

META-ANALYSIS OF DIAGNOSTIC TEST: A NEW BAYESIAN APPROACH

Pablo E. Verde

Coordination Centre for Clinical Trials
Heinrich-Heine Universität Düsseldorf

ISCB-2009
26-August-2009

[<pabloemilio.verde@uni-duesseldorf.de>](mailto:pabloemilio.verde@uni-duesseldorf.de)

Diagnostic test data

Test results of the study i ($i = 1, \dots, N$) are summarized in a 2×2 table giving the number of positive and negative test results for patients with and without disease:

		Patient status	
		With disease	Without disease
Test outcome	+	tp_i	fp_i
	-	fn_i	tn_i
Sum:		$n_{i,1}$	$n_{i,2}$

- Diagnostic summaries:
 - True positive rate or sensitivity: $\widehat{TPR}_i = tp_i/n_{i,1}$
 - True negative rate or specificity: $\widehat{TNR}_i = tn_i/n_{i,2}$
 - False positive rate: $\widehat{FPR}_i = fp_i/n_{i,2}$, etc.

Question: How can we combine and summarize these information ?

Meta-Analysis of Diagnostic Data Special Features

Meta-analysis of diagnostic tests data implied the synthesis of complex and imperfect evidence.

- **Dependency between diagnostic summaries:**
 - Sensitivity and specificity are linked together by the diagnostic threshold used.
- **Different sources of heterogeneity:**
 - Different study designs, population characteristics, diagnostic set-up, etc.
- **Reporting deviations:**
 - The way that results are published differ, some relevant information may be incomplete.

There is a large amount of methodological work (SROC, HSROC, Bivariate random effects, etc.) in general they do not reflex the complexity of the problem, little or no model checking at all.

A New Bayesian Bivariate Model

The model has three components:

- A data model for the observed tables
- A structural distribution:
 - Based on scale mixing distributions
 - Study accuracy effects
 - Study relevance (new concept)
 - Meta-regression and structural systematic dispersion (not presented here, see the original paper)
- Priors based on weakly informative distributions

We follow a [Bayesian Data Analysis](#) approach with extensive application of predictive posteriors for [model checking](#) and meta-analysis summaries.

A New Bayesian Bivariate Model

Data model:

Let tp_i and fp_i be the true positive and false positive results for study i ($i = 1, \dots, N$), we model these quantities with binomial distributions

$$tp_i \sim \text{Bin}(\text{TPR}_i, n_{i,1}), \quad fp_i \sim \text{Bin}(\text{FPR}_i, n_{i,2}),$$

where TPR_i and FPR_i are the probabilities to observe a positive test result in the disease and non-disease population respectively.

Structural distribution: study accuracy effects

We model the variability between studies with **differences** and the **sums**

$$D_i = g(\text{TPR}_i) - g(\text{FPR}_i), \quad S_i = g(\text{TPR}_i) + g(\text{FPR}_i) \quad (1)$$

with a *scale mixture of bivariate normal distributions*

$$(D_i, S_i) \sim \mathcal{N}(\mu, \Psi_i), \quad i = 1, 2, \dots, N, \quad (2)$$

$$\Psi_i = w_i \times \Lambda, \quad (3)$$

$$w_i \sim p(w_i), \quad (4)$$

where the link $g(\cdot)$ transforms the N pairs of probabilities TPR_i and FPR_i to a scale where they are defined in the range $(-\infty, \infty)$, Λ is the precision matrix, i.e., $\Lambda = \Sigma^{-1}$ and $p(w_i)$ a scale mixing density.

Structural distribution: Scale Mixing Densities

Depending on the functional form $p(w_i)$ we have different marginal distributions of (D_i, S_i) . Two particular cases are:

- The bivariate Normal corresponds to $w_i = 1$ and $p(w_i)$ is degenerated at one.
- The bivariate t -distribution corresponds to $w_i \sim \Gamma(\nu/2, \nu/2)$ with ν degrees of freedom.

The mixture weights w_i are just computational parameters to construct heavy tailed distributions. However, in this framework we are going to take them seriously!

Structural distribution: The Study Relevance

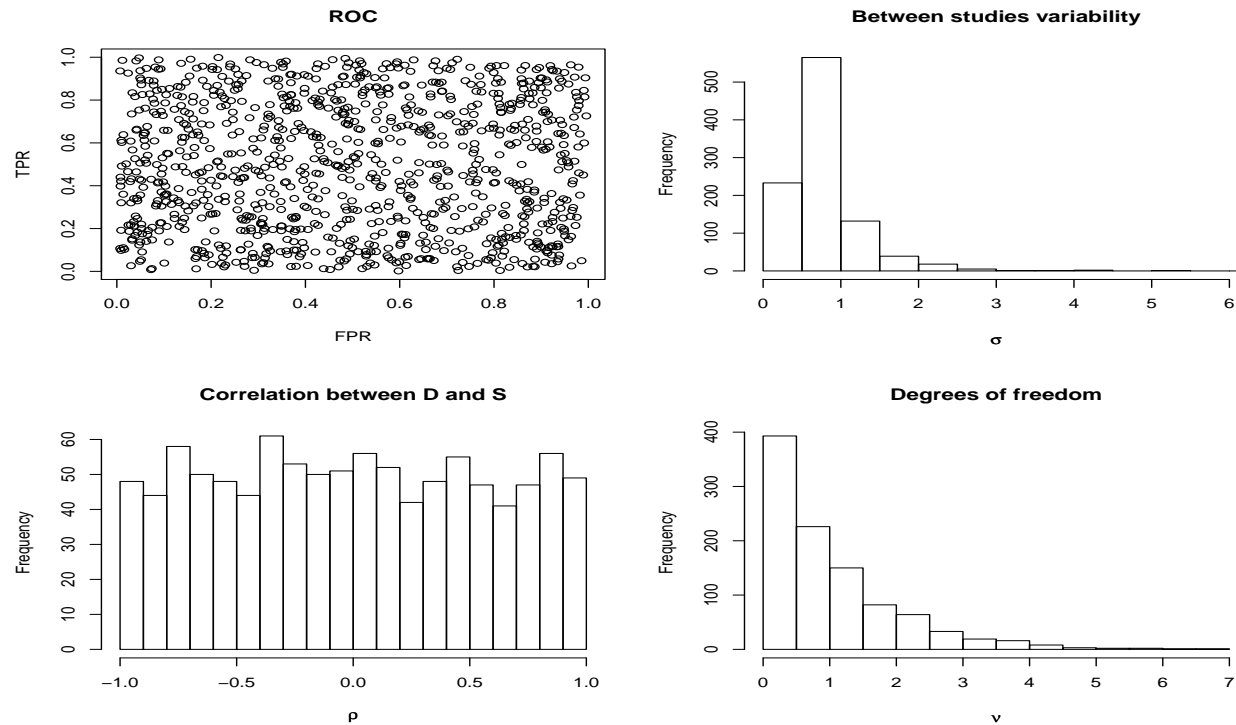
We give direct interpretation to the mixture weights w_i . We called this weights *The Study Relevance*.

- We say that a study is relevant to the systematic review, if its results do not substantially deviate from the rest of the studies, i.e. $E(w_i | Data) \approx 1$
- All studies are relevant a-priory with $E(w_i) = 1$
- Less relevant studies will concentrate their posterior distribution of w_i with values substantially less than 1, say $w_i < 0.7$

Clearly if all $w_i \approx 1$ a bivariate normal is an appropriate model. If some of these w_i s are lower than 1 then the effect of these studies will be down-weighted resulting in a [robust inferential method](#).

Weakly Informative Priors

$\mu \sim N(0, 0.25 I)$, $\Lambda \sim \text{Wishart}(I, 3)$ and $\nu \sim \text{Exp}(1)$. Some prior predictions:



Summary quantities of interest

- **Pooled summaries** at the group level: The posterior distribution of

$$\text{TPR} = g^{-1}[(\mu_D + \mu_S)/2], \quad \text{FPR} = g^{-1}[(\mu_D - \mu_S)/2]. \quad (5)$$

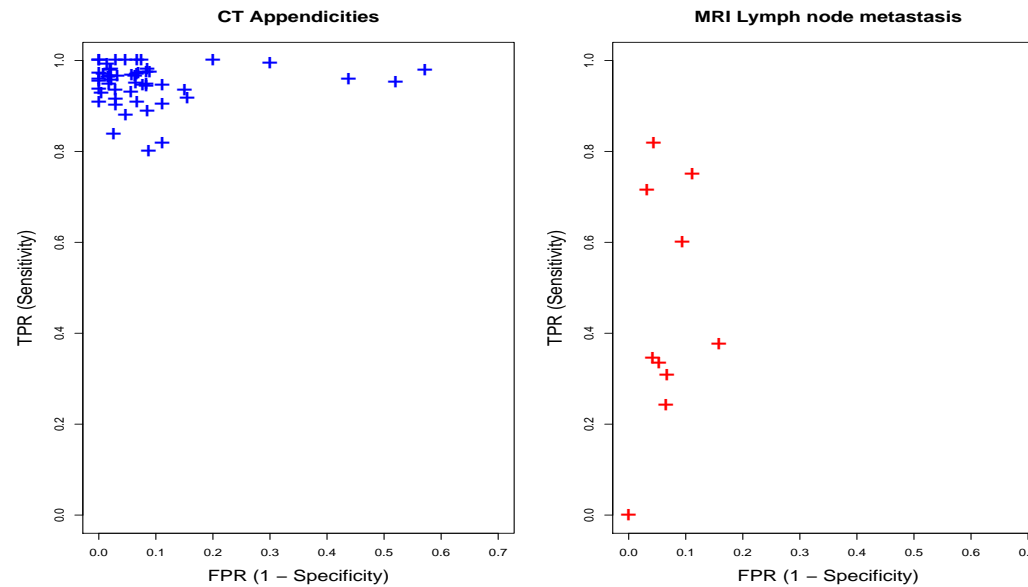
- **Predictive summaries** for a study not included in the review:

$$p(\text{FPR}^{pred}, \text{TPR}^{pred} | Data). \quad (6)$$

- **Other summaries** are possible: summaries at study level, Bayesian versions of SROC, etc.

All of these posteriors are analytical non-tractable, all calculations are based on Markov Chain Monte Carlo ([MCMC](#)) computations implemented in BUGS and R.

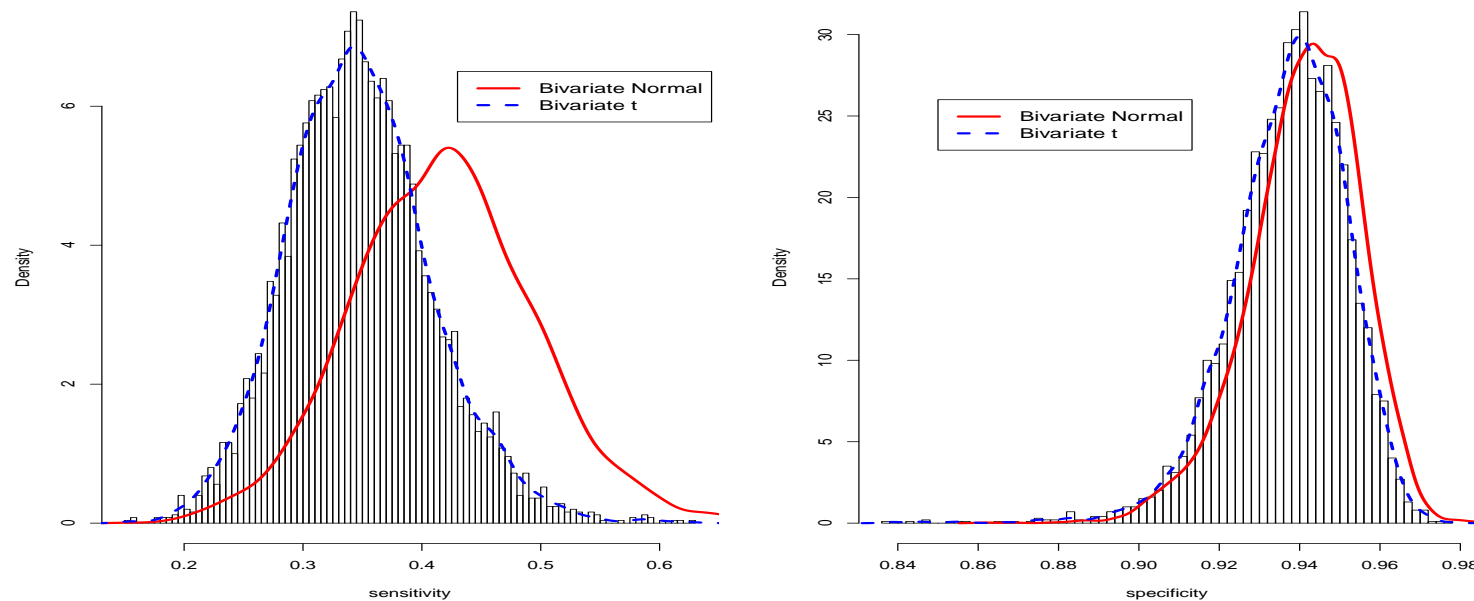
Two Running Examples



Left panel: 52 studies, computer tomography (CT) scans in the diagnosis of appendicitis (Ohmann et al. 2006).

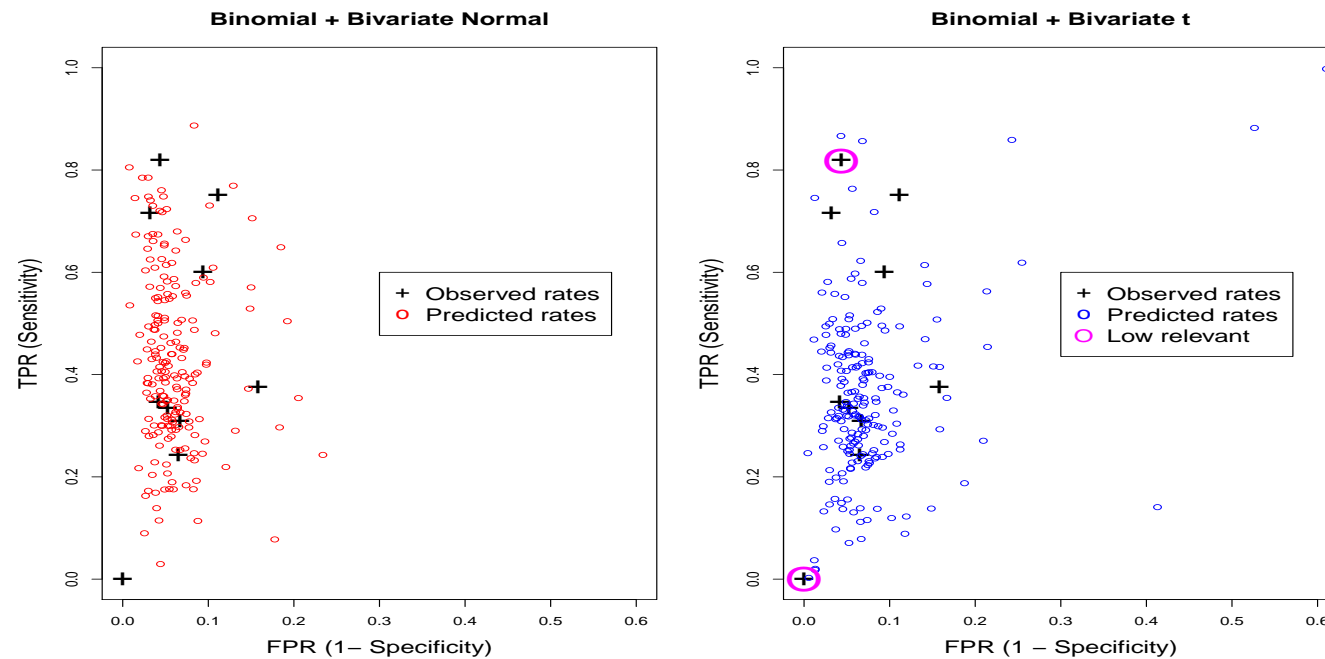
Right panel: 10 studies, magnetic resonance imaging (MRI) for the diagnosis of lymph node metastasis in women with cervical cancer (Scheidler et al. 1997).

Results for MRI: Pooled Summaries



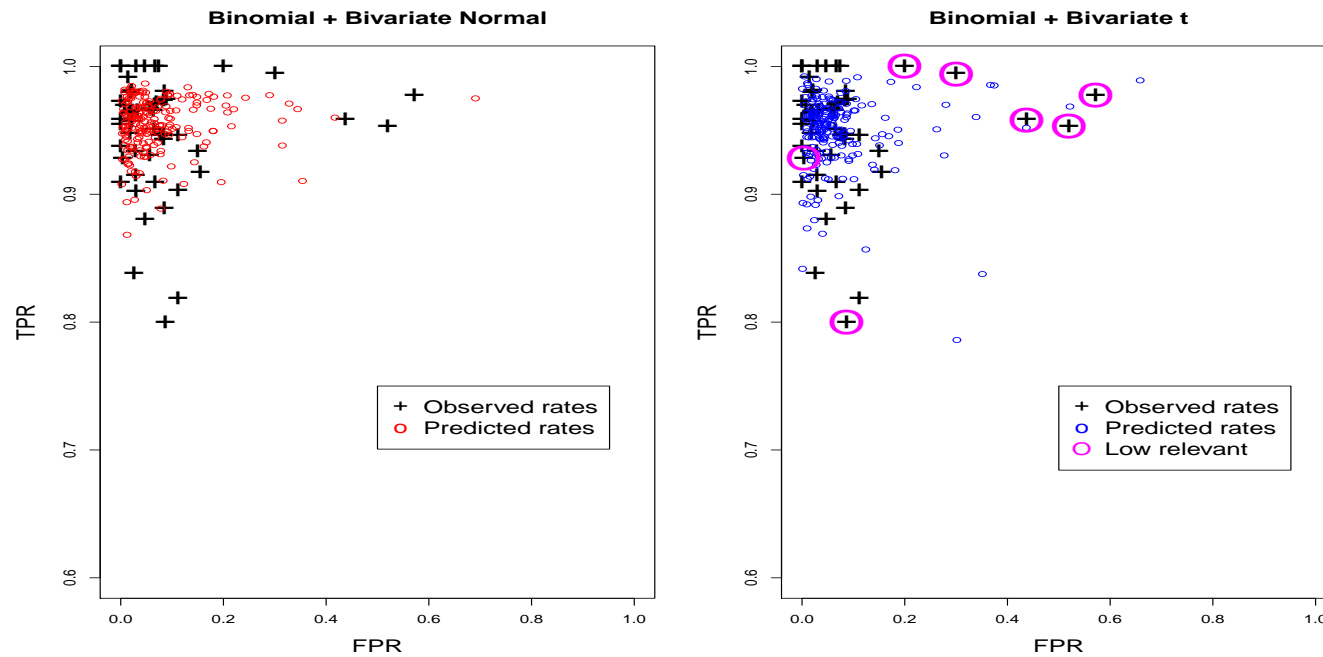
Results with logistic link. *Bivariate Normal*: DIC = 82.4 *Bivariate t-distribution*: DIC = 79.8. Estimated df $\nu = 2.634$. The model with bivariate *t*-distribution gives 30% narrow posteriors for the pooled sensitivity.

Results for MRI: Predictive model checking



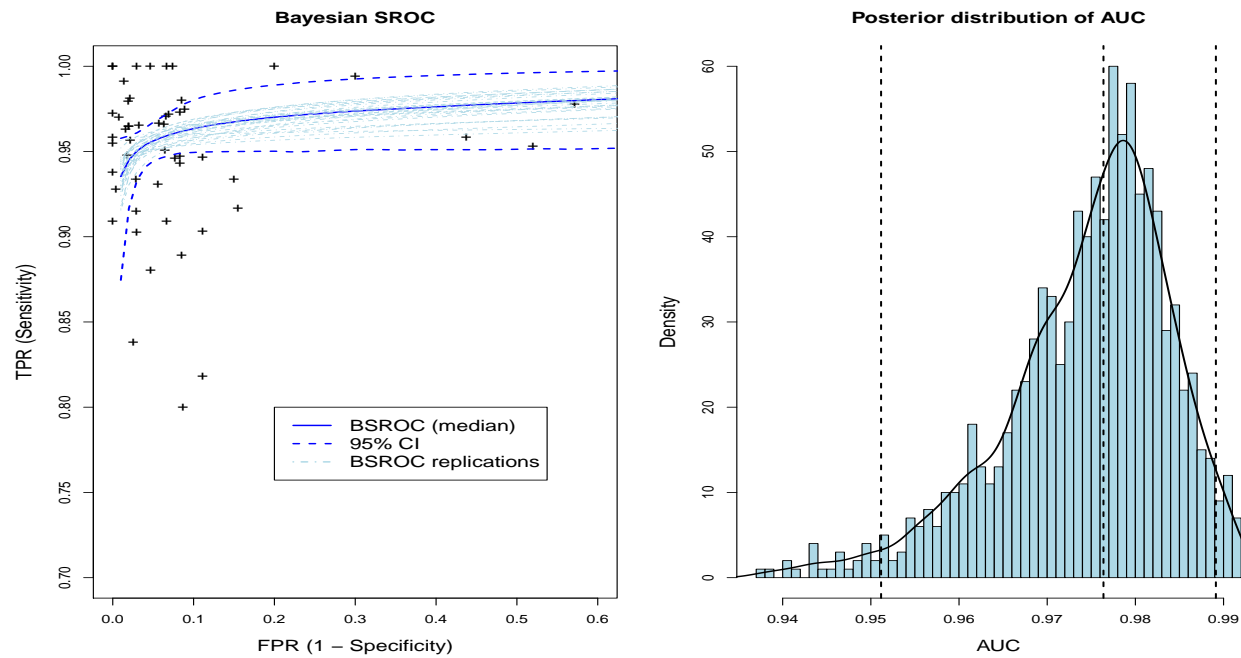
There are 2 studies with *low relevance*, one reported 0% sensitivity and 100% specificity and the other a big study with 272 patients with a high imbalance between disease and non-disease groups.

Results for CT: Predictive model checking



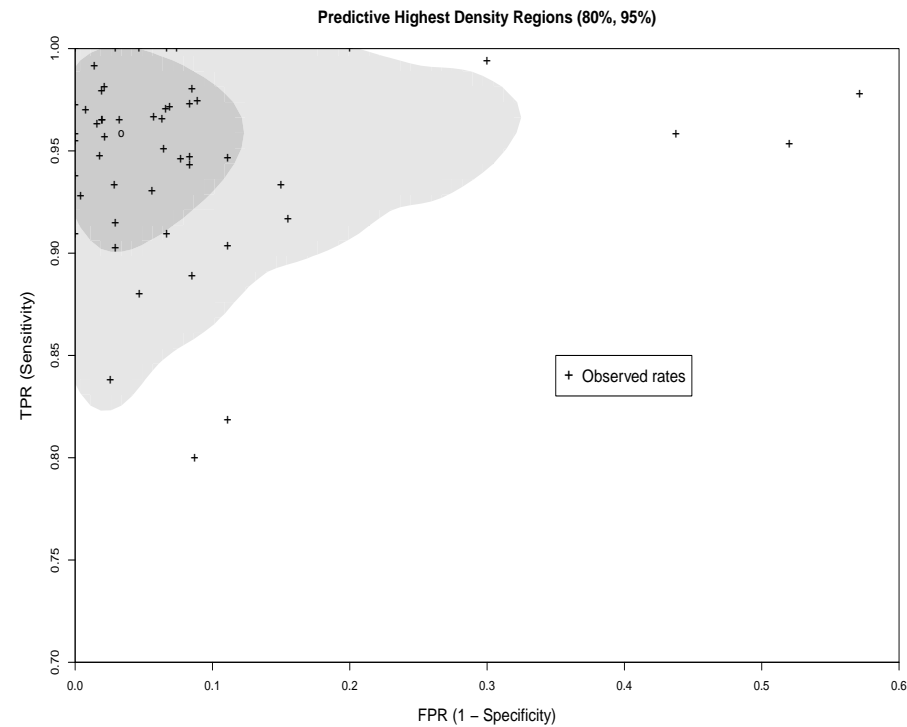
Bivariate Normal: DIC = 426. *Bivariate t-distribution*: DIC = 407. $\nu = 3.675$, 7 studies with *low relevance*!

Pooled summaries: Bayesian SROC and AUC



The **BSROC** and the **AUC** are *ad hoc* summaries. They do not represent the problem at hand.

Predictive summary: The Highest Density Region



The predictive HDR is the **more informative** summary for meta-analysis of diagnostic test data!

Conclusions

The Bayesian bivariate model allows us to realistically model the multiple of uncertainty presented in the data. In particular:

- For model building, prediction and summary we found the bivariate approach more appropriate than a ROC approach.
- The bivariate *Binomial+Normal* model may be too simple for the data we face and it may be too sensible to outliers.
- A bivariate *Binomial+ t* model is more efficient than the Normal in terms of estimation and more flexible in terms of prediction.
- Summaries based on predictive quantities (e. g. HDR) which go hand in hand with model checking are more appropriate than ROC curves and other summaries (e.g. AUC).

Further extensions and applications: a new R package called [bmeta](#) will soon be available and a paper with more details will be submitted to *Statistics in Medicine*.

References

- Arends, L. R., Hamza TH, Houwelingen JCv, Heijenbrok MH, Hunink MGM, Stijnen T. (2008) Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making*, **28**,621-638.
- Chappell FM, Raab GM and Wardlaw (2009) When are summary ROC curves appropriate for diagnostic meta-analyses? *Statistics in Medicine*; **28**: 2653-2668.
- Chu, H. and Cole, S. R. (2006) Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*. **59**, 1331-1332.
- Chu, H., Nie, L, Cole, S. R. and Poole C (2009) Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parametrization and model selection. *Statistics in Medicine*. Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/sim.3627.
- Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P. and Sterne, J. A. (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*.**8**,239-251
- Ohmann, C., Verde, P. E., Gilbers, T., Franke, C., Fuerst, G., Sauerland, S. and Boehner, H. (2006) Systematic Review of CT Investigation in Suspected Acute Appendicitis. Technical report *Coordination Centre for Clinical Trials, Heinrich-Heine University, Duesseldorf, Germany*

References

- Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Scholten, R. J, Bossuyt, P. M. and Zwinderman, A. H. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*. **58**, 982-90
- Riley RD, Dodd SR, Crain JV Thompson JR and Williamson (2008) Meta-analysis of diagnostic test studies using individual patient data and aggregated data. *Statistics in Medicine*. **27** 6111-61336.
- Rutter, C. M. and Gatsonis, C. A. (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*. **20**, 2865-84.
- Scheidler J., Hricak H, Yu KK, Subak L, and Segal MR (1997) Radiological evaluation of lymph node metastases in patients with cervical cancer: a meta-analysis. *JAMA* **278**:1096-1101.
- Verde, P. E. (2006). Generalized Evidence Synthesis for Diagnostic Test Data. *IceBUGS 2006*. Presentation available at: <http://www.math.helsinki.fi/openbugs/IceBUGS/IceBUGSAbstracts.html>
- Verde, P. E. (2007). Modern meta-analysis: a case study in combining results of diagnostic test data. Invited presentation at The Department of Statistics, Stanford University. Available at: <http://www-stat.stanford.edu/seminars/stat/>
- Verde, P. E. (2008). Meta-Analysis of Diagnostic Test Data: Modern Statistical Approaches. *Dissertation. Deutsche Nationalbibliothek*. urn:nbn:de:hbz:061-20080715-114548-5.
- Verde, P. E. (2009). Meta-Analysis of Diagnostic Test: a New Bayesian Approach. *Submitted to Statistics in Medicine*.
- Zwinderman AH and Bossuyt PM. (2009) We should not pool diagnostic likelihood ratios in systematic reviews. *Statistics in Medicine*. **27**: 687-697.