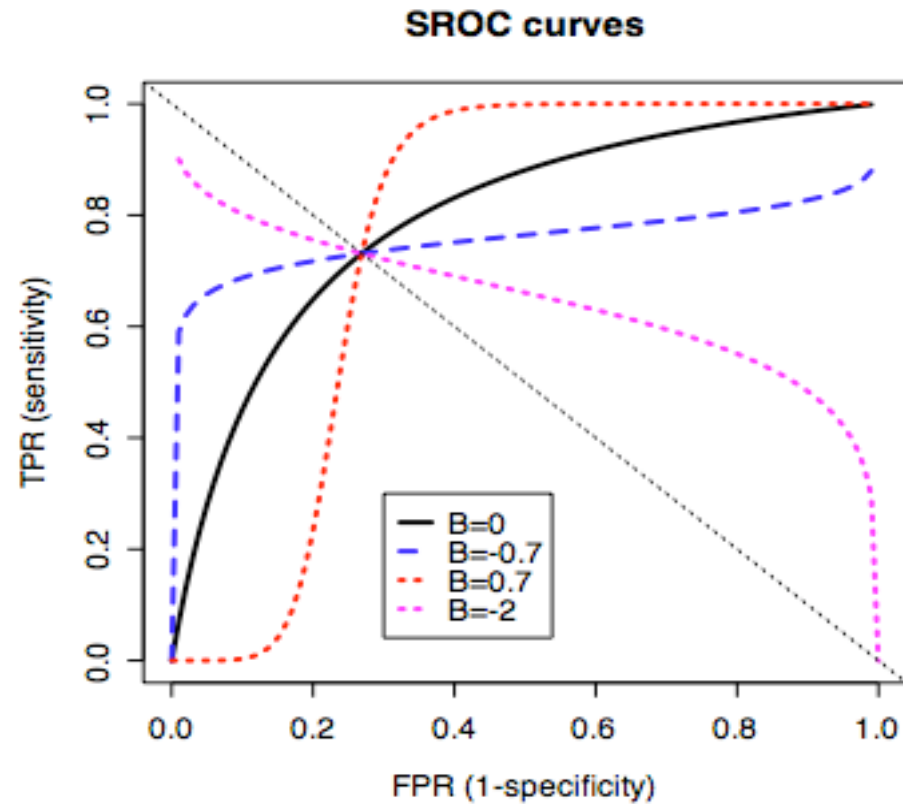
- 2) Calculate:  $\log(\text{DOR}) = D = U - V$  and  $S = U + V$

- 3) Fit a straight line:  $D = A + B \cdot S$

- 4) Reverse the transformations and deduce the relationship between TPR and FPR:

$$\text{TPR} = \frac{\exp(-A/(1-B)) \times (\text{FPR}/(1-\text{FPR}))^{(1+B)/(1-B)}}{\left[ 1 + \exp(-A/(1-B)) \times (\text{FPR}/(1-\text{FPR}))^{(1+B)/(1-B)} \right]}$$

# SROC curve's behavior



The inflexion point is at  $se + sp = 1$ .

## Limitations of the SROC curve

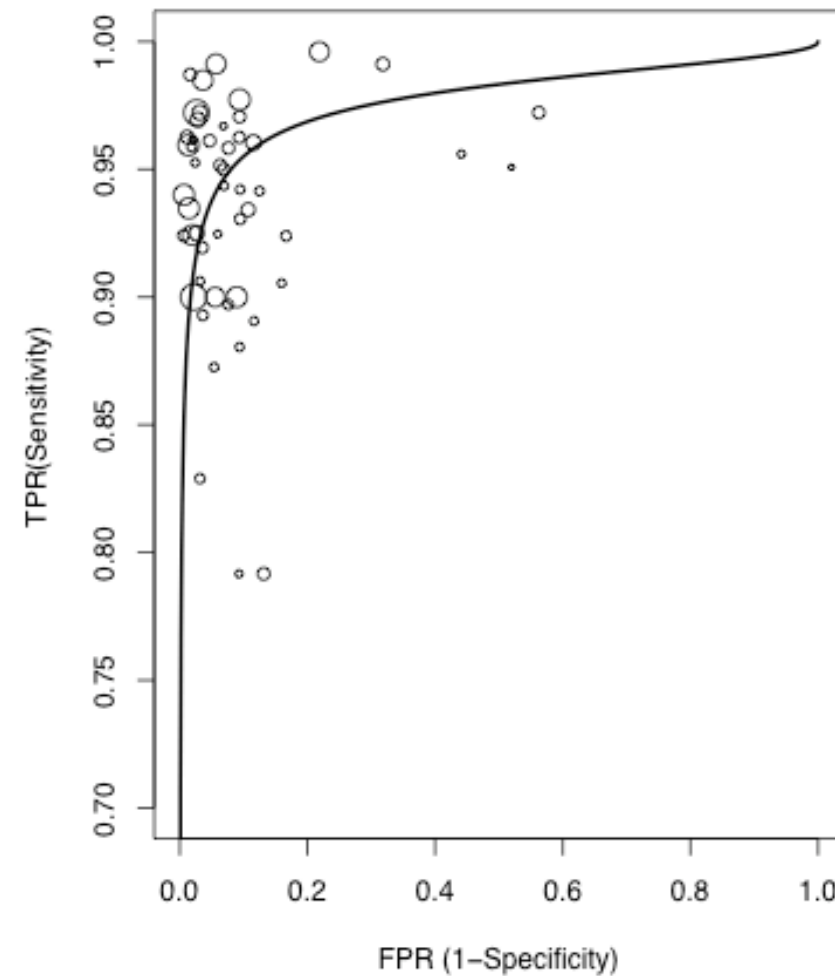
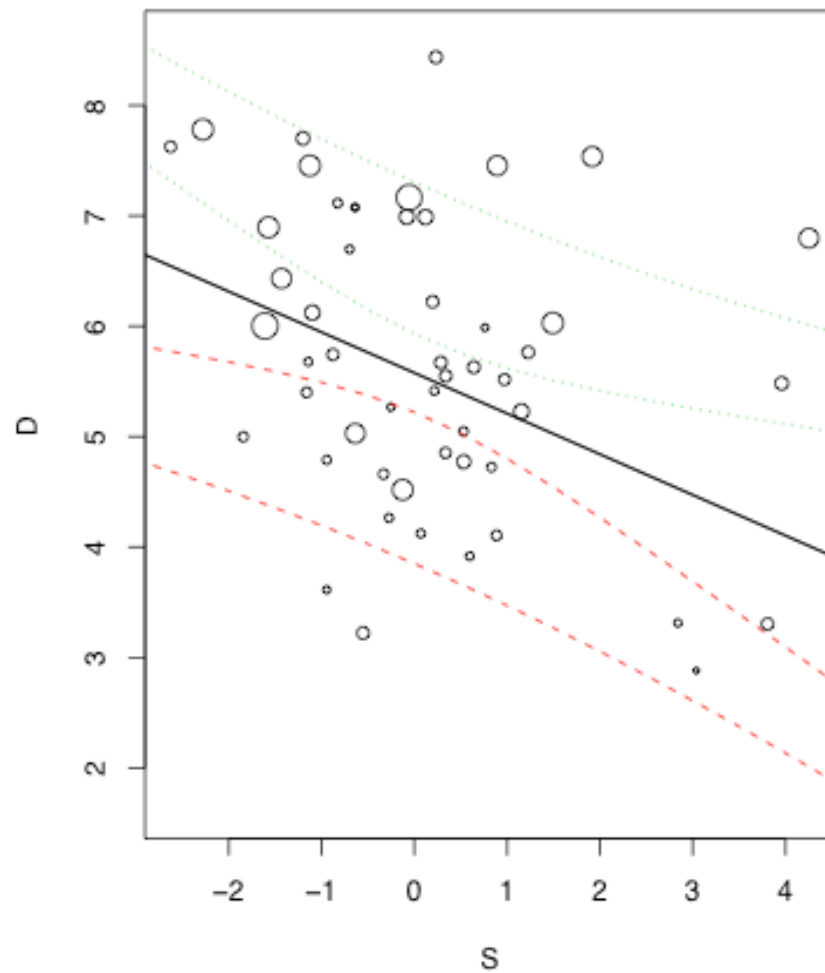
It has been essentially developed as a graphical device :

- It is not useful to make predictions of future studies
- S is assumed fixed, which is NOT.
- Difficult to link explanatory variables.
- General functional statistics, e.g., the area under the SROC curve:

$$\text{AUC} = \int_0^1 \frac{\exp(-A/(1-B)) \left( \frac{x}{1-x} \right)^{(1+B)/(1-B)}}{\left[ 1 + \exp(-A/(1-B)) \left( \frac{x}{1-x} \right)^{(1+B)/(1-B)} \right]} dx$$

is only analytically tractable for  $B=0$ .

# The SROC curve analysis CT- appendicitis



## Design study quality assessment



*The Cochrane Methods Group on Systematic Review of Screening and Diagnostic Test: Recommended Methods".*

*Some quality assessment items:*

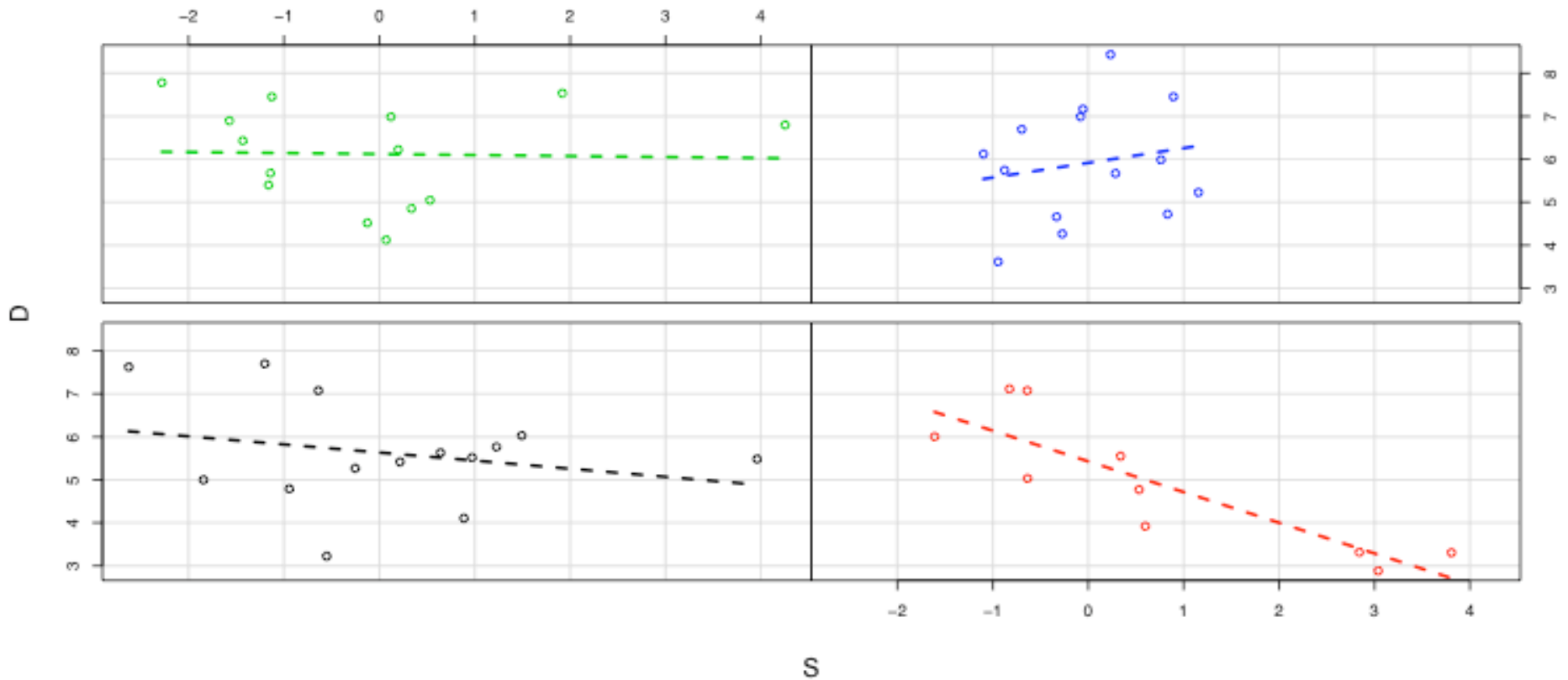
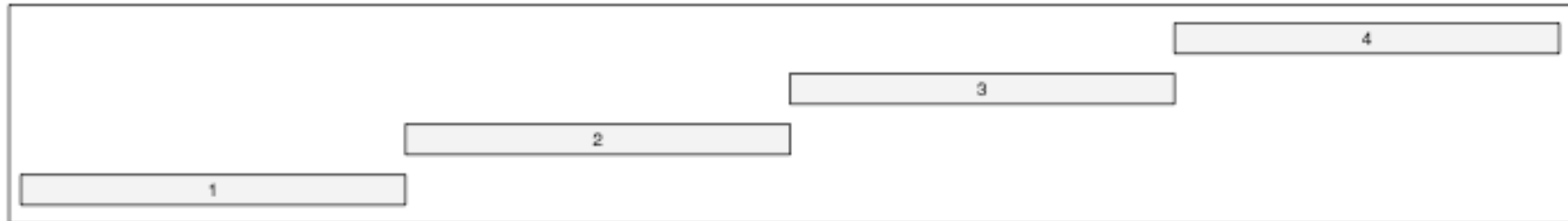
- Was the CT compared with a valid reference standard?
- Was the selection of the sample adequate?
- Were indeterminate CT-findings included in the analysis?
- Was the study design adequate?

e.g. make a quality score from 1 to 4, with 4 as the maximum score.



# Variability: Study Quality Score

Given : factor(score)



# Generalized Evidence Synthesis Modeling



We can model the number of true positives  $a_{ik}$  and false positives  $b_{ik}$  for a study  $i$  ( $i = 1, \dots, m_k$ ) of type  $k$  ( $k=1, \dots, K$ ) as follows:

$$a_{ik} \sim \text{Bin}(\text{TPR}_{ik}, n_{ik,1}), b_{ik} \sim \text{Bin}(\text{FPR}_{ik}, n_{ik,2}), n_{ik,} = n_{ik,1} + n_{ik,2}$$

with

$$\text{logit}(\text{TPR}_{ik}) = (D_{ik} + S_{ik})/2, \text{logit}(\text{FPR}_{ik}) = (D_{ik} - S_{ik})/2,$$

$$(D_{ik}, S_{ik}) \sim \text{MN}(\mu_k, \Lambda_k), \Lambda_k^{-1} = \Sigma_k$$

and between studies type variability modeled as

$$\mu_k \sim \text{MN}(\mu, \Gamma), \Gamma^{-1} = \text{diag}(\sigma_D^2, \sigma_S^2).$$

Note that we have  $2K+2$  component of variance and  $K$  covariances.

## Prior specification

The variance components  $\sigma_{D,1}, \sigma_{S,1}, \dots, \sigma_{D,k}, \sigma_{S,k}, \sigma_D, \sigma_S$  follows a half-Cauchy( $\lambda_j$ ) with a pick at 0 and scale parameter  $\lambda_j$  estimated from the data with hyper prior  $\lambda_j \sim \text{unif}(0, \text{Scale})$ .

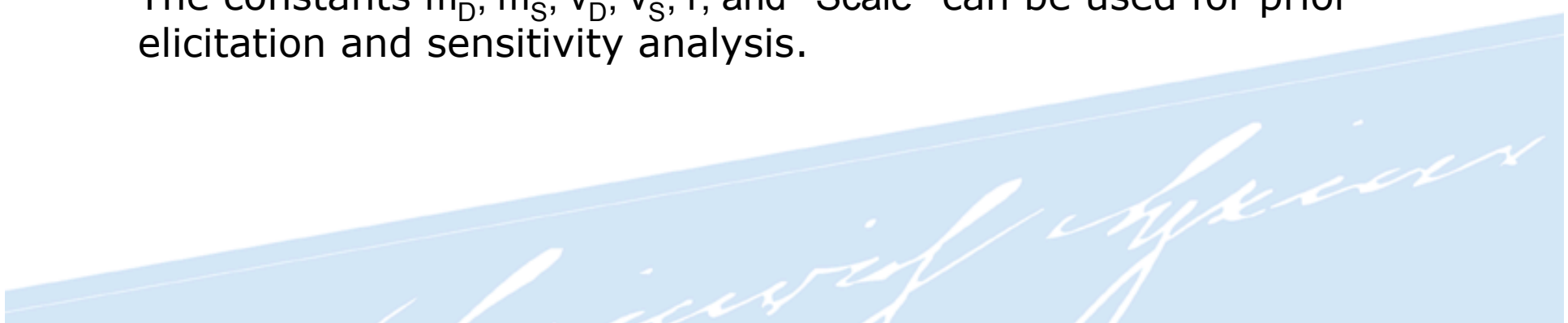
The covariance components

$$\sigma_{D,S,k} = \sigma_{S,D,k} = \rho_k \sigma_{D,k} \sigma_{S,k}, \quad \rho_k \sim \text{Uniform}(-r, r).$$

Independent Normal distributions for the components  $\mu = (\mu_D, \mu_S)$ :

$$\mu_D \sim N(m_D, v_D), \quad \mu_S \sim N(m_S, v_S).$$

The constants  $m_D, m_S, v_D, v_S, r$ , and "Scale" can be used for prior elicitation and sensitivity analysis.



# Inferential Statements

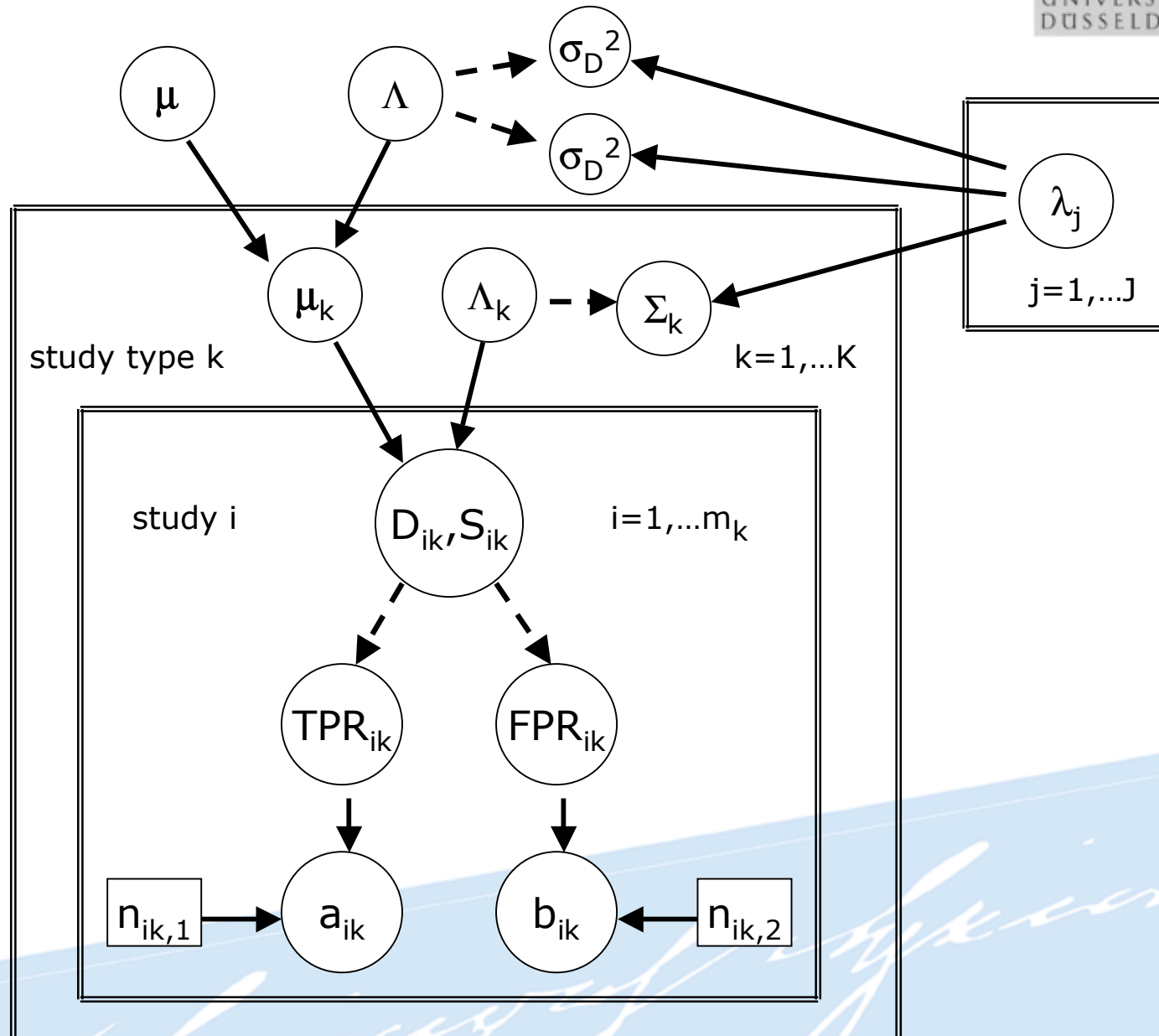
- Observed quantities, i.e., data  $y$
- Unknown quantities are represented by  $\theta$  (parameters, missing data, ...)
- Inference is based on the posterior distribution

$$p(\theta | y) \propto p(\theta) p(y | \theta).$$

- Inference on functional parameters, say  $g(\theta)$ , is based on the marginal posterior  $p(g(\theta) | y)$ .
- Predictions of new data  $y^*$  are based on the marginal predictive  $p(y^* | y)$

Those posteriors are empirically approximated using MCMC (a Gibbs sampler). Computations based on WinBUGS 1.4.1

# DAG for Generalized Evidence Synthesis



## Recovering the SROC

From the distribution of  $(D_{ik} | S_{ik} = s)$  we can recover the SROC as follows:

$$E(D_{i,k} | S_{i,k} = s) = A_k + B_k (s - \mu_{S,k}),$$

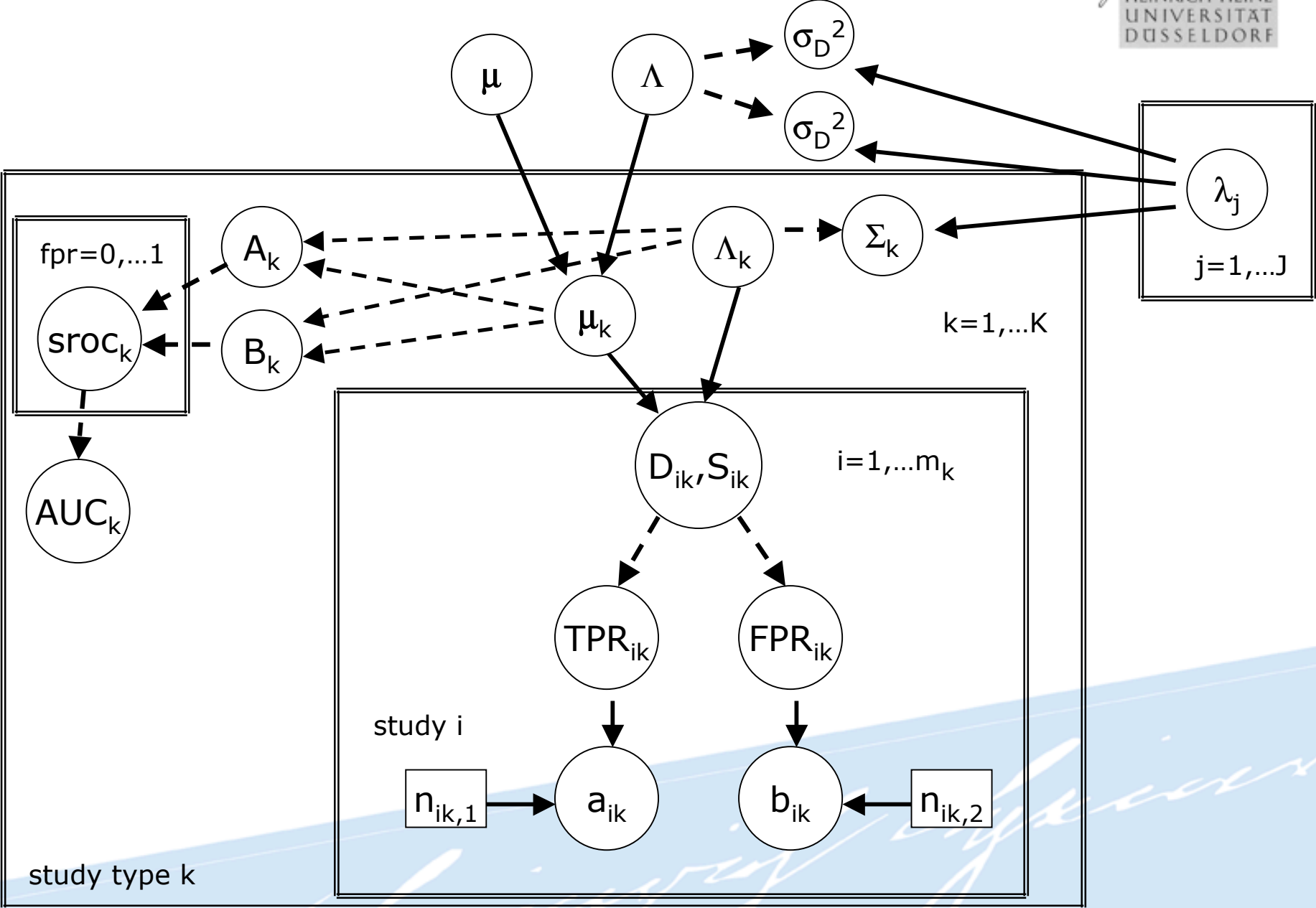
$$A_k = \mu_{D,k} - \mu_{S,k} \rho_k \sigma_{D,k} / \sigma_{S,k} \quad \text{and} \quad B_k = \rho_k \sigma_{D,k} / \sigma_{S,k}$$

$$\text{var}(D_{i,k} | S_{i,k} = s) = \sigma_{D,k}^2 (1 - \rho_k^2).$$

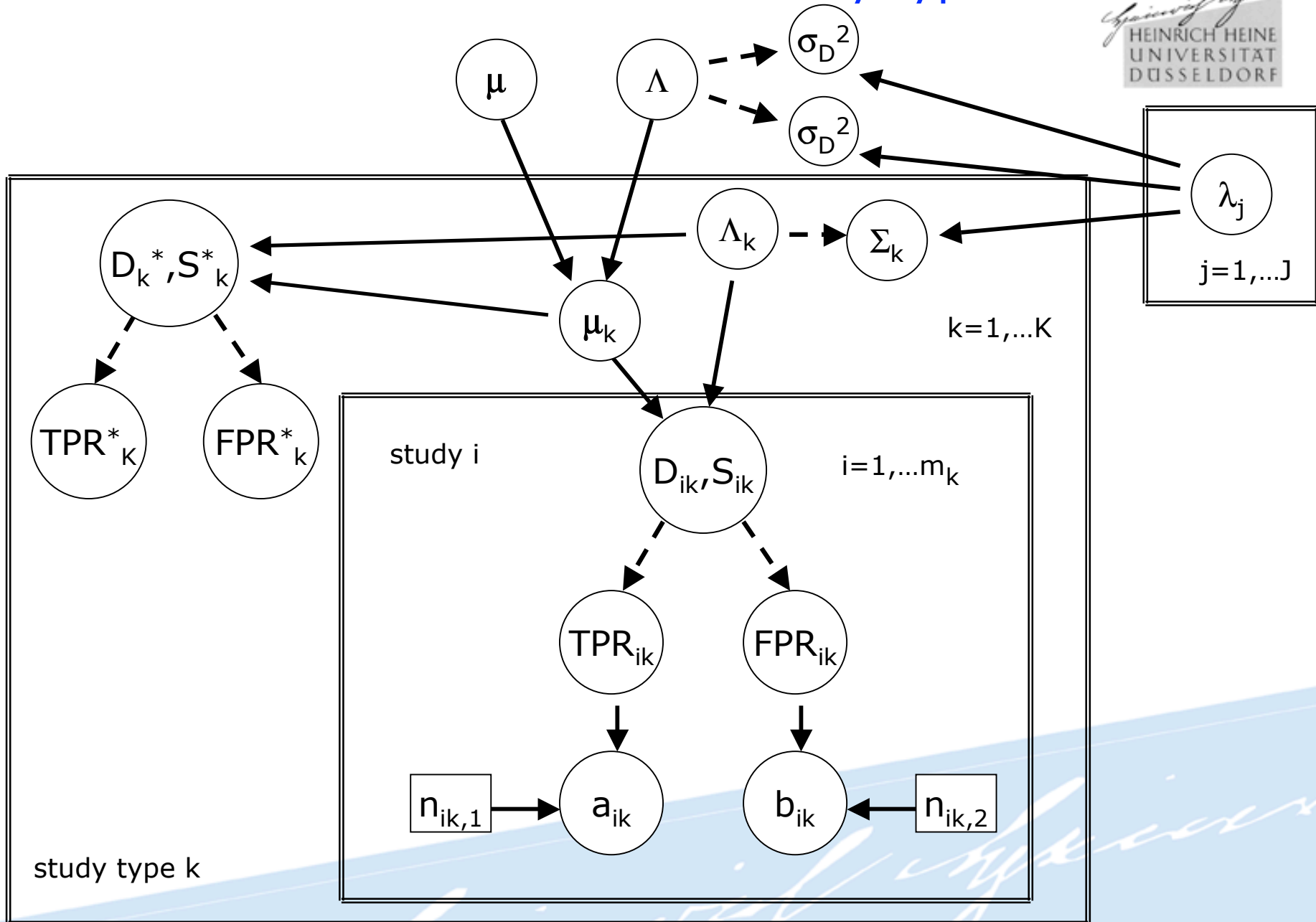
- Reverse the transformations and deduce the relationship between TPR and FPR.
- The AUC can be calculated from SROC numerically.

Operationally, we calculate the required functions (A, B, SROC, AUC) as *logical nodes* at each iteration of the MCMC and we summarise posterior samples.

# DAG for the SROC curve



# DAG for simulations at study type level

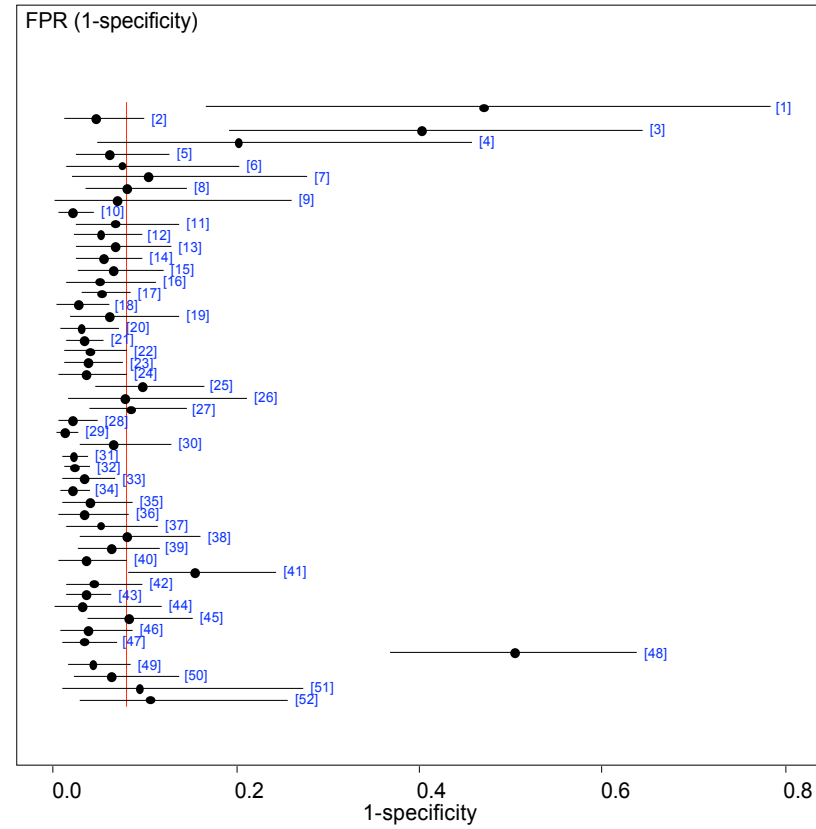
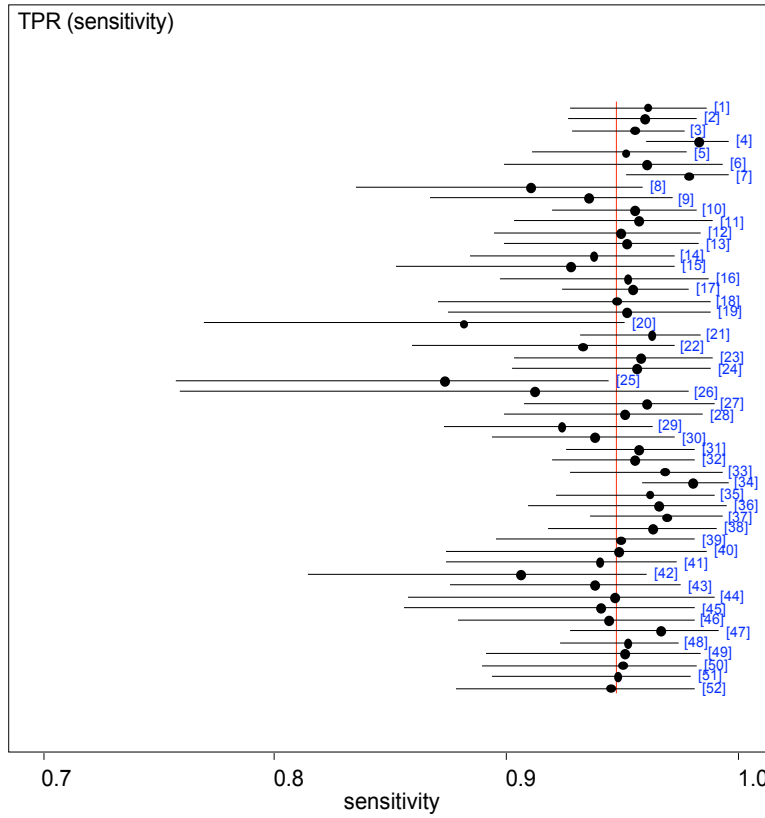


# MCMC Analysis

## Computational specification:

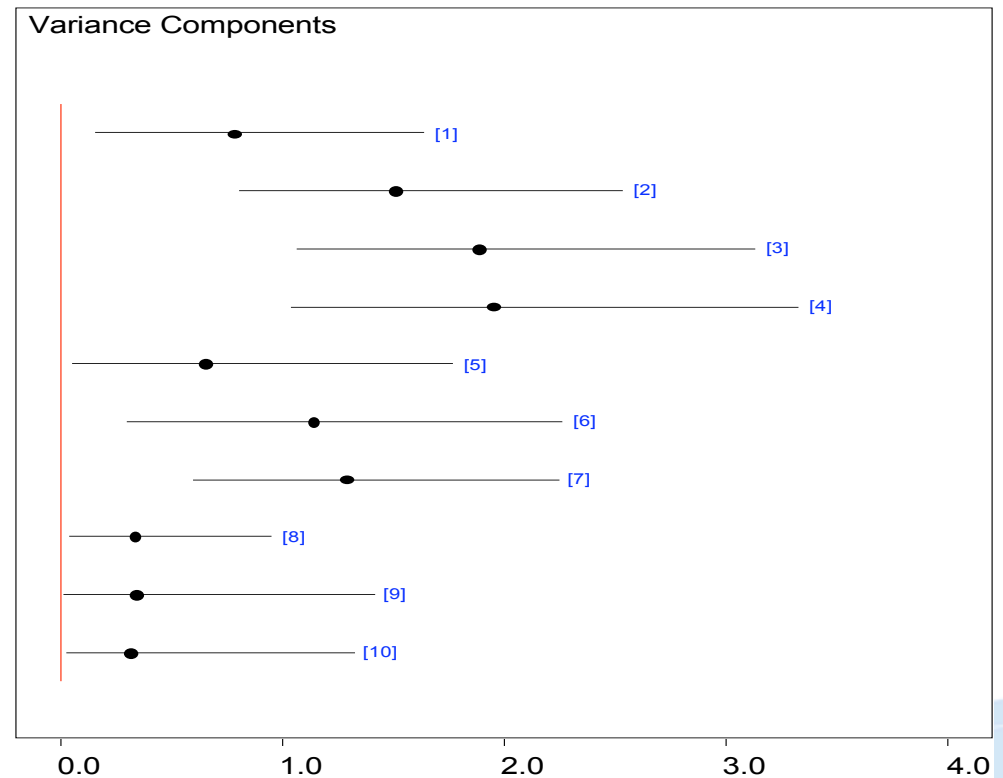
- We take the hyper-prior constants:  
 $m_D=5$ ,  $m_S=0$ ,  $v_D=0.05$ ,  $v_S=0.05$ , and  $r=1$ .
- MCMC set up: We run a single chain of length 10,000. We take the last 5000 iterations for inference.
- Convergence diagnostics were done graphically. No major convergence problems were presented. The variance and covariance components mixed fast.
- Sensitivity analysis:
  - take "Scale" = 3, 5, 6, 7 on  $\lambda_j \sim \text{unif}(0, \text{Scale})$
  - "Total.Var" =  $\sigma_{D,1} + \sigma_{S,1} + \dots + \sigma_{D,k} + \sigma_{S,k} + \sigma_D + \sigma_S$
  - "Total.Var" =  $10.06 \pm 1.353$ ,  $10.73 \pm 1.539$ ,  $9.87 \pm 1.73$ ,  $10.51 \pm 1.769$

# Pr( $\text{TPR}_i|y$ ) and Pr( $\text{FPR}_i|y$ )

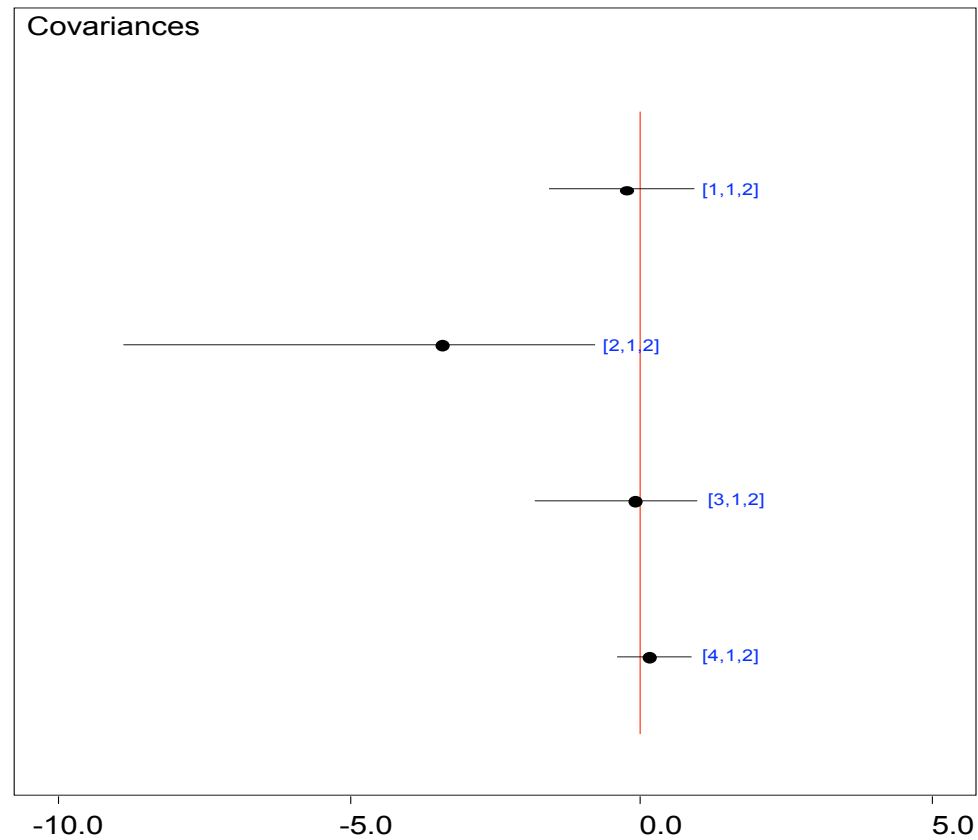


# Pr(variance components|y)

	<b>mean</b>	<b>95% CI</b>
D: sigma.star[1]	0.789	[0.150, 1.653]
S: sigma.star[2]	1.524	[0.805, 2.559]
D: sigma.star[3]	1.855	[1.040, 3.066]
S: sigma.star[4]	1.906	[1.009, 3.231]
D: sigma.star[5]	0.577	[0.056, 1.636]
S: sigma.star[6]	1.116	[0.295, 2.236]
D: sigma.star[7]	1.300	[0.630, 2.241]
S: sigma.star[8]	0.330	[0.035, 0.937]
D: sigma.star[9]	0.361	[0.015, 1.478]
S: sigma.star[10]	0.334	[0.024, 1.395]

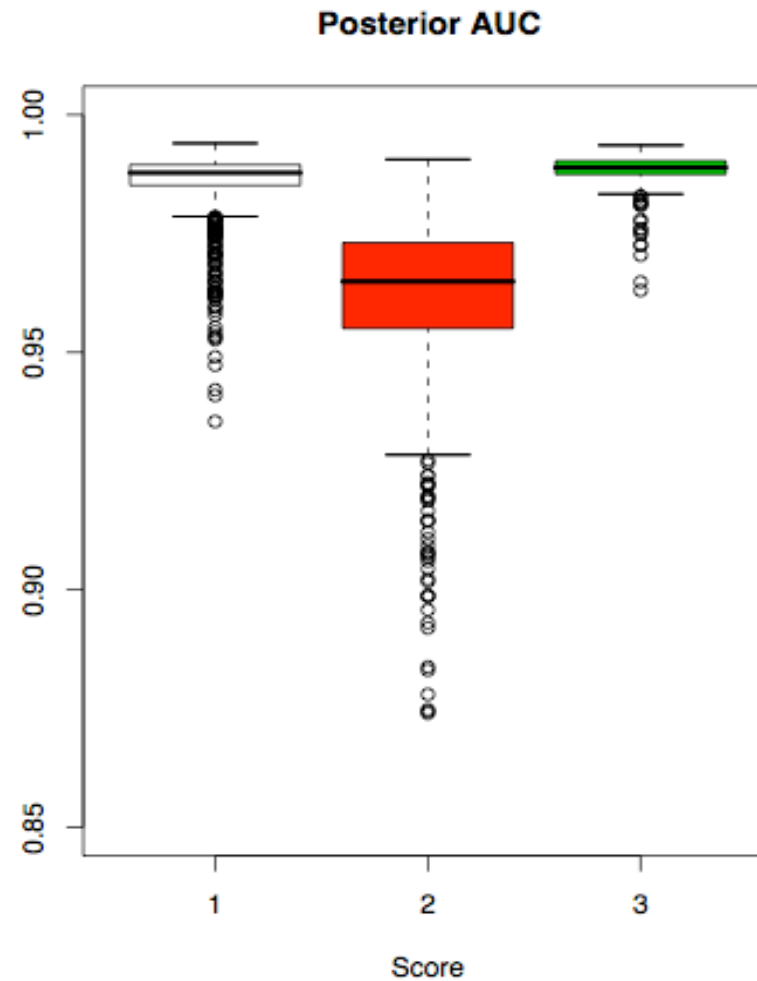
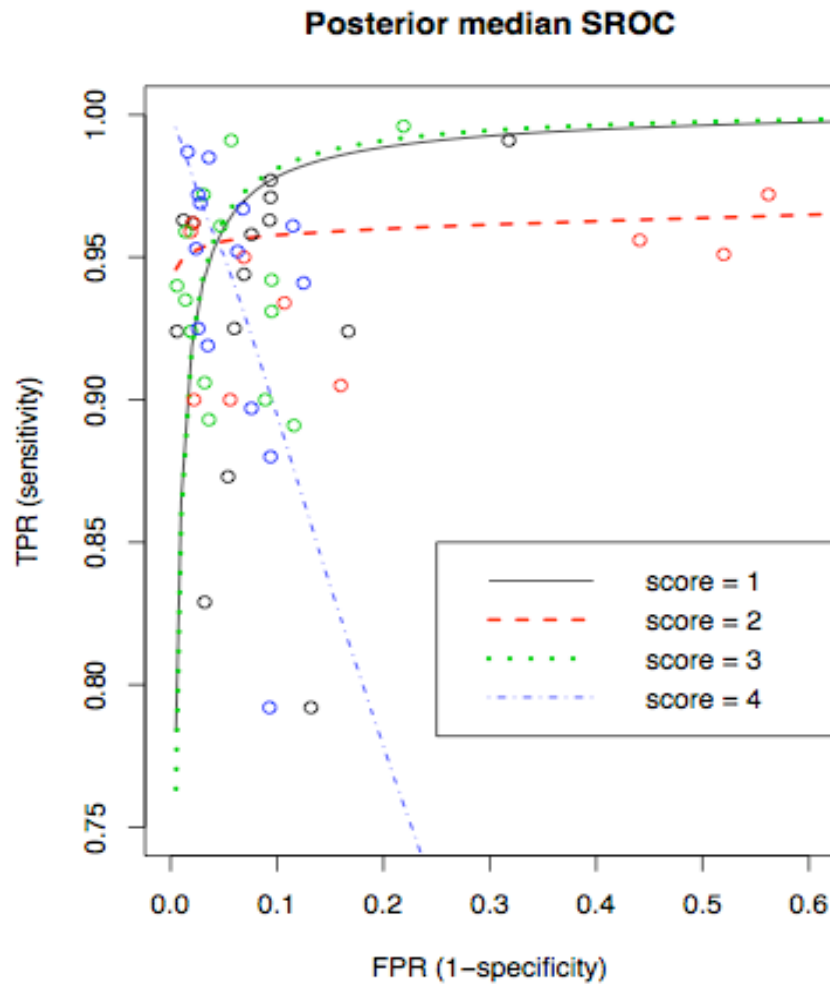


# Pr(covariance components|y)



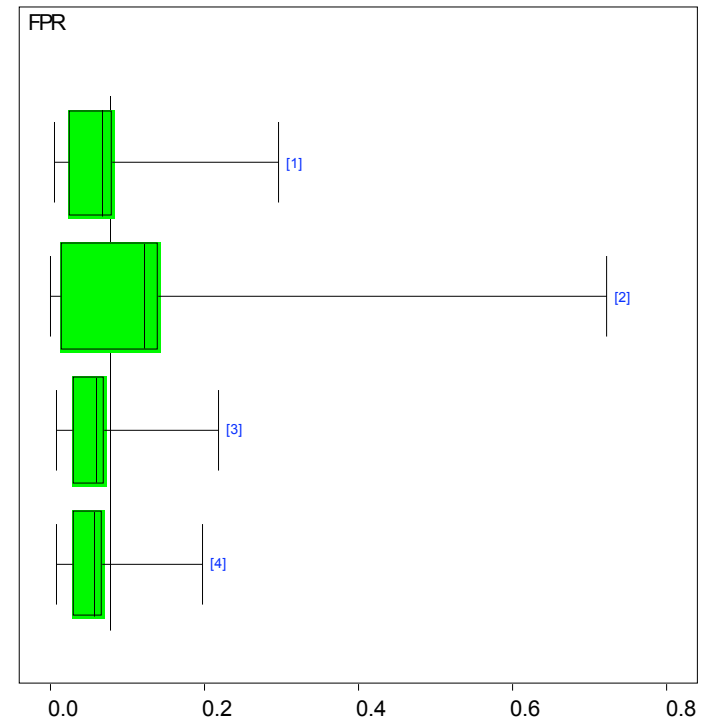
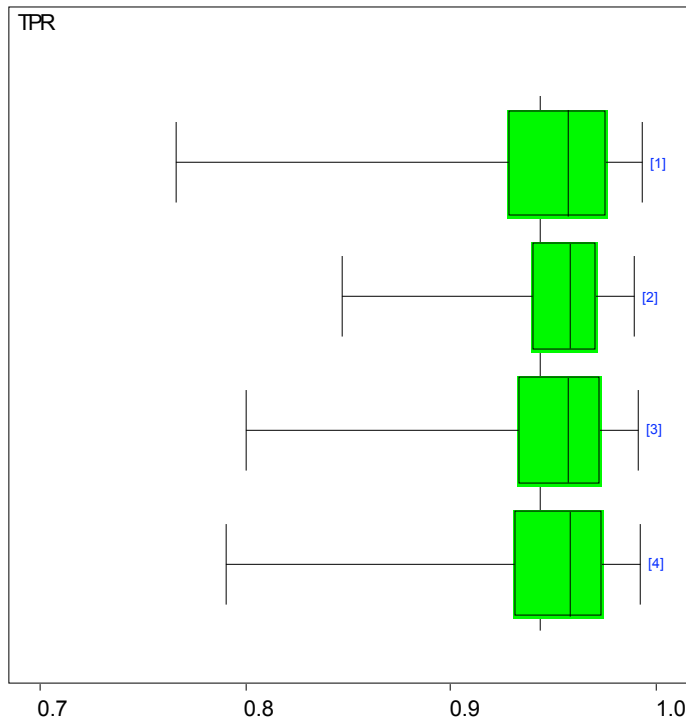
	<b>mean</b>	<b>95% CI</b>
$\sigma[1,1,2]$	-0.2637	[-1.624, 0.911]
$\sigma[2,1,2]$	-3.259	[-8.523, -0.778]
$\sigma[3,1,2]$	-0.108	[-1.366, 0.851]
$\sigma[4,1,2]$	0.1131	[-0.431, 0.841]

# $\Pr(\text{SROC} | y)$ and $\Pr(\text{AUC} | y)$



The SROC for the group with score=4 degenerate and the AUC is not applicable as summary statistics.

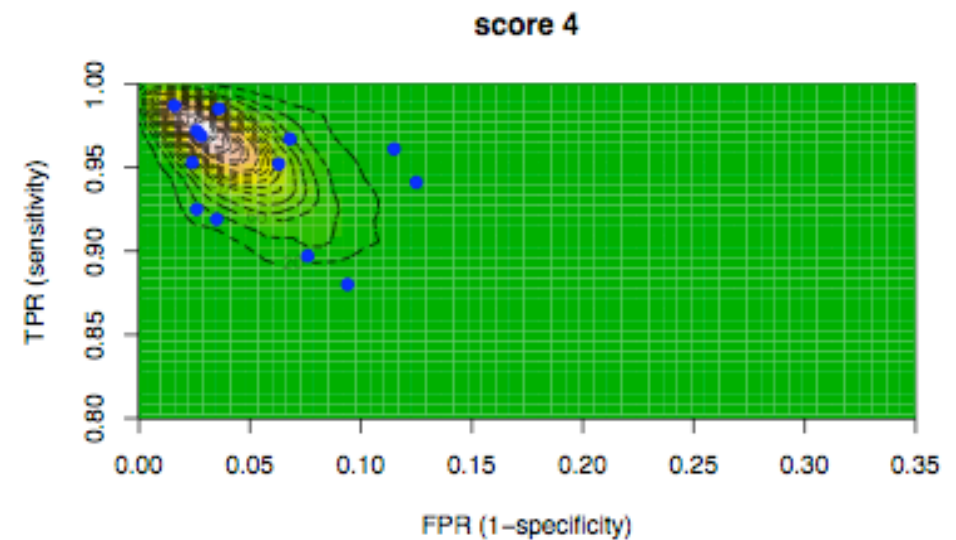
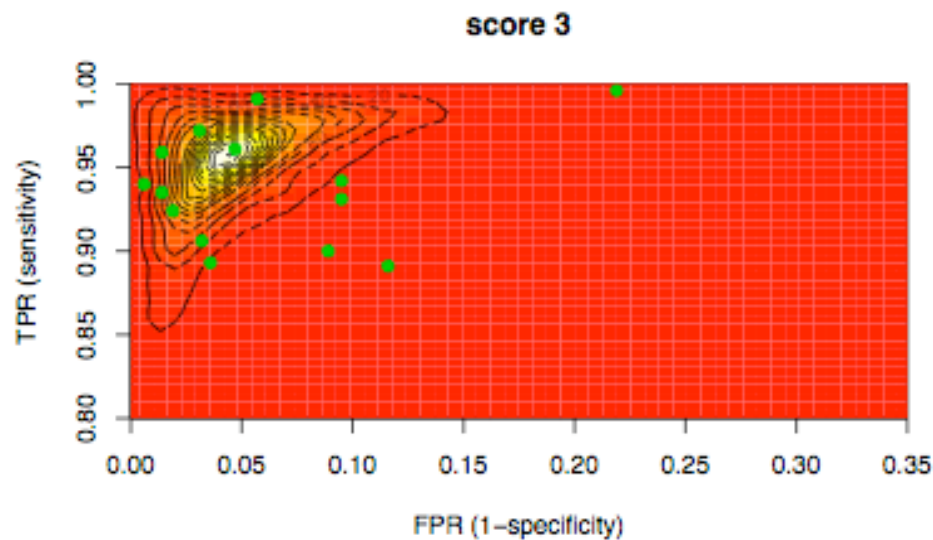
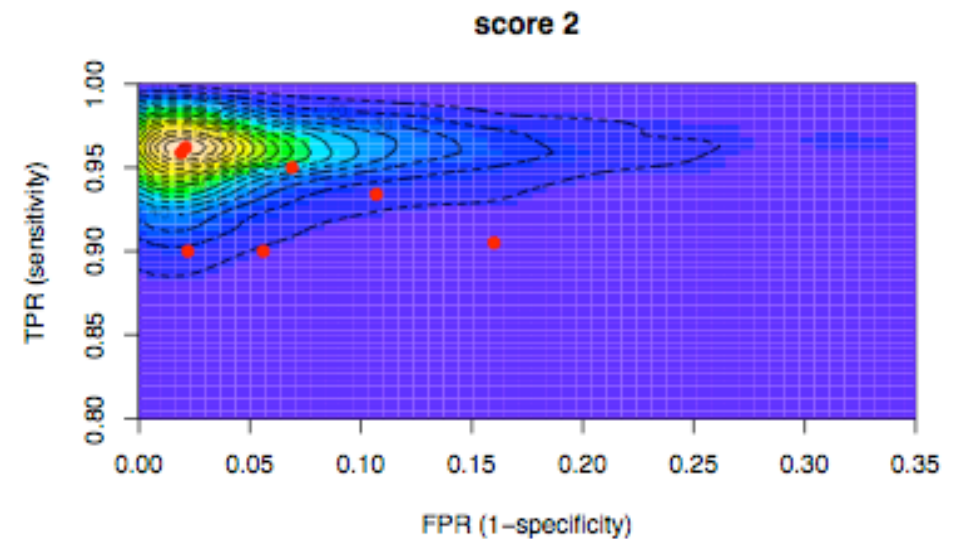
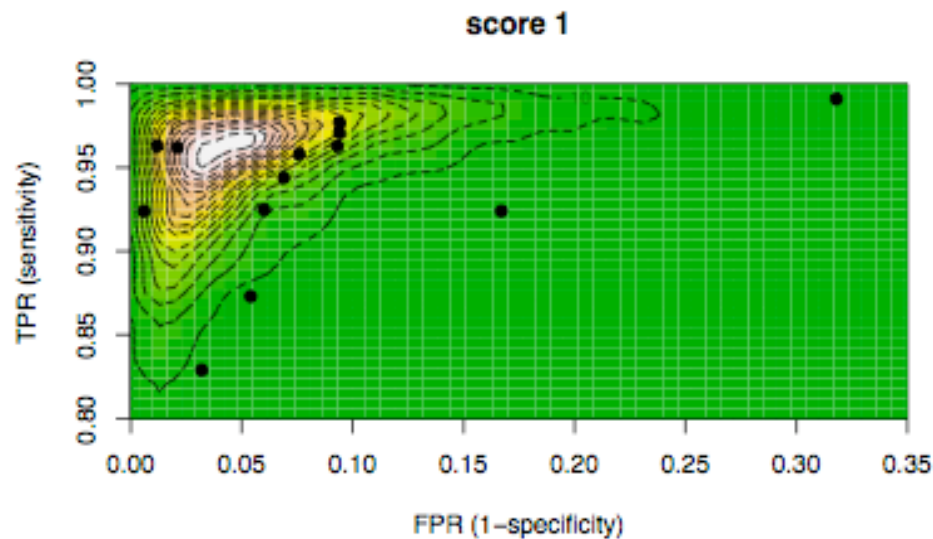
# Predictions: $\Pr(\text{TPR}^* | y)$ and $\Pr(\text{FPR}^* | y)$



	mean	95% CI
tprstar[1]	0.9395	[0.7569, 0.9932]
tprstar[2]	0.9485	[0.8487, 0.9898]
tprstar[3]	0.9438	[0.8089, 0.9908]
tprstar[4]	0.9417	[0.7866, 0.9925]

	mean	95% CI
fprstar[1]	0.0701	[0.0052, 0.3031]
fprstar[2]	0.1194	[0.0010, 0.7088]
fprstar[3]	0.0604	[0.0080, 0.2136]
fprstar[4]	0.0583	[0.0093, 0.2029]

# Predictive Summary Surface: $\Pr(\text{TPR}^*, \text{FPR}^* | y)$



## Conclusions / To do list / Open ...



- Analytical strategy shift:
  - From classical meta-regression to multiple variance components modelling
- Diagnostic test meta-analysis:
  - Benefits in promoting more flexible statistical summaries, e.g., predictions from the model.
  - Investigate the problem of indirect comparisons
  - Nothing is done in publication bias
- Knowledge transfer:
  - SROC is a successful story in meta-analysis!The challenge, now, is to communicate new technical possibilities.