

***Méta-analyses des études des technologies
diagnostiques***

Intérêts de la méta-analyse pour les tests diagnostiques

- Permet de gérer
 - la multiplicité des études
 - La « disparité » des résultats

- Permet de faire une synthèse
 - Non arbitraire
 - Sans sélection en fonction des résultats
 - Quantitative
 - Estimation globale de la performance diagnostique du test

Motivation : forte variabilité des résultats d'une étude à l'autre

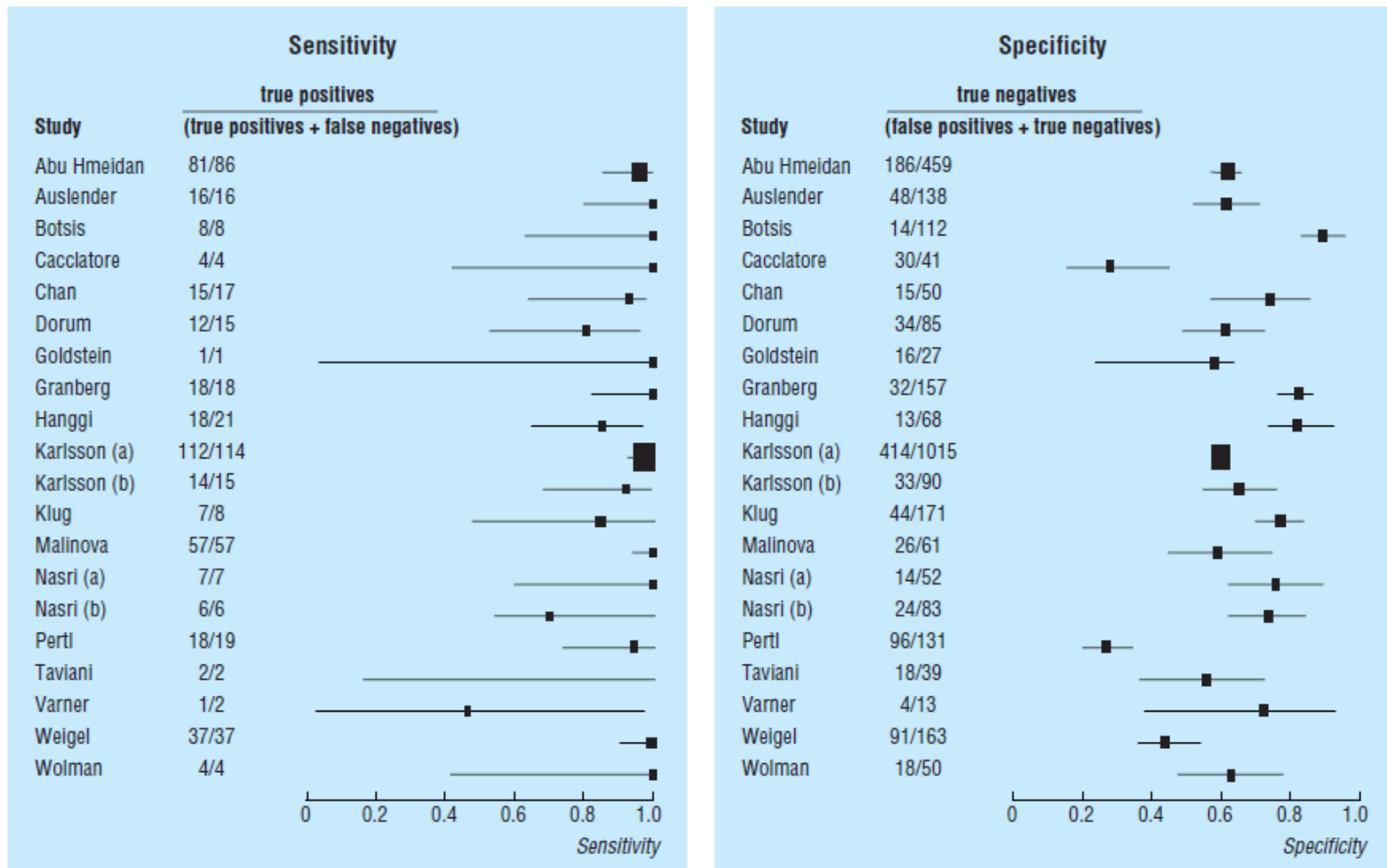
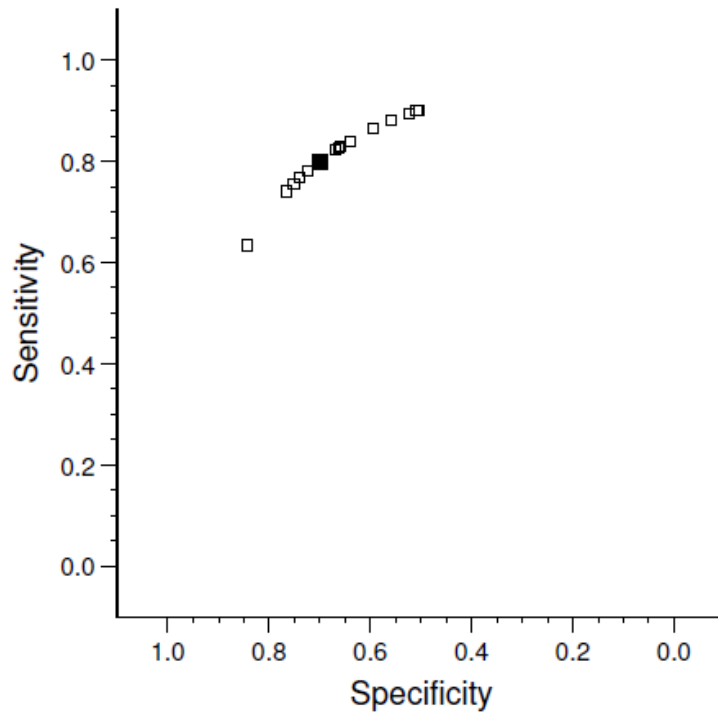


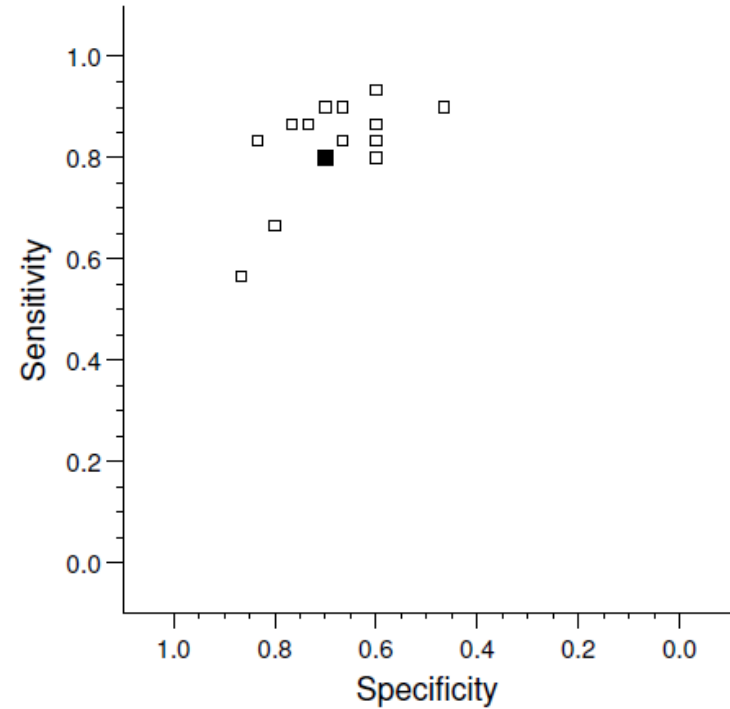
Fig 2 Estimates from 20 studies of sensitivity and specificity of measurement of endometrial thicknesses of more than 5 mm using endovaginal ultrasonography for detecting endometrial cancer.¹⁵ Points indicate estimates of sensitivity and specificity. Horizontal lines are 95% confidence intervals for estimates. Size of points reflects total sample size

Origines de l'hétérogénéité des résultats

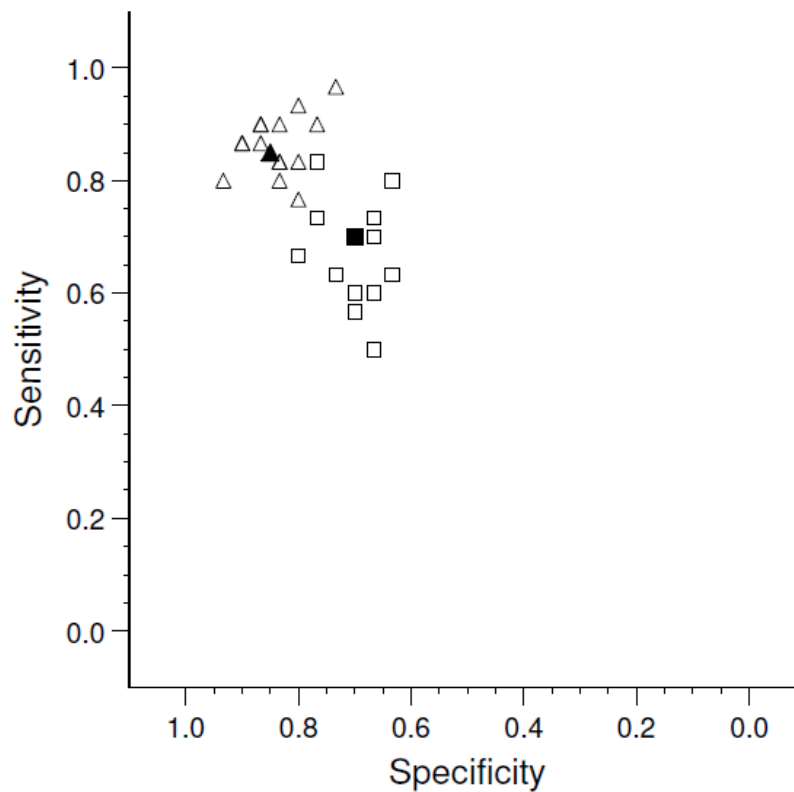
Variation du seuil, sans fluctuation aléatoire



Variation du seuil + fluctuation aléatoire



Hétérogénéité des performances réelles en fonction des patients, des contextes, de la méthodologie (biais)



Méthode

- Similaire à celle de la méta-analyse des essais thérapeutiques
- Question précise
- Recherche exhaustive des études
 - Publiées et non publiées
- Sélection
 - Thématique : adéquation à la question posée
 - Méthodologique : étude le plus possible à l'abris des biais
- Agrégation des résultats
 - Indices de performance diagnostique globale
 - Exploration/explication de l'hétérogénéité

Évaluation de la qualité méthodologique

■ Objectif

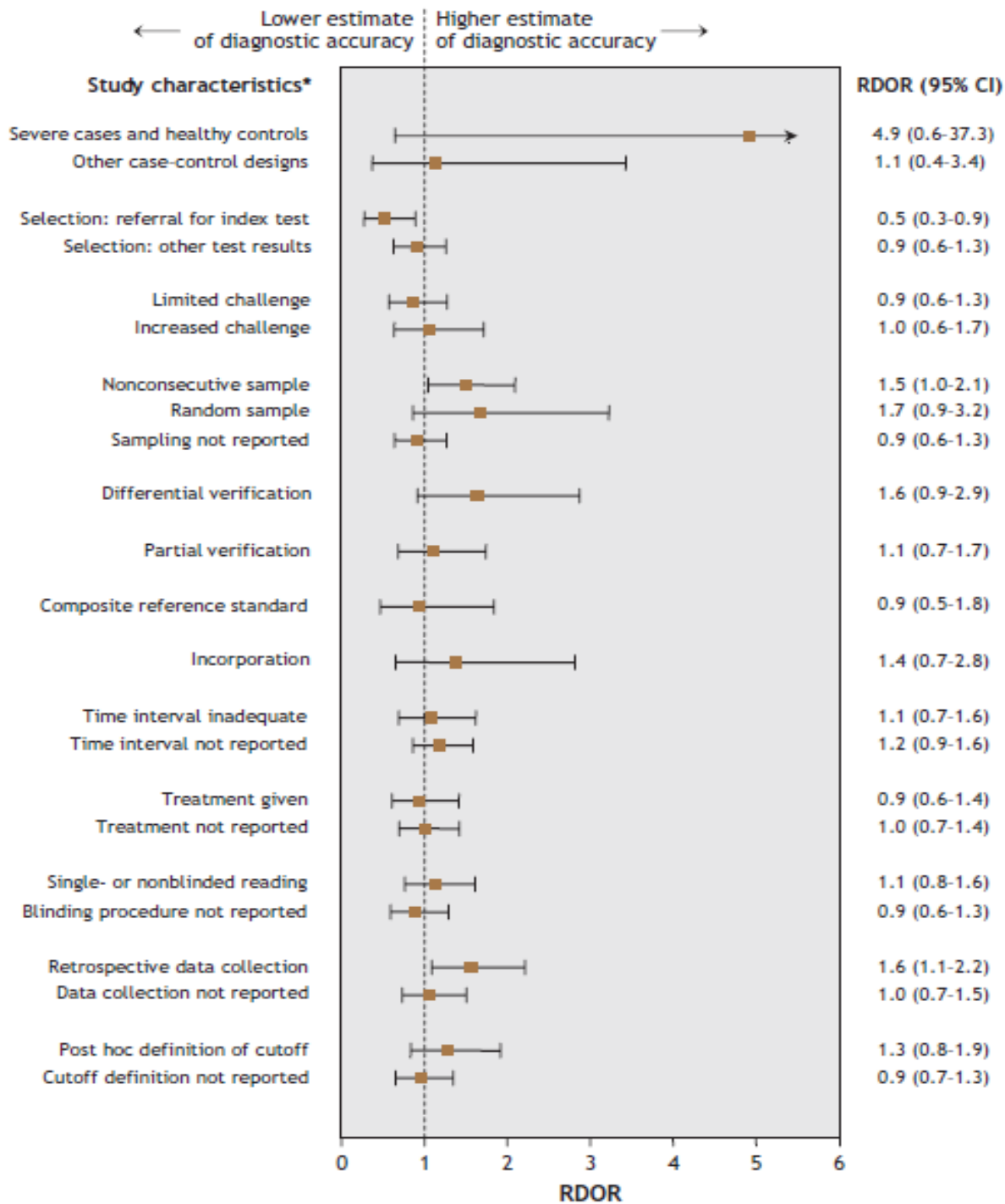
- Éviter les études biaisées
- Limiter l'hétérogénéité des résultats

■ Outils actuel : QUADAS 2

- Permet de classer les études en fonction de leur risque de biais
 - Haut risque de biais
 - Faible risque de biais
 - Risque de biais inévaluable

■ Élaboré à partir des résultats des études de méta-épidémiologie

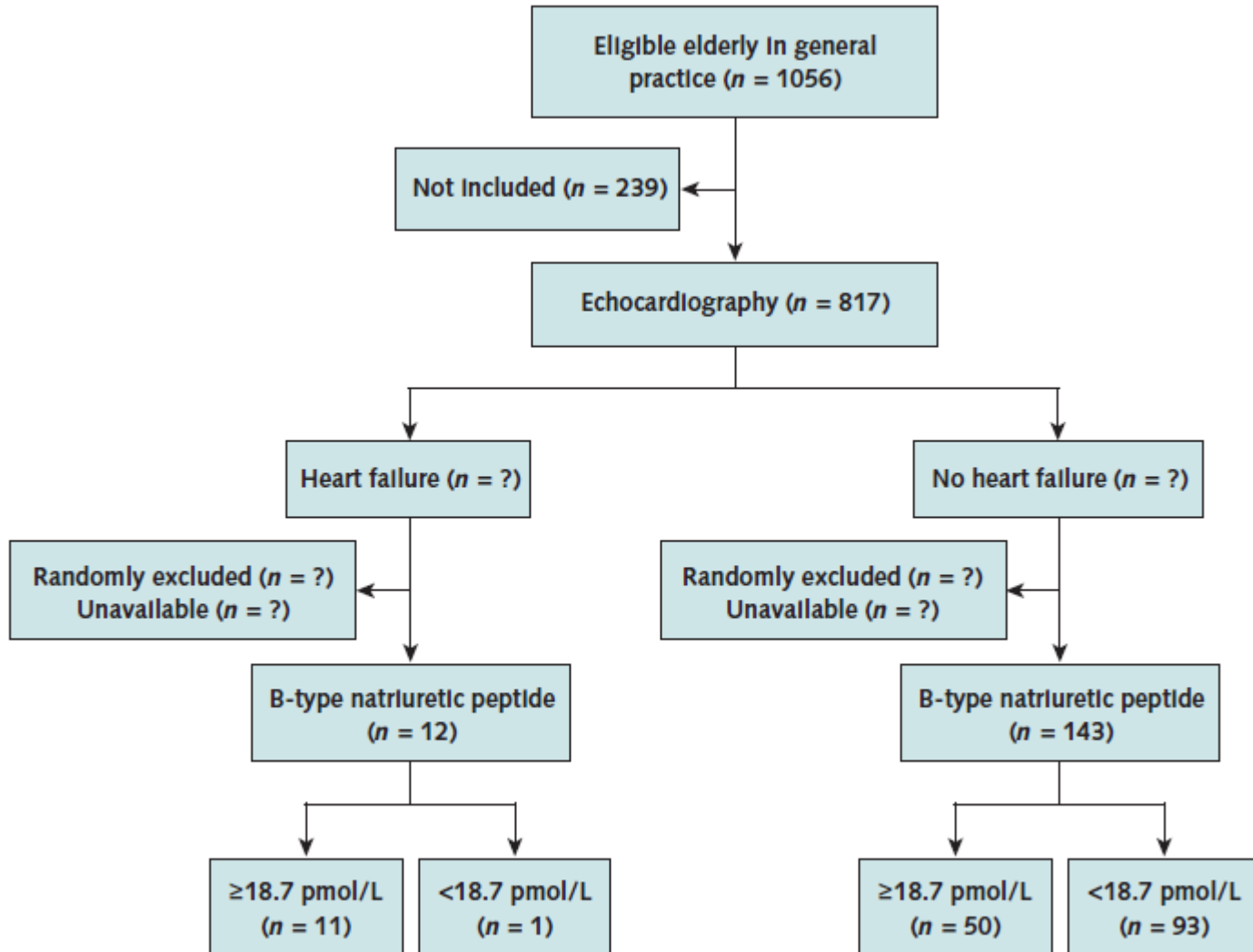
- Identification empirique des facteurs de risque de biais



Lijmer JG, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282(11):1061-6.

Structure de la grille QUADAS 2

- 3 phases
 - Phase 1 : élaborer la question de la revue
 - Phase 2 : flow-chart
 - Phase 3 : analyse de la méthodologique (risque de biais et applicabilité) suivant 4 domaines
- 4 domaines pour l'évaluation du risque de biais et de la pertinence
 - Sélection des patients
 - Le test index (évalué)
 - Le gold standard
 - Les aspect longitudinaux (flux des patients et chronologie)



Domaine 1 - Sélection des patients

- Risque de biais : Est-ce que la sélection est susceptible d'être à l'origine d'un biais ?
 - patients consécutifs ou échantillon aléatoire
 - Absence d'exclusions excessive
 - patients difficile à diagnostiquer -> surestimation des performances diagnostiques
 - Le design n'est pas une étude cas contrôle

- Applicabilité
 - Différences entre ce qui a été fait et la question du lecteur
 - Démographie, comorbidité, contexte, place dans le processus diagnostique

Domaine 2 – Le test évalué (Index test)

■ Risque de biais

- Est-ce que l'interprétation du test évalué a été effectuée sans avoir connaissance du test de référence ?
 - Aveugle ?
- Seuil a-t-il été pré-spécifié ?
 - L'ajustement du seuil pour optimiser Se/Sp entraîne un biais de surestimation -> Seuil data driven

■ Applicabilité, pertinence

- Différences entre ce qui a été fait et la question du lecteur
- Technologie, exécution, interprétation du test

Domaine 3 - Procédure de référence

■ Risque de biais

- La référence est-elle apte à classer correctement les patients
 - $Se=Sp=100\%$
- A-t-elle été interprétée de façon indépendante du test évalué
 - aveugle

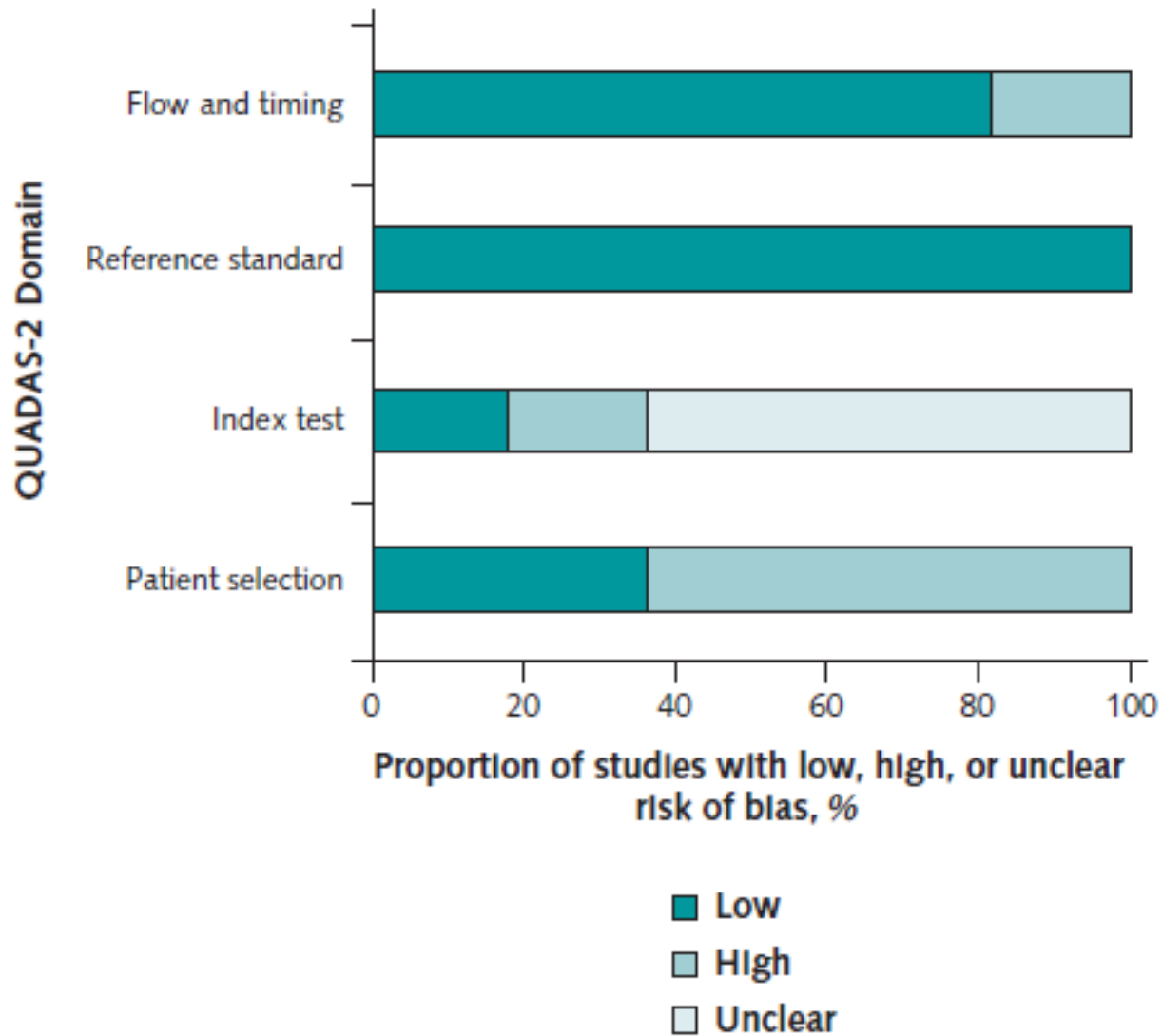
■ Applicabilité

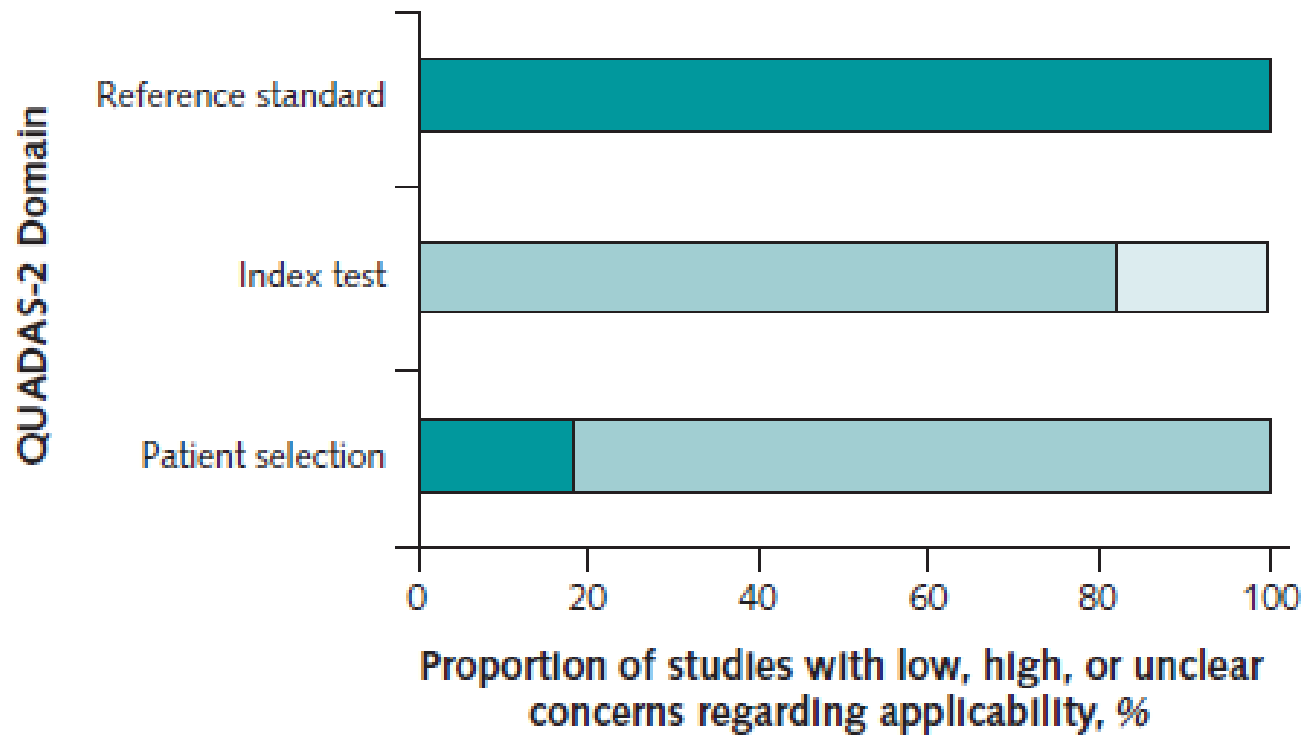
- Définition de la maladie
- (seuil)

Domaine 4 – Flux et timing (aspect longitudinaux)

■ Risque de biais

- Le délais entre test évalué et test de référence est-il approprié
- Tous les patients ont-ils eu le test de référence
- Les patients ont-ils tous eu le même test de référence
- Est-ce que tous les patients ont-ils été inclus dans l'analyse





Analyse statistique, calcul de méta-analyse

■ But

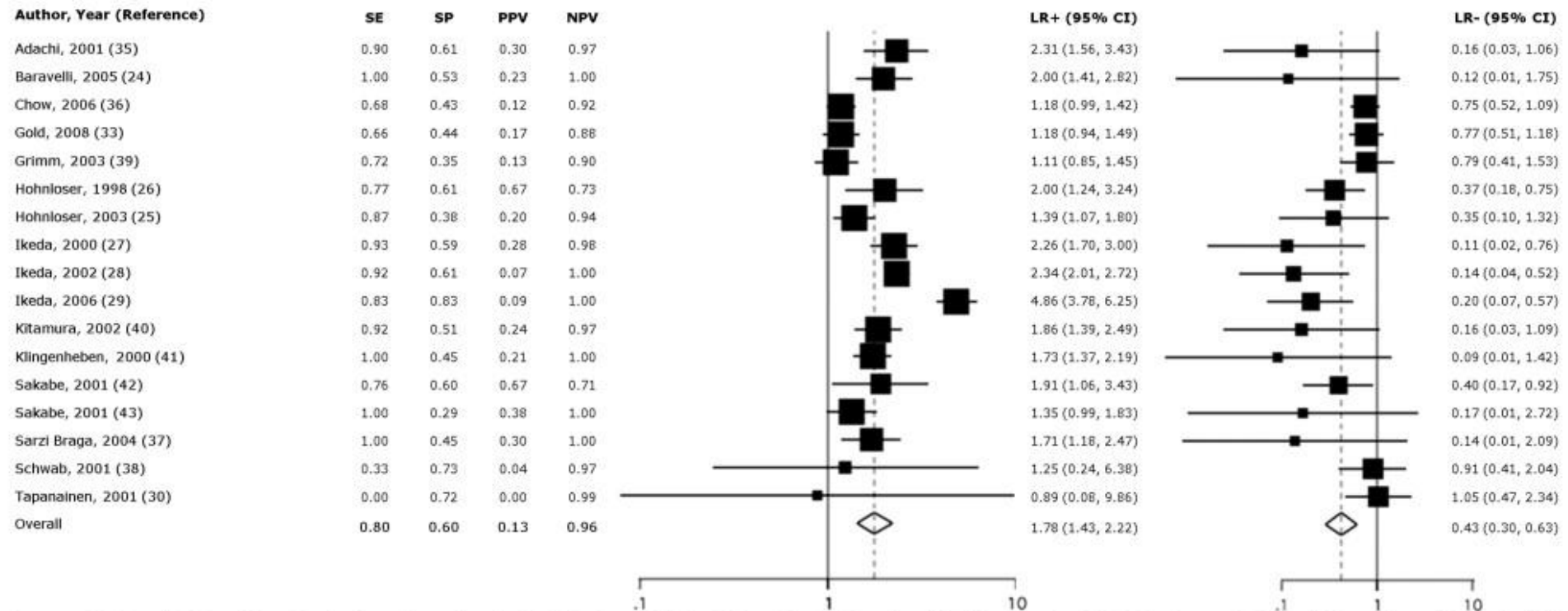
- Synthétiser les résultats des études disponibles
- Pour obtenir une quantification unique de sa performance diagnostique

■ Difficulté

- La quantification de la performance diagnostique est un problème à 2 dimensions : Se et Sp

Méta-analyse des Se et Sp ou RV+ et RV-

Paired forest plot of LR+ and LR- excluding indeterminate MTWA results



Summary estimates and 95% confidence intervals for positive and negative MTWA test results (that is, indeterminate MTWA results excluded from the analysis). Depicted are sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and risk ratio (RR), as well as forest plots for positive (LR+) and negative (LR-) likelihood ratios.

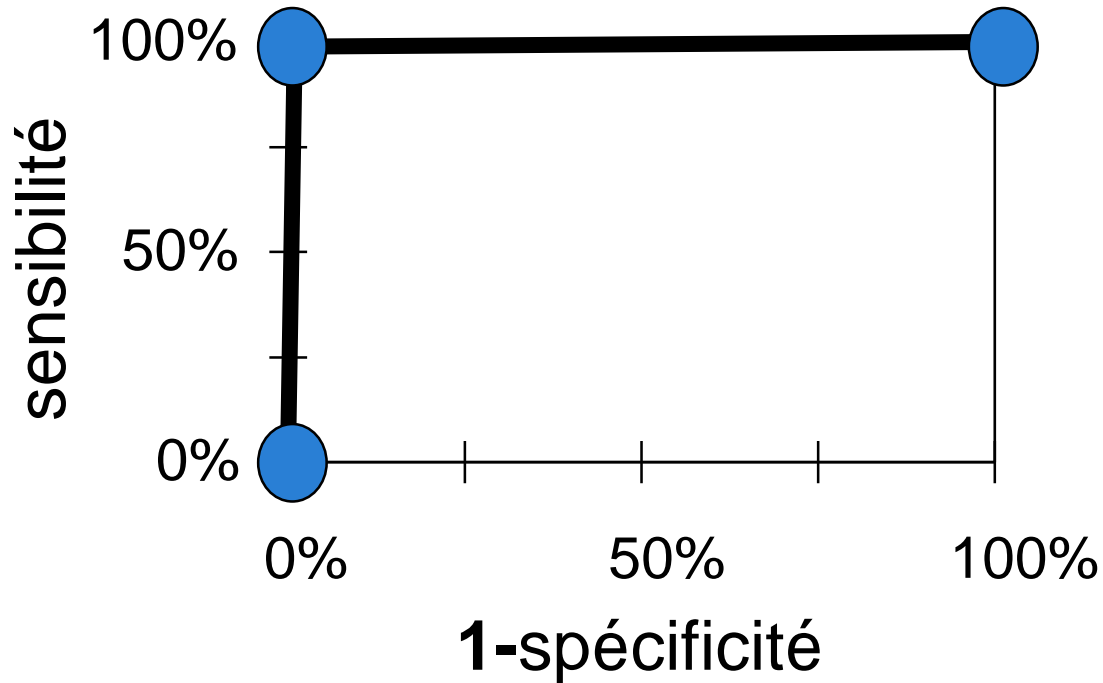
Estimation globale de la performance diagnostique

■ Moyennes des Se et Sp

	sensibilité	spécificité
étude 1	100%	0%
étude 2	99%	99%
étude 3	0%	100%
<hr/>		
moyenne	66%	66%

- Les moyennes des Se et Sp suggèrent une performance diagnostique modeste, **MAIS...**

Courbe ROC



- Performance diagnostique maximale !
- Se et Sp ne sont pas indépendantes
 - Notion de seuil implicite ou explicite

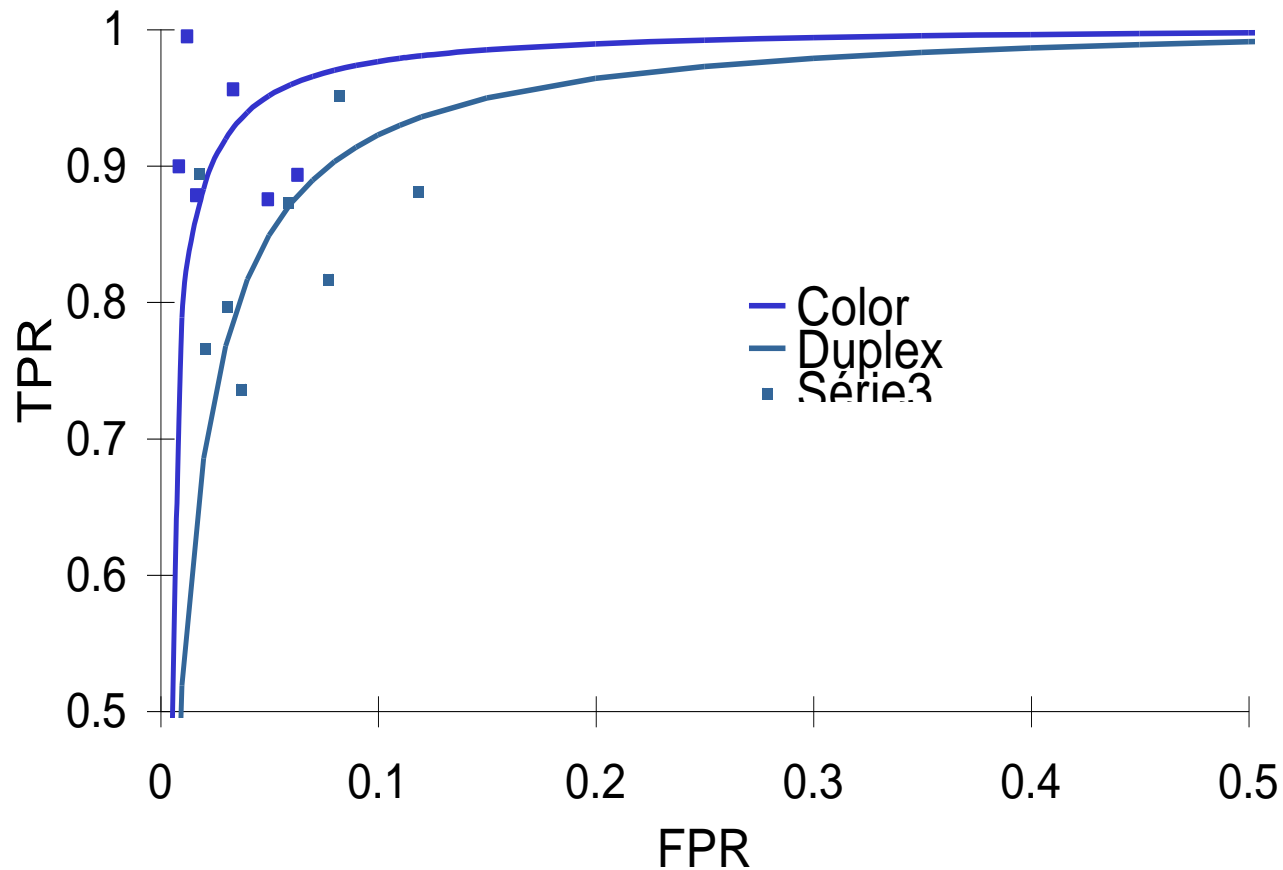
Méthodes statistiques développées pour prendre en compte cette corrélation négative structurelle

- Summary ROC (Moses, Littenberg)
- Bivariate random
- Hierarchical summary ROC

Courbe « Summary ROC » (Moses, Littenberg)

Moses Stat Med 1993

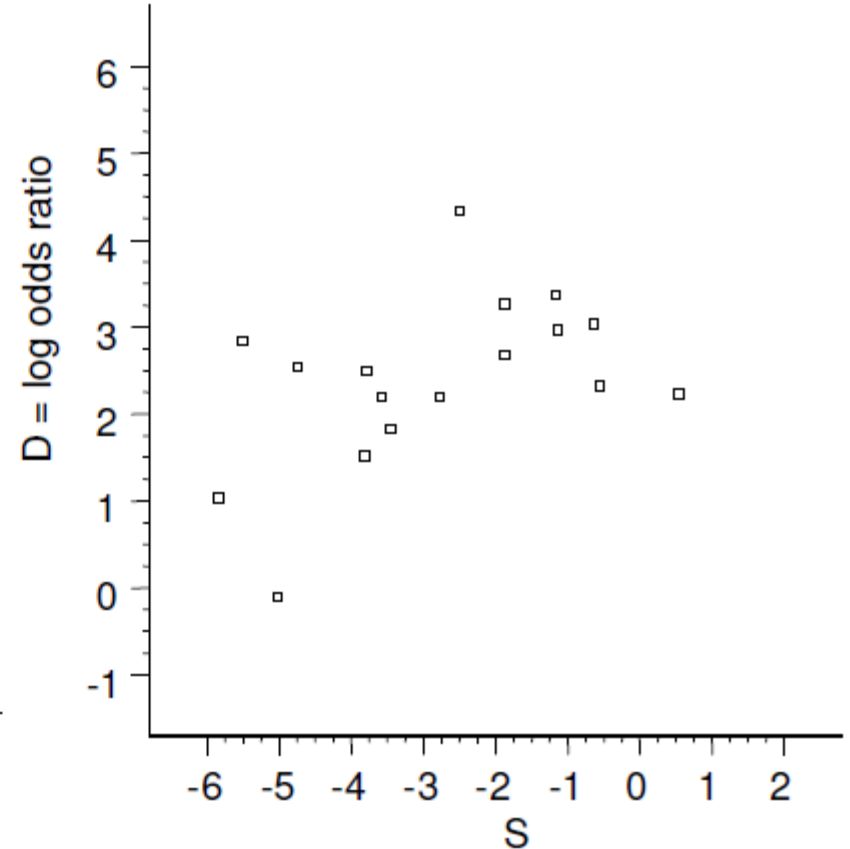
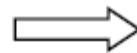
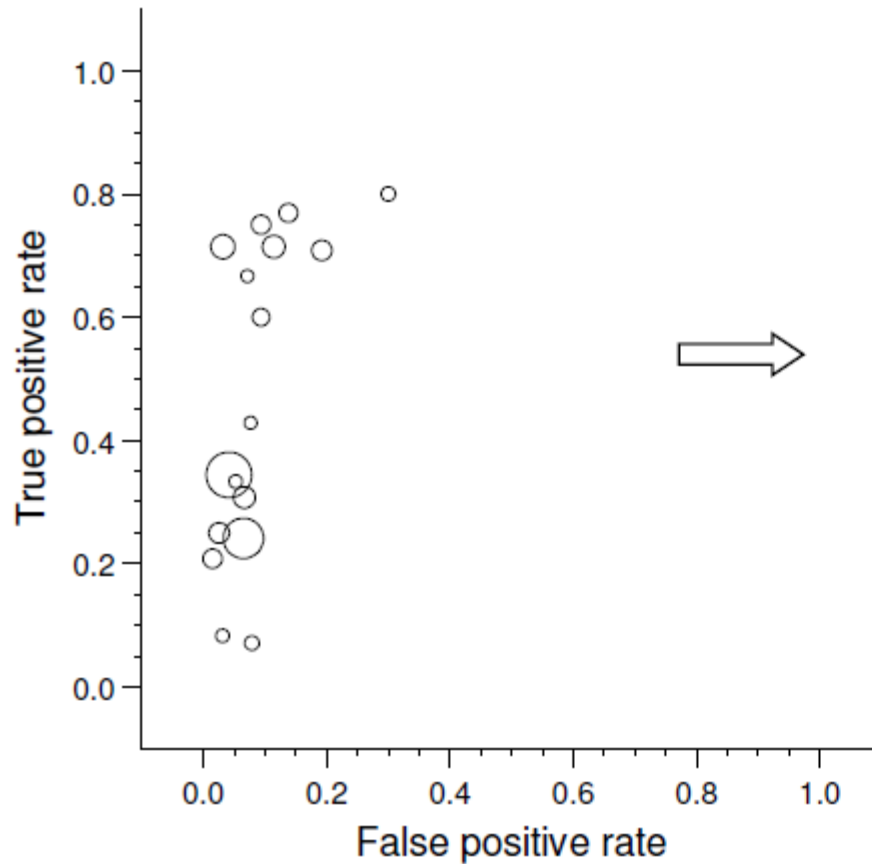
- Ajustement d'une courbe pseudo ROC sur les points représentatifs des études



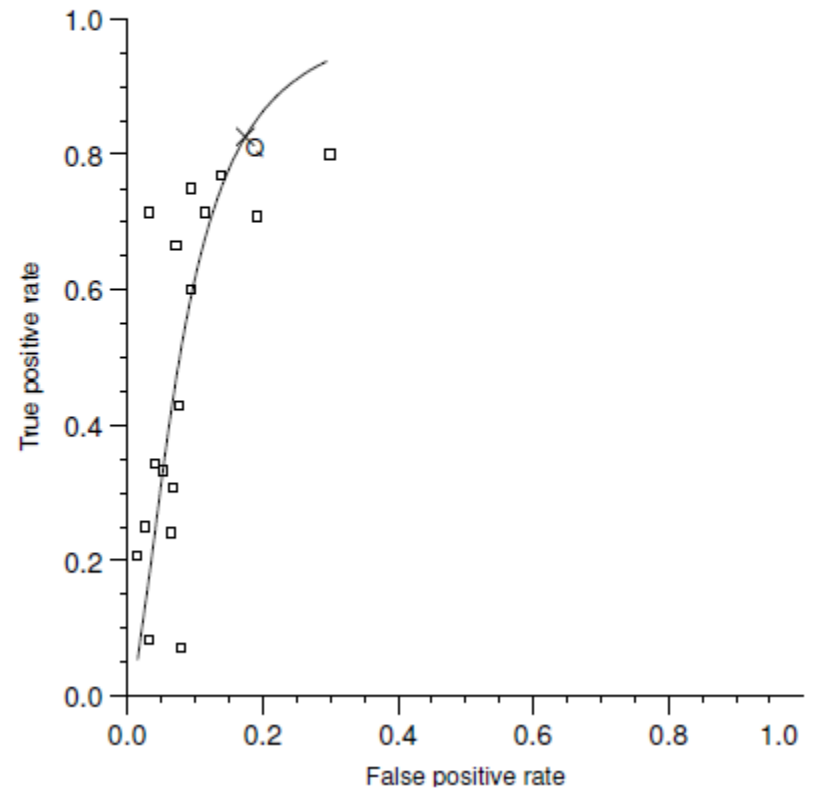
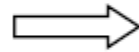
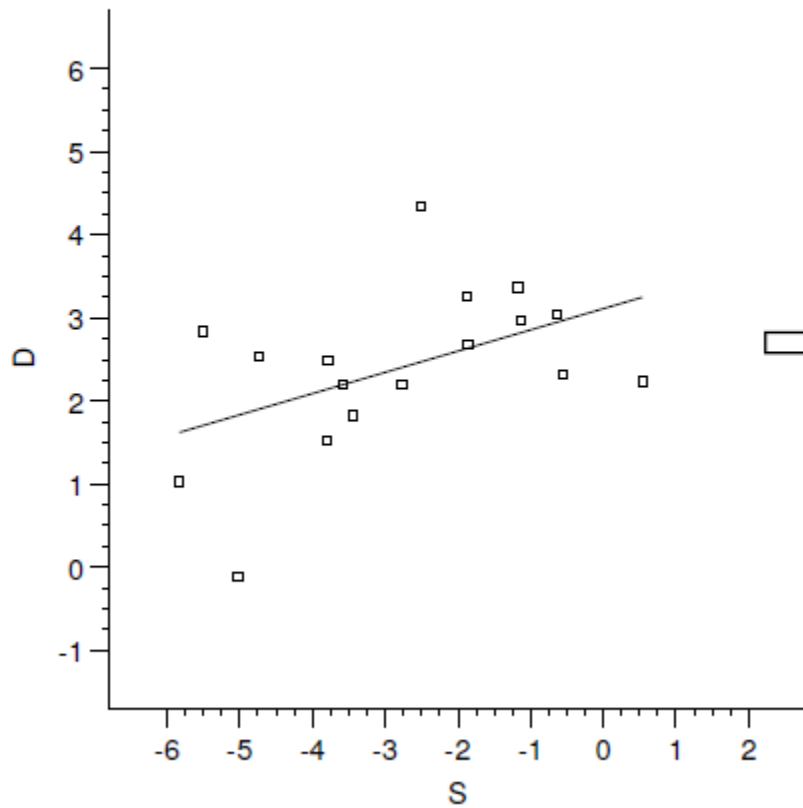
Transformation des variables

$$D = \text{logit}(V+) - \text{logit}(F+)$$

$$S = \text{logit}(V+) + \text{logit}(F+)$$



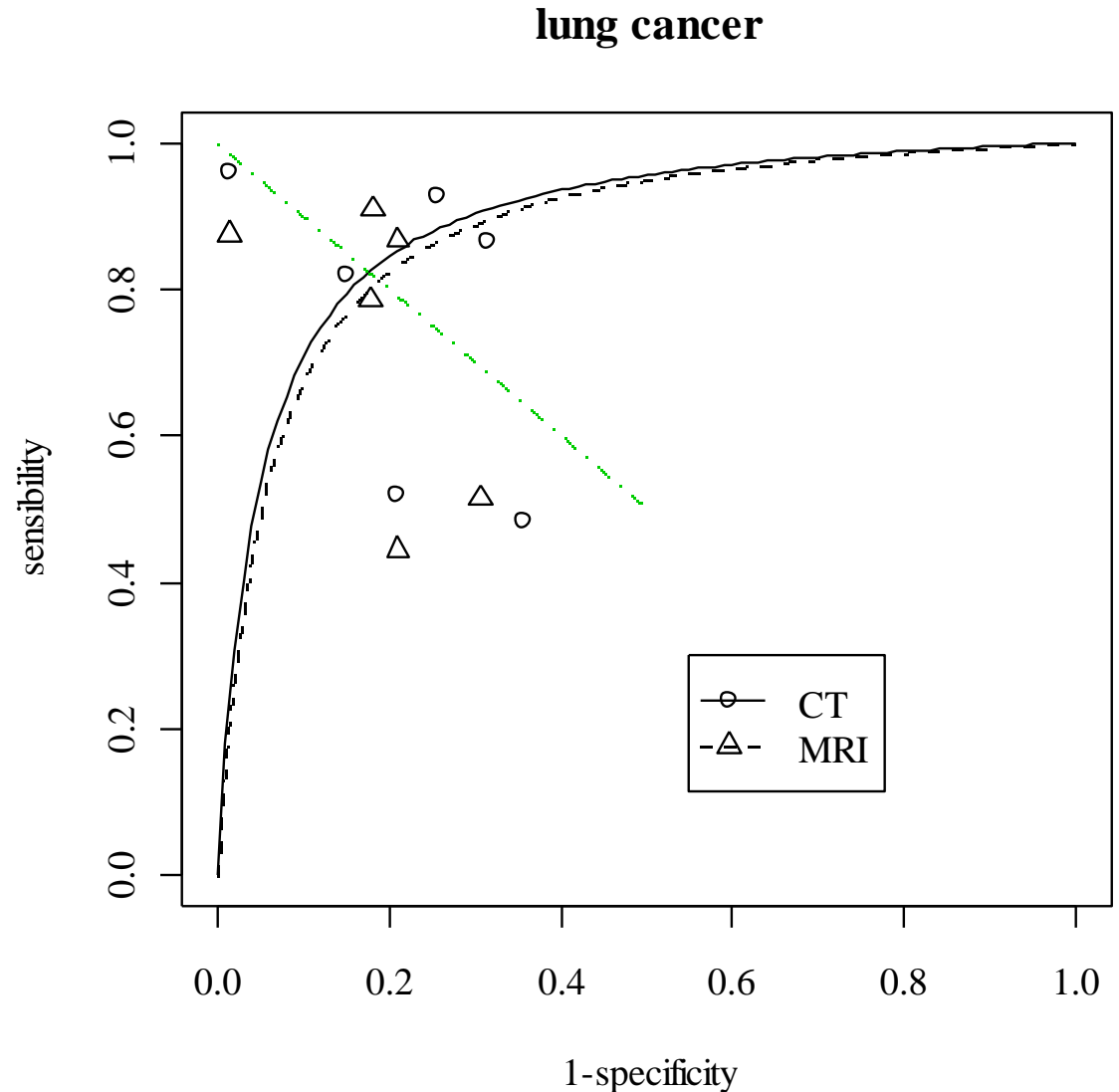
Régression linéaire et transformation inverse



Interprétation des courbes summary ROC

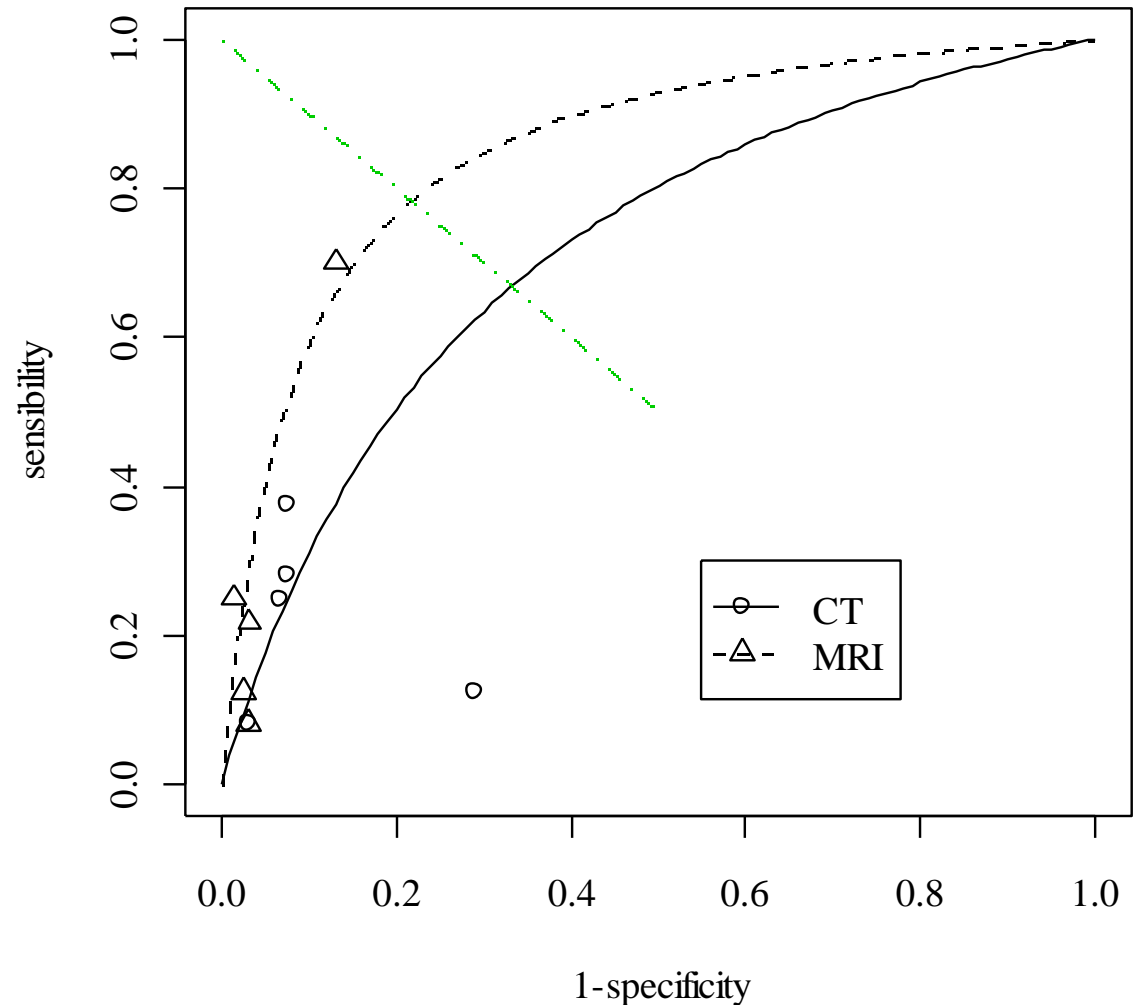
Extrapolation d'une sensibilité et d'une spécificité à partir du point Q^*

$Q^* : Se=Sp$



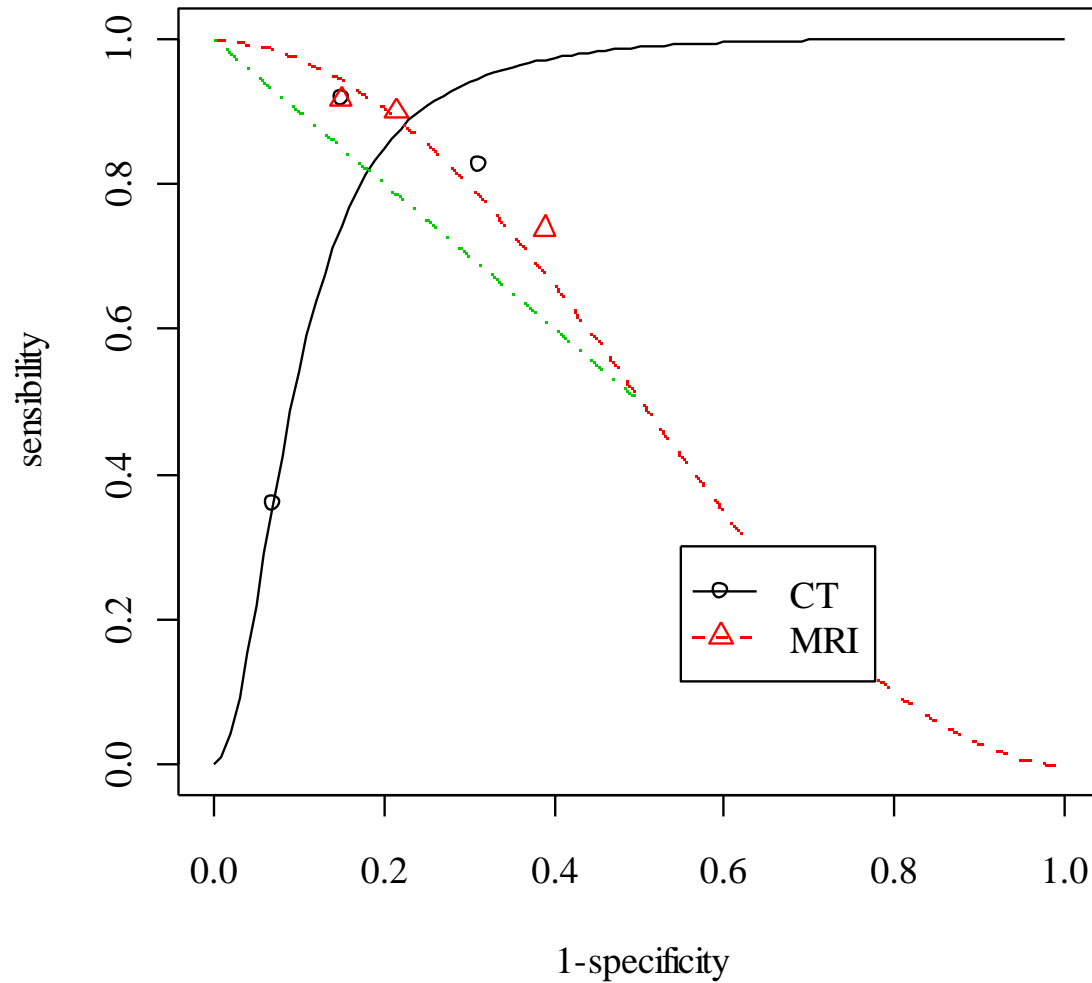
Limites de l'approche - 1

gynecologic cancer



Limites de l'approche - 2

Neck and head cancer



Limite de cette approche

- Ne prends pas correctement en compte la précision de chaque étude
- Ne permet pas d'évaluer l'hétérogénéité
- P values et IC sont erronés

- Ne permet pas d'estimer la sensibilité et la spécificité moyenne

Méthodes modernes - à fondement théorique solide

■ Deux approches

- hierarchical summary ROC model (HSROC)

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in medicine 2001; 20(19): 2865-84.

- Régression bivariée des Se et Sp (bivariate)

Reitsma JB et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982-90.

■ Les 2 approches sont en fait identiques en l'absence de covariables

Harbord et al. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2006;1:1-21

Méthodes modernes - caractéristiques

- **Modèle hiérarchique (aléatoire)**
 - Permet de prendre en compte la variabilité intra et inter études
- **Méthode bivariante**
 - Permet de prendre en compte la corrélation négative entre S_e et S_p
- **Distribution binomiale**
 - Prendre en compte de manière correcte la nature des résultats
 - Pas besoin de faire des ajustement pour les zéros
- **Méthode de régression**
 - Permet de prendre en compte des covariables
 - Permet d'examiner les sources d'hétérogénéité

■ Random bivariate

- Modélisation des logit de la sensibilité et de la spécificité par une loi normale bi-variée (ou directement par une binomiale bivariée)
- Prise en compte de la corrélation intrinsèque entre Se et Sp
- Estimation de l'hétérogénéité (variation non expliquée) et de la corrélation entre Se et Sp

$$\begin{pmatrix} \mu_{A,i} \\ \mu_{B,i} \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma \right] \text{ with } \Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

$$tp_i \sim \text{binomial}(tp_i + fn_i, \mu_{A,i})$$

$$tn_i \sim \text{binomial}(tn_i + fp_i, \mu_{B,i})$$

■ HSROC

- Modèle aléatoire hiérarchique
- Le seuil implicite et la performance diagnostique (InDOR) sont aléatoires (entre les études)

■ Estimation de Se et de la Sp moyenne, et ROC moyenne

- Pour les 2 méthodes
- Meta-paramètres

Calcul

- Nécessite une inférence Bayésienne
- Estimation par Markov Chain (MCMC)
 - WinBug
 - Méthode délicate à mettre en oeuvre
 - Diagnostic de convergence
 - Choix d'a-priori non informatif
 - Analyse de sensibilité

	Estimate	2.5 %	97.5 %
Sens	0.8555863	0.7003428	0.9375720
FPR	0.1738679	0.1010711	0.2826126
mu1	1.7791048	0.8489307	2.7092790
mu2	-1.5584592	-2.1853794	-0.9315390
Sigma11	1.1595810	-0.6178446	2.9370066
Sigma22	0.5905219	-0.1995998	1.3806436
Sigma21	0.7712834	-0.3004160	1.8429828

Résultat méthode bivariée

Parameter	Estimate	95% CI
Mean sensitivity	0.74	0.67 to 0.80
Mean specificity	0.75	0.64 to 0.84
Between-study variation in logit sensitivity	0.13	
Between-study variation in logit specificity	1.12	
Correlation between logit sensitivity and specificity	-0.45	
Mean DOR	8.78	3.84 to 13.7

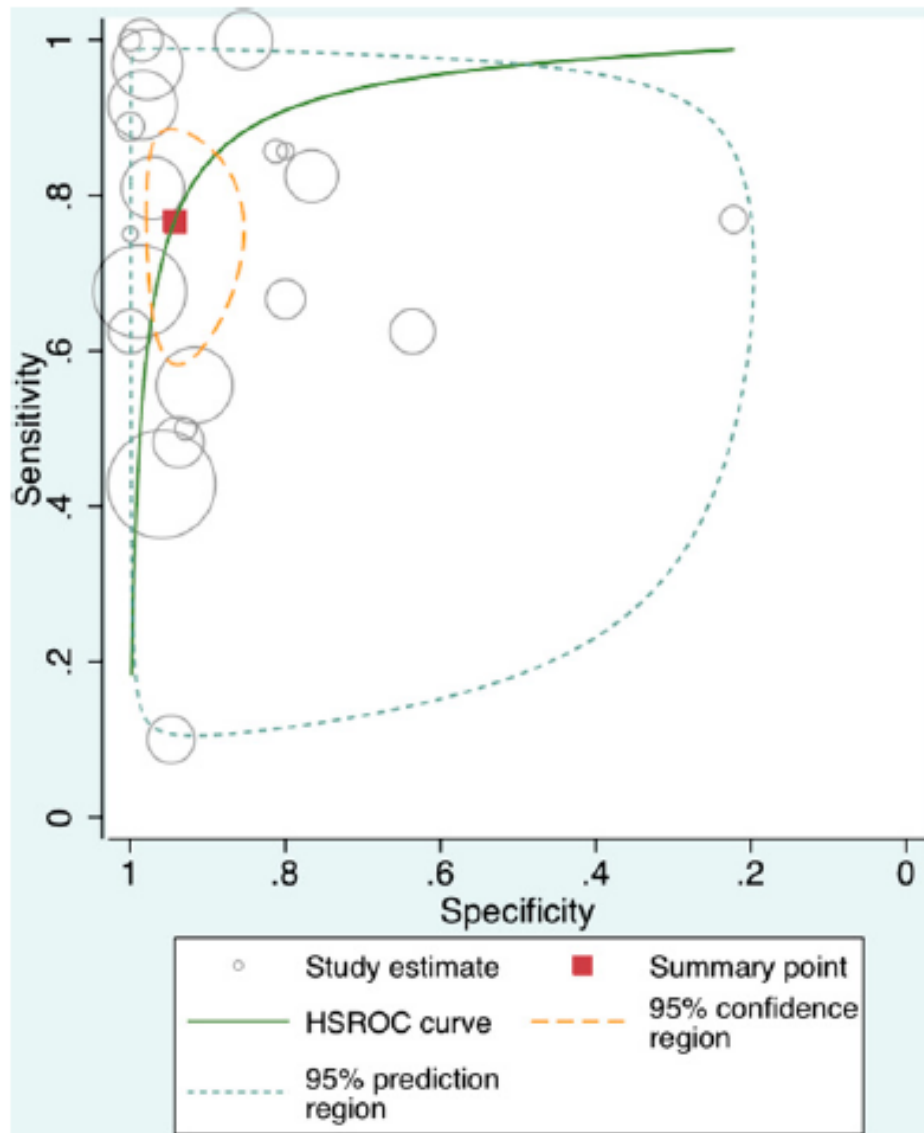
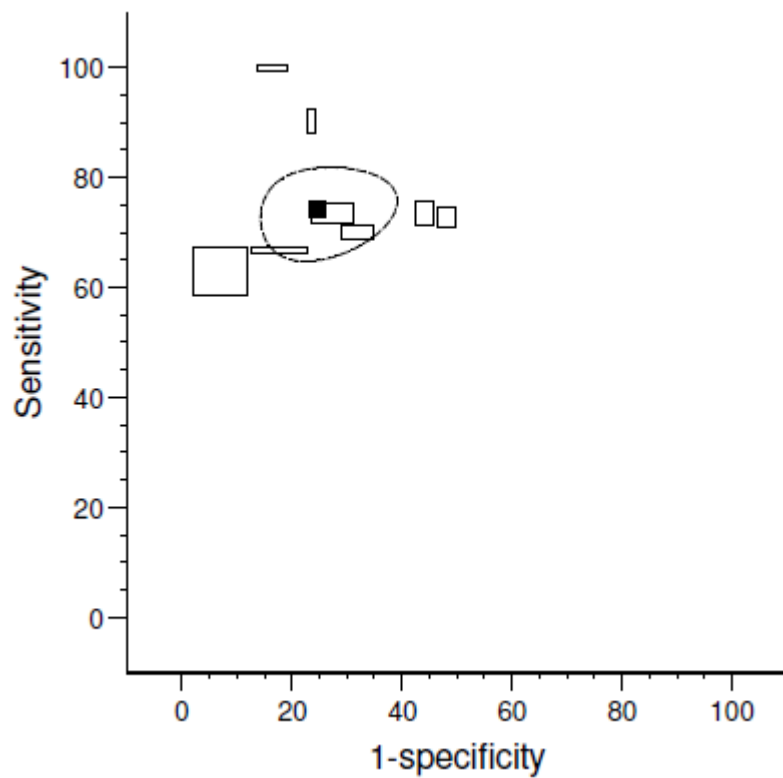
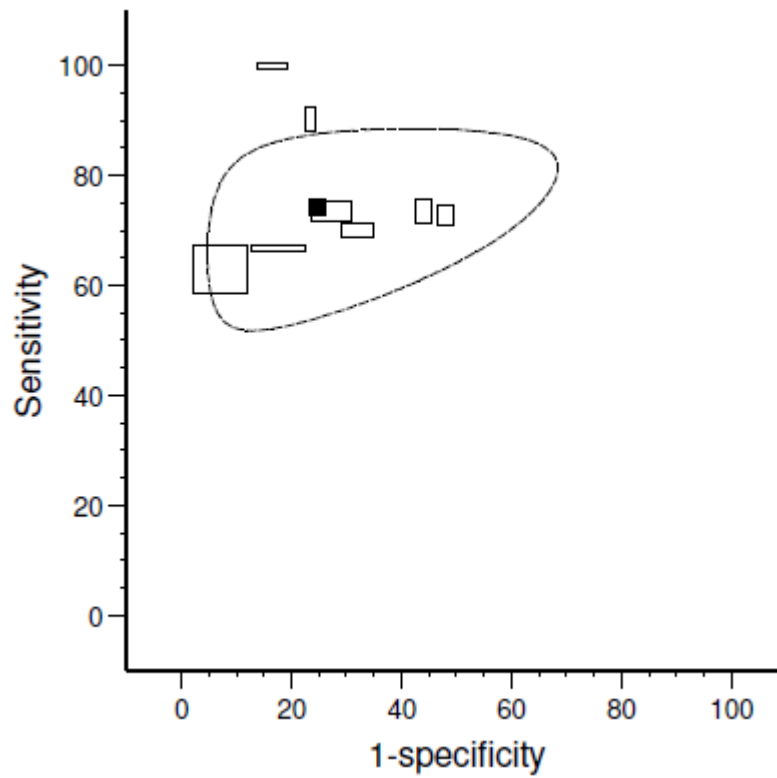


Figure 1 Plot comparing sensitivity and specificity of ultrasound vs CT, illustrating weighted point estimates of individual studies, pooled point estimate from bivariate meta-analysis, hierarchical summary ROC curve and elliptical 95% confidence interval of pooled estimate.

Intervalle de confiance - Précision



Intervalle de prédiction - Prédiction



Outils

- RevMan
- MetaDisc (2004)

- SAS
 - METADAS (Cochrane)
- STATA
- R
 - Bamdit
 - HSROC
 - MADA

