Meta-Analysis of Paired-Comparison Studies of Diagnostic Test Data: A Bayesian Modeling Approach

Pablo E. Verde
pabloemilio.verde@uni-duesseldorf.de

Coordination Center for Clinical Trials
University of Duesseldorf
Germany

Thursday 23 of May 2013
BAYES 2013
Rotterdam

Comparison of Medical Diagnostic Technologies

Paired-comparison diagnostic studies

- Two or more diagnostic tests are applied to the same group of patients
- Assess diagnostic performance between tests
- Compare pros and cons (e.g. invasive procedures versus noninvasive ones)

Issues in Meta-Analysis

- Correlated outcomes within and across studies
- Imperfect evidence, e.g. relevant data are not reported
- Common practice: use simple techniques and ignore problems

Running example

RAPT (Review of abdominal pain tools, Liu et al. 2006)

- Diagnosis of acute abdominal pain
  - Test 1: doctors using common medical practice (UD)
  - Test 2: doctors aided by decision tools (DT)
  - Decision tools are: classification statistical models (logistic regression, neural networks, naive Bayesian, etc.).
- N=9 studies reported paired-comparison between DT and UD

Results of Liu et al. 2006

- No difference in sensitivity between DT and UD
- The specificity of DT is better than the specificity of UD

Pieces of evidence of diagnostic test accuracy

Test results of the study i ($i = 1, \ldots, N$) are summarized in two 2×2 tables:

Results for Test 1

		Patient status: With disease	Patient status: Without disease
Test 1 outcome	+	$tp_{i,1}$	$fp_{i,1}$
	-	$fn_{i,1}$	$tn_{i,1}$
Sum:		$n_{i,1}$	$n_{i,2}$

Results for Test 2

		Patient status: With disease	Patient status: Without disease
Test 2 outcome	+	$tp_{i,2}$	$fp_{i,2}$
	-	$fn_{i,2}$	$tn_{i,2}$
Sum:		$n_{i,1}$	$n_{i,2}$

Review of abdominal pain tools (Liu et al. 2006)

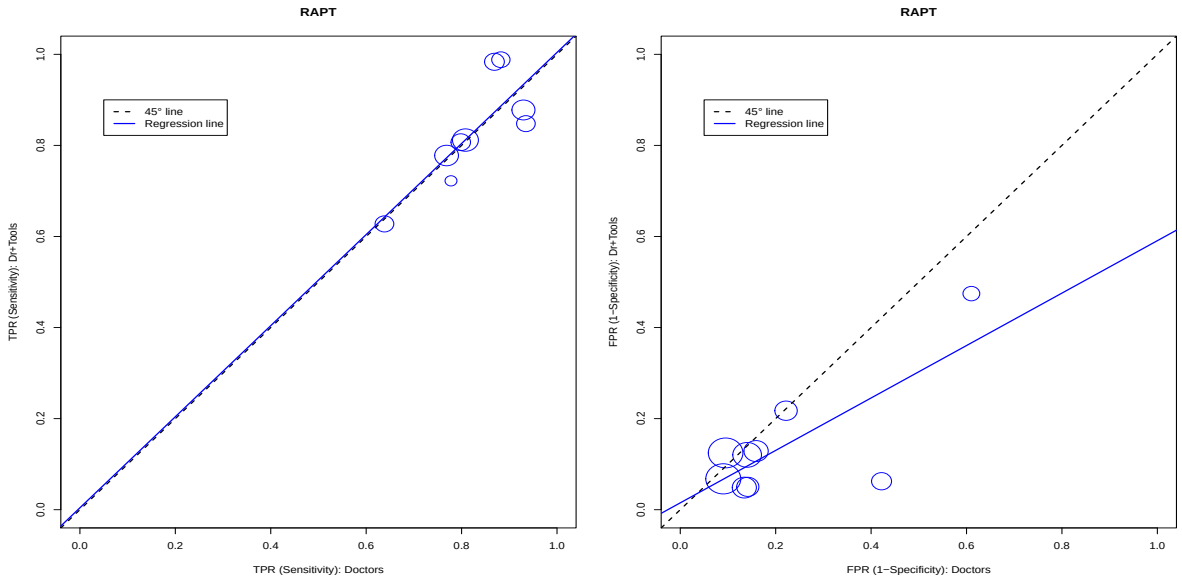


Figure : RAPT: Diagnostic of acute abdominal pain. Doctors aided by decision tools (DT) vs. unaided doctors (UD). Left panel: TPRs DT vs UD. Right panel: FPRs DT vs UD. (N=9)

Partially observed tables: indirect pieces of evidence

Patients status: with disease

		Test 2 outcome +	Test 2 outcome -	
Test 1 outcome	+	$y_{i,1}$	$tp_{i,1} - y_{i,1}$	$tp_{i,1}$
	-	$y_{i,2}$	$fn_{i,1} - y_{i,2}$	$fn_{i,1}$
Sum:		$tp_{i,2}$	$fn_{i,2}$	$n_{i,1}$

Patients status: without disease

		Test 2 outcome +	Test 2 outcome -	
Test 1 outcome	+	$y_{i,3}$	$fp_{i,1} - y_{i,3}$	$fp_{i,1}$
	-	$y_{i,4}$	$tn_{i,1} - y_{i,4}$	$tn_{i,1}$
Sum:		$fp_{i,2}$	$tn_{i,2}$	$n_{i,2}$

Table : The marginals are fixed and $y_{i,1}$ $y_{i,2}$ $y_{i,3}$ and $y_{i,4}$ are unobserved.

Accounting Lemma of Partial Observed Tables

Unobserved rates:

- $p_{i,1} = Pr(y_{i,1} = 1|\text{Test 1 tp})$ and $p_{i,2} = Pr(y_{i,2} = 1|\text{Test 1 fn})$
- $p_{i,3} = Pr(y_{i,3} = 1|\text{Test 1 fp})$ and $p_{i,4} = Pr(y_{i,4} = 1|\text{Test 1 tn})$

Lemma

The accounting relationships between the observed and unobserved diagnostic rates are given by:

$$\widehat{TPR}_{i,2} = p_{i,1}\widehat{TPR}_{i,1} + p_{i,2}(1 - \widehat{TPR}_{i,1}) \quad (1)$$

and

$$\widehat{FPR}_{i,2} = p_{i,3}\widehat{FPR}_{i,1} + p_{i,4}(1 - \widehat{FPR}_{i,1}) \quad (2)$$

Some remarks

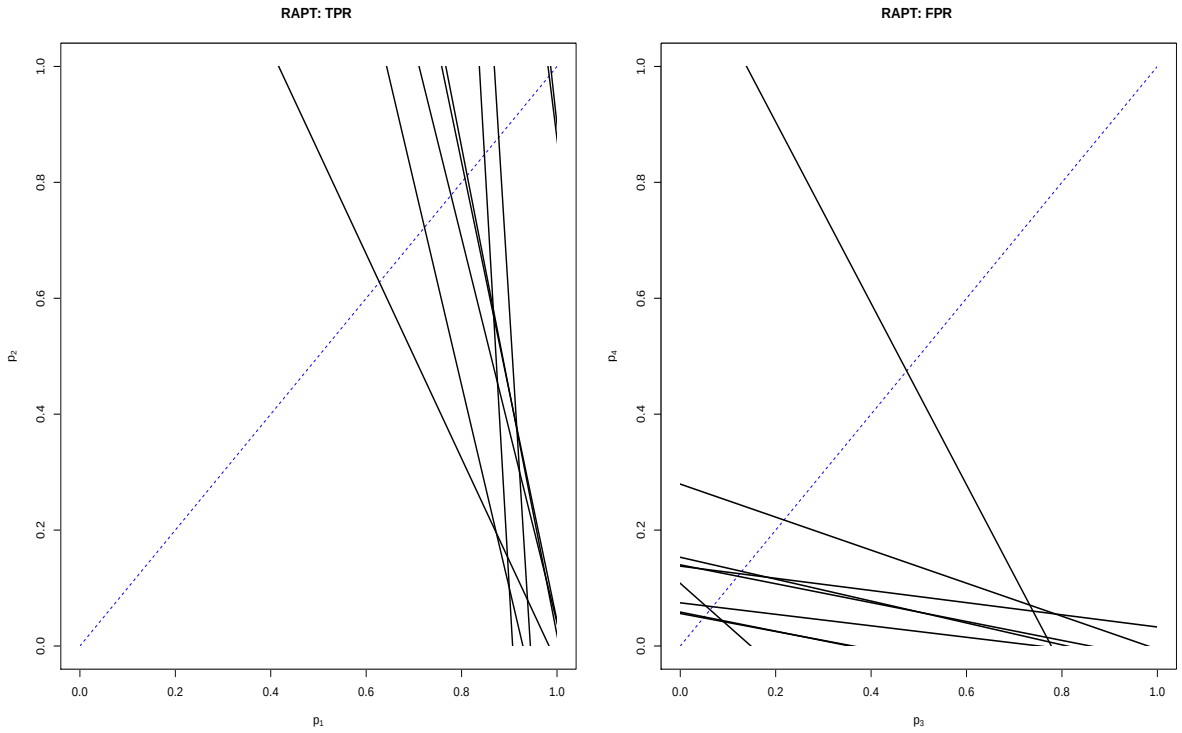
- Equations (1) and (2) are undetermined with four unknowns
- Unexpected solutions are possible (e.g $p_{i,1} = p_{i,2}$)
- They impose a deterministic data truncation constrains
- To display indirect evidence of the $p's$ we can plot:

$$p_{i,2} = \frac{\widehat{TPR}_{i,2}}{1 - \widehat{TPR}_{i,1}} - \frac{\widehat{TPR}_{i,1}}{1 - \widehat{TPR}_{i,1}} p_{i,1}, \quad (3)$$

and

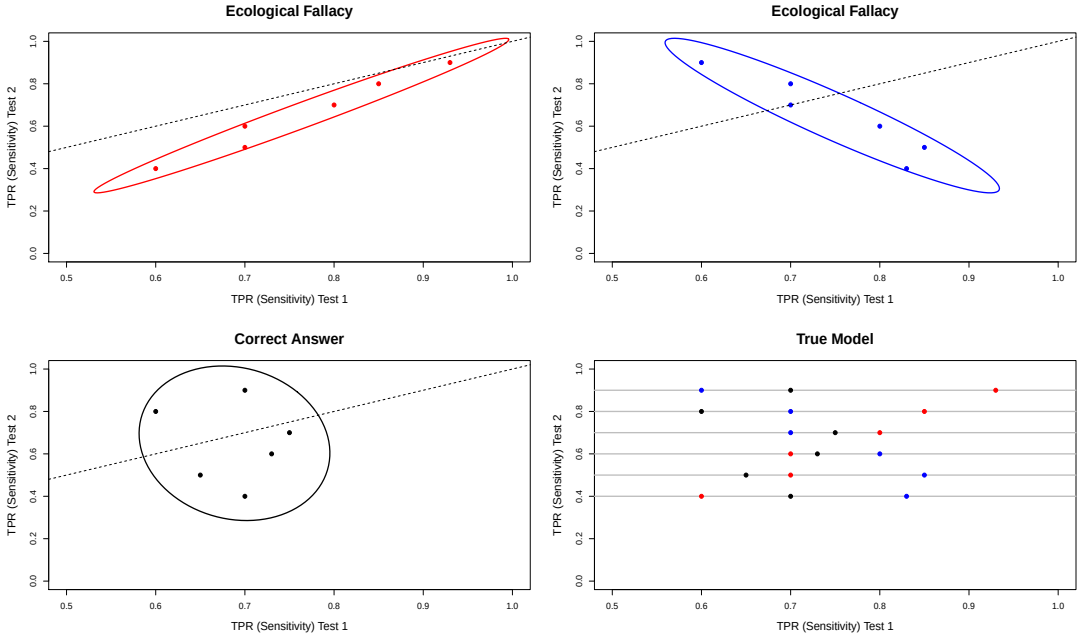
$$p_{i,4} = \frac{\widehat{FPR}_{i,2}}{1 - \widehat{FPR}_{i,1}} - \frac{\widehat{FPR}_{i,1}}{1 - \widehat{FPR}_{i,1}} p_{i,3}. \quad (4)$$

Displaying indirect evidence: RAPT



The ecological fallacy of two diagnostic tests

Ignoring these data structures may end in an ecological fallacy



Learning from evidence at face value

- Data on true positive results : $(tp_{i,1}, tp_{i,2}, n_{i,1})$
- The unobserved data are modeled as:

$$y_{i,1}|tp_{i,1} \sim Binomial(p_{i,1}, tp_{i,1}) \quad (5)$$
$$y_{i,2}|fn_{i,1} \sim Binomial(p_{i,2}, fn_{i,1}) \quad (6)$$

- Then $tp_{i,2} = y_{i,1} + y_{i,2}$ follows a convolution of these two binomial distributions with likelihood contribution:

$$L_{i,tp} = \sum_{k=\max(0,tp_{i,2}-fn_{i,1})}^{\min(tp_{i,1},tp_{i,2})} \binom{tp_{i,1}}{k} \binom{fn_{i,1}}{tp_{i,2}-k} p_{i,1}^{k}(1-p_{i,1})^{(tp_{i,1}-k)} \times p_{i,2}^{tp_{i,2}}(1-p_{i,2})^{tp_{i,1}-tp_{i,2}+k},$$

- The false positive tables $(fp_{i,1}, fp_{i,2}, n_{i,2})$ are modeled in similar way with likelihood contributions $L_{i,fp}$

Combining multiple sources of evidence

Study effects

We model the variability between studies with a scale mixture of normal distributions (Verde, 2010):

$$g(p_{i,j}) = \theta_{i,j} \sim N(\mu_j, w_i\lambda_j) \quad (7)$$
$$w_i \sim \Gamma(\nu/2, \nu/2), \quad (8)$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, 4$, where $g(\cdot)$ is a link function, λ_j are precision parameters and w_i mixture weights.

Between populations correlation

We model the correlation between disease and non-disease populations by

$$cor(\theta_{i,1}, \theta_{i,3}) = \rho. \quad (9)$$

Interpretation of the mixture weights

We use the posterior distribution of w_i to identify studies with unusual heterogeneity

- A-priory all studies included have a mean of $E(w_i) = 1$
- Studies which are unusual heterogeneous will have posteriors with values substantially less than 1, say $w_i < 0.7$
- Clearly if all $w_i \approx 1$ a multivariate Normal is an appropriate model
- If some w_i are lower than 1 then the effect of these studies will be down weighted resulting in a **robust inferential method**

Further modeling details

Hyper-parameters priors

We use independent and weakly informative priors for hyper-parameters:

$$\mu_j \sim N(0, .1), \quad \lambda_j \sim \Gamma(1, 0.1), \tag{10}$$

$$\nu \sim Exp(1), \quad logit((\rho + 1)/2) \sim N(0, 1). \tag{11}$$

Remarks in computations

- $L_{i,tp}$ and $L_{i,fp}$ are approximated by normal likelihoods (Wakefield 2004)
- Statistical computations are implemented in BUGS and R
- Most of the stochastic nodes in the model use conditional conjugate, so Gibbs sampling is straightforward

Posteriors of parameters of interest

Examples of parameters of interest

- Relationship between TPR_1 and TPR_2 we can use:

$$\Pr\left(p_2 + (p_1 - p_2)TPR_1 | \text{Data}, TPR_1 \in (0,1)\right) \tag{12}$$

- Study effects parameters:

$$\Pr\left(p_{i,j} | \text{Data}\right) \tag{13}$$

- Predictive posteriors and model checking parameters:

$$\Pr\left(y_1^{new}, y_2^{new} | \text{Data}\right) \quad \Pr\left(w_i < 1 | \text{Data}\right) \tag{14}$$

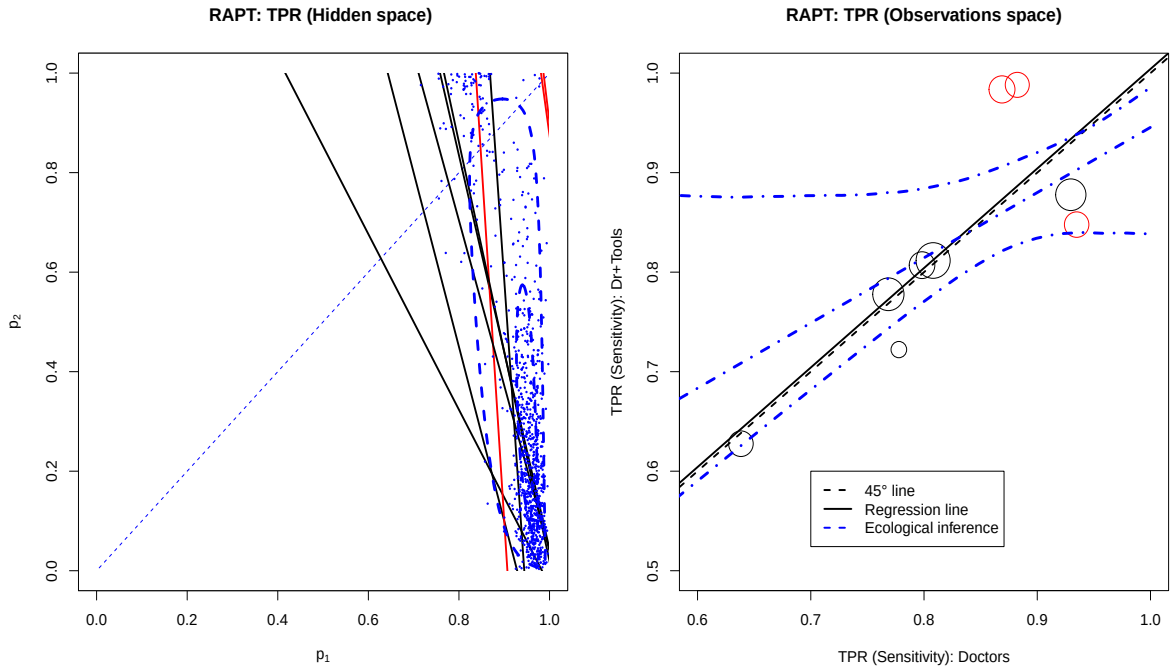


Figure : RAPT (TPR). Left panel: tomography lines and posterior surface. Right panel: Observed rates and regression lines ($\nu = 2.59$).

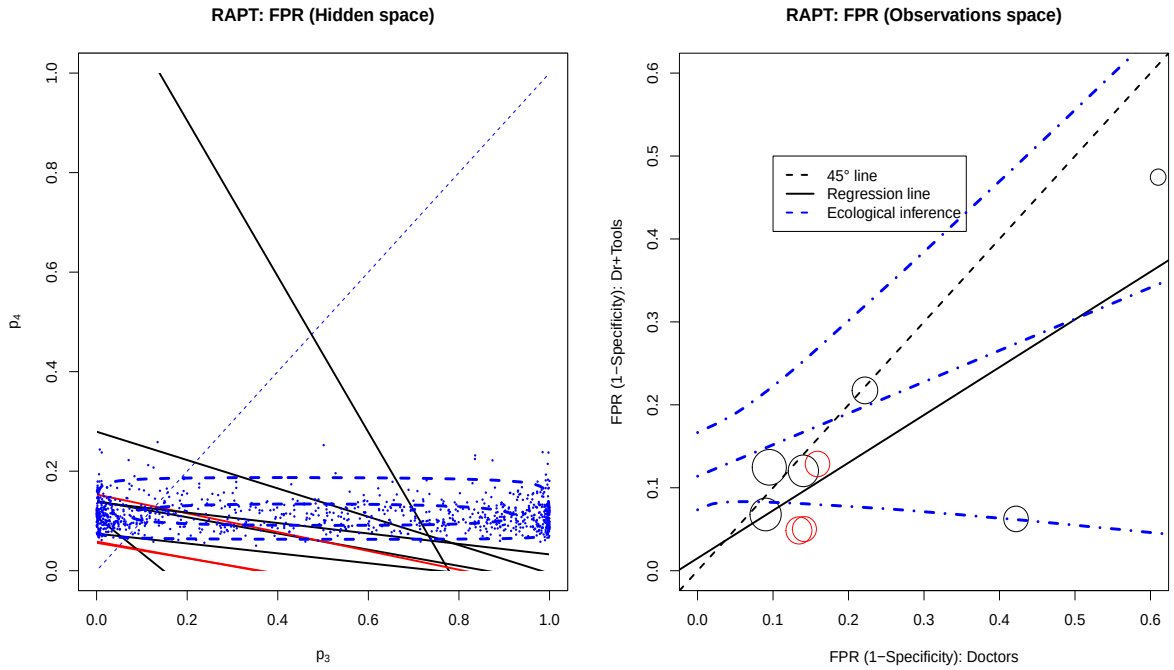


Figure : RAPT (FPR). Left panel: tomography lines and posterior surface. Right panel: Observed rates and regression lines ($\nu = 2.59$).

Summary

- **Direct modeling** observed rates could be misleading in paired-comparison studies of diagnostic test data
- **An indirect** approach seams to be more appropriate in this type of meta-analysis
- In practice systematic reviews **combine two types of evidence**: studies with paired-comparison design and studies with evidence on single diagnostic test. How to combine these two types of evidence remains open

- Birim O, Kappetein AP, Stijnen T, Bogers AJ., *Meta-analysis of positron emission tomographic and computed tomographic imaging in detecting mediastinal lymph node metastases in nonsmall cell lung cancer.* Ann Thorac Surg. 2005, Jan;79(1), 375-82.
- Liu JL, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, Ohmann C, Wellwood J, Dawes M, Altman DG., *Systematic reviews of clinical decision tools for acute abdominal pain*, Health Technol Assess., 2006, Nov;10(47):iii-iv.,1-167.
- Verde, P. E., *Meta-Analysis of Diagnostic Test Data: a Bivariate Bayesian Modeling Approach*, Statistics in Medicine, 2010, 29, 3088–3102.
- Wakefield, J., *Ecological inference for 2× 2 tables (with discussion)*, Journal of the Royal Statistical Society: Series A (Statistics in Society),2004, 167, 3, 385–445.