Meta-Analysis and Meta-Modelling for Diagnostic Problems

Dankmar Böhning

Southampton Statistical Sciences Research Institute
Mathematics and Medicine, Southampton, UK

Maha Sarakham, Thailand, May 2013

joint work with:

- **Heinz Holling**, Statistics and Quantitative Methods
  Faculty of Psychology and Sport Science
  University of Münster, Germany
- **Suphada Charoensawat**, Health Sciences Program Udon Thani Rajabhat University, Udon Thani, Thailand

principles of research

research should be

- **excellent**
- **novel**
- **relevant**
- **impact**

introduction and background of diagnostic setting

SROC-Diagram and a new measure

a mixed model approach

case studies

- MMSE and Dementia/MCI
- MOOD and depressive disorders
- MRS and prostate cancer

often studies are done in medicine or psychology to determine:

discriminatory ability of a diagnostic test to separate people

- **with** a specific disease (or condition)
- from those **without**

measures of diagnostic accuracy

- **Specificity**: $P(T-|D-) = q$
  Probability of a negative test result for a healthy person

- **Sensitivity**: $P(T+|D+) = p$
  Probability of a positive test result for a diseased person

estimating diagnostic accuracy

- **Specificity**: $P(\widehat{T-|D-}) = \hat{q} = \frac{x}{n}$
  where x is the number of true-negatives out of n healthy individuals, $n - x$ are the false-positives

- **Sensitivity**: $P(\widehat{T+|D+}) = \hat{p} = \frac{y}{m}$
  where y are the number of true-positives out of m healthy individuals, $y - m$ are the false-negatives

frequently available:

- a variety of diagnostic studies
- providing diagnostic measures

$$x_i, n_i \text{ (specificity)}$$

$$y_i, m_i \text{ (sensitivity)}$$

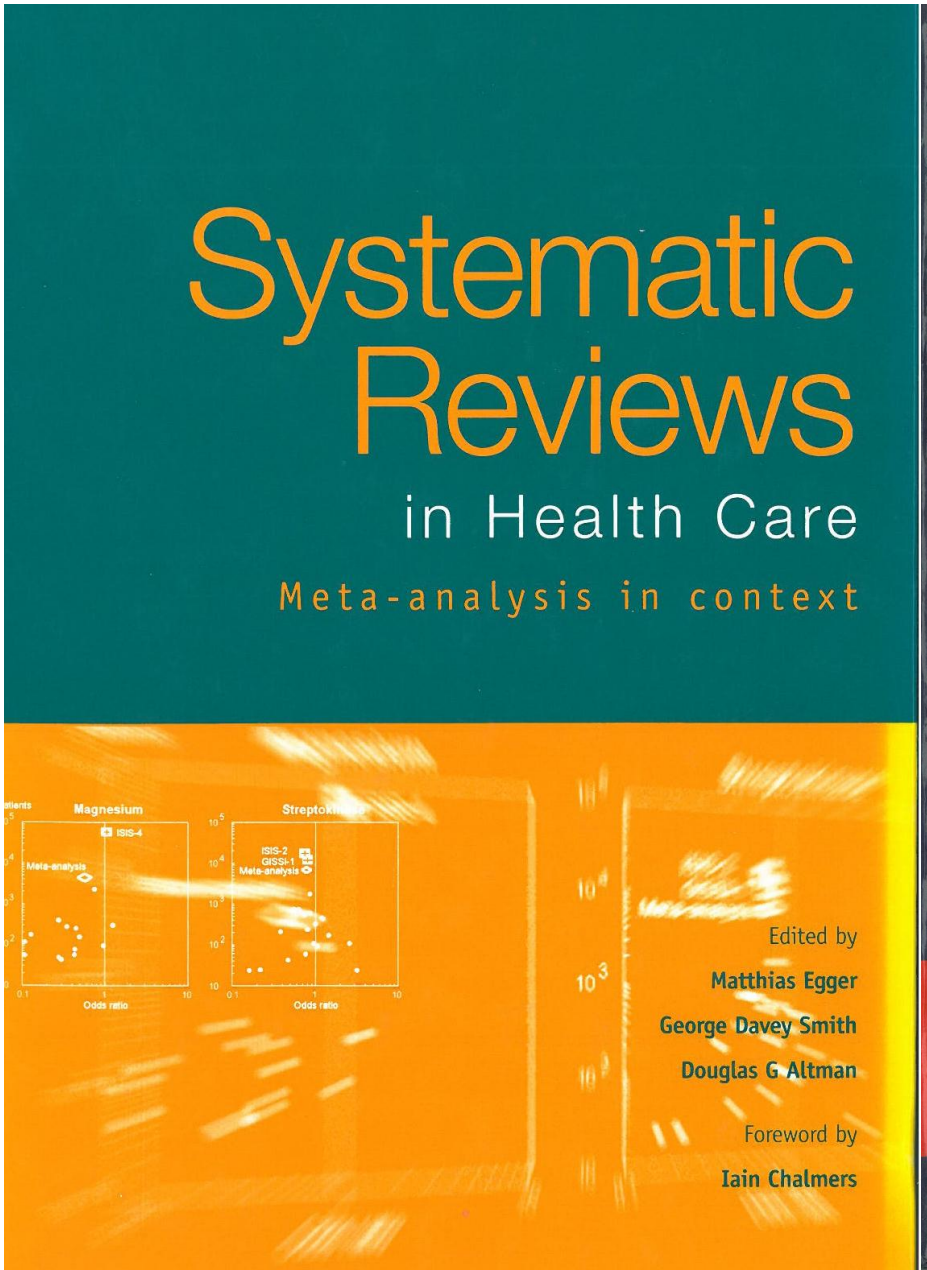
- for $i = 1, ..., k$
- leading to the field of **meta-analysis**

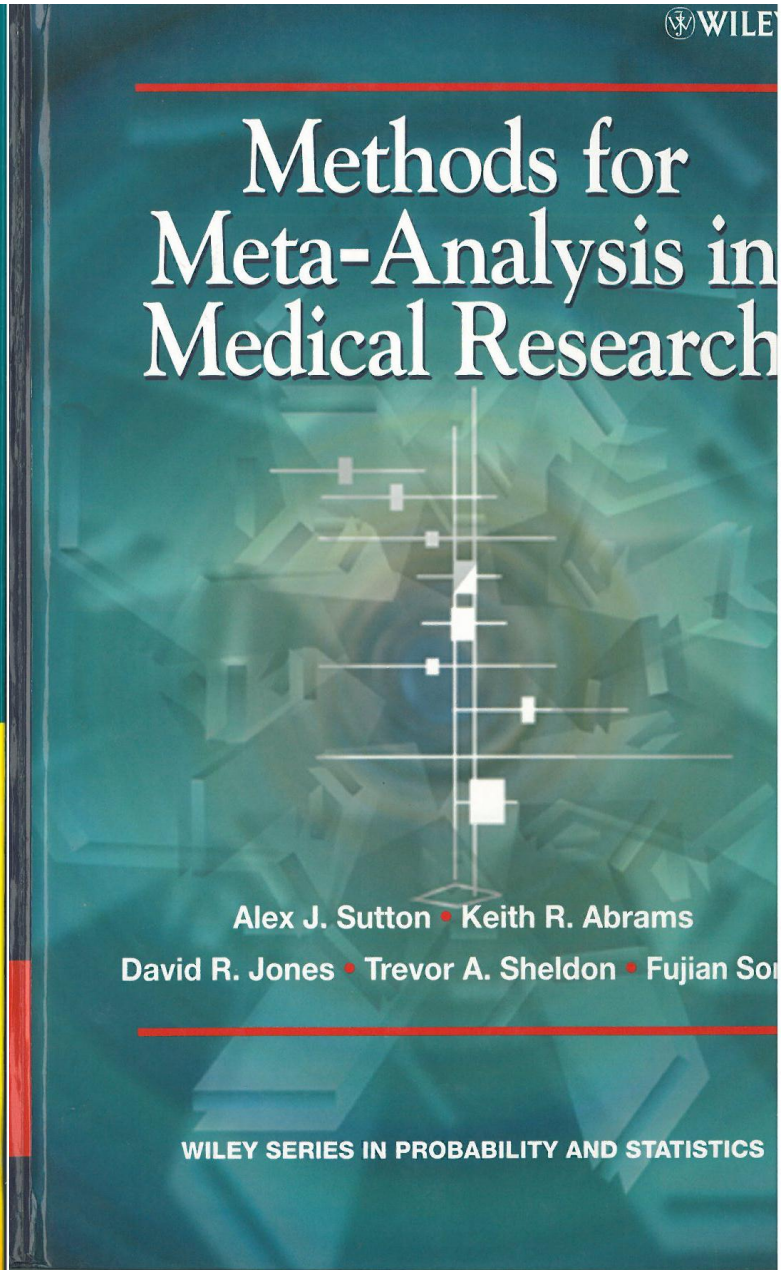
an example: meta-analysis of diagnostic accuracy of natriuretic peptides for heart failure

- diagnosis of heart failure is difficult
- overdiagnosis and underdiagnosis is occurring
- natriuretic peptides have been proposed as a diagnostic test
- meta-analysis provided by Doust *et al.* (2004) for brain natriuretic peptide (BNP)
- restriction on studies that use left ventricular ejection fraction of 40% or less as gold standard

Data on meta-analysis of diagnostic accuracy of natriuretic peptides for heart failure

	diseased		healthy		
study	y(TP)	$m - y$(FN)	x(TN)	$n - x$(FP)	$n + m$
Bettenc. 2000	29	7	46	19	101
Choy 1994	34	6	22	13	75
Valli 2001	49	9	78	17	153
Vasan 2002a	4	6	1612	85	1707
Vasan 2002b	20	40	1339	71	1470
Hutcheon 2002	29	2	102	166	299
Landray 2000	26	14	75	11	126
Smith 2000	11	1	93	50	155


Systematic Reviews
in Health Care
Meta-analysis in context
Magnesium
ISIS-4
Meta-analysis
Odds ratio
ISIS-2
GISSI-1
Meta-analysis
Odds ratio
Edited by
Matthias Egger
George Davey Smith
Douglas G Altman
Foreword by
Iain Chalmers


WILEY
Methods for
Meta-Analysis in
Medical Research
Alex J. Sutton • Keith R. Abrams
David R. Jones • Trevor A. Sheldon • Fujian So
WILEY SERIES IN PROBABILITY AND STATISTICS

a possible strategy:

compute a summary measure for each study:

- **Youden index**

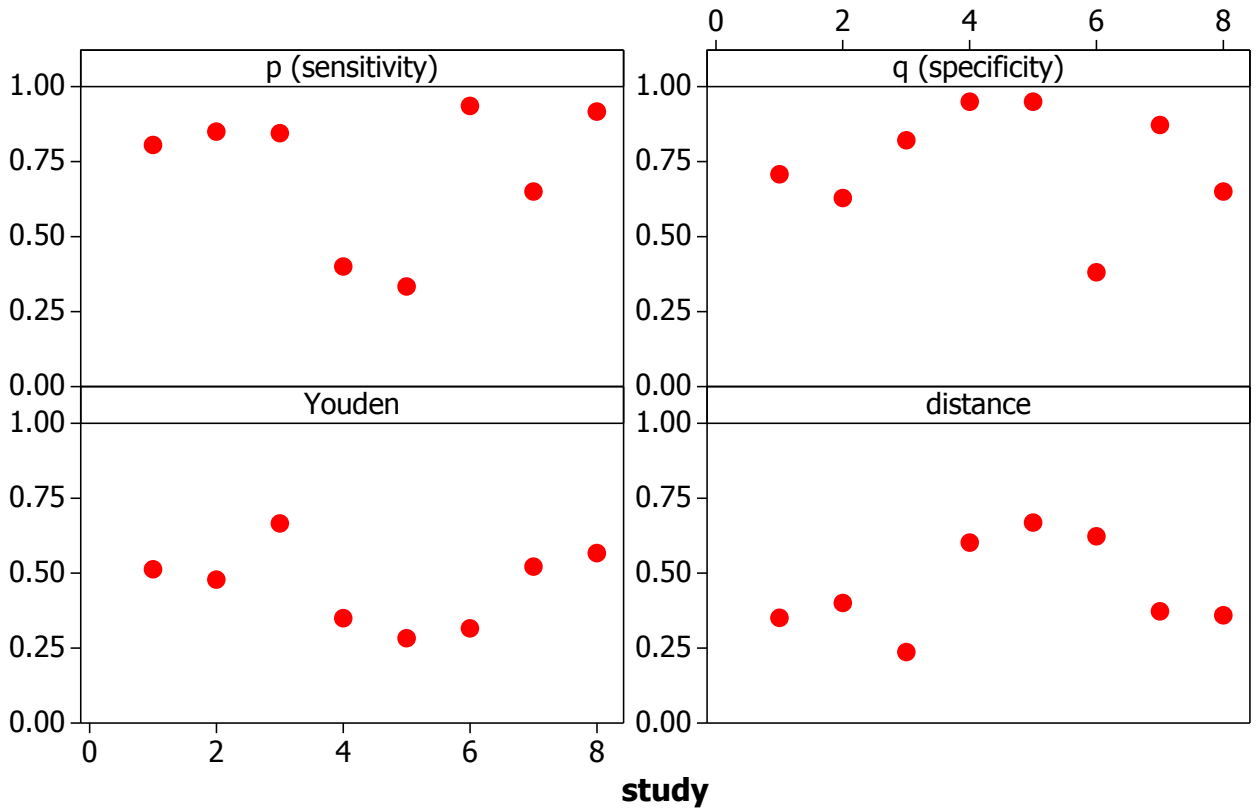
$$J_i = p_i + q_i - 1$$

- **Euclidean distance** (to the point of perfect separation)

$$E_i = \sqrt{(1 - p_i)^2 + (1 - q_i)^2}$$

- ... many others (Xinhua Liu 2012 *SiM*)

considerable variation across studies:

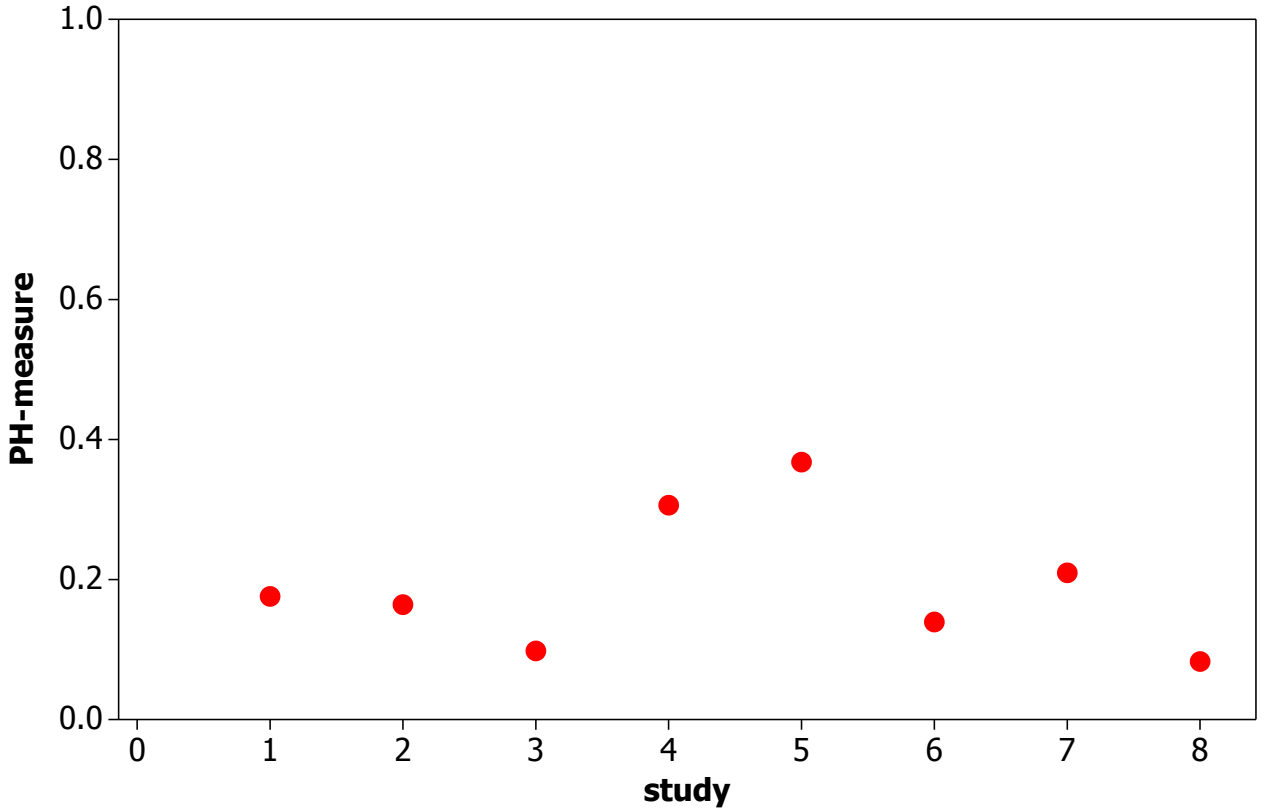


a new measure

proportional hazards (PH) measure

$$\theta = \frac{\log p}{\log(1-q)}$$

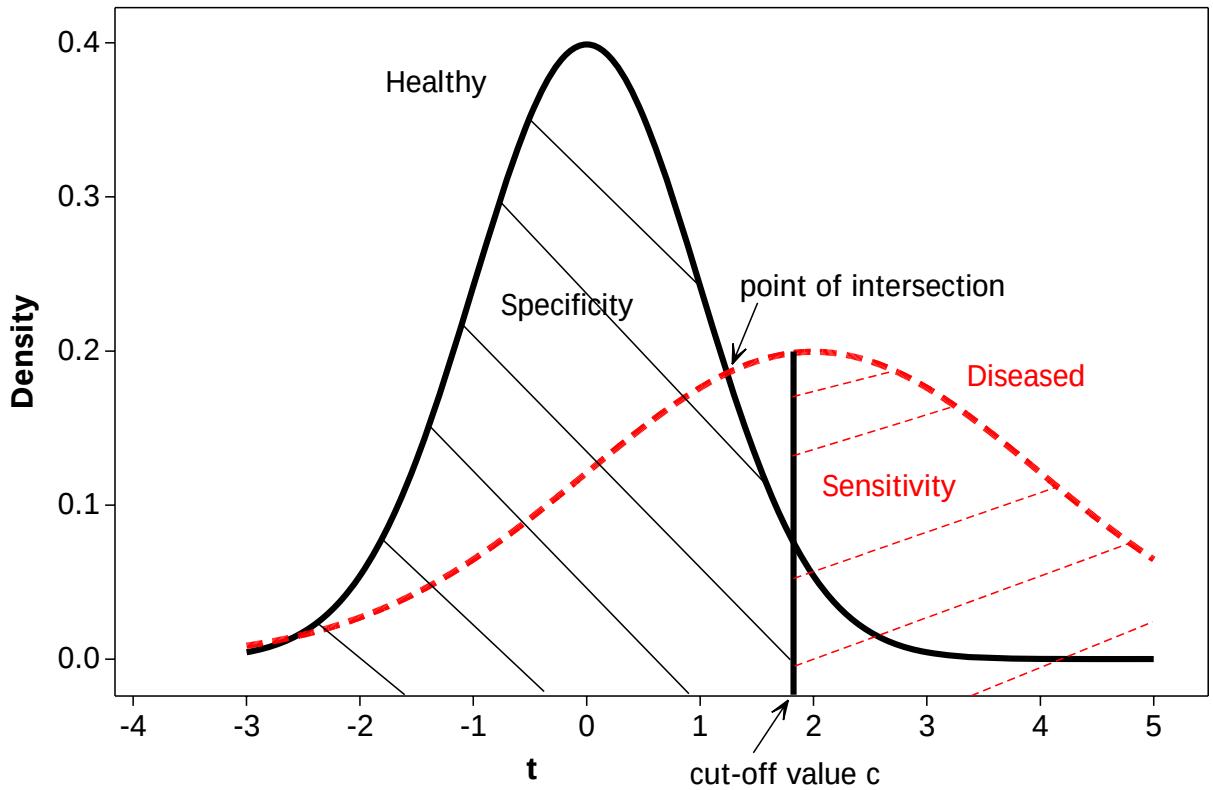
relating log-sensitivity to log-false positive rate


1.0
0.8
0.6
0.4
0.2
0.0
PH-measure
0
1
2
3
4
5
6
7
8
study

the cut-off value problem

- Why not proceed with the **available armada of meta-analysis methods**?
- continuous or ordered categorical test uses **cut-off value**

Illustration of the cut-off value problem:



the cut-off value problem

- sensitivities and specificities from different studies **not comparable**
- different values for sensitivity and specificity might be due to different **diagnostic accuracy** or **different cut-off value**
- cut-off problem introduces **bias of unknown direction and size**

The SROC-diagram for meta-analytic situations

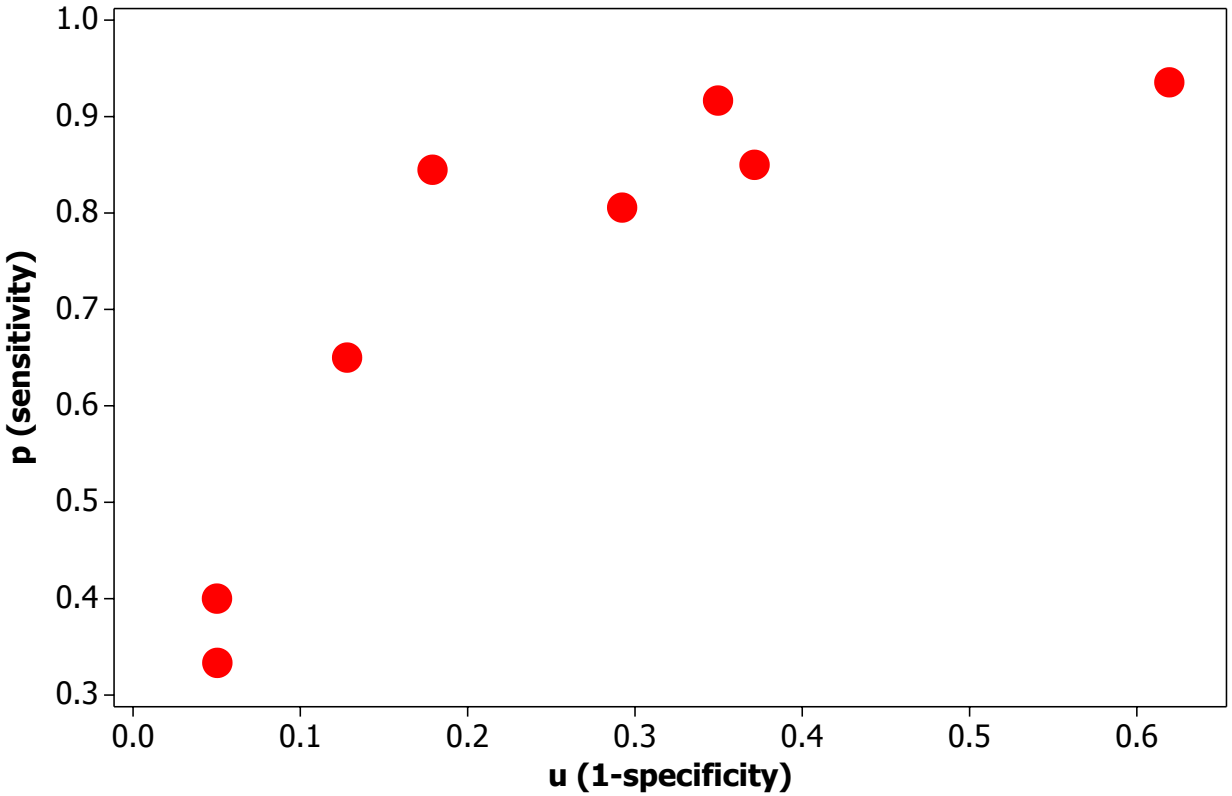
- Consider the pairs (sensitivity, 1-specificity) estimated by

$$(\hat{p}_i, 1 - \hat{q}_i) = (y_i/m_i, 1 - x_i/n_i)$$

  for $i = 1, ..., k$

- include them in a **Receiver Operating Characteristic (ROC)** diagram
- called **summary** ROC because the points relate to **different studies** with potentially **different cut-off values**

SROC-diagram for MA of BNP and heart failure

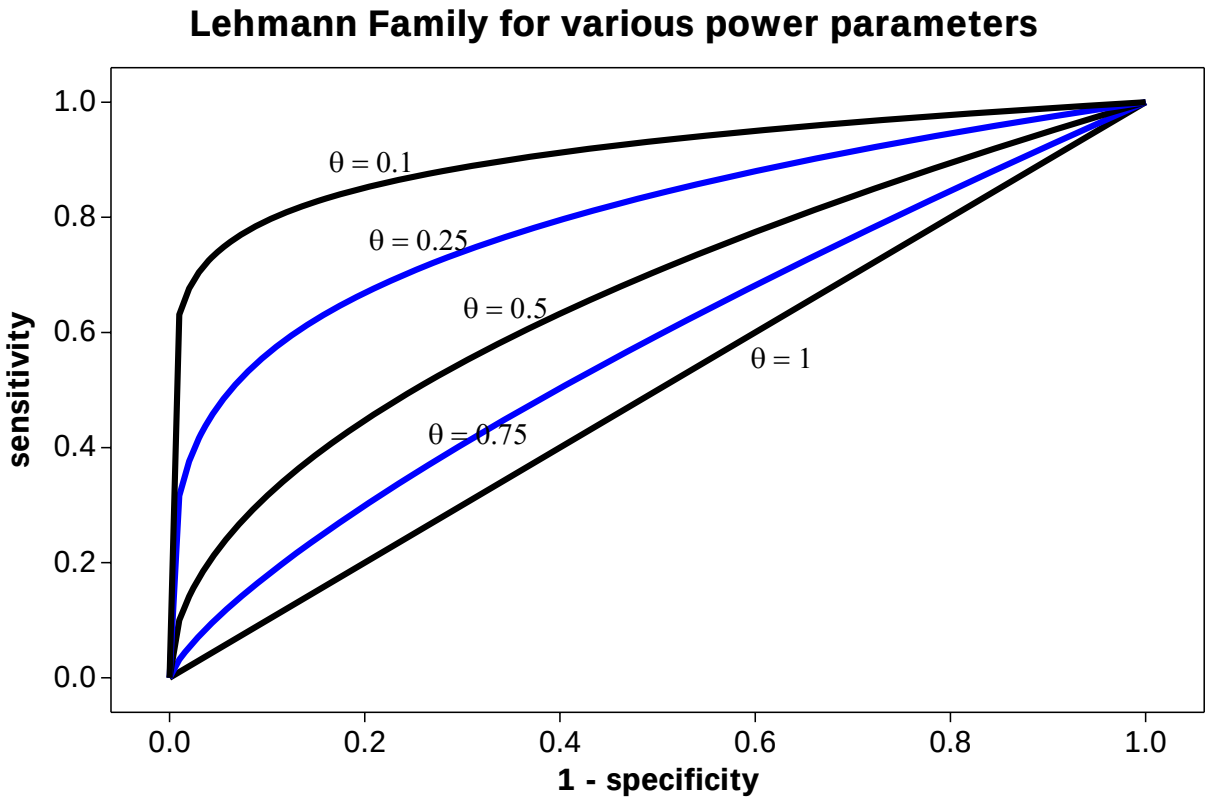


Modelling of the SROC-diagram

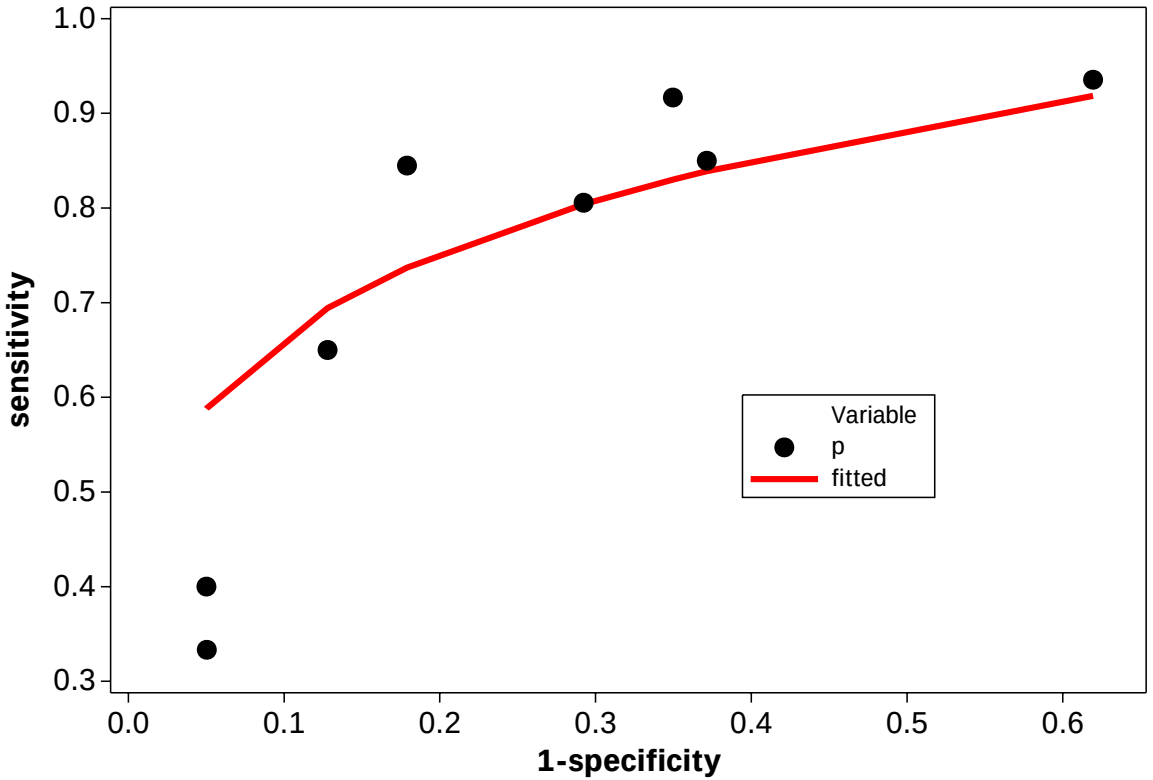
- Consider the **Lehmann family** for $\theta > 0$ fixed (Le 2006):

$$p = (1 - q)^{\theta}$$

- note that θ represents the **diagnostic power**


Lehmann Family for various power parameters
sensitivity
1 - specificity
0.0
0.2
0.4
0.6
0.8
1.0
θ = 0.1
θ = 0.25
θ = 0.5
θ = 0.75
θ = 1

instead of constructing average SROC model ...



Modelling of the SROC-diagram

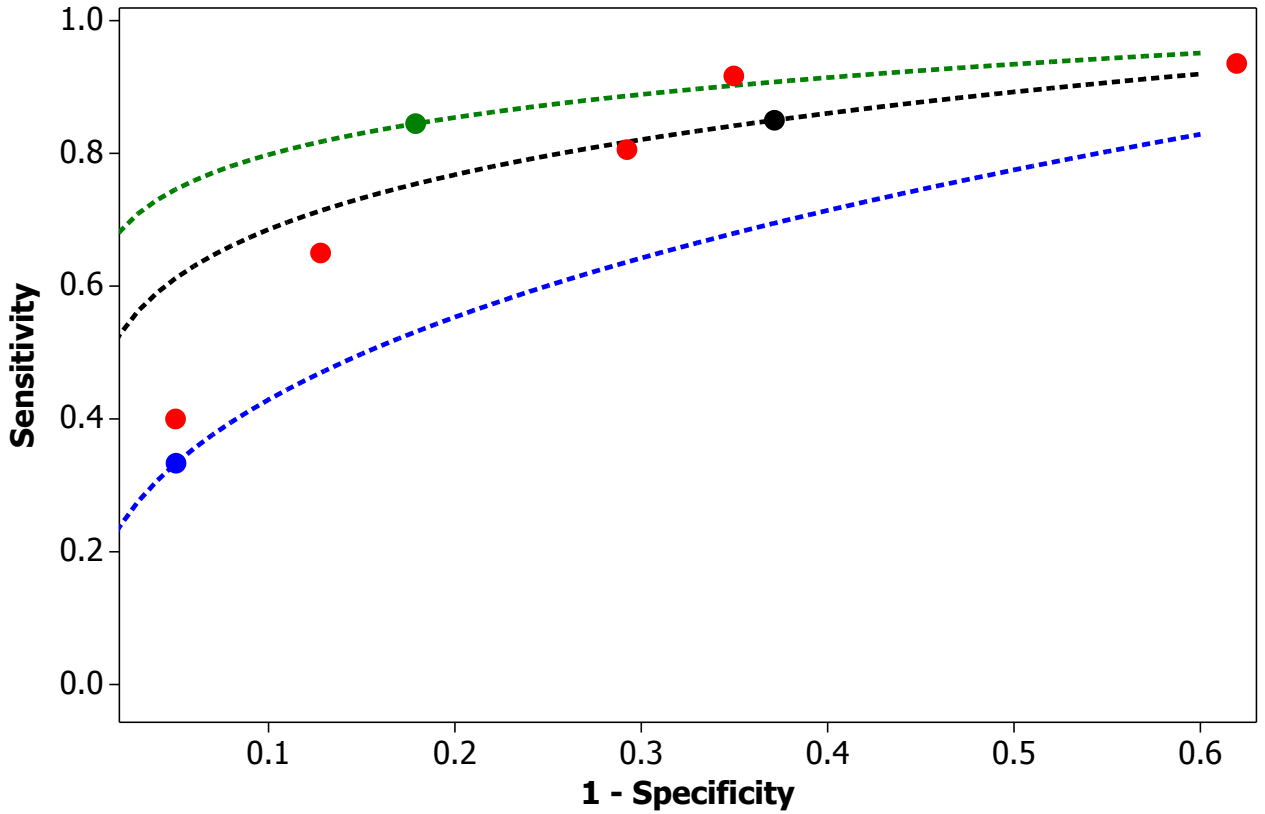
- consider the **study-specific Lehmann curves** :

$$p = (1 - q)^{\hat{\theta}_i} \tag{1}$$

  where $\hat{\theta}_i = \frac{\log \hat{p}_i}{\log(1-\hat{q}_i)}$

- so that (1) goes **exactly through the point**

$$(\hat{p}_i, 1 - \hat{q}_i)$$

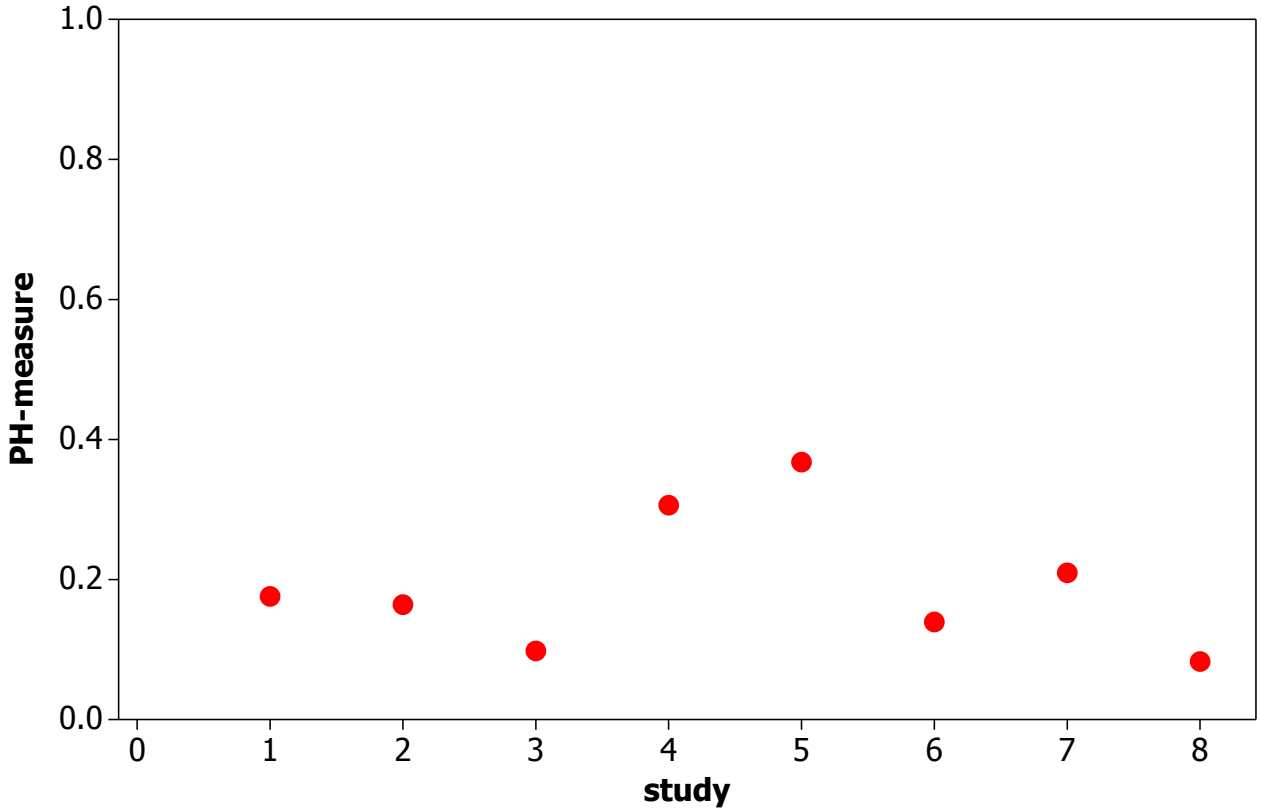

1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
0.1
0.2
0.3
0.4
0.5
0.6
1 - Specificity

explaining the reduced variation of the PH measure

proportional hazards (PH) measure for study i

$$\theta_i = \frac{\log \hat{p}_i}{\log(1 - \hat{q}_i)}$$

relating log-sensitivity to log-false positive rate

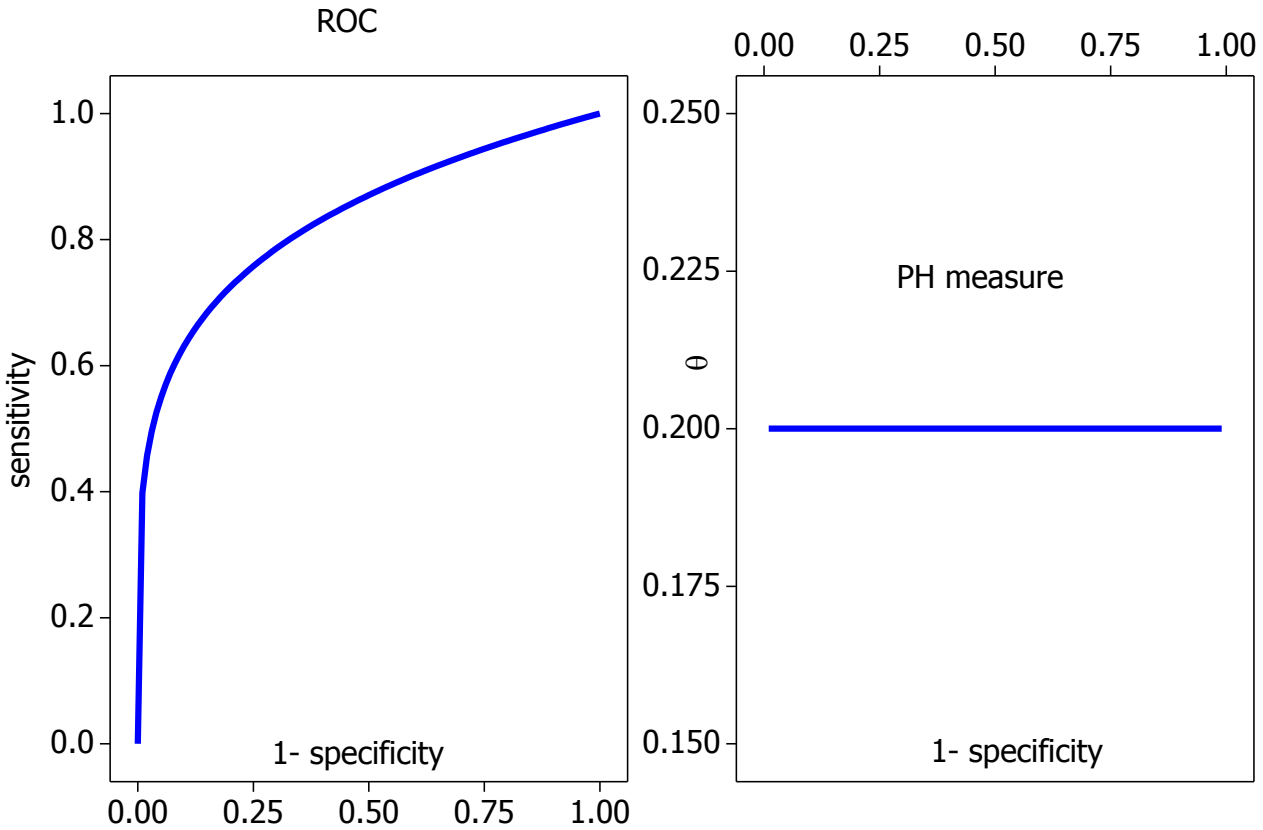

1.0
0.8
0.6
0.4
0.2
0.0
PH-measure
0
1
2
3
4
5
6
7
8
study

explaining the reduced variation of the PH measure

- Youden index, euclidean distance, and others measure

  diagnostic accuracy + something else

- whereas the PH measure focuses more **diagnostic accuracy**


ROC
sensitivity
1- specificity
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.25
0.50
0.75
1.00
PH measure
θ
0.150
0.175
0.200
0.225
0.250
0.00
0.25
0.50
0.75
1.00
1- specificity

a mixed model approach

k studies available with diagnostic accuracies $\hat{\theta}_1, \cdots, \hat{\theta}_k$ where

$$\hat{\theta}_i = \frac{\log \hat{p}_i}{\log(1 - \hat{q}_i)}$$

- **linear mixed model**

$$\log \hat{\theta}_i = \beta^{\mathbf{T}} \mathbf{x}_i + \delta_i + \epsilon_i$$

a mixed model approach

k studies available with diagnostic accuracies $\hat{\theta}_1, \cdots, \hat{\theta}_k$ where

$$\hat{\theta}_i = \frac{\log \hat{p}_i}{\log(1 - \hat{q}_i)}$$

- **linear mixed model**

$$\log \hat{\theta}_i = \beta^{\mathbf{T}} \mathbf{x}_i + \delta_i + \epsilon_i$$

- $\mathbf{x}_i$ is a known covariate vector in study i

a mixed model approach

k studies available with diagnostic accuracies $\hat{\theta}_1, \cdots, \hat{\theta}_k$ where

$$\hat{\theta}_i = \frac{\log \hat{p}_i}{\log(1 - \hat{q}_i)}$$

- **linear mixed model**

$$\log \hat{\theta}_i = \beta^{\mathbf{T}} \mathbf{x}_i + \delta_i + \epsilon_i$$

- $\mathbf{x}_i$ is a known covariate vector in study i
- $\delta_i \sim N(0, \tau^2)$, τ^2 **unknown** and $\epsilon_i \sim N(0, \sigma_i^2)$ with **known** variance σ_i^2

the within study variance of

$$\log \hat{\theta} = \log(-\log \hat{p}) - \log(-\log(1-\hat{q})$$

using the $\delta-$method $(VarT(X) \approx T'(EX)^2 Var(X))$

- $$Var \log(-\log \hat{p}) \approx \frac{\hat{p}(1-\hat{p})/m}{\hat{p}^2(\log \hat{p})^2}$$
- $$Var \log(-\log 1 - \hat{q}) \approx \frac{\hat{q}(1-\hat{q})/n}{(1-\hat{q})^2(\log \hat{p})^2}$$

leads to the within study variance

$$\sigma_i^2 = \frac{m_i - y_i}{m_i y_i (\log \hat{p}_i)^2} + \frac{x_i}{n_i(n_i - x_i)(\log 1 - \hat{q}_i)^2}$$

a mixed model approach

- has great flexibility and embeds conventional approaches

a mixed model approach

- has great flexibility and embeds conventional approaches
- "fixed" effect model

$$\log \hat{\theta}_i = \beta_0 + \epsilon_i$$

a mixed model approach

- has great flexibility and embeds conventional approaches
- "fixed" effect model

$$\log \hat{\theta}_i = \beta_0 + \epsilon_i$$

- "random" effect model

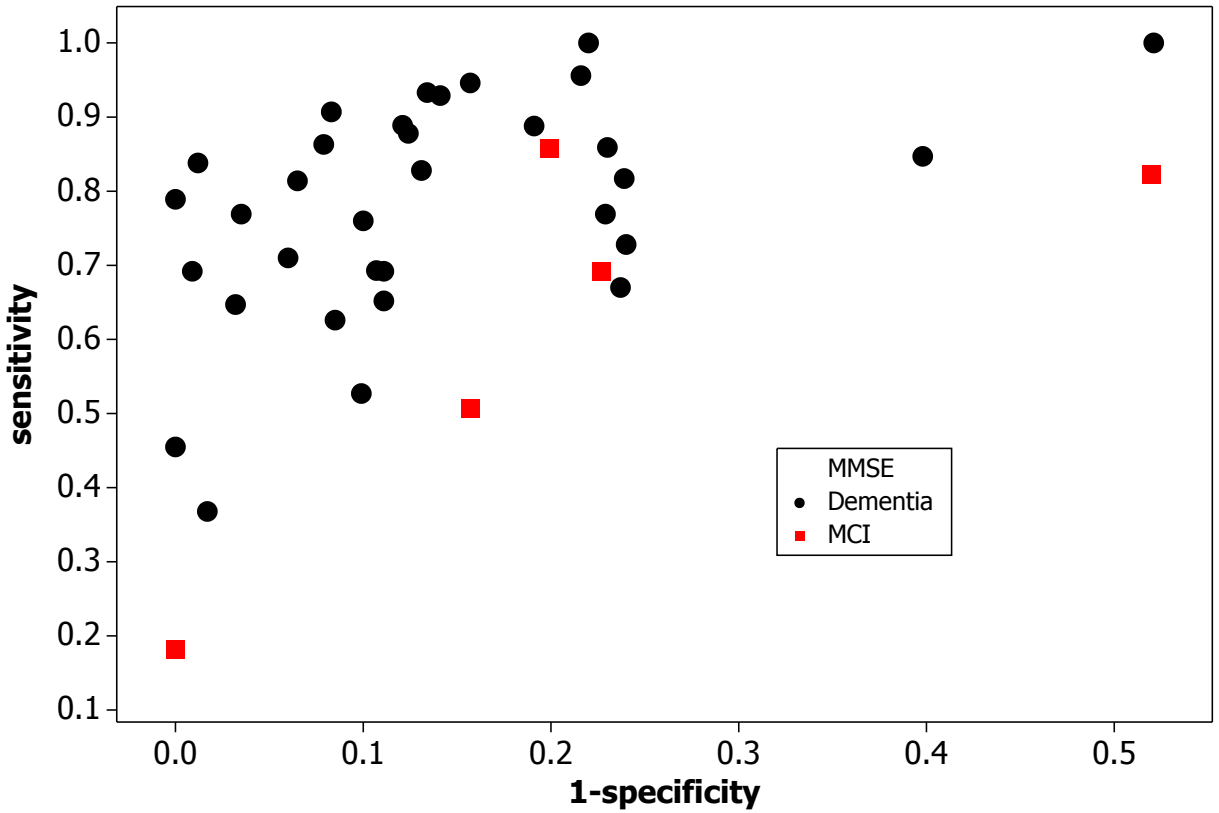
$$\log \hat{\theta}_i = \beta_0 + \delta_i + \epsilon_i$$

MA of MMSE and Dementia/MCI

- **mini-mental state examination (MMSE)** as a diagnostic test for the detection of **dementia** and, more recently, **mild cognitive impairment (MCI)**
- MA by Mitchell (2009, *J Psychiatr Res*) included **38 studies**

MA of MMSE and Dementia/MCI

study	author(s)	condition	TP	FN	FP	TN
1	Belle *et al.*, 2000	Dem	65	3	240	870
2	Borson *et al.*, 2000	Dem	117	12	10	110
3	Brayne *et al.*, 1989	Dem	24	5	44	292
4	Brodaty *et al.*, 2002	Dem	67	15	48	153
...	...	...	...	...	...	
36	Borson *et al.*, 2005	MCI	37	36	22	118
37	Kalbe *et al.*, 2004	MCI	67	30	22	75
38	Nasred. *et al.*, 2005	MCI	17	77	0	90


sensitivity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.1
0.2
0.3
0.4
0.5
1-specificity
MMSE
Dementia
MCI

proc mixed in SAS

```
proc mixed data=MMSE method=ml covtest;
class study condition;
model logtheta = condition /s;
* weight is inverse variance
weight w;
random study(condition);
* do NOT estimate residual variance component
parms (1) (1) /hold=2;
run;
```

solution for fixed effects

effect	parameter	SE	Z-value
Intercept	-2.2878	0.1208	-18.94
condition	0.8605	0.3187	2.70

associated SROC curves

dementia:

$$p = (1 - q)^{\exp(-2.2878)}$$

MCI:

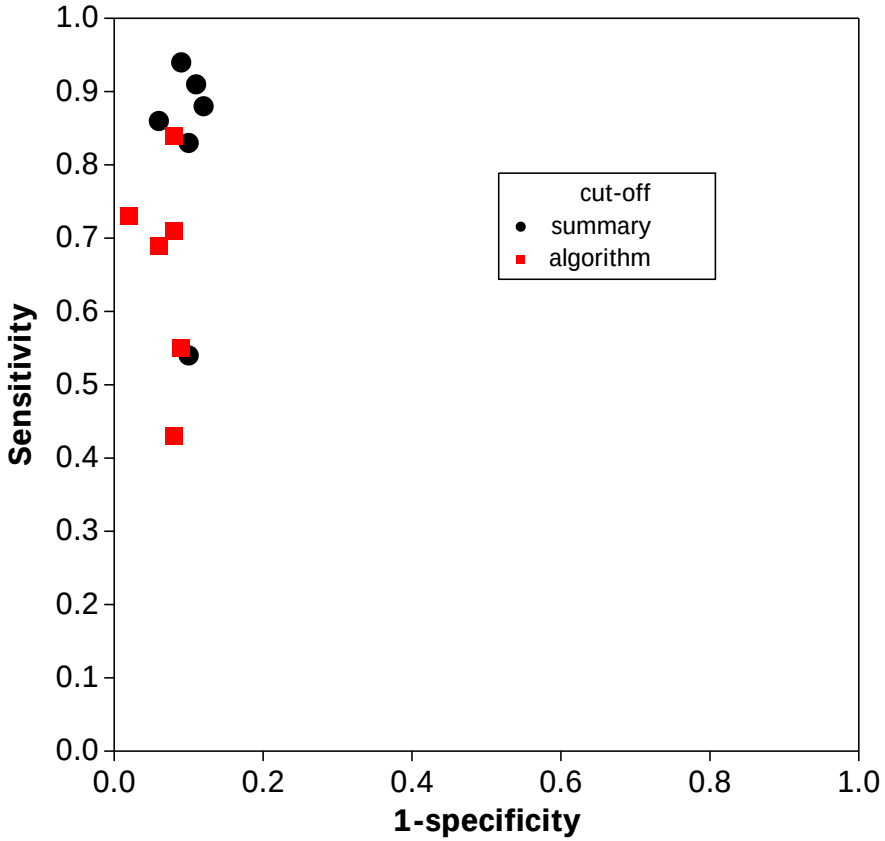
$$p = (1 - q)^{\exp(-2.2878+0.8605)}$$

MA of MOOD and depressive disorder

- nine-item **MOOD** module of the Patient Health Questionnaire (PHQ-9) developed to screen and to diagnose patients in primary care with **depressive disorders**
- MA by Wittkampf *at al.* (2007, *General Hospital Psychiatry*) included **12 studies**
- the instrument consists of 9 questions each could receive 0-3 points
- hence the total score ranges from 0 to 27
- the studies used either a cut-off of 10 (summary score) or a more complex evaluation algorithm (algorithm)

MOOD and depressive disorder

study	1st author	cut-off	TP	FN	FP	TN
1	Corapcioglu 2004	algorithm	65	26	104	1192
2	Diez-Quevedo 2001	algorithm	70	13	74	846
3	Grafe 2004	sum score	62	10	27	429
4	Kroenke 2001	sum score	36	5	65	474
5	Lowe 2004	sum score	55	11	43	392
6	Mazzotti2003	algorithm	6	8	12	144
7	McManus2005	sum score	121	103	80	720
8	Persoons 2003	algorithm	11	5	5	76
9	Picardi 2005	algorithm	6	5	0	3
10	Spitzer 1999	algorithm	85	31	9	460
11	Watnick 2005	sum score	15	1	4	42
12	Williams 2005	sum score	96	10	23	187


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1-specificity
cut-off
summary
algorithm

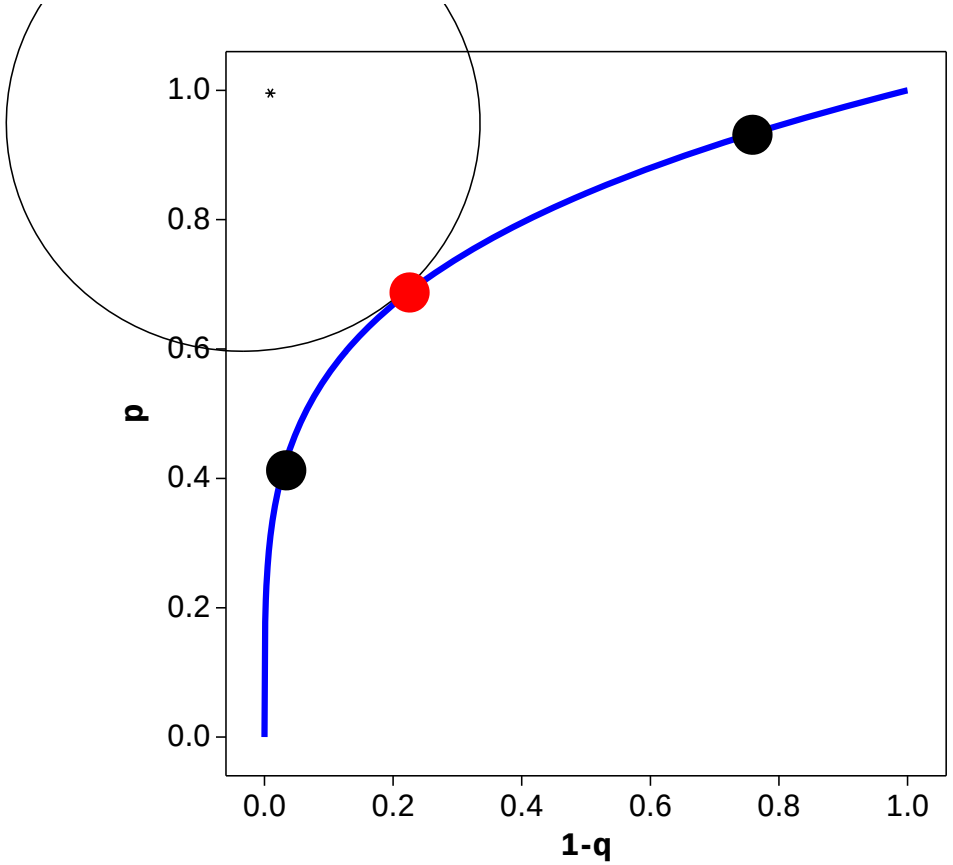
solution for fixed effects

effect	parameter	SE	Z-value
Intercept	-2.5332	0.2817	-8.99
cut-off	0.4804	0.3966	1.21

different criterion?

better use Euclidean distance:

$$E = (1-p)^2 + (1-q)^2$$


1.0
0.8
0.6
0.4
0.2
0.0
p
0.0
0.2
0.4
0.6
0.8
1.0
1-q

variance computation:

$$\hat{E} = (1 - \hat{p})^2 + (1 - \hat{q})^2$$

where $\hat{p} = y/m$ and $\hat{q} = x/n$

$$Var\hat{E} \approx 4(1 - \hat{p})^2 \hat{p}(1 - \hat{p})/m + 4(1 - \hat{q})^2 \hat{q}(1 - \hat{q})/n$$

used $\delta-$method:

$$VarT(X) \approx T'(EX)^2 Var(X)$$

solution for fixed effects using Euclidean distance

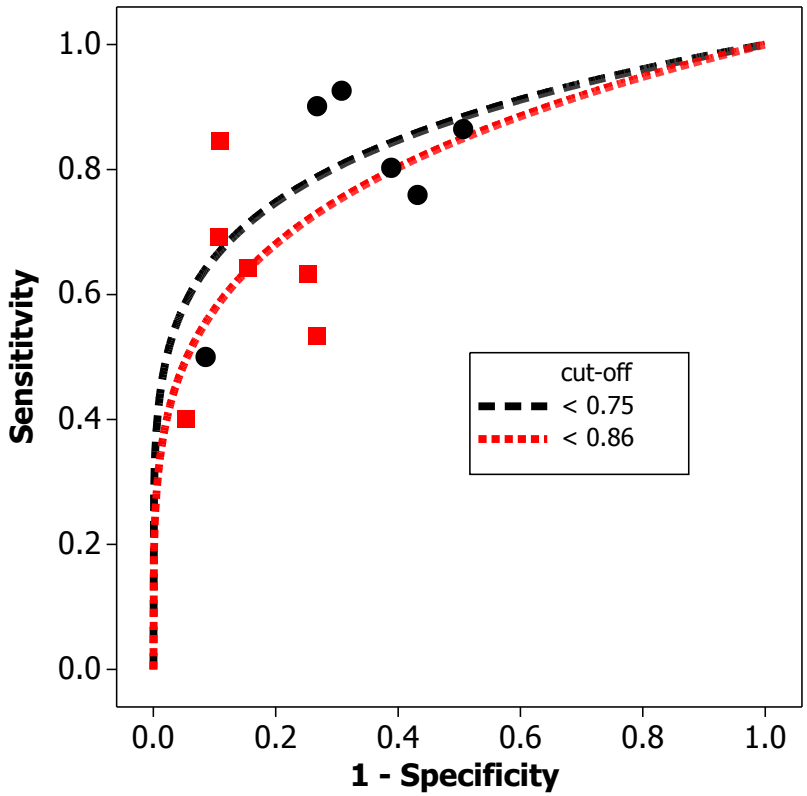
criterion	effect	parameter	SE	Z-value
PH model	cut-off	0.4804	0.3966	1.21
Euclidean distance	cut-off	0.05629	0.04297	1.31

MA of MRS and prostate cancer

- **magnetic resonance spectroscopy** has ability to discriminate prostate cancer from benign prostatic hyperplasia based on reduced citrate and elevated choline in the cancerous region
- test works on a voxel of signal intensity ratios of (choline+creatine)/citrate
- two cut-off points are in use: < 0.75 and < 0.86
- MA by Wang *et al.* (2008) including 12 studies

MA of MRS and prostate cancer

study	1st author	cut-off	TP	FN	FP	TN
1	Ullrich	0.75	122	30	35	55
2	Juyoung I	0.75	73	8	80	219
3	Juyoung II	0.75	75	6	92	207
4	Wiefer	0.75	123	39	38	50
5	Juergen	0.75	134	21	40	39
6	Kyle	0.75	12	12	7	75
7	Ullrich	0.86	81	71	24	59
8	Juyoung I	0.86	56	25	32	267
9	Juyoung II	0.86	52	29	20	59
10	Scheidler	0.86	98	57	20	59
11	Yuen	0.86	6	9	15	266
12	Prando	0.86	44	8	32	264


1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity
cut-off
< 0.75
< 0.86

solution for fixed effects using PHM and Euclidean distance

criterion	effect	parameter	SE	Z-value
PH model	cut-off	**0.2049**	0.3516	0.58
Eucliden distance	cut-off	**-0.02119**	0.05730	-0.37

some conclusions

- benefit of the approach: a **bivariate** problem is reduced to a **univariate** one
- this is **not unique**: $\log p = \theta \log(1-q)$ could be replaced by $\log p = \theta + \log(1-q)$ or

$$p = \exp\theta(1-q)$$

- or

$$\exp\theta = \frac{p}{1-q} = \textbf{likelihood ratio positive}$$

recent work

- Holling, H., Böhning, W., and Böhning, D. (2012). Likelihood based clustering of meta-analytic SROC curves. *Psychometrika* **77**, 106-126.
- Holling, H., Böhning, W., and Böhning, D. (2012). Meta-analysis of diagnostic studies based upon SROC-curves: a mixed model approach using the Lehmann family. *Statistical Modelling - an International Journal* **12**, 347-375.
- Doebler, P., Holling, H., and Böhning, D. (2012). A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. *Psychological Methods* **17**, 418-436.
- www.personal.soton.ac.uk/dab1f10/home.htm