



Sciences Economiques et Sociales de la Santé  
& Traitement de l'Information Médicale

[sesstim.univ-amu.fr](http://sesstim.univ-amu.fr)

**Fabien SUBTIL**

*Service de Biostatistique des HCL, Equipe biostatistiques-santé,  
Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Lyon 1*

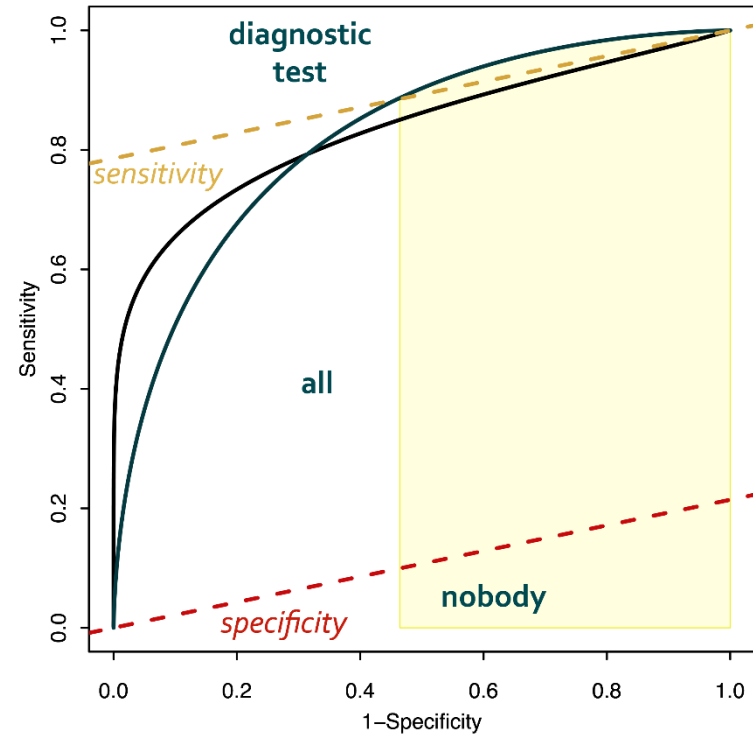
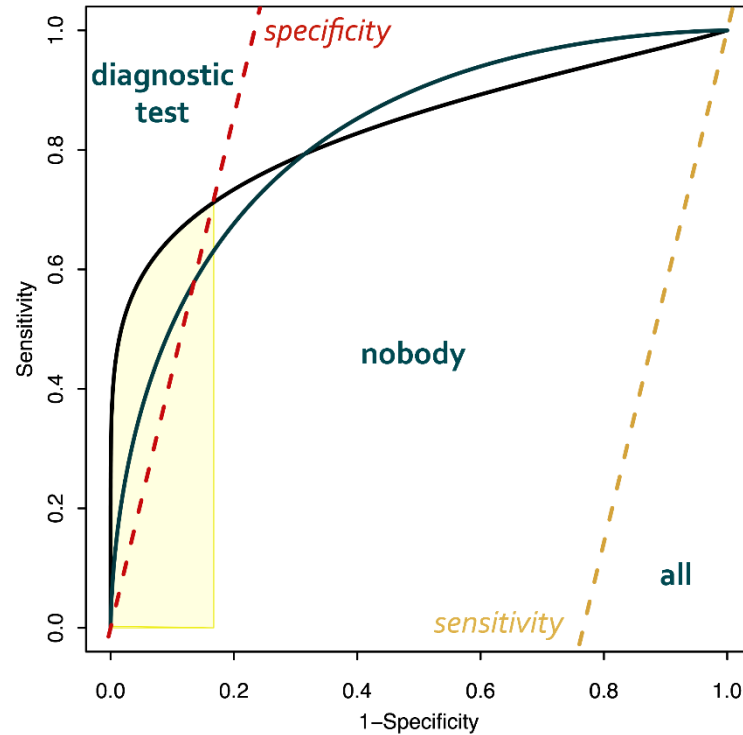
**Evaluer et comparer des biomarqueurs diagnostiques quantitatifs  
en tenant compte de leur utilité clinique**

**Mai 2018**



**Cliquez ici pour voir l'intégralité des ressources associées à ce document**

# Evaluer et comparer des biomarqueurs diagnostiques quantitatifs en tenant compte de leur utilité clinique



Fabien Subtil

Service de Biostatistique des HCL

Laboratoire de Biométrie et Biologie Evolutive – Université Claude Bernard Lyon 1

# Biomarqueurs diagnostiques quantitatifs

- Troponine I et T pour détecter un infarctus du myocarde
- D Dimères pour détecter une maladie thromboembolique
- C Reactive Protein pour détecter une maladie inflammatoire
- PSA pour détecter un cancer de la prostate
- Antigène CA 19-9 et anticorps CA 125 pour détecter un cancer du pancréas

# Performances diagnostiques

$Y$  : valeur prise par le marqueur –  $c$  : seuil de positivité

- Sensibilité

$$P(Y > c | M)$$

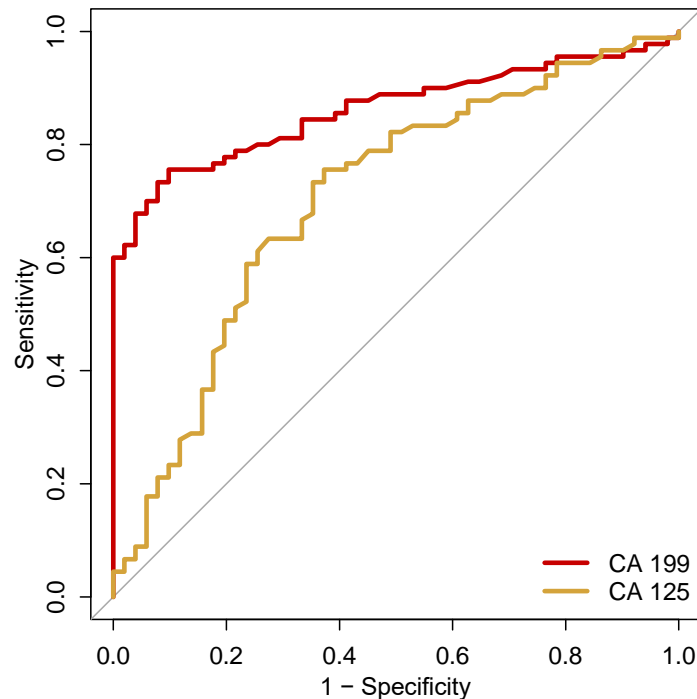
- Spécificité

$$P(Y \leq c | \bar{M})$$

- Lorsque le seuil de positivité n'est pas encore défini, évaluation globale sur l'ensemble des seuils possibles (courbe ROC)

# Exemple : détection du cancer du pancréas

- Distinguer la pancréatite du cancer du pancréas à l'aide du CA 19-9 ou du CA 125



■ AUC CA 19-9 : 0,86

■ AUC CA 125 : 0,71

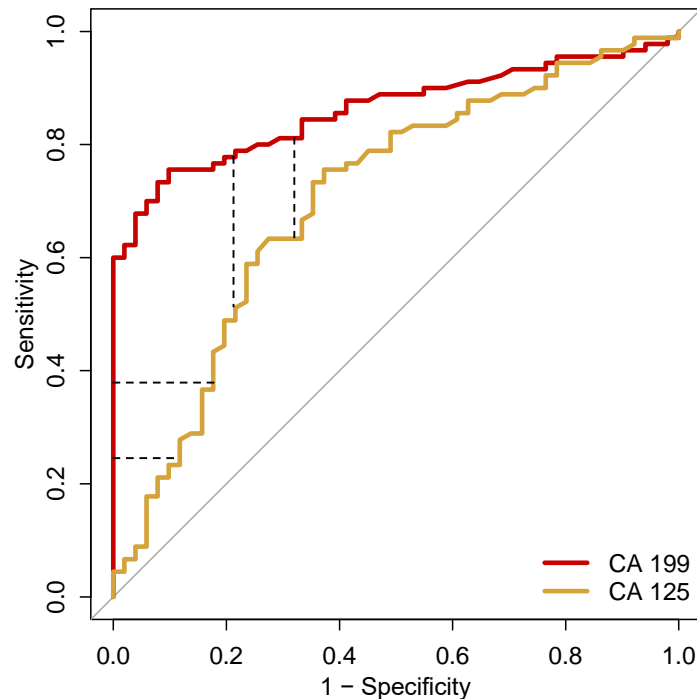
P-value = 0,006

Intérêt de l'AUC : test global

- Conclusion : privilégier le CA 19-9

# Evaluation globale des performances diagnostiques

- Evaluation globale des performances diagnostiques possible car les courbes ROC sont **emboîtées**

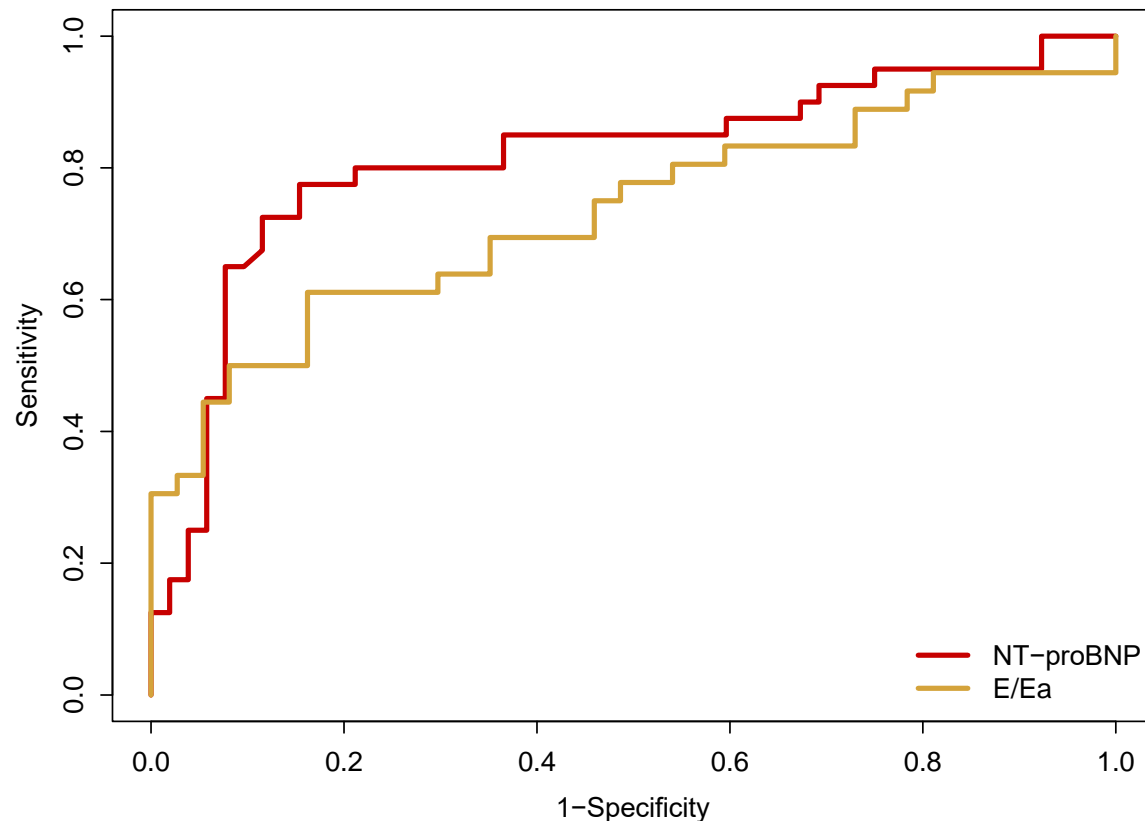


- Quelle que soit la sensibilité désirée :  
spécificité<sub>CA 19-9</sub> ≥ spécificité<sub>CA125</sub>
- Quelle que soit la spécificité désirée :  
sensibilité<sub>CA 19-9</sub> ≥ sensibilité<sub>CA125</sub>

# Diagnostic d'hypertrophie ventriculaire gauche (HVG)

- 93 patients consécutifs consultant à l'hôpital de la Croix-Rousse pour hypertension
- Gold standard pour le diagnostic d'HVG : échocardiographie (41 HVG)
- Deux biomarqueurs :
  - concentration de NT-proBNP dans le plasma
  - ratio E/Ea

# Diagnostic d'hypertrophie ventriculaire gauche (HVG)



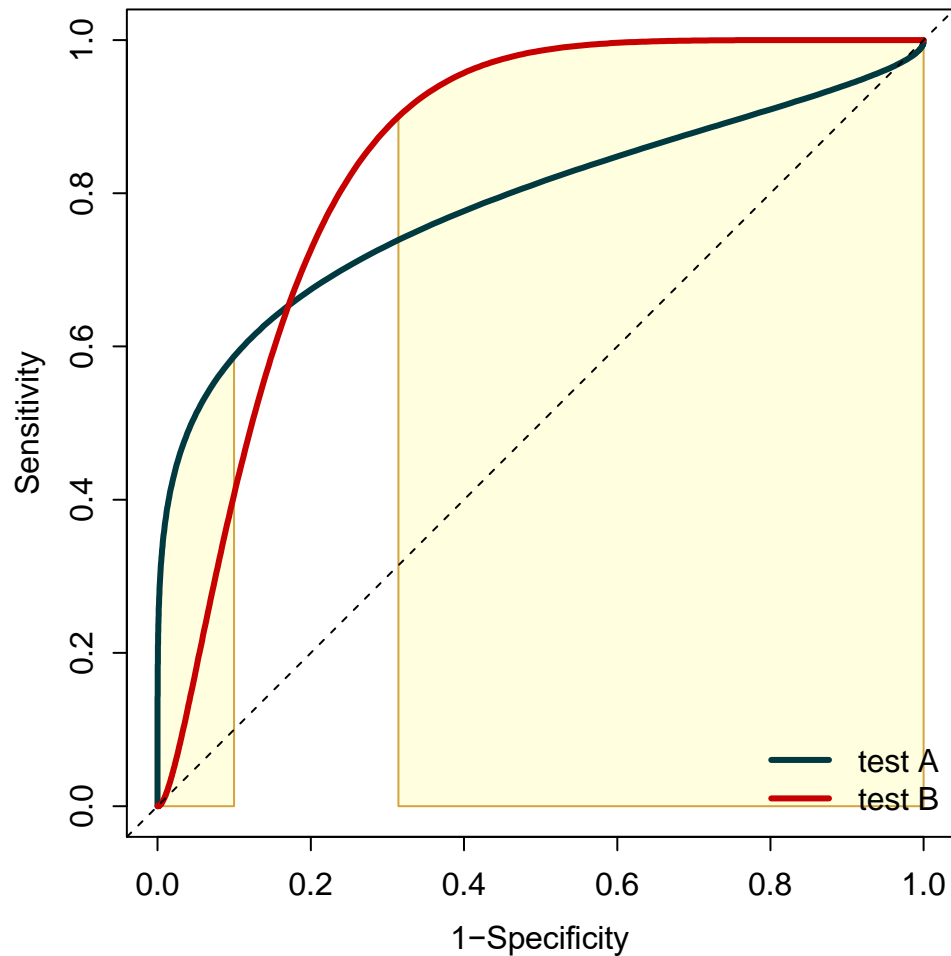
- Pour une bonne sensibilité : privilégier CA 19-9
- Pour une bonne spécificité : privilégier CA 125

Impossible de comparer « globalement » les performances diagnostiques des deux biomarqueurs



# Aire sous la courbe partielle

- Aire sous la courbe ROC dans une zone sans croisement



# Aire sous la courbe partielle

- Comment déterminer la zone d'intérêt ?
- Privilégier la sensibilité ou la spécificité ?
- A partir de quelle valeur de sensibilité / spécificité le test est-il utile ?
- Quelle est la sensibilité / spécificité minimale nécessaire ?

# Aire sous la courbe partielle

- Passage de la notion de **performances diagnostiques** (propres au test)
- A la notion d'**utilité du test dans une population** (spécifique d'une population)

# Prise en compte de la prévalence

- Définition d'une fonction d'**utilité** du test au sein de la population
- Exemple : proportion de patients correctement classés (pour un seuil  $c$  du test)

	Malade	Non Malade
Test positif	$Sen(c) \times \pi$	$(1 - Spe(c)) \times (1 - \pi)$
Test négatif	$(1 - Sen(c)) \times \pi$	$Spe(c) \times (1 - \pi)$

Fonction d'utilité

$$U(c) = Sen(c) \times \pi + Spe(c) \times (1 - \pi) \quad (\neq \text{indice de Youden})$$

$\pi$  : prévalence de la maladie dans la population étudiée

## Balance coûts / bénéfiques

- Traiter à tort un patient non malade  $\neq$  ne pas traiter un patient malade
- Définition de l'utilité associée à chacune des quatre situations possibles

	Malade	Non Malade
Test positif	$U_{VP}$	$U_{FP}$
Test négatif	$U_{FN}$	$U_{VN}$

- Utilité en termes d'état de santé associé à la situation (qualité de vie, quantité de vie, ...)
- Nouvelle fonction d'utilité :

$$U(c) = Sen(c) \times \pi \times U_{VP} + (1 - Sen(c)) \times \pi \times U_{FN} \\ + (1 - Spe(c)) \times (1 - \pi) \times U_{FP} + Spe(c) \times (1 - \pi) \times U_{VN}$$

# Quantification des coûts et bénéfices

La fonction d'utilité peut-être ré-écrite sous la forme :

$$\tilde{U}(c) = Sen(c) \times \pi + (1 - Spe(c)) \times (1 - \pi) \times \frac{U_{VN} - U_{FP}}{U_{VP} - U_{FN}}$$
$$\frac{U_{VP} - U_{FN}}{U_{VN} - U_{FP}} = \frac{BN}{CN} = \frac{1 - r}{r}$$

- **BN/CN** : ratio bénéfice net sur coût net  
s'interprète comme le nombre d'individus que le clinicien accepte de traiter à tort pour ne pas manquer de traiter un patient réellement malade
- **r** : seuil de traitement  
risque de maladie à partir duquel un clinicien accepte de traiter

# Pertinence des valeurs de sensibilité et de spécificité

Est-il pertinent dans l'AUC d'intégrer les valeurs de sensibilité sur l'ensemble des valeurs de  $1 - \text{spécificité}$  ?

- Sensibilité 98 % : n'est-t-il pas plus pertinent de traiter tous les patients ?
- Spécificité 98 % : n'est-t-il pas plus pertinent de ne traiter personne ?

# Stratégies limites

La stratégie : « traiter en fonction des valeurs du test diagnostique » s'inscrit entre **deux stratégies limites** :

- ne traiter personne
- traiter tout le monde

Conditions limites

- $BN/CN$  très élevé, ou  $\pi$  très élevée → traiter tous les patients
- $BN/CN$  très faible, ou  $\pi$  très faible → ne traiter personne



# Stratégies limites

$$\tilde{U}(c) = Sen(c) \times \pi + (1 - Spe(c)) \times (1 - \pi) \times \frac{CN}{BN}$$

- Utilité de ne traiter personne :

$$\tilde{U}_{nobody} = 0$$

- Utilité de traiter tous les patients :

$$\tilde{U}_{all} = \pi + (1 - \pi) \times \frac{CN}{BN}$$

# Zone d'intérêt du test diagnostique

- Stratégie de traitement reposant sur le test diagnostique plus utile que celle ne traitant personne

$$\tilde{U}(c) \geq \tilde{U}_{nobody} \quad \rightarrow \quad Sen(c) \geq (1 - Spe(c)) \times (1 - \pi) / \pi \times CN / BN$$

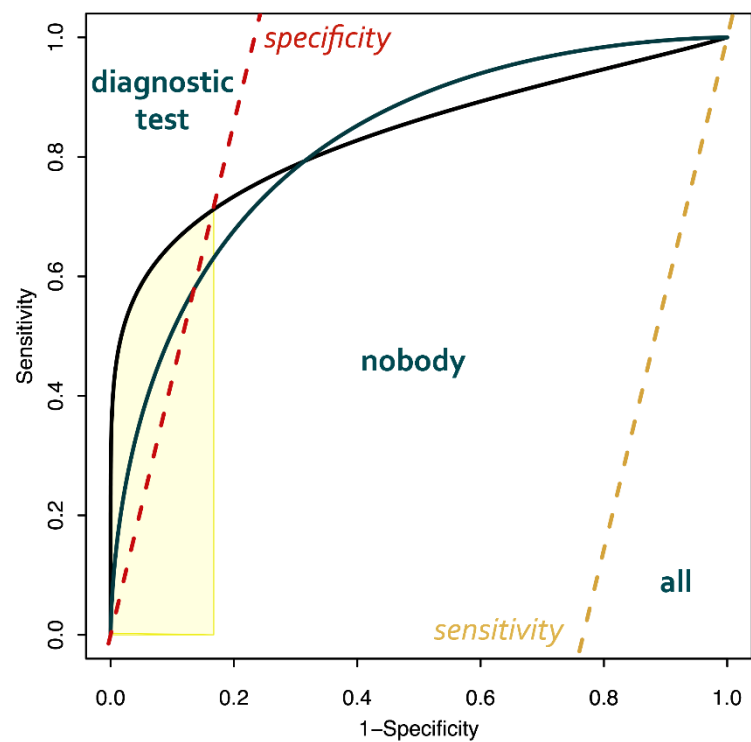
droite de spécificité

- Stratégie de traitement reposant sur le test diagnostique plus utile que celle traitant tout le monde

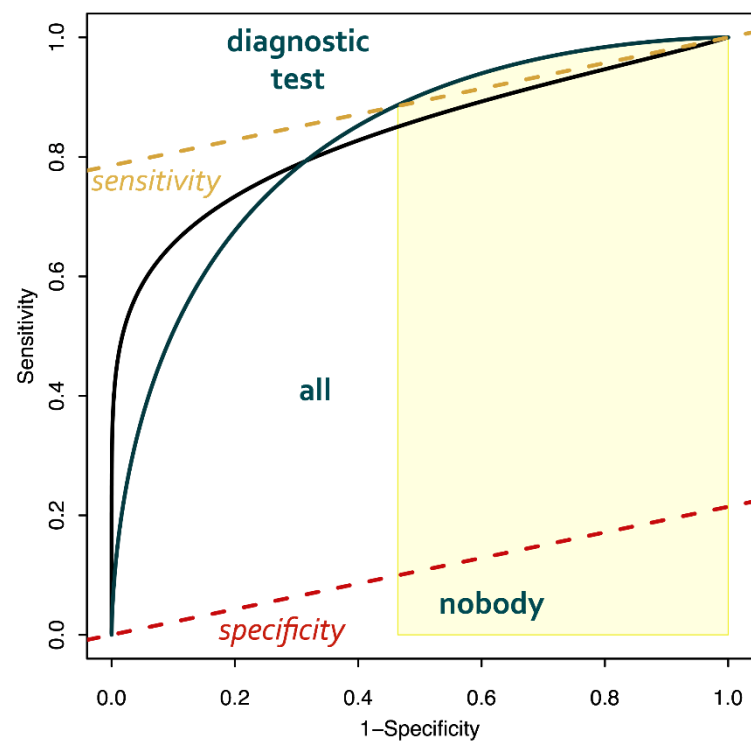
$$\tilde{U}(c) \geq \tilde{U}_{all} \quad \rightarrow \quad Sen(c) \geq 1 - Spe(c) \times (1 - \pi) / \pi \times CN / BN$$

droite de sensibilité

# Zone d'intérêt du test diagnostique



$BN/CN = 0.1$

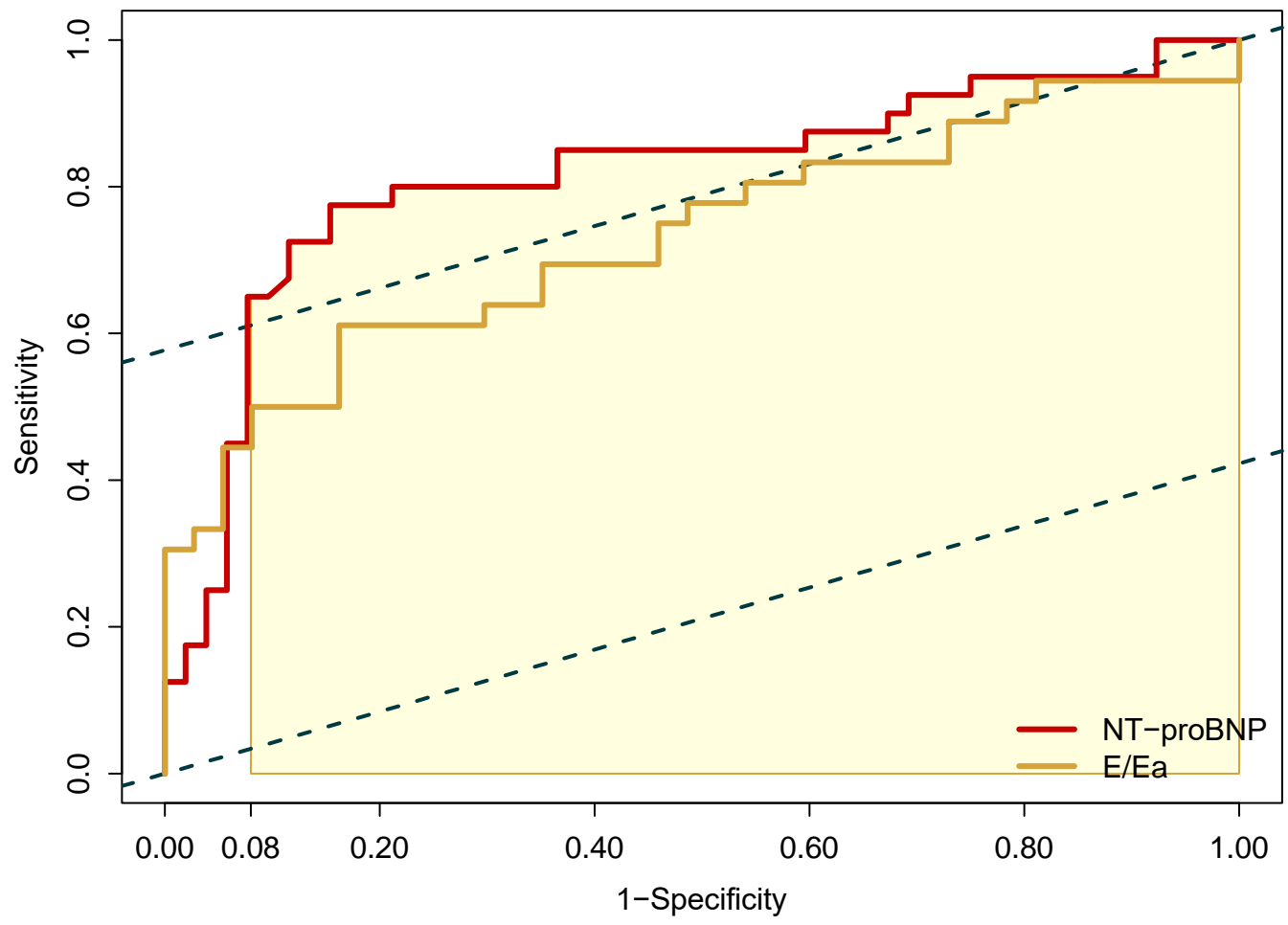


$BN/CN = 2$

$\pi = 0.7$

# Diagnostic d'HVG

$\pi = 44\%$        $BN/CN = 3$  (au moins 3)



# Comparaison des aires sous la courbe ROC partielles

	pAUC	IC 95 %
NT-proBNP	0.85	[0.72, 0.91]
E/Ea	0.74	[0.61, 0.95]

p-value = 0.19

# Comparaison des approches

- Courbes de décision de Vickers  
stratégie optimale (évaluée en un point)
  - ❑ pour une valeur de prévalence  $\pi$
  - ❑ pour une gamme de valeurs de  $BN/CN$nécessite l'estimation du seuil optimal pour chaque valeur de  $BN/CN$
- Courbes ROC  
stratégie optimale (évaluée sur une gamme de points)
  - ❑ pour une valeur de prévalence  $\pi$
  - ❑ pour une valeur de  $BN/CN$approche non paramétrique

# Éléments de bibliographie

- Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006; 26(6):565-74.
- Sox CH, Higgins MC, Owens DK. *Medical Decision Making*. Wiley & Sons. 2013.
- Subtil F, Rabilloud M. An enhancement of ROC curves made them clinically relevant for diagnostic-test comparison and optimal-threshold determination. *J Clin Epidemiol*. 2015; 68(7):752-9.