

# Elaboration/ Evaluation d'un Test Diagnostique

Philippe Montravers

DAR

CHU Jean Verdier

Bondy

# 5 critères de qualité d'un test

- Bonne qualité technique
- Interpretation reproductible
- Information valide
- Utile à la prise de décision
- Efficace pour le malade et la population

# 5 critères de qualité d'un test

- Bonne qualité technique
- Interpretation reproductible
- Information valide
- Utile à la prise de décision
- Efficace pour le malade et la population

# Bonne qualité technique

## Domaine technique

- La technique mesure-t-elle ce qu'elle est censée mesurer ?
- Est-elle exacte (absence d'erreur systématique) ?
- Est-elle précise (faible variabilité) ?
- L'examen est-il facile à exécuter correctement ?

# 5 critères de qualité d'un test

- Bonne qualité technique
- **Interprétation reproductible**
- Information valide
- Utile à la prise de décision
- Efficace pour le malade et la population

# Interprétation reproductible

- Reproductibilité de l'examen
    - Au même endroit ou dans des lieux différents
  - Reproductibilité du jugement
    - Variabilité inter et intra-observateur
- Coef ficient d'agrément Kappa

# Coef f i c i e n t d' a g r é m e n t K a p p a

- I n d e x d' a g r é m e n t e n t r e d e u x o b s e r v a t e u r s
- V a r i e d e - 1 à + 1
- V a u t 0 l o r s q u e l a c o n c o r d a n c e d e r é p o n s e e s t d u e a u h a s a r d
- T e n d v e r s - 1 : d i s c o r d a n c e d' i n t e r p r é t a t i o n  
T e n d v e r s + 1 : c o n c o r d a n c e d' i n t e r p r é t a t i o n

# Coef f i c i e n t d' a g r é m e n t K a p p a

Obs e r v 2

		Pr é s e n t	A b s e n t	T o t a l
Obs e r v 1	Pr é s e n t	a	b	$p1=a+b$
	A b s e n t	c	d	$q1=c+d$
	T o t a l	$p2=a+c$	$q2=b+d$	$n=a+b+c+d$

$p0$  = pr o p o r t i o n d' a g r é m e n t o b s e r v é =  $(a + d) / n$

$pe$  = pr o p o r t i o n d' a g r é m e n t a t t e n d u =  $[(p1 \times p2) + (q1 \times q2)] / n$

Kappa =  $(p0 - pe) / (1 - pe)$

I n c o n v é n i e n t = d é p e n d d e l a p r é v a l e n c e d e l a m a l a d i e

# 5 critères de qualité d'un test

- Bonne qualité technique
- Interpretation reproductible
- **Information valide**
- Utile à la prise de décision
- Efficace pour le malade et la population

# Information valide

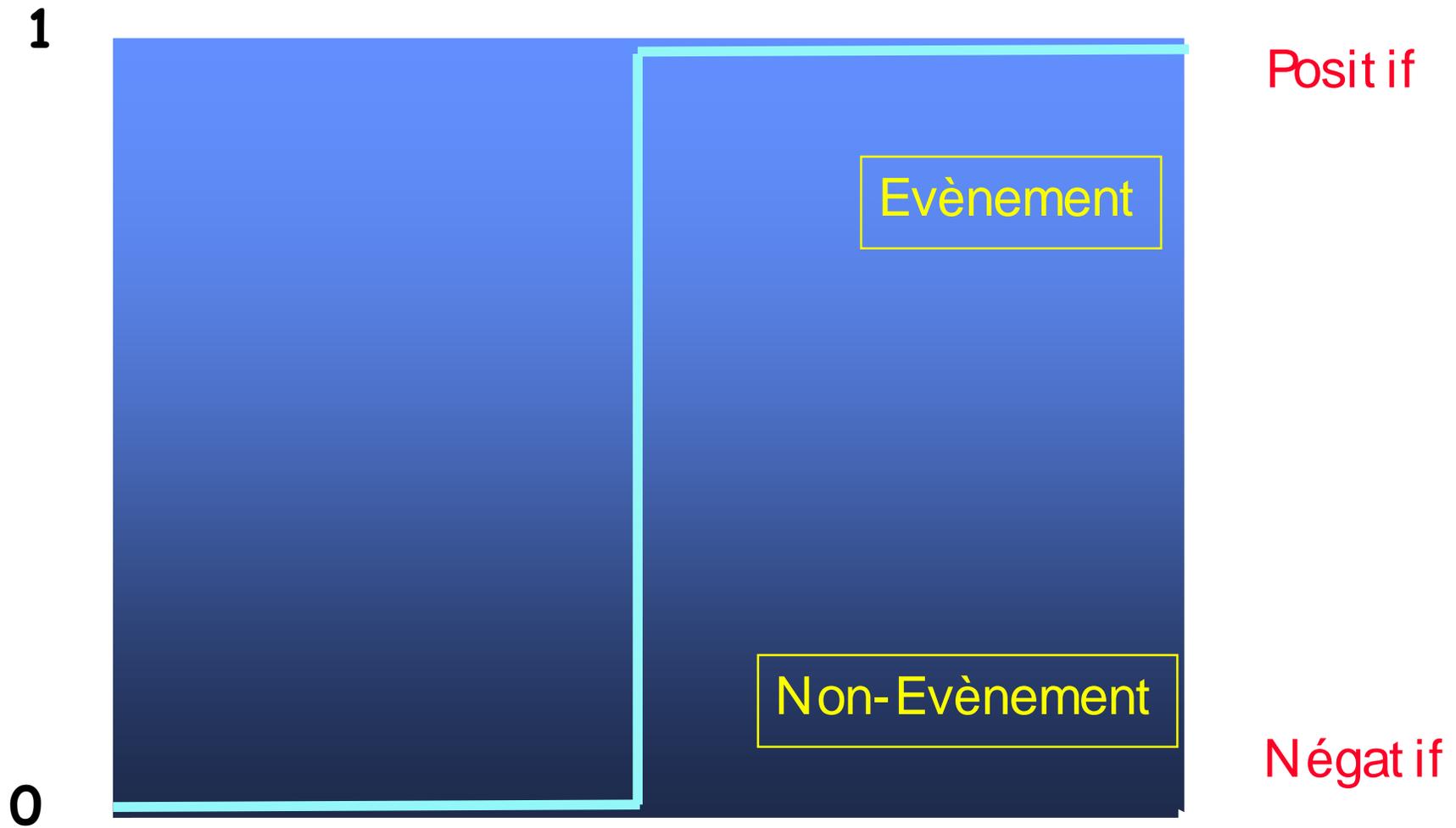
- Problème de la valeur-seuil d'un examen
- Examen clinique donne une estimation des "chances" qu'a le sujet d'être atteint de la maladie
- Examen complémentaire : modifie les probabilités afin d'estimer en fonction des résultats le risque d'être malade ou pas
  - Dépend des qualités intrinsèques du test (sensibilité et spécificité)
  - Dépend de la prévalence de la maladie (probabilité primaire ou à priori)

# Information valide

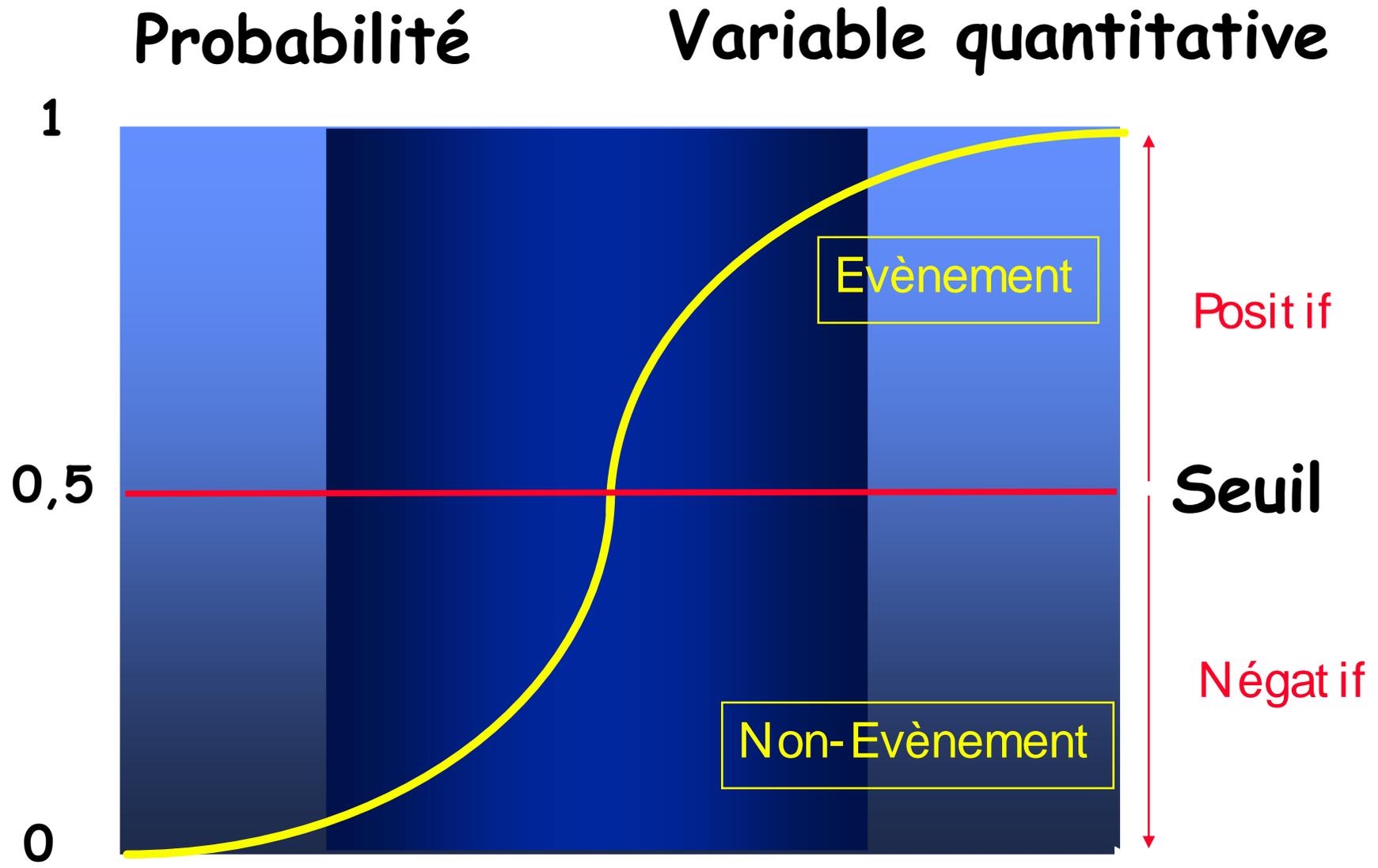
- D'où la nécessité d'être très précis sur
  - la définition du signe
  - La définition des patients
  - La définition des malades et non-malades

# Définition du signe

## Variable qualitative



# Définition du signe



# Définition des patients

- Population cible

= Population susceptible de bénéficier de cet examen

# Définition des malades-non malades

- Examen comparé avec le diagnostic réel  
Gold standard , diagnostic bét on...
- Examen de référence ou anapath ou consensus!!!
- L'examen de référence peut lui même être entaché d'erreur
- L'examen de référence doit être indépendant de l'examen évalué et le résultat non connu

# Expression des résultats d'une évaluation

Signe binaire (variable qualitative)

Sensibilité

Spécificité

Signe à valeur continue (variable quantitative)

Courbes ROC

# Tableau de contingence

	<b>Malades</b>	<b>Sains</b>
<b>Test +</b>	<b>a</b>	<b>b</b>
<b>Test -</b>	<b>c</b>	<b>d</b>

Prévalence de la maladie : Malades/ Cohorte

$$p = a + c / a + b + c + d$$

# Tableau de contingence

	Malades	Sains
Test +	VP	FP
Test -	FN	VN

**Sensibilité =  $VP / (VP + FN)$**

**prop de résultats anormaux chez malades**

**Spécificité =  $VN / (VN + FP)$**

**prop de résultats normaux chez non-malades**

# Se Sp Avant ages/ inconvénient s

$$\text{Sensibilité} = \frac{VP}{VP + FN}$$

$$\text{Spécificité} = \frac{VN}{VN + FP}$$

Mesures indépendantes de la prévalence

Utilisables pour dépistage (prévalence faible)

ou diagnostique (prévalence élevée)

Pas utilisable en tant que tel

Le résultat de l'examen modifie-t-il la prob de maladie ?

Donc utiliser des valeurs prédictives

# Tableau de contingence

	Malades	Sains
Test +	VP	FP
Test -	FN	VN

$$VPP = VP / VP + FP$$

Prob qu'un sujet soit malade si examen +

$$VPN = VN / VN + FN$$

Prob qu'un sujet soit non-malade si examen -

$$VPP = VP / VP + FP$$

$$VPP = Se / [Se + (1 - Sp) \times (1 - p) / p]$$

$$VPN = VN / VN + FN$$

$$VPN = Sp / [Sp + (1 - Se) \times p / (1 - p)]$$

Dépendent de  $Se$ ,  $Sp$  et  $p$  de la maladie

Valeurs prédictives calculées valables que dans un environnement où la prévalence est identique à celle de l'étude

# ROLE DE LA PREVALENCE

TEST     $Se = 90 \%$ ;  $Sp = 90 \%$

Pr évalence

VPP

VPN

9 %

47 %

99 %

50 %

90 %

90 %

90 %

99 %

99 %

# Intervalles de confiance (I C)

Tests réalisés sur des échantillons de population et non pas sur la population totale

Établir I C en choisissant un risque d'erreur que la valeur réelle dans la population ne soit pas dans I C de la valeur déterminée sur l'échantillon .

Par convention, un intervalle de confiance à 95% est utilisé

# Intervalles de confiance (IC)

$$\text{IC}_{95\%(\text{Se})}: \text{Se} \pm 1.96 * \sqrt{(\text{Se} * (1 - \text{Se}) / m)}$$

$$\text{IC}_{95\%(\text{Sp})}: \text{Sp} \pm 1.96 * \sqrt{(\text{Sp} * (1 - \text{Sp}) / n)}$$

m= nombre de malades

n= nombre de non-malades

IC 99% remplacer 1,96 par 2,576

IC 99,9% remplacer 1,96 par 3,290

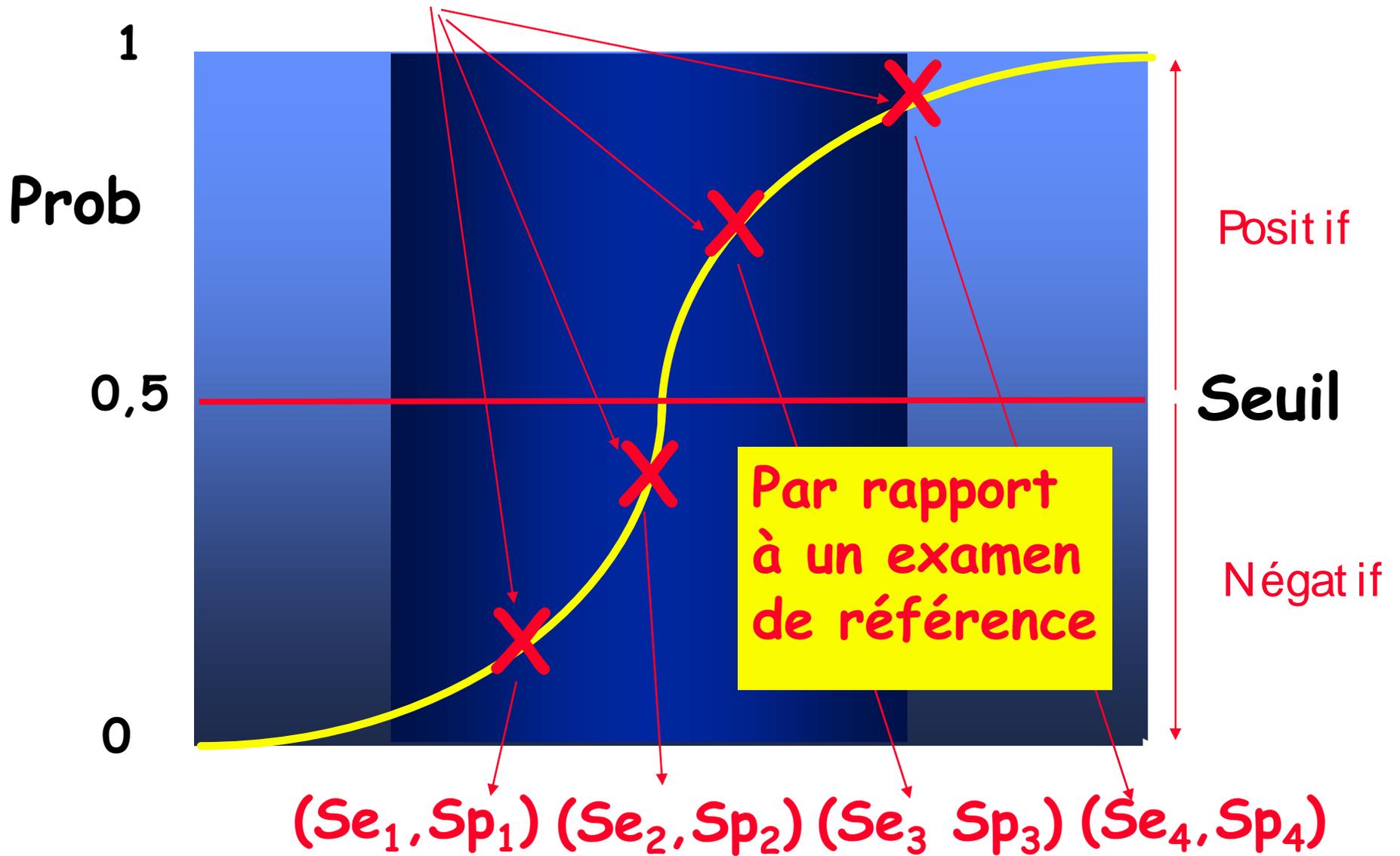
malades/ Total	Se (%)	IC 95 %
10/ 20	90 %	71,4-100
100/ 200	90 %	84,1-95,9
1000/ 2000	90 %	88,1-91,9

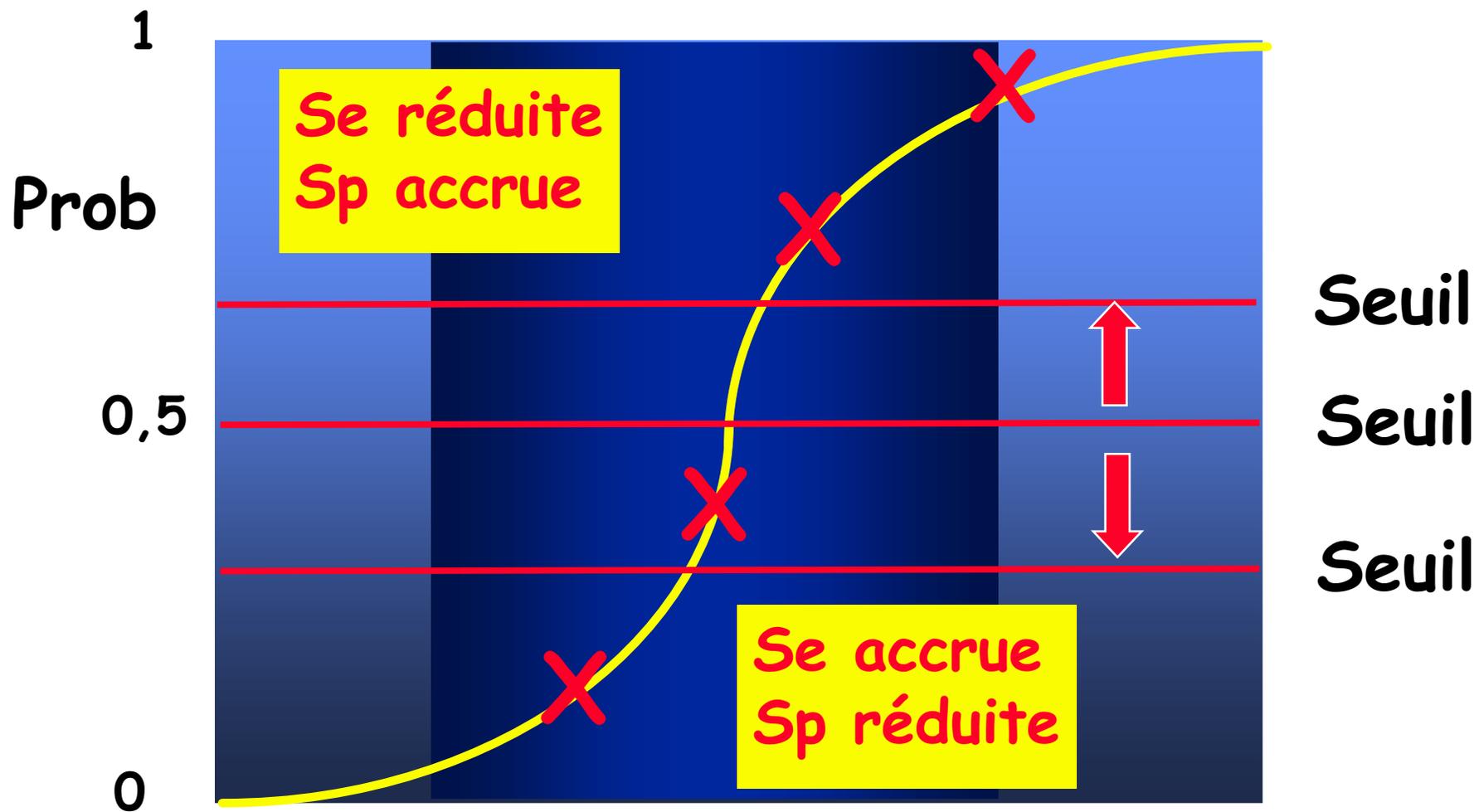
Cas des variables quantitatives

COURBES ROC

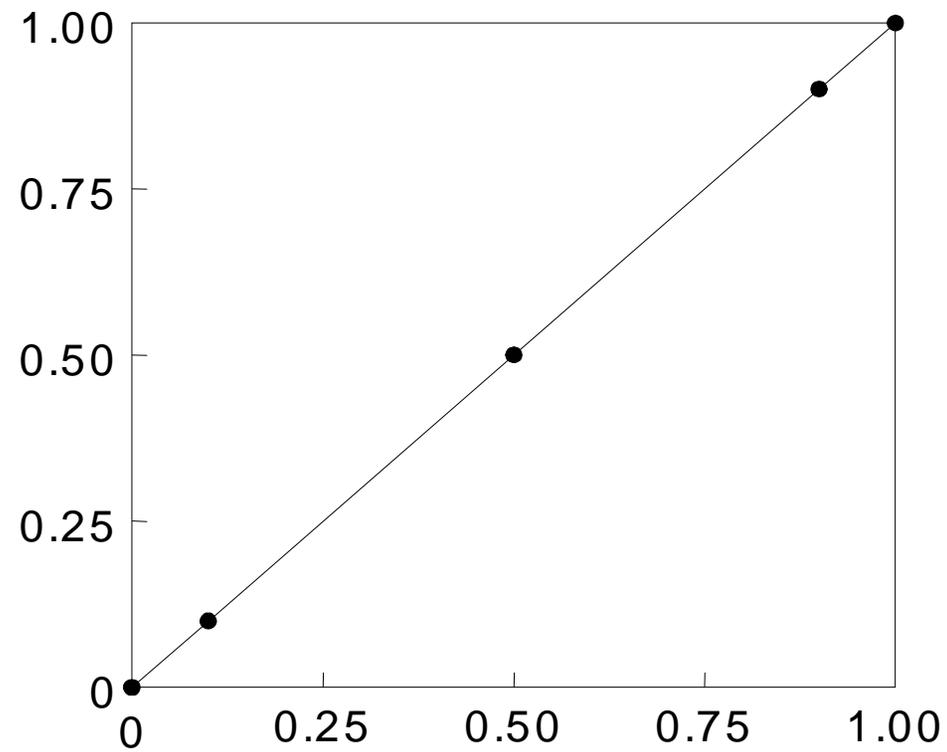
Receiver Operating Characteristic Curve

# Valeurs seuils arbitraires

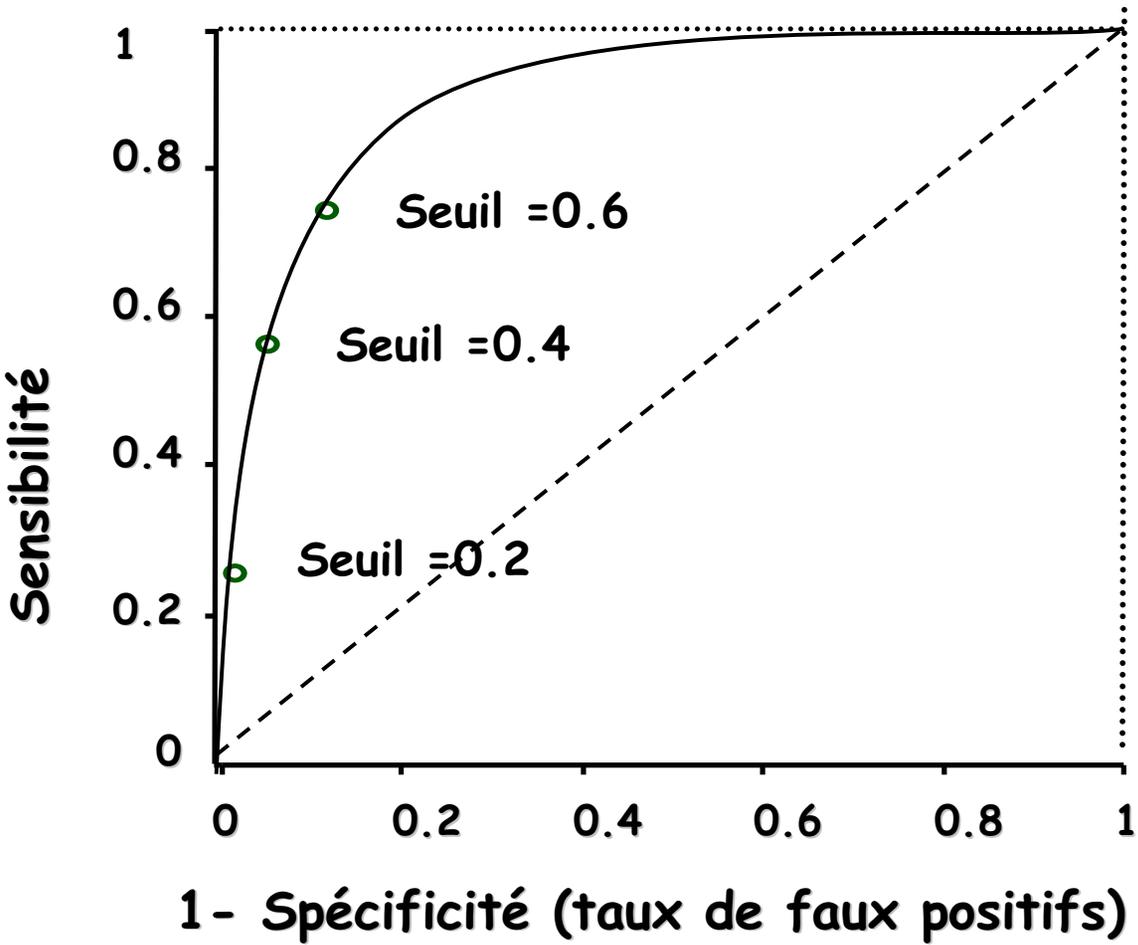




Sensibilit é (Vrais posit if s)



1-Spécif icit é (Faux posit if s)



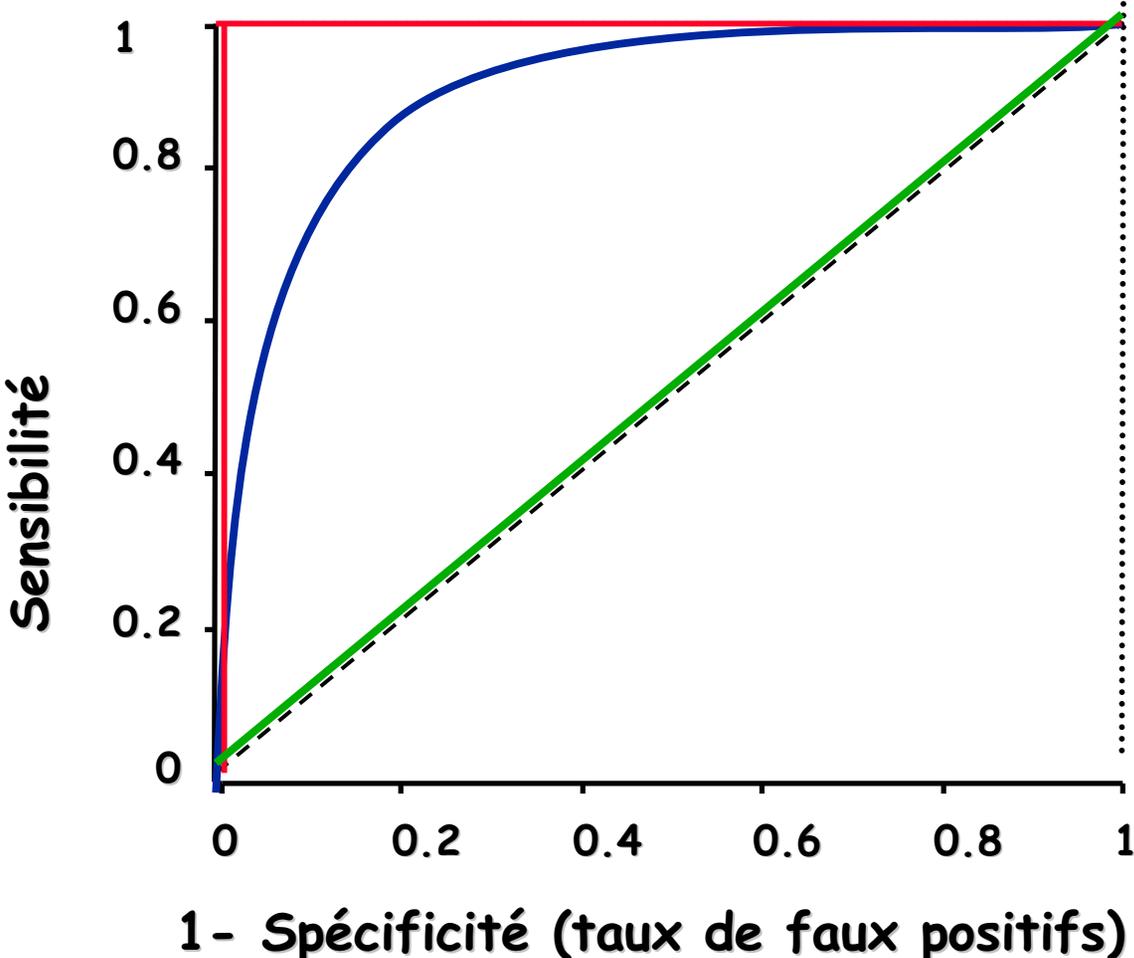
	M+	M-	
$p_1 > S$	VP	FP	Seuil = 0.2
$p_1 \leq S$	FN	VN	
	M+	M-	
$p_1 > S$	VP	FP	Seuil = 0.4
$p_1 \leq S$	FN	VN	
	M+	M-	
$p_1 > S$	VP	FP	Seuil = 0.6
$p_1 \leq S$	FN	VN	

# Avantages des courbes ROC

Pas d'hypothèse sur le seuil

Analyse globale des VP et VN

Index habituel d'évaluation du test AUC



Discrimination parfaite

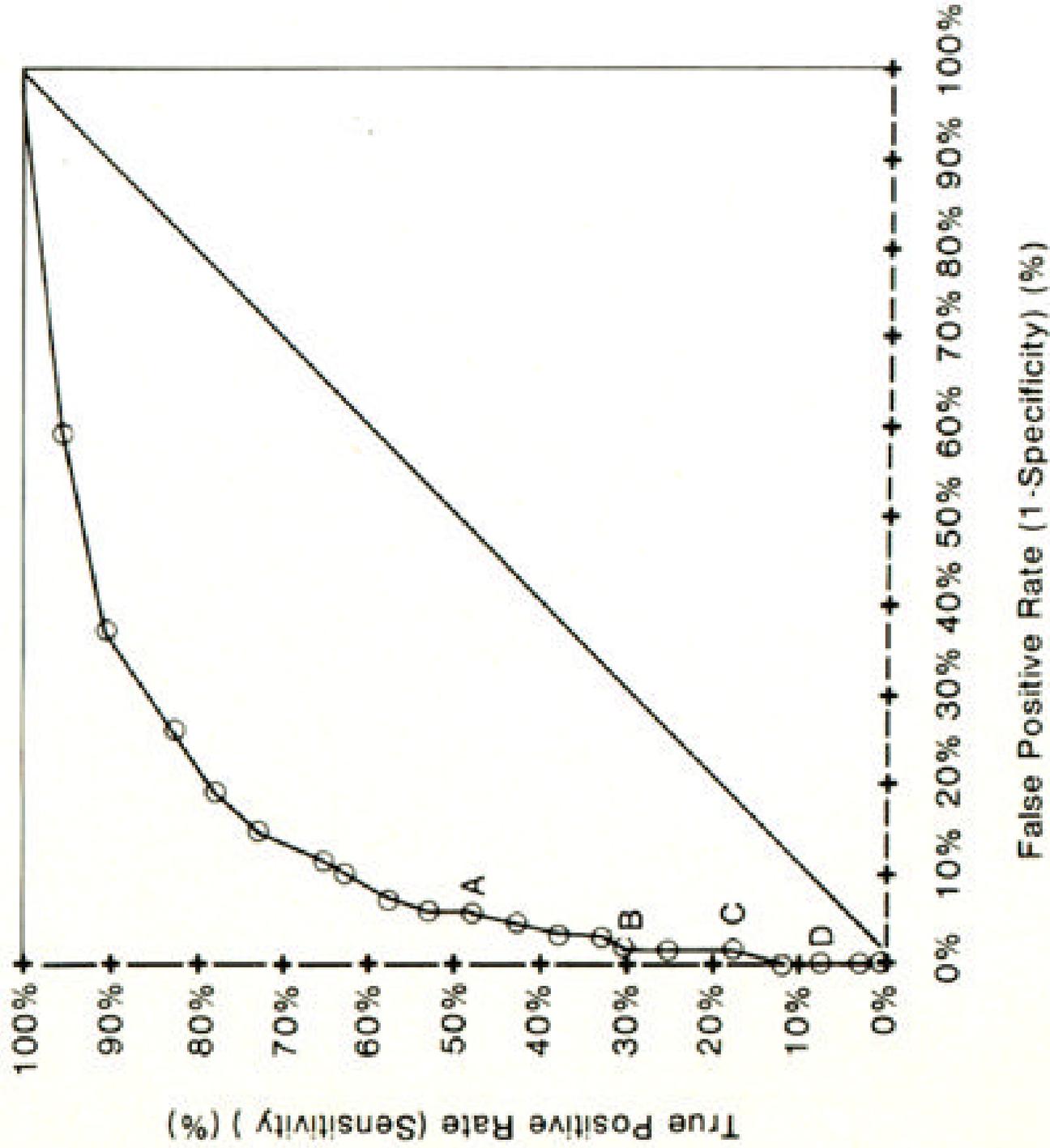
Pratique clinique

Pas de discrimination

Plus l'aire est grande

Plus le test apporte  
d'information

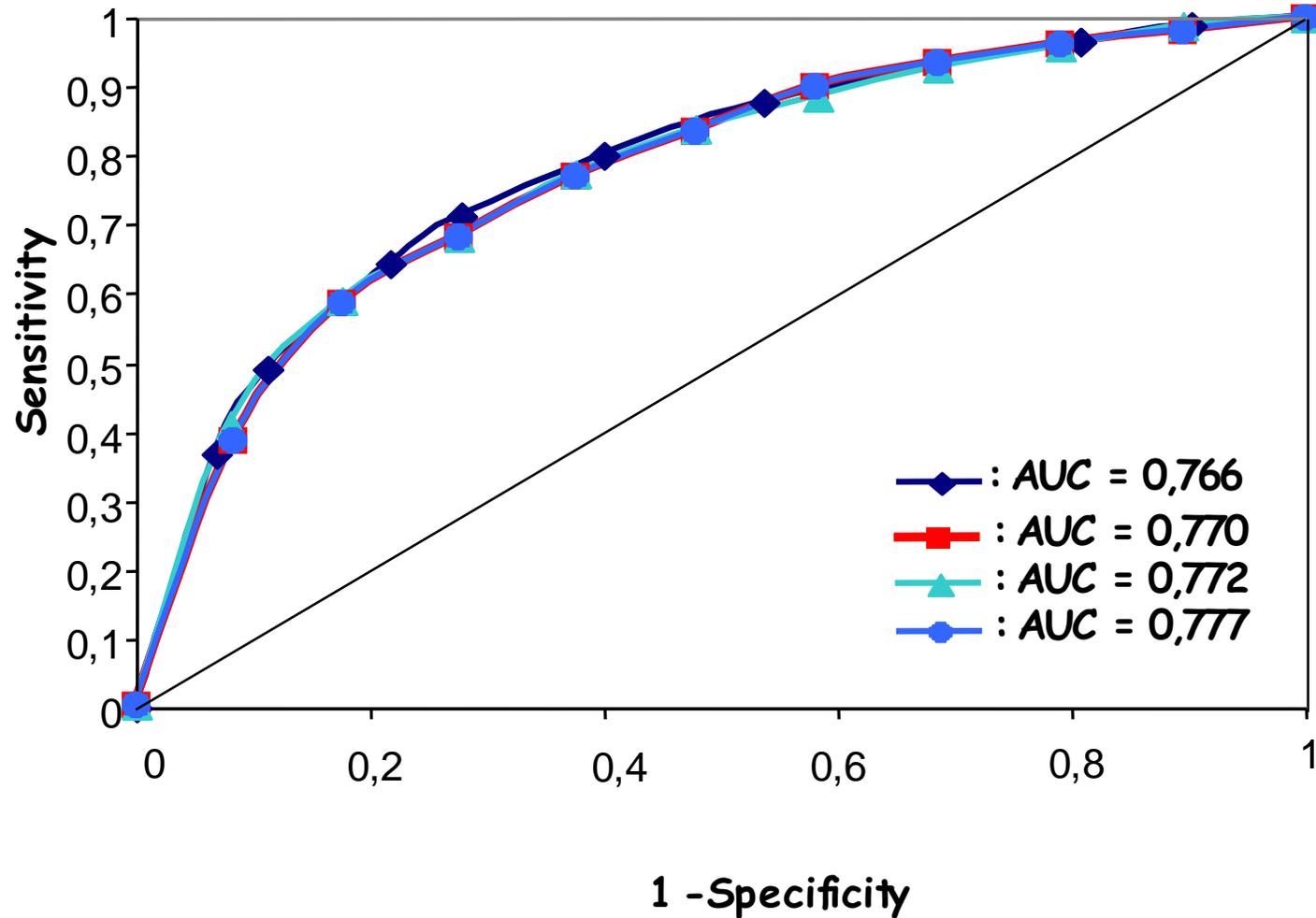
Quelque soit le seuil choisi



A method of comparing the areas under receiver operating characteristics curves derived from the same cases

Hanley JA, McNeil BJ Radiology 1983; 148: 838-43

## Predictive performance of three methods on validation sample



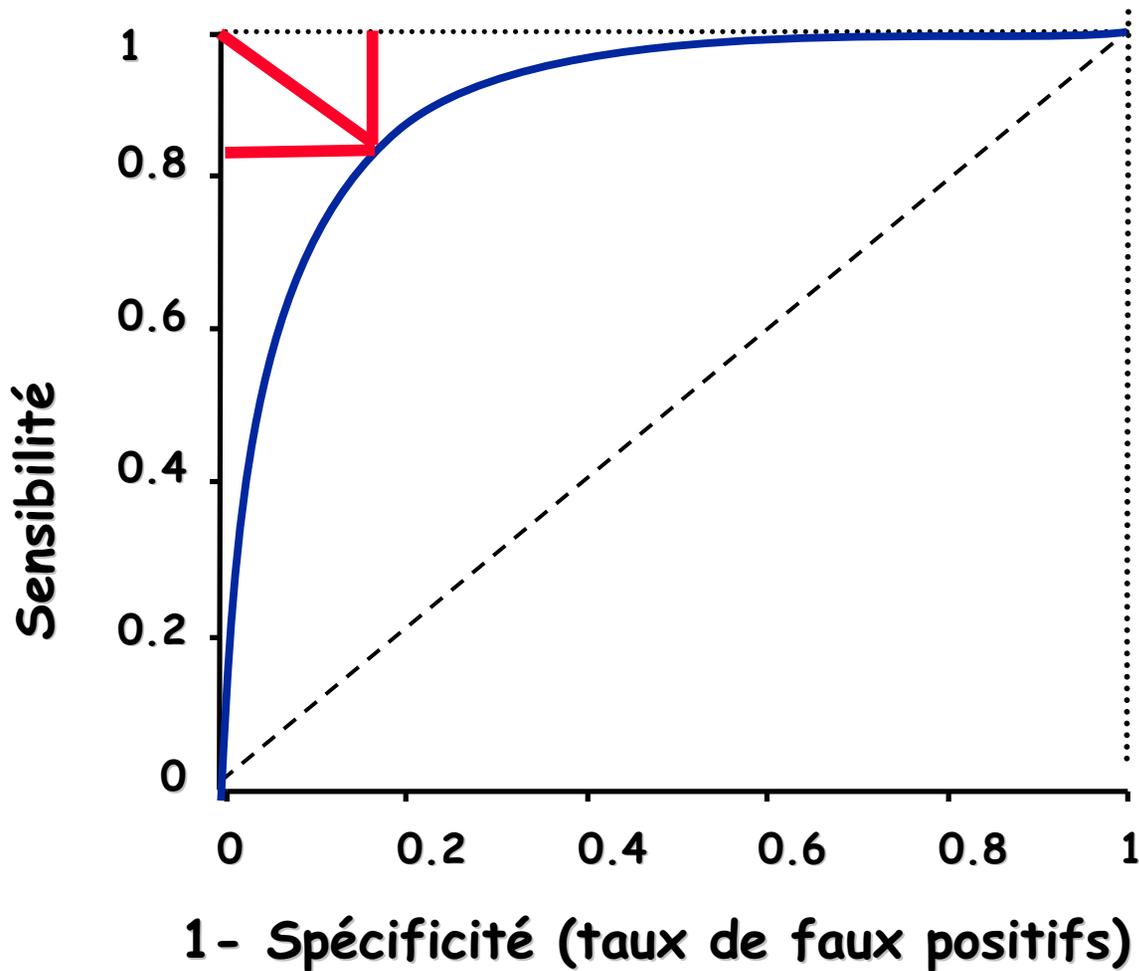
# Choix du seuil

Dépend de l'utilisation de l'examen

Privilégier la sensibilité = Dépistage

Privilégier la spécificité = Diagnostic de certitude

# Choix du seuil



Minimiser la distance par rapport au point idéal:  $Se=Sp=1$

# 5 critères de qualité d'un test

- Bonne qualité technique
- Interprétation reproductible
- Information valide
- Utile à la prise de décision
- Efficace pour le malade et la population

# QUELLE PRIORITE ?

Dépister les patients malades (Sensibilité)

Exclure les patients non malades certitude diagn  
(Spécificité)

Exactitude de la mesure (Bien classés)

# Indices de syntaxe

Visent à établir la valeur d'un test :

Exactitude

Test de Youden

Rapport de vraisemblance

# Exactitude (E)

	Malades	Sains
Test +	VP	FP
Test -	FN	VN

**Exactitude = Bien classés =  $VP + VN / VP + VN + FP + FN$**

% résultats exacts, cad VP et VN par rapport à la cohorte

Varie de 0 à 1

1 correspondant à une absence de FP et de FN

Un tirage au sort correspond à 0,5

## Indice de Youden (J)

Addition des 2 qualités du test : Se et Sp

$$J \text{ (en \%)} = Se + Sp - 100$$

Varie de -1 à 1

0 = pas d'orientation diagnostique

Cet indice est peu utilisé

# Rapport de vraisemblance (LR)

Les **rapports de vraisemblance** ("*likelihood ratios*") sont des indices qui facilitent le calcul de la probabilité post test de maladie.

Ils correspondent, pour un résultat d'examen donné, au rapport de la probabilité d'être malade sur la probabilité de ne pas l'être.

Il varie donc selon que le résultat de l'examen diagnostique est positif ou négatif.

## 28 ans, douleur thoracique

Pr é-t est	Scint igr aphie	Post -t est
0,20	Pr obabilit é élevée	0,82
0,20	Pr obabilit é int er médiair e	0,23
0,20	Pr obabilit é f aible	0,08
0,20	Nor male	0,02

# 78 ans, douleur thoracique PO

Pr é-t est	Scint igr aphi e	Post -t est
0,70	Pr obabilit é é levée	0,97
0,70	Pr obabilit é int er médi air e	0,74
0,70	Pr obabilit é f aible	0,46
0,70	Nor male	0,19

# Rapport de vraisemblance

L'ampleur du changement de la probabilité de maladie que l'on avait avant le test par rapport à celle que l'on obtient après le test varie selon le rapport de vraisemblance.

Plus le rapport de vraisemblance d'un test positif est grand et plus le rapport de vraisemblance d'un test négatif est petit, plus le test est utile cliniquement.

# Rapport de vraisemblance

Le rapport de vraisemblance d'un test positif

$$L = \text{Se} / (1 - \text{Sp})$$

Varie de 0 (Se nulle) à l'infini (Sp tend vers 1)

Plus LR s'écarte de la valeur 1, (plus le rapport L est élevé, plus il nous permet de confirmer la maladie)

# Rapport de vraisemblance

Le rapport de vraisemblance d'un test négatif

$$\text{LR}_- = (1 - \text{Se}) / \text{Sp}$$

Rapport négatif nul quand  $\text{Se} = 1$

Test totalement spécifique (pathognomonique),  $\text{LR}_-$  est égal à  $1 - \text{Se}$

Plus  $\text{LR}_-$  est petit, plus il permet d'exclure la maladie.

# Rapport de vraisemblance

---

L	—	Changement entre la probabilité est et post test d'avoir la maladie
$\geq 10$	$< 0,1$	Important
5-10	0,1-0,2	Modéré
2-5	0,2-0,5	Faible mais parfois significatif
1-2	0,5-1	Faible rarement significatif
1	1	Aucun

---

Pour résumer, la valeur diagnostique d'un test est d'autant plus grande que

- L'indice de Youden est proche de 1
- L'apport diagnostique d'un résultat positif du test est d'autant plus grand que  $L$  est plus élevé
- L'apport diagnostique d'un résultat négatif est d'autant plus grand que  $\bar{L}$  est petit et proche de 0

Quel est le meilleur examen ?

Sensibilité = 50 %; Spécificité = 50%

Sensibilité=10%; Spécificité = 90%

Sensibilité=90%; Spécificité=10%

# Quel est le meilleur examen ?

Se=50 %Sp=50%

	<b>Malades</b>	<b>Sains</b>
<b>Test +</b>	<b>50</b>	<b>50</b>
<b>Test -</b>	<b>50</b>	<b>50</b>

E      Y      L       $\bar{\phantom{L}}$   
0,5    0      1      1

# Quel est le meilleur examen ?

Se=10% Sp=90%

	<b>Malades</b>	<b>Sains</b>
<b>Test +</b>	10	10
<b>Test -</b>	90	90

E      Y      L       $\bar{\phantom{L}}$   
0,5    0      1      1

# Quel est le meilleur examen ?

Se=90 %Sp=10%

	<b>Malades</b>	<b>Sains</b>
<b>Test +</b>	90	90
<b>Test -</b>	10	10

E      Y      L       $\bar{\phantom{L}}$   
0,5    0      1      1