Meta-Analysis for Diagnostic Test Data: a Bayesian Approach

Pablo E. Verde

Coordination Centre for Clinical Trials

Heinrich Heine Universität Düsseldorf

Preliminaries: motivations for systematic reviews

- To generalized a medical result in order to apply it to other populations, subgroups, etc.

- To collect background information to design a new clinical trial.

- Costs: it is substantially cheaper to analyze published information than carry out a new clinical trial.

- Health Care Evaluation: to evaluate critically published information.

Running Example

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) has been advocated to be of high potential diagnostic benefit in suspected appendicitis.

Question: Which diagnostic performance has this technology ?

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
Cole2001	USA	96	40	5	4	43	89	91
D'Ippolito1998	Brazil	52	40	4	0	8	91	100
Hershko2002	Israel	197	67	5	7	118	93	94

Pooled results with a random effects model:

Sensitivity: 94.5% [94.1, 95.8]
Specificity: 94.1% [91.8, 94.5]

Particular features of data coming from systematic reviews

- Studies included in a systematic review are unique, i.e., they are NOT a random sample of studies.

- The data collection is vulnerable to multiple sources of bias:

  study internal bias
  bias to external validity
  bias to inclusion criteria
  publication bias

- Usually an small number of studies are included in the review.

Bayesian methods are particularly well-suited to such scenario.

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

The Diagnostic Test Data Summaries

- 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)
- The false negative rate: FNR= c/(a+c)
- The Positive Predictive Value: PPV = a/(a+b)
- The Negative Predictive Value: NPV = d/(c+d)
- Likelihood ratio for a positive test: LR_+ = Sensitivity / (1- Specificity)
- Likelihood ratio for a negative test: LR_- = (1-Sensitivity)/Specificity
- Diagnostic Odds Ratio DOR= (a×d)/(b×c)

$$DOR = \frac{\left(\dfrac{Sensitivity}{1\text{-}Sensitivity}\right)}{\left(\dfrac{1\text{-}Specificity}{Specificity}\right)} = \frac{LR_+}{LR_-}$$

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 tread off between TPR and FPR.

The construction of diagnostic test introduces in general strong correlation between TPR and FPR.
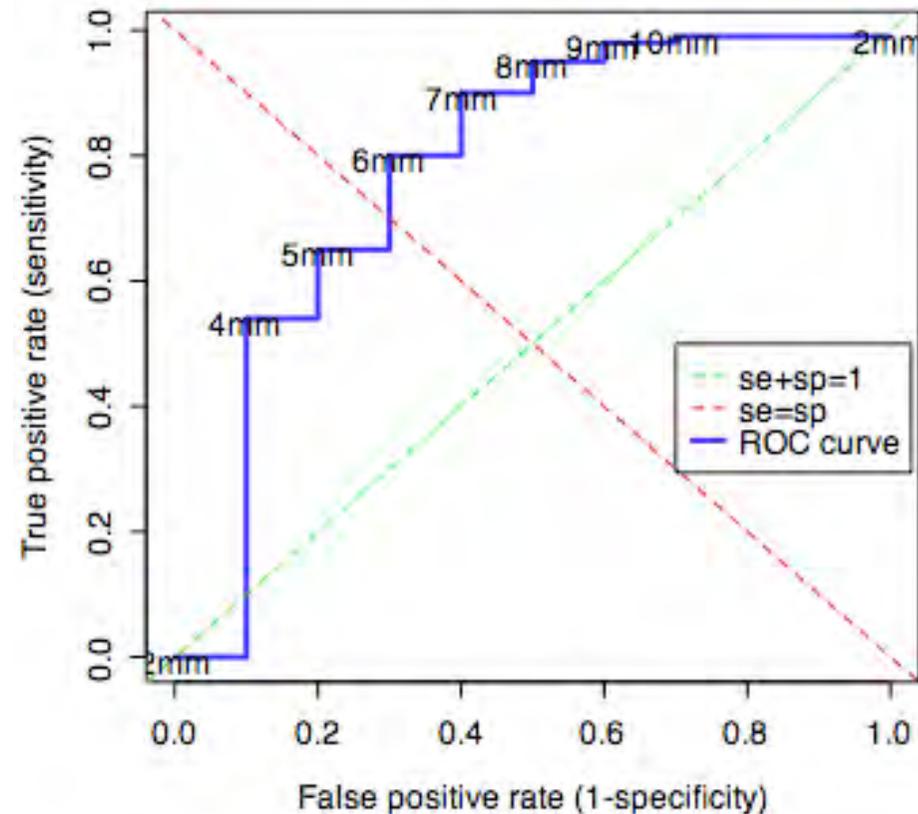
The Summary ROC curve (SROC)

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!).

It can be achieved by using a meta-regression model (Moses et al. 1993) that under a random effect model is:

$$D_i \sim \text{Normal}(A + BS_i, \sigma^2_{Di} + \tau^2),$$

$D_i = \text{logit}(TPR_i) - \text{logit}(FPR_i)$, $S_i = \text{logit}(TPR_i) + \text{logit}(FPR_i)$ and σ_{Di} is the asymptotic standard error of D_i.

Reverse the transformations and deduce the relationship between TPR and FPR:

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

The SROC curve analysis
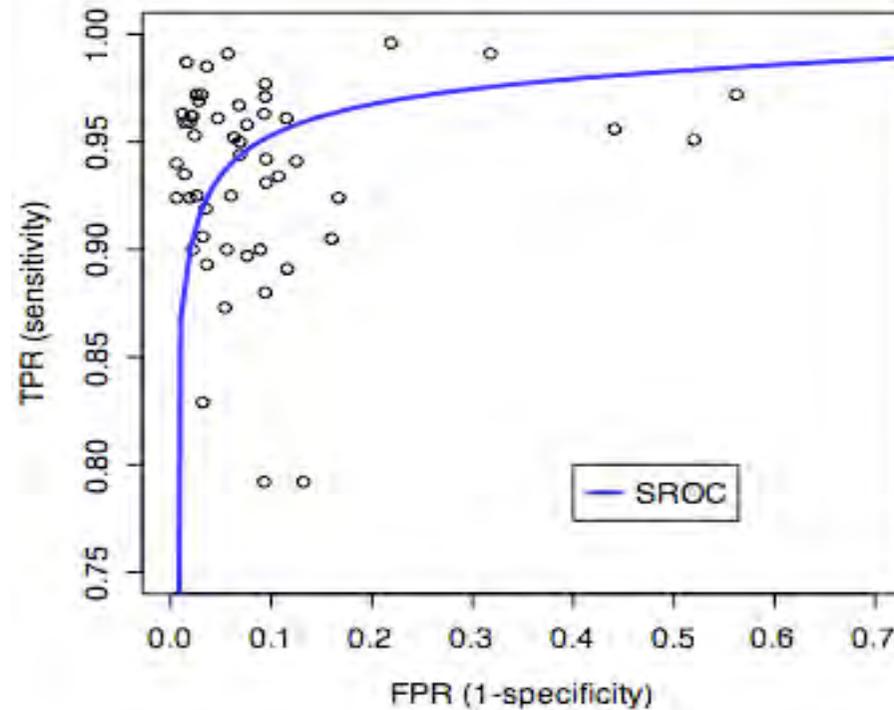

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.5783     0.1768  31.554  < 2e-16 ***
S            -0.3683     0.1227  -3.001  0.00419 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Limitations of the SROC curve

It has been developed, essentially 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:

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

is only analytically tractable for B=0.

A practical Bayesian approach may complement these issues.

A Basic Bayesian Model

Given the data "y" of "m" studies, we can model the number of true positives a_i and false positives b_i directly by a GLMM as follows:

$$a_i \sim \text{Binomial}(TPR_i, n_{i,1}), \quad b_i \sim \text{Binomial}(FPR_i, n_i - n_{i,1}),$$

with

$$\text{logit}(TPR_i) = (D_i + S_i) / 2, \quad \text{logit}(FPR_i) = (D_i - S_i) / 2,$$

and

$$(D_i, S_i) \sim \text{MNormal}(\mu, \Lambda), \Lambda = \Sigma^{-1}.$$

Note: the $\text{logit}(\cdot)$ transformation can be replaced by other link function (e.g. probit or complementary log logistic, etc.). The multivariate Gaussian assumption can, also, be replaced by a multivariate t or other density.

Prior specification

Independent Normals for the components of $\mu = (\mu_D, \mu_S)$:

$$\mu_D \sim N(m_D, v_D), \quad \mu_S \sim N(m_s, v_s) .$$

Independent Uniforms for the variance covariance matrix of D_i and S_i, $\Lambda^{-1} = \Sigma$:

$$\sigma^2_{D,D} \sim Uniform(0, k_D), \quad \sigma^2_{S,S} \sim Uniform(0, k_S),$$

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

The constants m_D, m_s, v_D, v_s, k_D, k_S and r can be use for prior elicitation and sensitivity analysis.

Inferential Statements

- Observed quantities, i.e., data y

- Unknown quantities are represented by θ (parameters, missing data, ...)

- Inference is based on the posterior distribution

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta).$$

- Inference on functional parameters, say $g(\theta)$, is based on the marginal posterior $p(g(\theta) \mid y)$.

- Predictions of new data y^* are based on the marginal predictive $p(y^* \mid y)$

Those problems are no analytical tractable, we based our inference on empirical approximations using MCMC (a Gibbs sampler).
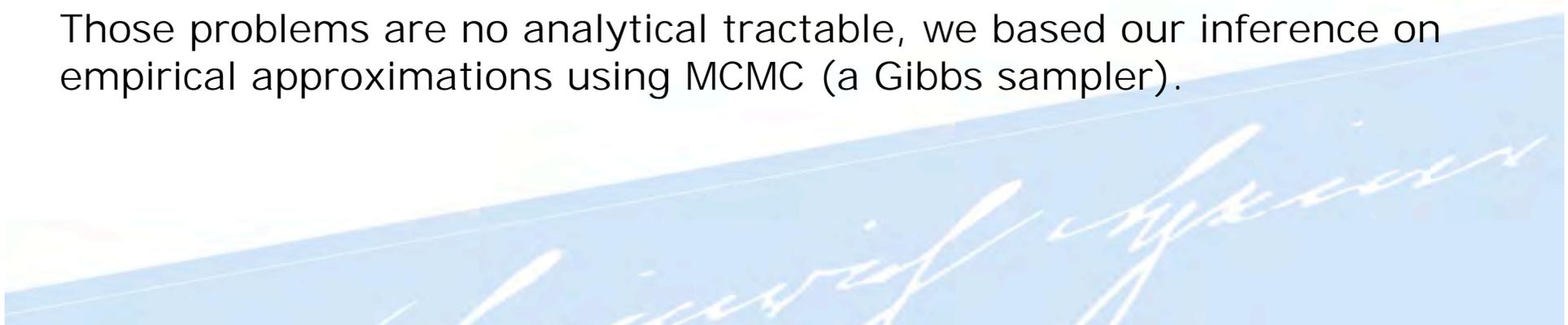
The directed acyclic graph (DAG)

The posterior density $p(\theta \mid y)$ is factorised using a "directed local Markov" property

$$p(\theta \mid y) = p(\theta_v \mid \text{parents}[v], y),$$

This decomposition:

- defines relationship between model quantities,

- programs the kernel of the Markov chain,

- gives a non-algebraic description of the model.

The DAG of the model
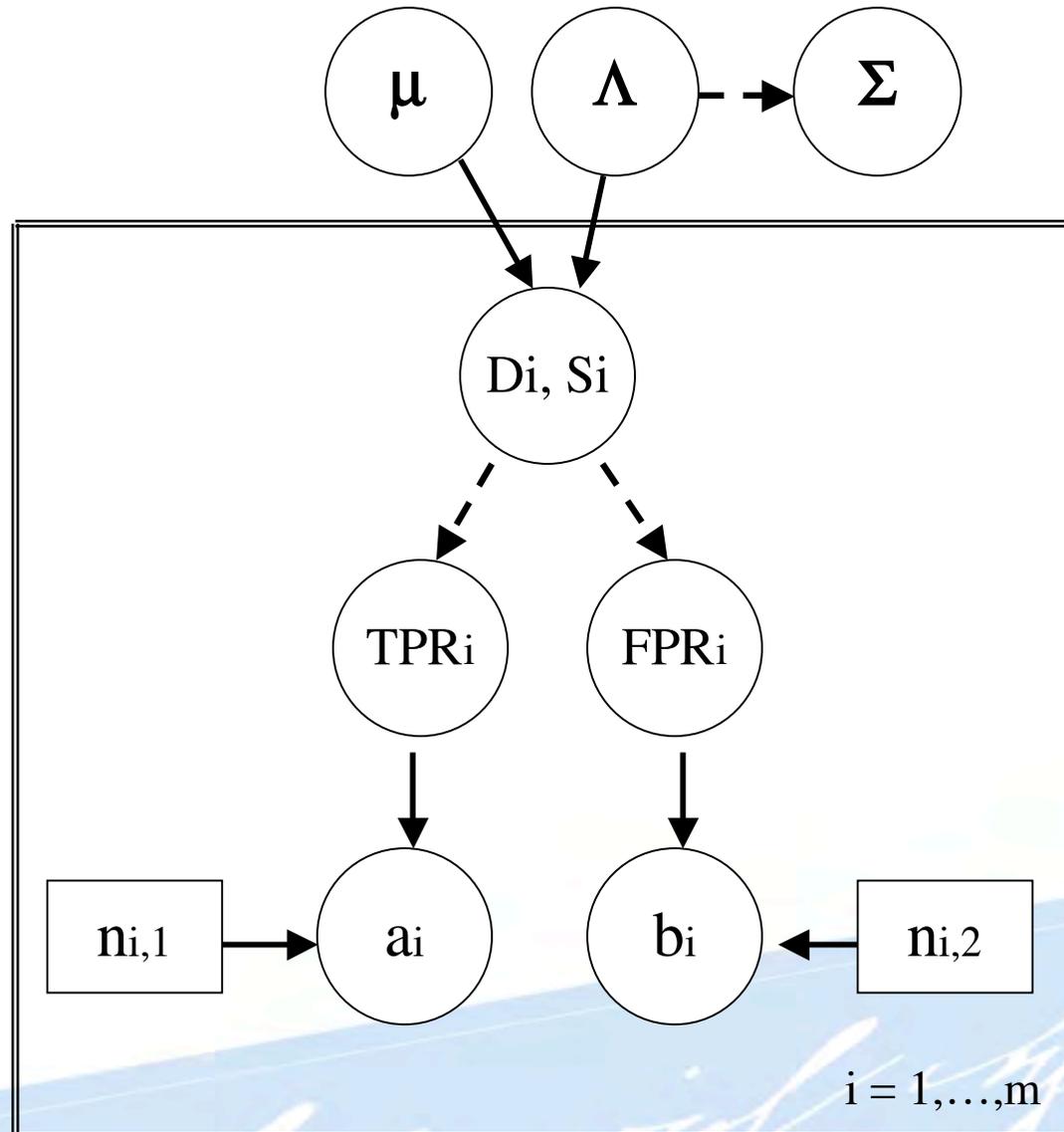
Recovering the SROC

From the distribution of $(D_i \mid S_i = s_i)$ we can recover the SROC as follows:

$$E(D_i \mid S_i = s_i) = A + B (s_i - \mu_S),$$

$$A = \mu_D - \mu_S \, \rho \, \sigma_{D,D} / \sigma_{S,S} \quad \text{and} \quad B = \rho \, \sigma_{D,D} / \sigma_{S,S}$$

$$\text{var}(D_i \mid S_i = s_i) = \sigma^2_{D,D} (1-\rho^2).$$

- 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 posteriors samples.

Extended DAG for the SROC curve
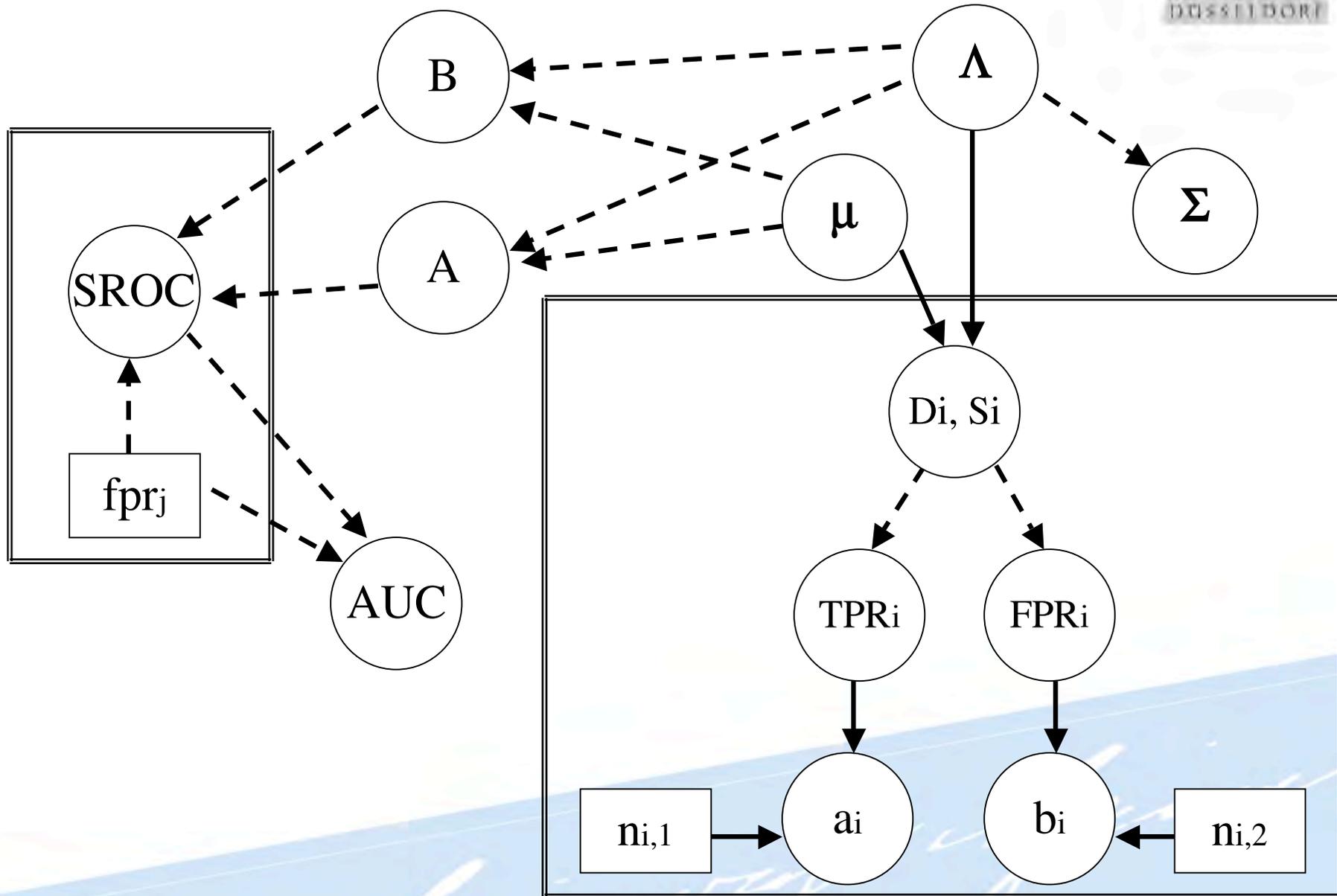
Making predictions and simulating data from the model

Reasons for predictions and data simulation:

- Predict the results of a new study.

- Assess the results of a study not included in the review.

- Proof-of-concepts and plan a new study.

- Model checking and model selection.

To predict a new study results we define an "*stochastic node child*" (D^*, S^*) of $p(\mu, \Lambda \mid y)$ and two logical nodes $TPR(D^*, S^*)$ and $FPR(D^*, S^*)$. These results in a predictive posterior of TPR^* and FPR^*.

Given the sample size of the study, new data is simulated as an "*stochastic node child*" of $Bin(TPR^*, n1)$ and $Bin(FPR^*, n2)$.

Extended DAG for predictions
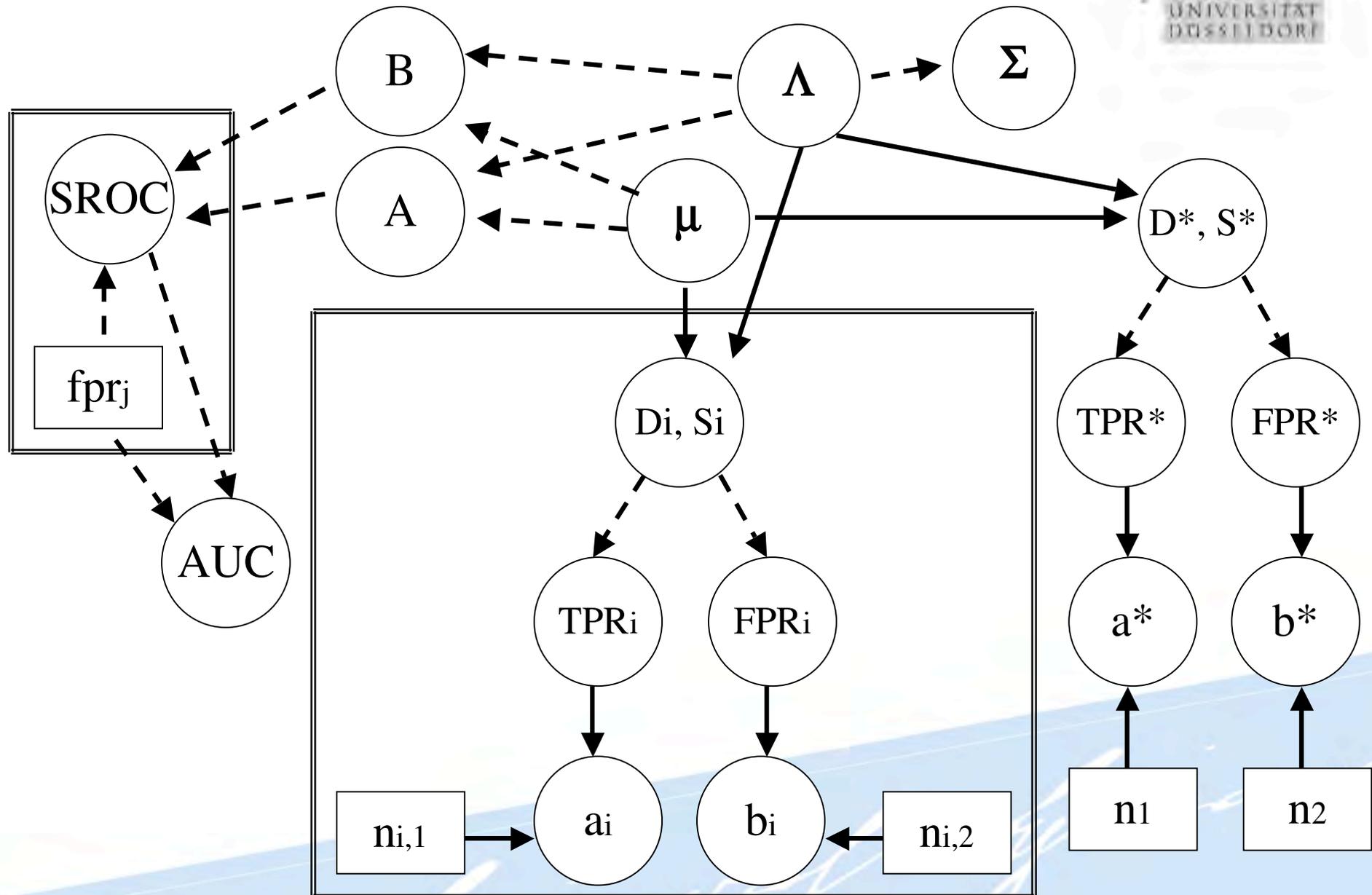
Analysis

Computational specification:

No prior elicitation was done. We take the hyper-prior constants:
$m_D=0$, $m_s=0$, $v_D=0.05$ $v_s=0.05$, $k_D=100$, $k_S=100$ and r=1.

MCMC set up: We run a single chain of length 20,000 we take one in 5 we discard the first 500 values. 3500 values to make inference.
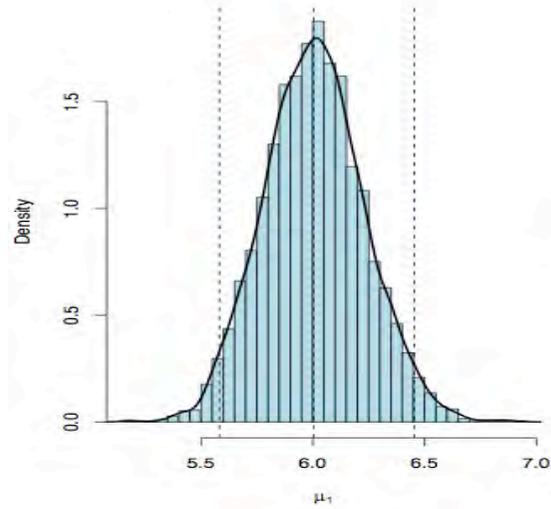
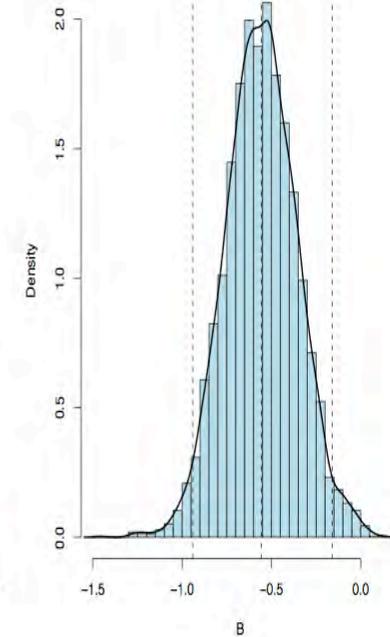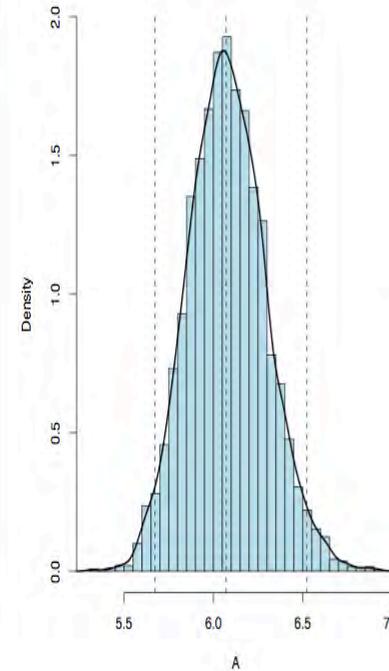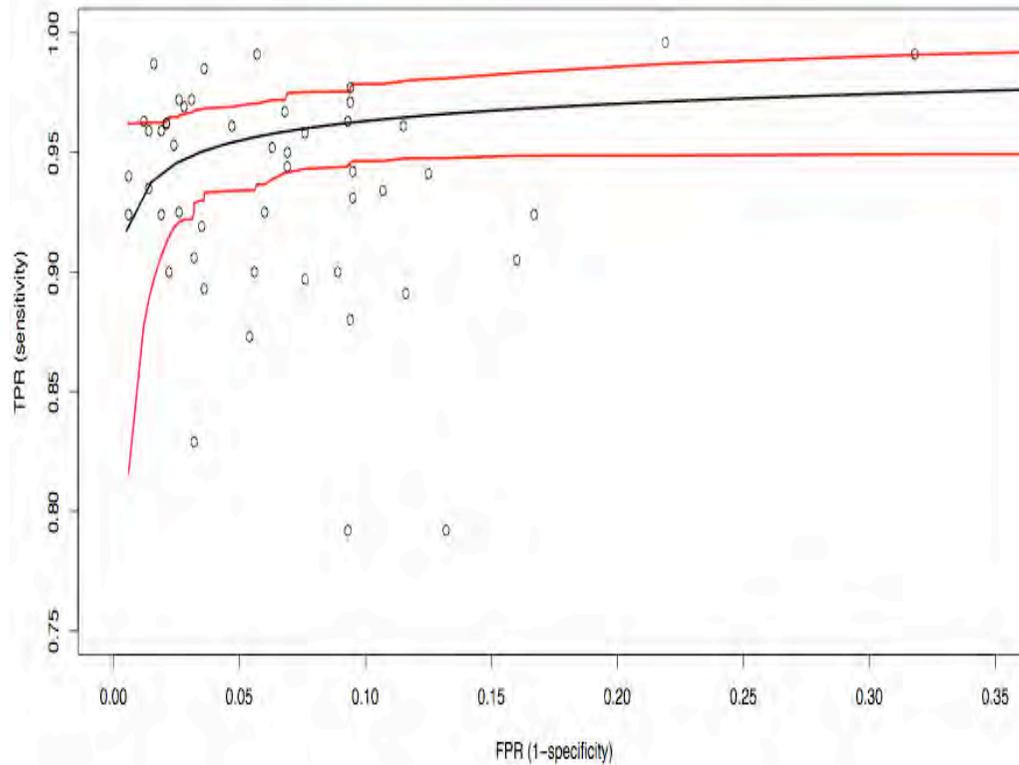Convergence diagnostics were done graphically. No mayor convergence problems were presented.

Pr(TPR$_i$|y) and Pr(FPR$_i$|y)

Pr(μ, Σ|y)
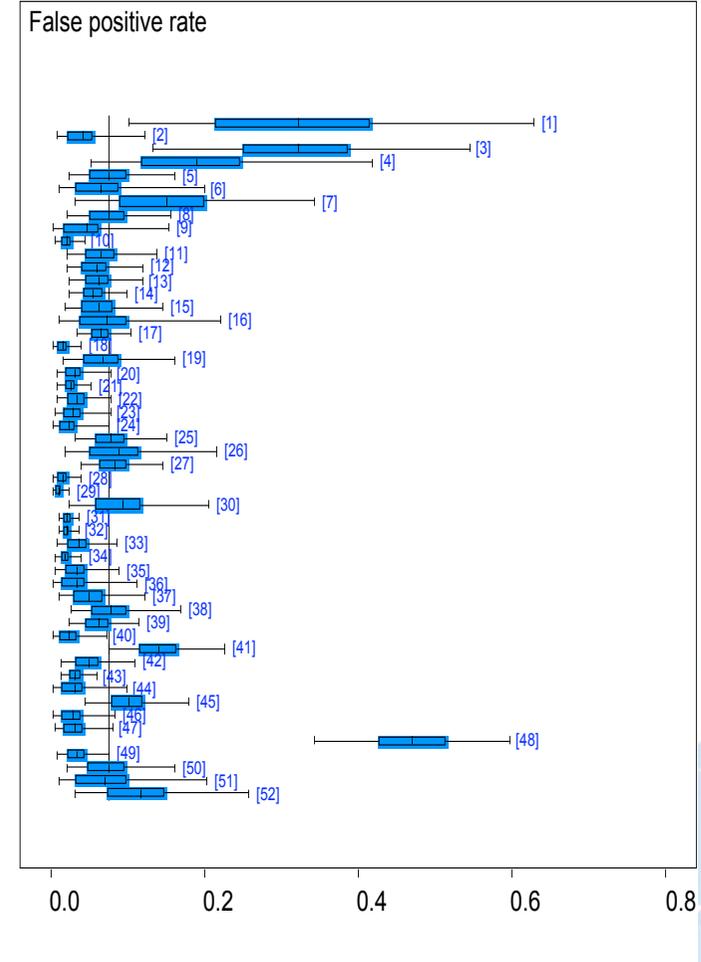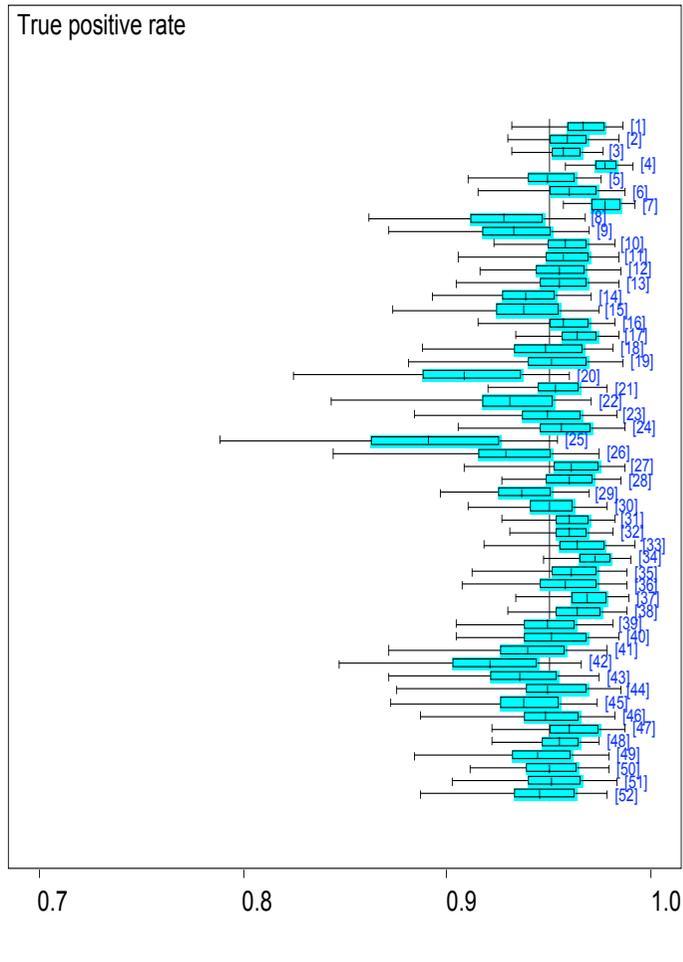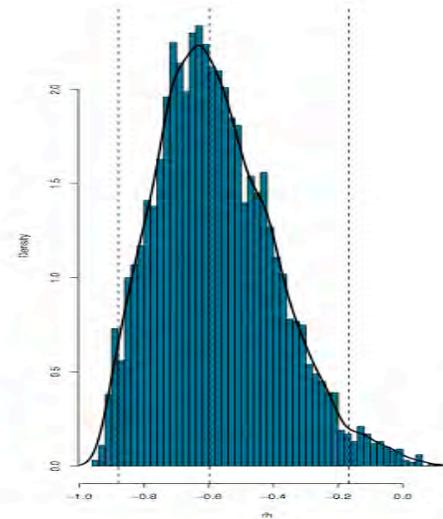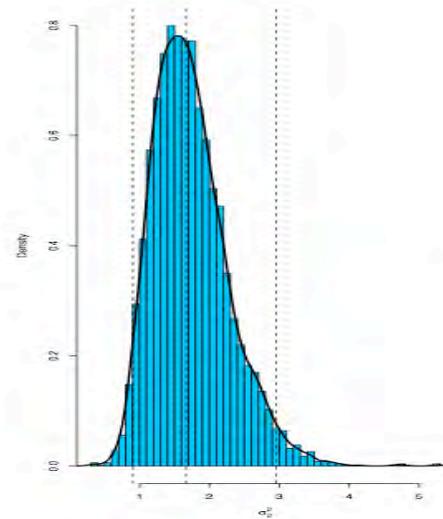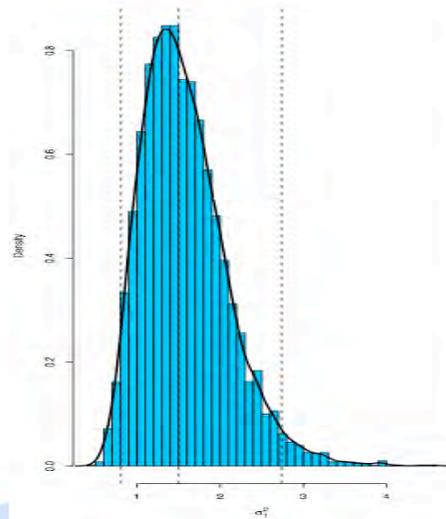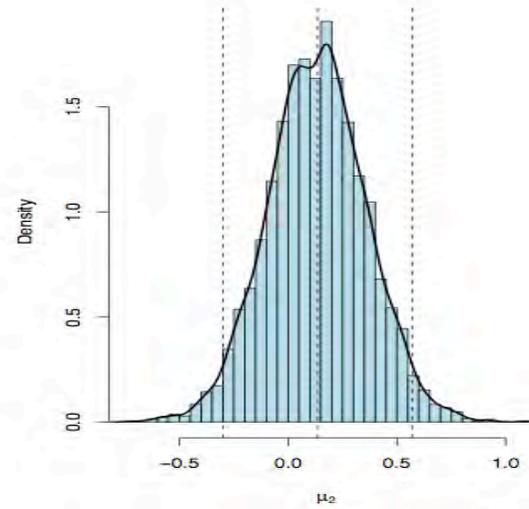
Pr(SROC| v). Pr(A|y) Pr(B|y)
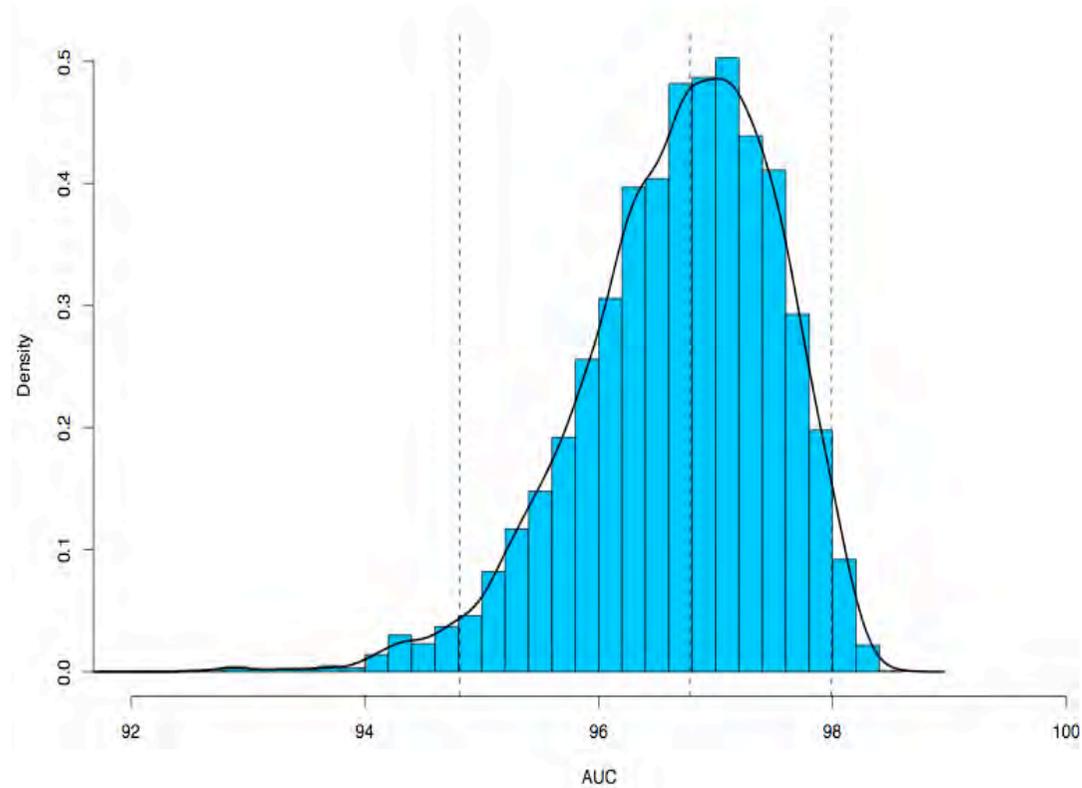

CI 95%:

A: 6.062 [5.654, 6.4775] the length is 25% wider than RE
B: -0.571 [-0.953, -0.167] here the length is 63% wider!

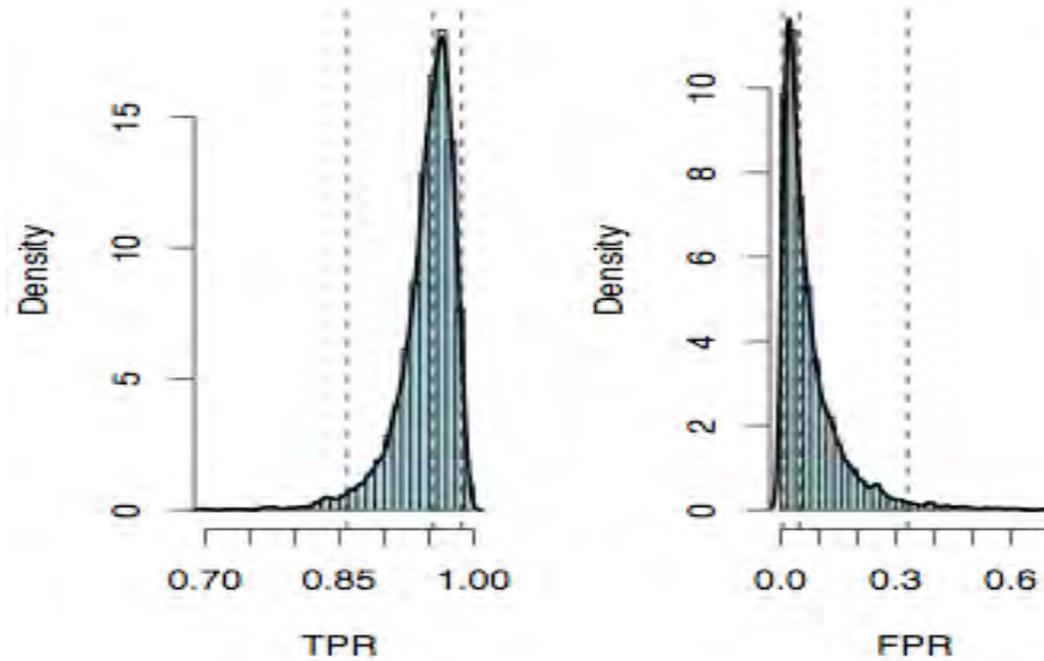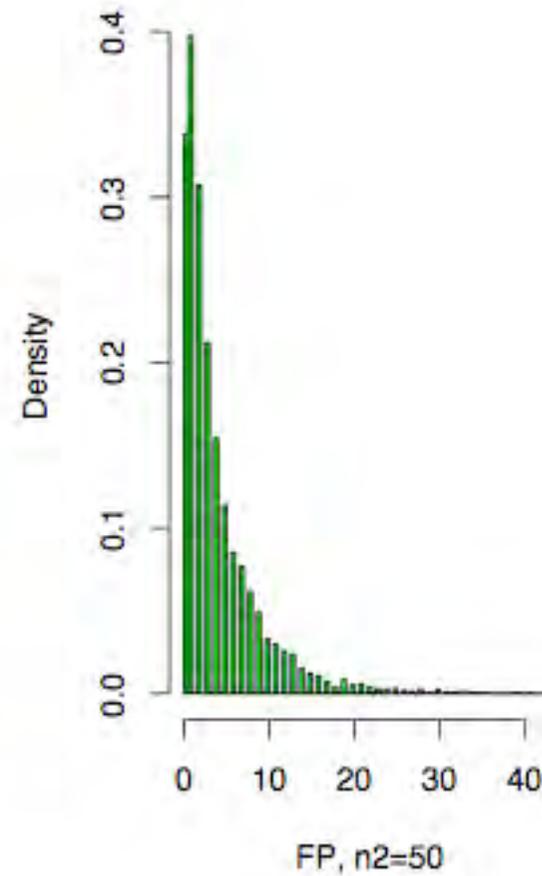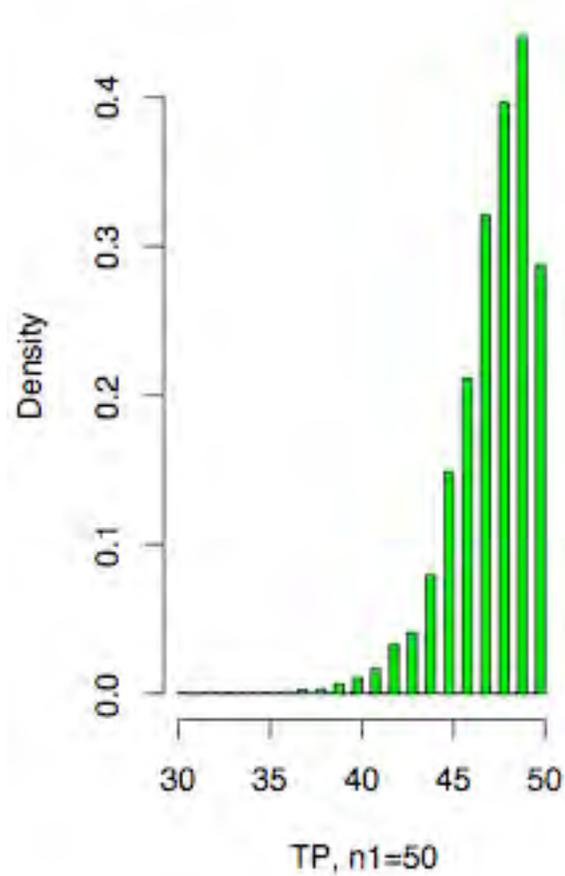# Pr(AUC|y)

CI 95%:
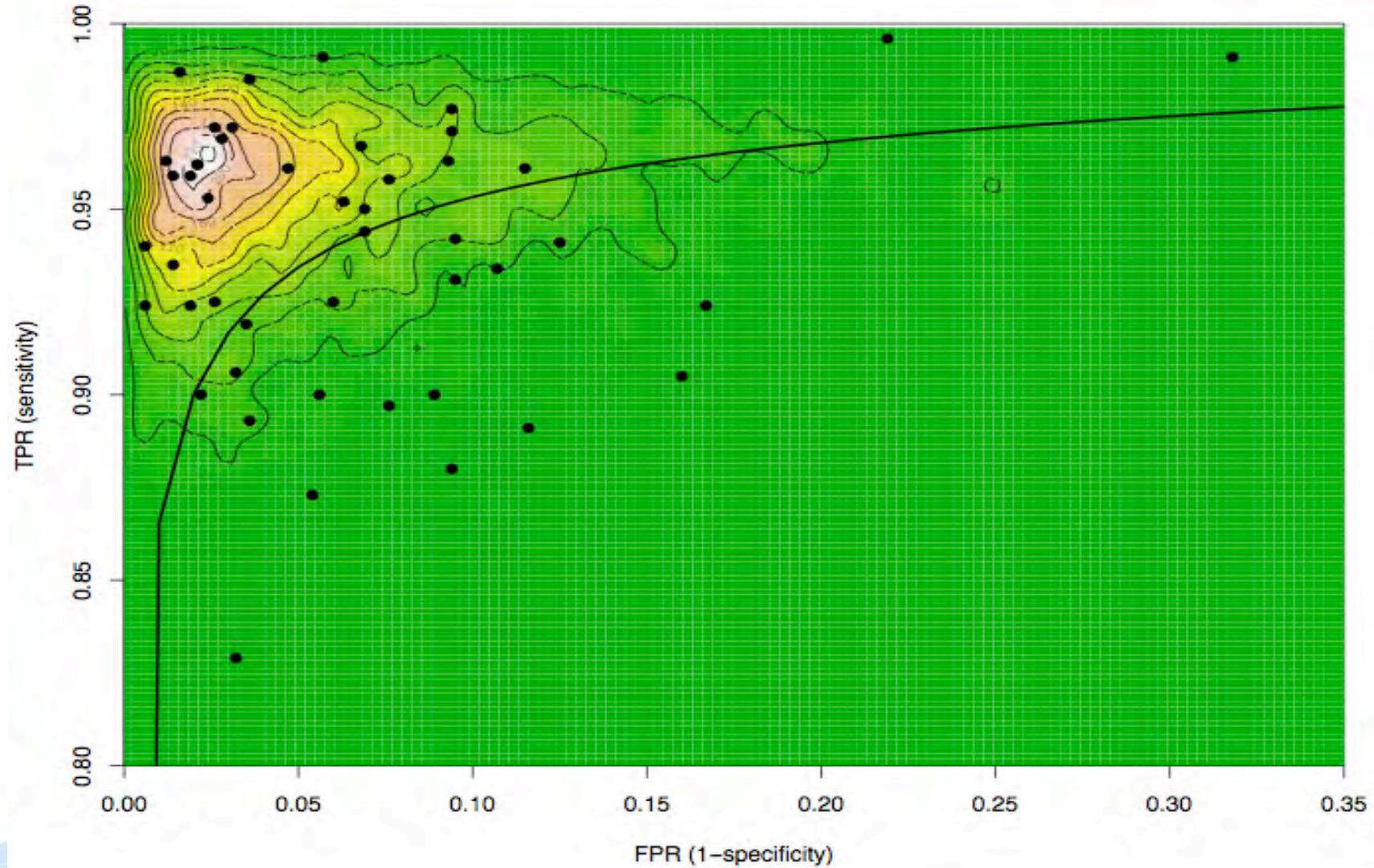
AUC: 96.73, [94.75, 97.94]

Predictions: Pr(TPR*|y) and Pr(FPR*|y)

CIs 95%

TPR*: 0.954 [0.867, 0.986] ;   1 - FPR*: 0.951 [0.670, 0.994]

Predictions: $\Pr(TP^*|y)$, $\Pr(FP^*|y)$, n=100

TP, n1=50   FP, n2=50

Predictive Summary Surface:
Pr(TPR*,FPR*|y)

Further work ...

- Model diagnostic: sensitivity to priors; residuals at different levels

- Generalized evidence synthesis:

  How to combine studies when they are different
  Link explanatory variables
  Model study quality

- Publication bias

Conclusions

[image removed]