

Evaluation d'un test diagnostique

P. Staccini

Objectifs pédagogiques

- ▶ Argumenter la valeur informationnelle d'un test ou d'un signe
- ▶ Définir et être en mesure de calculer la spécificité, la sensibilité, les valeurs prédictives positive et négative et les rapports de vraisemblance d'un test
- ▶ Savoir tracer et interpréter une courbe ROC
- ▶ S'initier à l'analyse critique d'un article d'évaluation d'un test diagnostique

PLAN

- ▶ **Introduction**
- ▶ Validité interne d'un test
 - ▶ Sensibilité / Spécificité
 - ▶ Courbe ROC
- ▶ Apport décisionnel d'un test
 - ▶ Valeurs prédictives
 - ▶ Indices de synthèse

Introduction

Introduction

- ▶ Raisonnement clinique : suite d'étapes (démarche clinique) aboutissant à l'élaboration d'une décision
- ▶ Démarche clinique : suite d'observations et d'évaluations (interprétations)
- ▶ Exemple : mesure de la flèche hépatique
 - ▶ *La percussion permet de localiser la limite supérieure du lobe hépatique droit, généralement située à hauteur de la cinquième côte, en dedans de la ligne médio-claviculaire. Le bord inférieur du foie passe par une ligne oblique unissant le neuvième cartilage costal droit au huitième gauche.*
 - ▶ *La flèche hépatique est la hauteur de la matité mesurée sur la ligne médio-claviculaire droite, sa valeur normale varie entre 8 et 10 cm (**donnée objective = mesure**). Pour des valeurs supérieures on parle d'hépatomégalie (**donnée inférée = interprétation**)*

Introduction

- ▶ Interprétation clinique d'une mesure
 - ▶ assez simple dans le cas de la flèche hépatique : comparaison à une valeur de référence obtenue par analyse statistique d'une distribution des valeurs (courbe de gauss)
 - ▶ plus difficile pour poser un diagnostic et décider d'une thérapeutique dans un contexte donné :
 - ▶ de nombreux signes cliniques, de résultats d'examens complémentaires...
- ▶ Objectif : choisir le(s) signe(s) qui m'apportent le plus (valeur informationnelle)

Exemples

- ▶ Streptotest pour angine à Strepto Bêta
- ▶ ECG pour accident coronarien aigu
- ▶ Radiographie pour fracture osseuse
- ▶ Hémocult et cancer colo-rectal
- ▶ A chaque fois, (observer / mesurer / interpréter) un signe direct ou indirect (ou une combinaison de signes) qui (évoque / oriente / prouve) la présence d'une pathologie sous-jacente sans recourir à des examens plus invasifs, plus coûteux, à risque pour le patient...

PLAN

- ▶ Introduction
- ▶ **Validité interne d'un test**
 - ▶ **Sensibilité / Spécificité**
 - ▶ **Courbe ROC**
- ▶ Apport décisionnel d'un test
 - ▶ Valeurs prédictives
 - ▶ Indices de synthèse
- ▶ Précision, reproductibilité et concordance
- ▶ Initiation à l'analyse critique d'article

Validité interne d'un test diagnostique

Evaluer un test

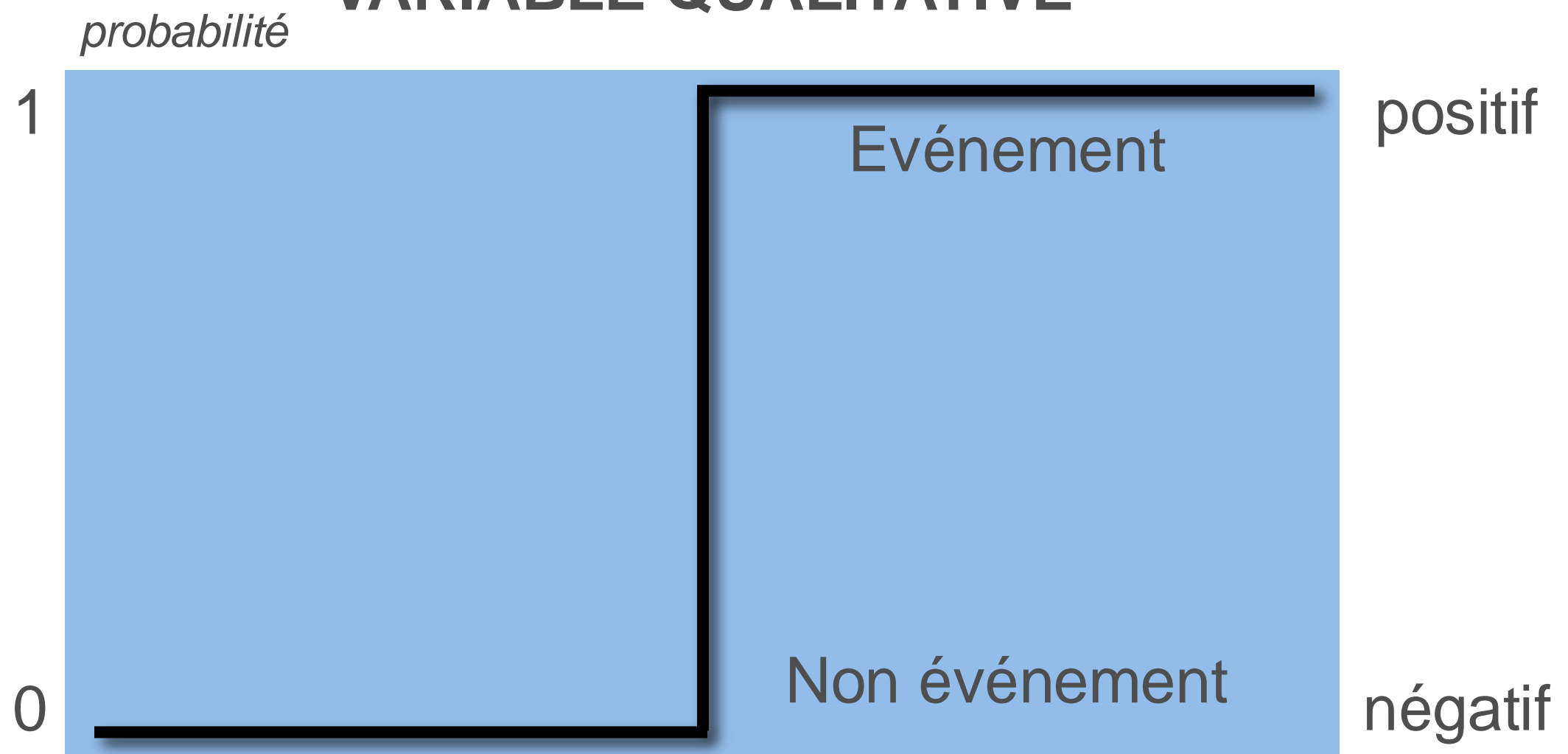
- ▶ Le test doit mesurer effectivement ce qu'on veut mesurer
- ▶ Il doit donc (identifier / différentier / discriminer) le mieux possible les malades et les sujets sains :
 - ▶ On vérifie d'abord que :
 - ▶ si un sujet est Malade → Test +
 - ▶ si un sujet est Sain → Test –
 - ▶ Pour pouvoir dire ensuite :
 - ▶ si le Test + → le sujet est Malade
 - ▶ si le Test - → le sujet est Sain

Evaluer un test

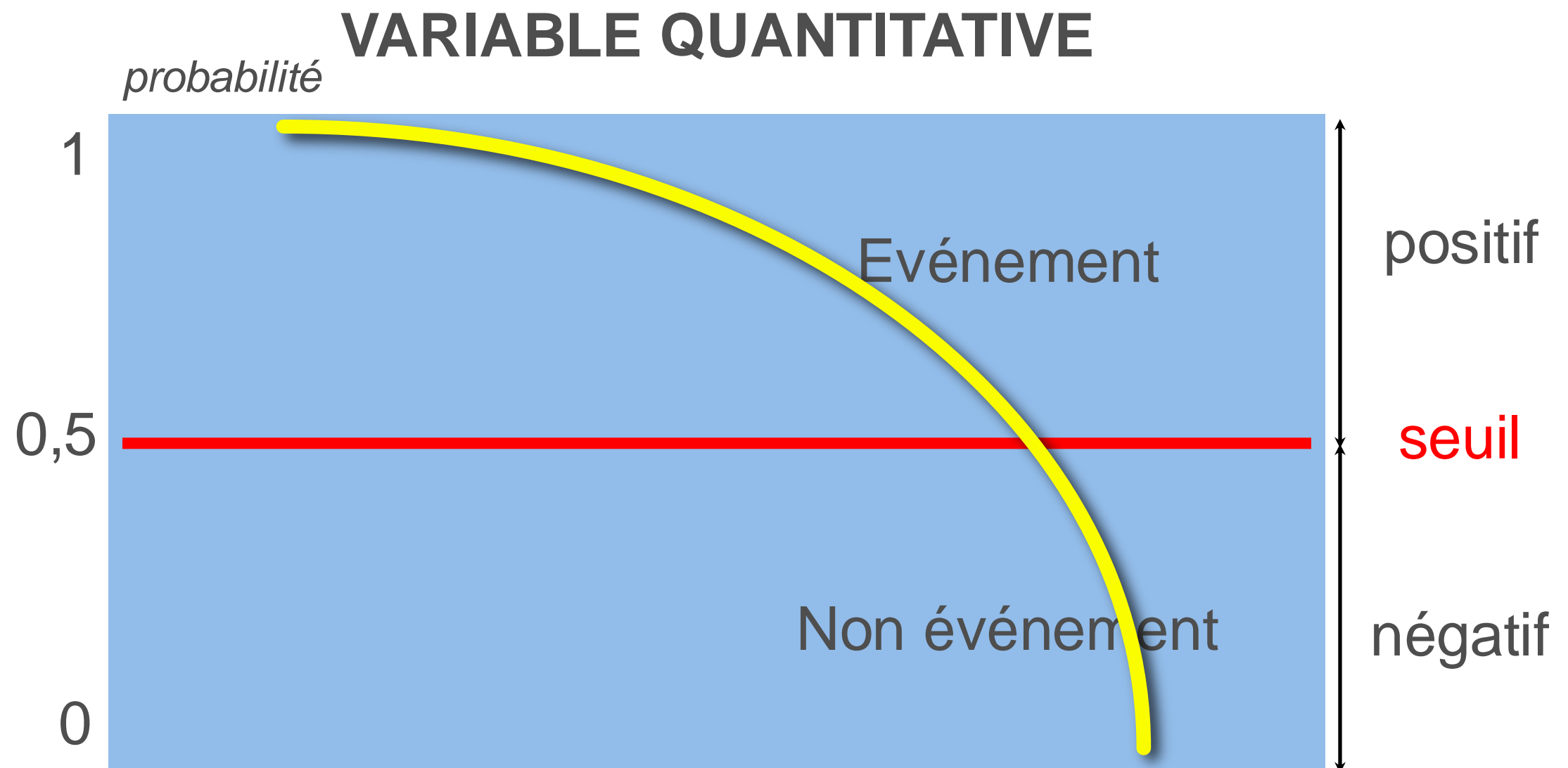
- ▶ Nécessité d'être très précis sur
 - ▶ La définition du signe
 - ▶ La définition des patients
 - ▶ La définition des malades et non-malades

Définition d'un signe

VARIABLE QUALITATIVE



Définition d'un signe



Expression des résultats

- ▶ Qualitatif
 - ▶ ordinal
 - ▶ absent / peu important / très important
 - ▶ binaire
 - ▶ absent / présent
 - ▶ positif / négatif
- ▶ Quantitatif (mesure, valeur continue)
 - ▶ x ml de ... / x g/l ...

Définition des patients

- ▶ Population cible = population susceptible de bénéficier de cet examen

Test de référence

- ▶ Résultat de (examen / test) comparé avec le diagnostic réel
 - ▶ “gold standard” (diagnostic béton)
- ▶ Examen de référence ou anatomo-pathologique ou consensus !!!
- ▶ L'examen de référence doit être indépendant de l'examen évalué et le résultat non connu (lorsqu'on analyse le résultat du test à évaluer)

*Cas d'un test
qualitatif*

Exemple

- ▶ Selvachandran SN, Hodder RJ, Ballal MS, Jones P, Cade D. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. *Lancet* 2002; 360: 278-83
- ▶ Population
 - ▶ Patients adressés pour une coloscopie suite à l'apparition de troubles colorectaux
- ▶ Test = questionnaire
 - ▶ Saignement rectal : absent / présent
- ▶ Gold-standard = examen histologique
 - ▶ Diagnostic de cancer suite à la coloscopie

Résultats bruts

		CANCER COLORECTAL (diagnostic histologique)	
		OUI	NON
SAIGNEMENT RECTAL	présent	82	1505
	absent	13	668

Questions

1. Quels sont les patients bien classés ? Quels sont ceux mal classés ?
2. Quelle est la fréquence de la maladie dans cette étude ?
3. Quelle est la proportion de patients malades qui ont un test positif ?
4. Quelle est la proportion de patients non malades qui ont un test négatif ?

Résultats bruts

		CANCER COLORECTAL (diagnostic histologique)	
		OUI	NON
SAIGNEMENT RECTAL	présent	82	1505
	absent	13	668

Réponses

1. Patients bien classés : 82 et 668 (diagonale principale du tableau)
2. Patients mal classés : 13 et 1505 (diagonale secondaire)
3. Proportion de patients bien classés : $(82+668) / (82+668+13+1505) = 33\%$

Résultats bruts

		CANCER COLORECTAL (diagnostic histologique)	
		OUI	NON
SAIGNEMENT RECTAL	présent	82 (VP) <i>vrais positifs</i>	1505 (FP) <i>faux positifs</i>
	absent	13 (FN) <i>faux négatifs</i>	668 (VN) <i>vrais négatifs</i>

Réponses

1. Patients bien classés : 82 (VP) et 668 (VN) (diagonale principale du tableau)
2. Patients mal classés : 13 (FP) et 1505 (FN) (diagonale secondaire)
3. Proportion de patients bien classés : **Exactitude** = 33%

Qualités du test

- ▶ Sensibilité (Se)
 - ▶ capacité à détecter tous les malades
 - ▶ probabilité d'avoir de test positif sachant que je suis malade
 - ▶ probabilité de tests positifs chez les malades
 - ▶ proportion de VP sur le nombre total de cas malades
 - ▶ nombre compris entre 0 et 1, exprimé en
- ▶ Qualité recherchée pour le dépistage

Qualités du test

▶ Spécificité (Sp)

- ▶ capacité à ne reconnaître que les non malades (capacité à écarter les sujets sains)
 - ▶ probabilité d'avoir le test négatif sachant que je suis sain
 - ▶ proportion de test négatifs chez les sujets sains
 - ▶ proportion de VN sur le nombre total de sujets sains
 - ▶ nombre compris entre 0 et 1, exprimé en pourcentage
- ▶ Qualité recherchée avant de débiter un traitement invasif

Résultats bruts

		CANCER COLORECTAL (diagnostic histologique)		TOTAL
		OUI	NON	
SAIGNEMENT RECTAL	OUI	82	1505	1587
	NON	13	668	681
TOTAL		95	2173	2268

Réponses

1. Prévalence de la maladie = $95 / 2268 = 4,2\%$

2. Proportion de patients malades qui ont un test positif = $82 / 95 = 86,3\% = \text{Se}$

• Proportion de patients non malades qui ont un test négatif = $668 / 2173 = 30,7\% = \text{Sp}$

Sensibilité - Spécificité

- ▶ Qualités / propriétés intrinsèques du test indépendantes du contexte d'application
- ▶ La sensibilité d'un examen pour une affection repose sur la définition de la population des "malades", et est donc caractéristique de la maladie et du signe.
- ▶ En particulier, elle n'est pas susceptible de varier d'un centre à l'autre (d'un service hospitalier spécialisé à une consultation de médecin généraliste, par exemple). Le même raisonnement peut s'appliquer à la spécificité, si on considère qu'elle repose aussi sur la définition de la maladie.

Exercice

- ▶ On veut évaluer les performances d'un test de dépistage de la toxoplasmose congénitale. On a effectué :
 - ▶ 58 prélèvements sur des enfants atteints de façon certaine de toxoplasmose. Parmi eux, le test a été positif dans 54 cas.
 - ▶ 125 prélèvements sur des enfants indemnes de façon certaine. Parmi eux, le test a été négatif dans 114 cas.
- ▶ Calculer la sensibilité et la spécificité du test.
- ▶ Calculer la prévalence de la pathologie dans cette étude.

Test parfait

		MALADIE	
		OUI	NON
SIGNE	présent	140	0
	absent	0	160

Signe pathognomonique

1. Sensibilité : 100% / détecte tous les malades (aucun faux négatif)
2. Spécificité : 100% / détecte tous les sains (aucun faux positif)

*Cas d'un test
quantitatif*

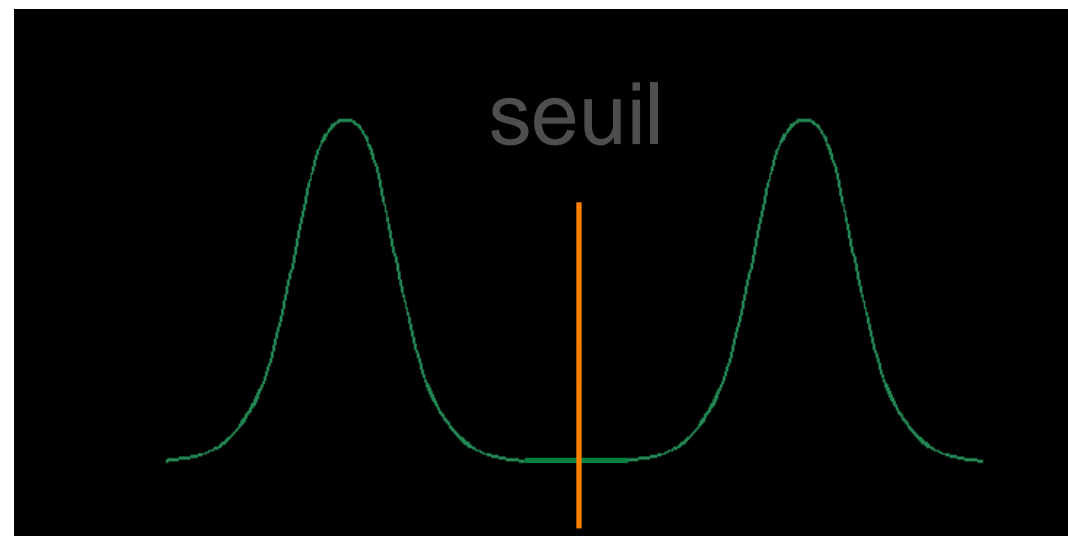
Problématique

- ▶ Résultat exprimé sous forme d'une valeur numérique.
- ▶ Valeurs différentes d'un sujet à l'autre.
- ▶ Nécessité de déterminer une valeur seuil permettant de classer les malades et les sujets sains.

Test discriminant

- ▶ distribution dans le groupe des cas bien séparée de distribution des valeurs dans un groupe de sujets sains.
- ▶ valeur seuil simple à déterminer.
- ▶ Mais situation rarement observée...

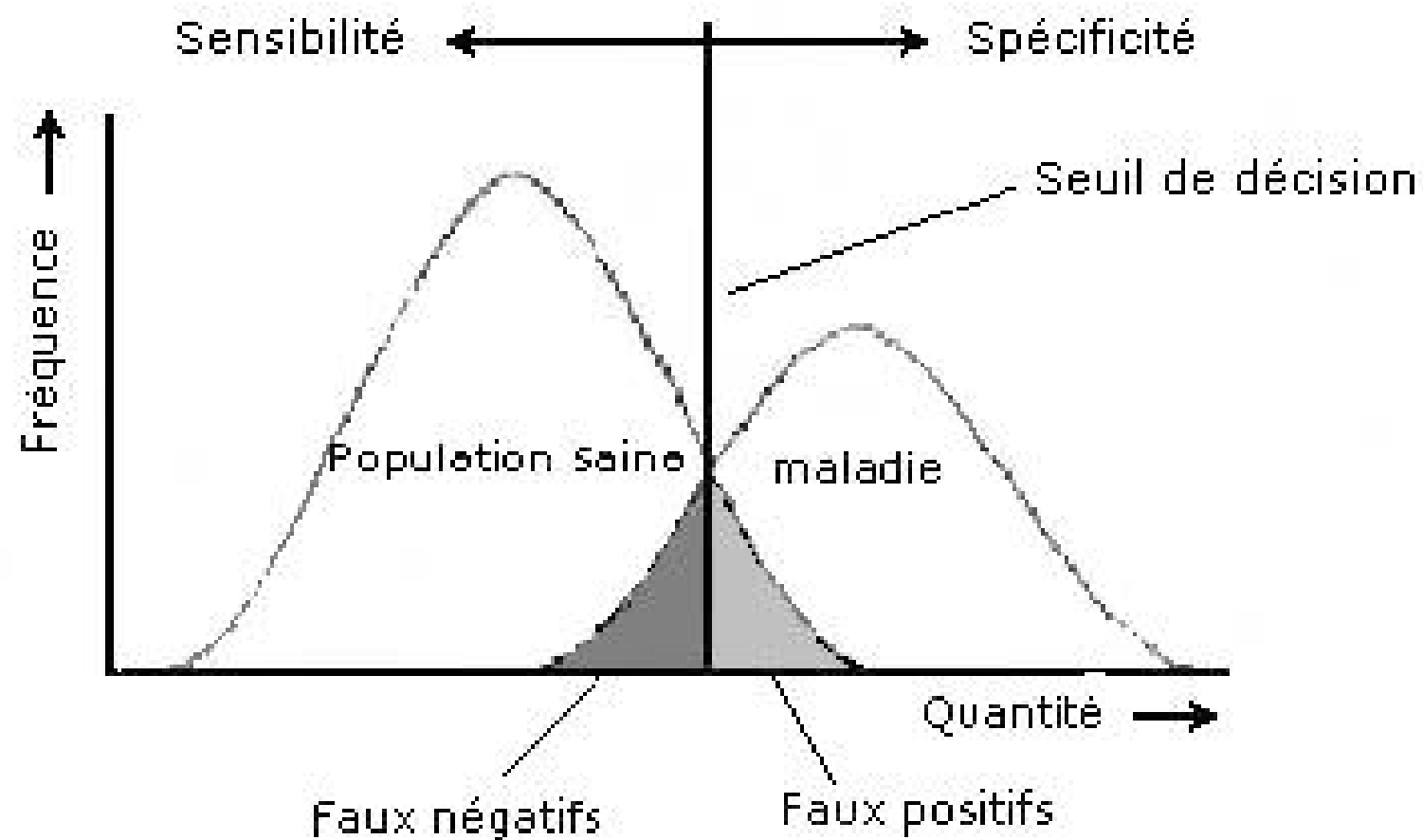
sains



malades

Le plus souvent...

- ▶ chevauchement des deux distributions



Choix du seuil

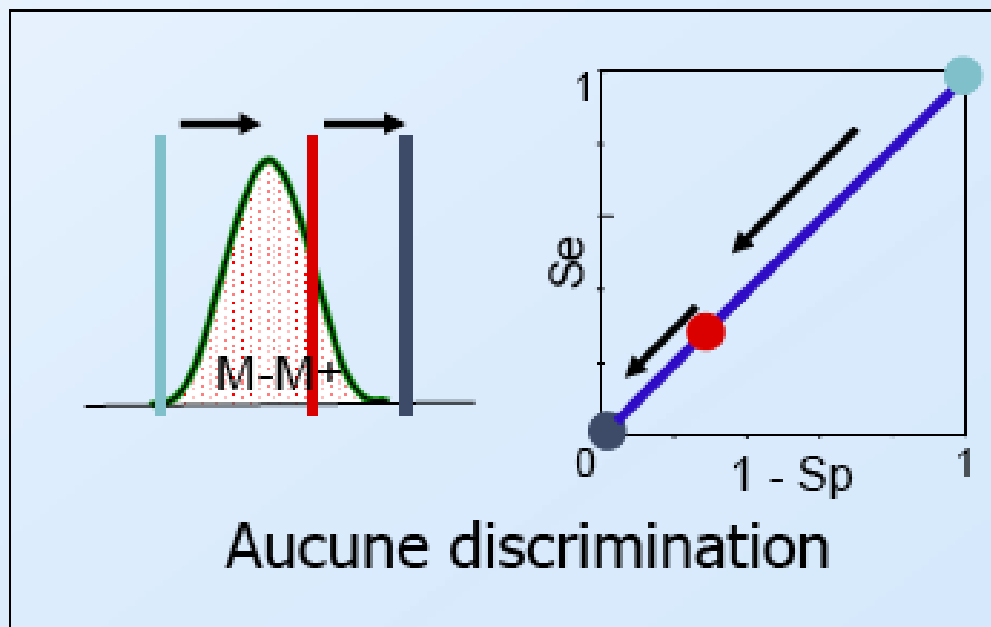
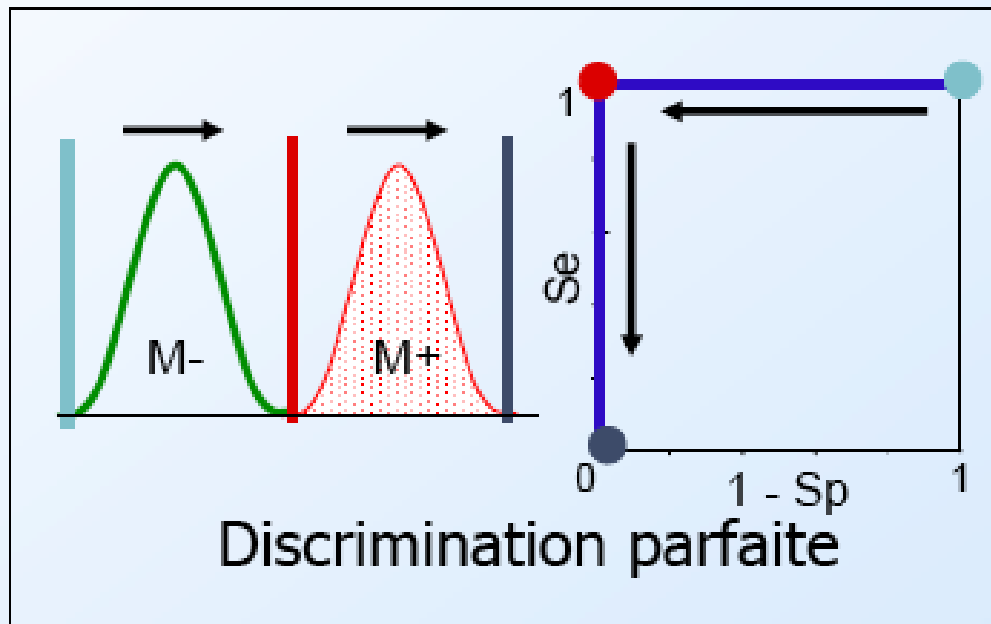
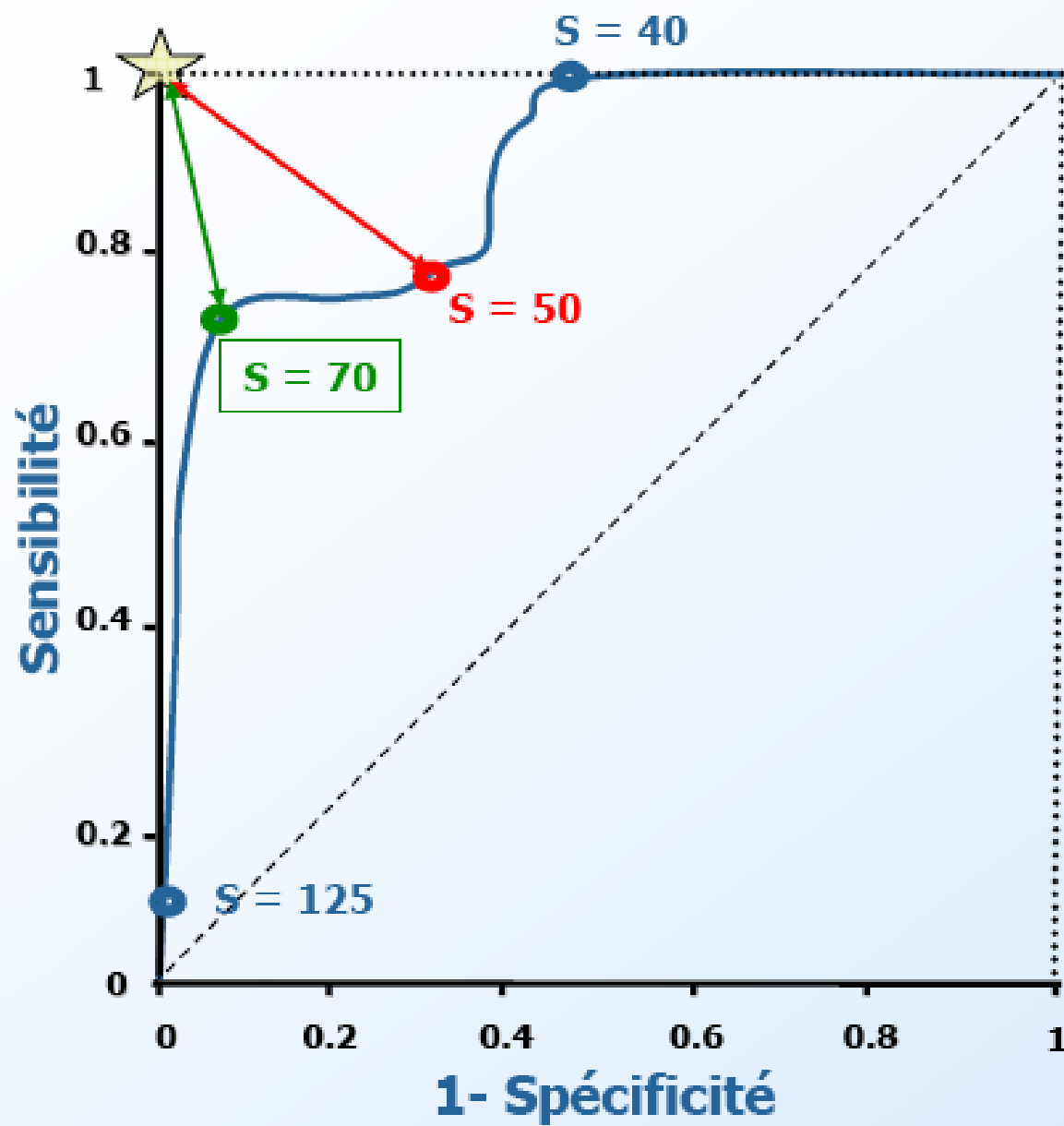
- ▶ Lorsque les erreurs par excès sont plus graves que les erreurs par défaut, on cherche à minimiser le nombre de faux positifs, donc privilégier la Sp.
- ▶ Il faut donc hausser le seuil de positivité.
- ▶ Exemple: dépistage anténatal de l'anencéphalie.
 - ▶ FP => conséquences lourdes (ITG)
 - ▶ FN => rattrapage ultérieur par échographie

Choix du seuil

- ▶ Lorsque les erreurs par défaut sont plus graves que les erreurs par excès, on cherche à minimiser le nombre de faux négatifs, donc privilégier la Se.
- ▶ Il faut donc baisser le seuil de positivité.
- ▶ Exemple : dépistage de la phénylcétonurie à la naissance
 - ▶ FN => développement de la maladie chez l'enfant
 - ▶ FP => prévention inutile qui pourra être corrigée ensuite

Courbe ROC

- ▶ Lorsque l'on cherche à fixer le seuil d'une méthode quantitative, on applique le test à un groupe de malades et un groupe de sujets sains.
- ▶ Pour chaque seuil possible, on calcule la Se et la Sp.
- ▶ On obtient donc une liste de couples Se-Sp.
- ▶ On peut ainsi dessiner un graphe appelé courbe ROC (*Receiver Operating Characteristics*), avec en ordonnée la sensibilité de chaque seuil (Se : vrais positifs) et en abscisse le pourcentage de faux positifs (1-Sp).

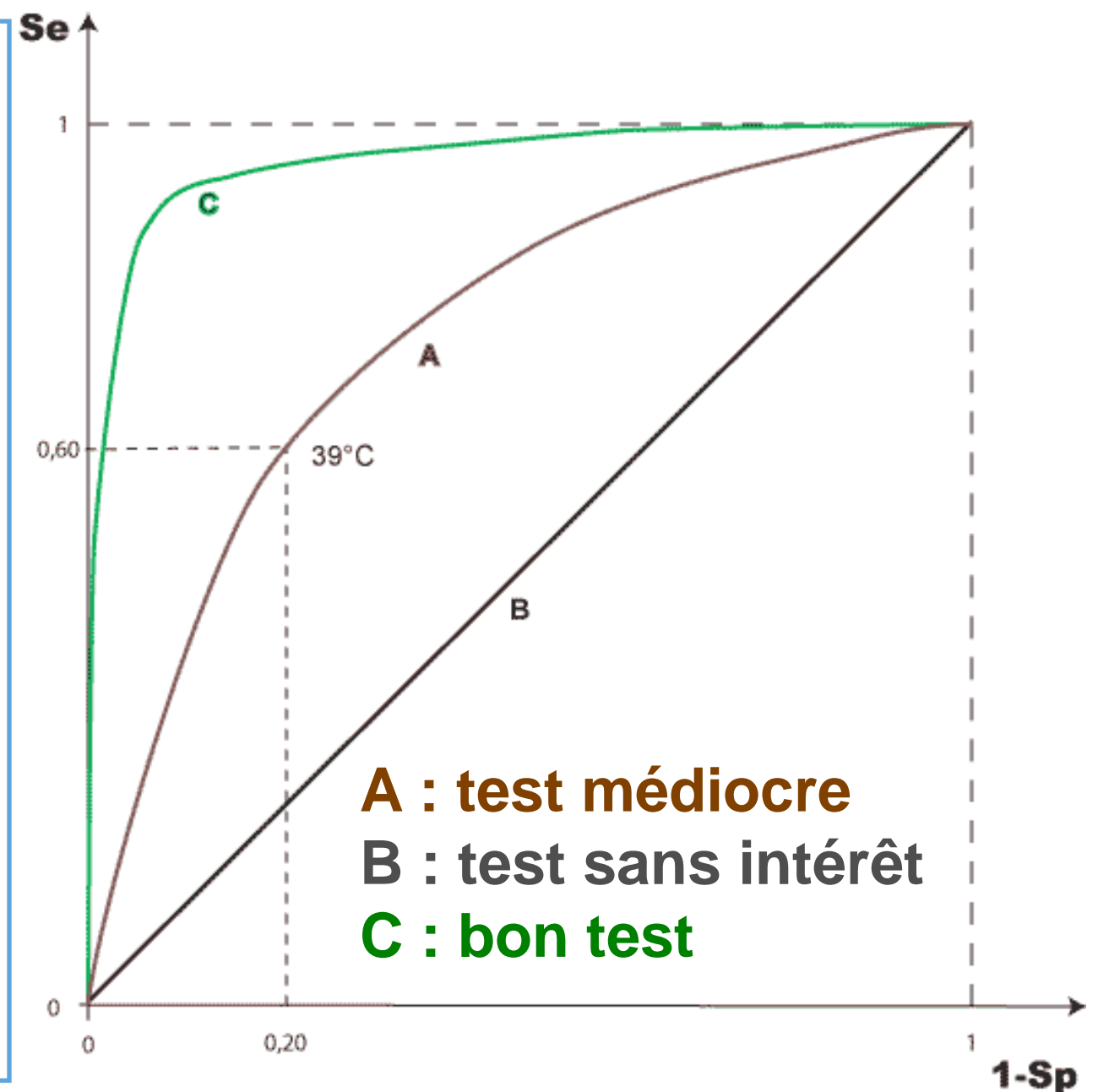


Courbe ROC

- ▶ L'aire sous la courbe ROC (ou AUC : *Area Under the Curve*) représente un indice global de la performance discriminante du test, quelque soit le seuil considéré.
- ▶ $AUC = 1$
 - ▶ Discrimination parfaite
- ▶ $AUC = 0,5$
 - ▶ Pas de discrimination
- ▶ Avantages des courbes ROC
 - ▶ Pas d'hypothèse sur le seuil
 - ▶ Analyse globale des VP et VN

Courbe ROC

- ▶ Une courbe ROC qui se rapproche de la diagonale illustre un test très peu discriminant, donc de qualité médiocre.
- ▶ Les Courbes ROC sont intéressantes pour comparer plusieurs techniques entre elles: la technique la plus discriminante est celle dont le graphe est le plus convexe dans le coin en haut à gauche.



Stratégies

- ▶ Dépend de l'utilisation de l'examen
 - ▶ Privilégier la sensibilité = Dépistage
 - ▶ Privilégier la spécificité = Diagnostic de certitude
- ▶ Quelle priorité ?
 - ▶ Dépister les patients malades (Sensibilité)
 - ▶ Exclure les patients non malades certitude diag (Spécificité)
 - ▶ Si on veut éviter d'inquiéter à tort les Faux Positifs (maladie grave) : privilégier Spécificité
 - ▶ Exactitude de la mesure (Bien classés)

PLAN

- ▶ Introduction
- ▶ Validité interne d'un test
 - ▶ Sensibilité / Spécificité
 - ▶ Courbe ROC
- ▶ **Apport décisionnel d'un test**
 - ▶ **Valeurs prédictives**
 - ▶ **Indices de synthèse**

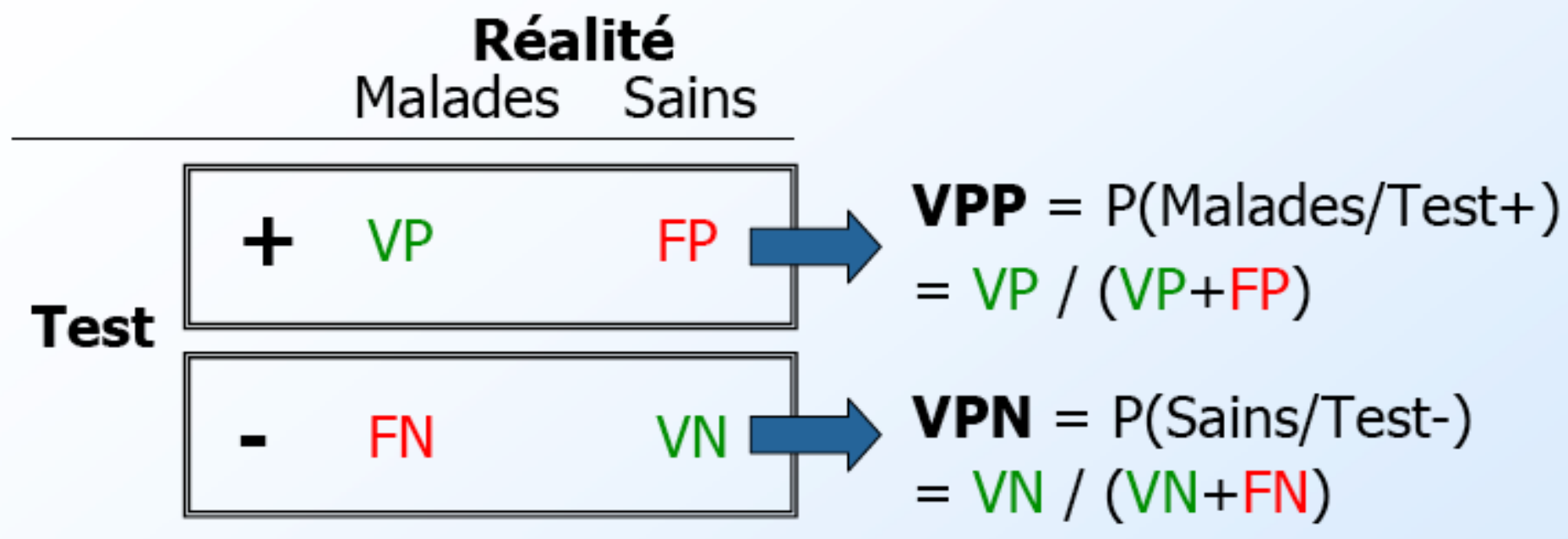
Apport décisionnel d'un test

Application au patient

- ▶ Sensibilité et Spécificité ne sont pas utilisables en tant que tel : ils définissent la probabilité d'avoir le test positif ou négatif selon le fait d'être malade ou pas.
- ▶ La question est donc inverse : si le test est positif, quelle est la probabilité que le patient soit atteint de la maladie ?

Valeurs prédictives

Valeurs prédictives



		Cancer	Sains	
Saignement rectal	+	82	1505	$\text{VPP} = 82 / (82 + 1505) = 5,2\%$
	-	13	668	$\text{VPN} = 668 / (668 + 13) = 98,1\%$

Valeurs prédictives

- ▶ $VPP = VP / (VP + FP)$
 - ▶ $VPP = Se / [Se + (1-Sp) \times (1-p) / p]$
- ▶ $VPN = VN / (VN + FN)$
 - ▶ $VPN = Sp / [Sp + (1-Se) \times p / (1-p)]$
- ▶ Valeurs prédictives dépendent de Se, Sp et p de la maladie
- ▶ Valeurs prédictives calculées ne sont valables / applicables que pour un patient extrait d'un environnement où la prévalence est identique à celle de l'étude

Rôle de la prévalence

Test	Se = 90%	Sp = 90%
Prévalence	VPP	VPN
9%	47%	99%
50%	90%	90%
90%	99%	99%

Rôle de la prévalence

N=2268
Prévalence
= 4,2%

		Cancer	Sains		
Saignement rectal	+	82	1505	➔	VPP = 5,2%
	-	13	668	➔	VPN = 98,1%
		↓	↓		
		Se = 86,3%	Sp = 30,7%		

N=2268
Pop. à risque
Prévalence
= 29,7%

		Cancer	Sains		
Saignement rectal	+	581	1105	➔	VPP = 34,5%
	-	92	490	➔	VPN = 84,2%
		↓	↓		
		Se = 86,3%	Sp = 30,7%		



Indices de synthèse

Indices de synthèse

Exactitude

Exactitude

- ▶ Exactitude = Bien classés
 - ▶ = $\frac{VP+VN}{VP+VN+FP+FN}$
 - ▶ % 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

Indices de synthèse

Indice de Youden

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

Indices de synthèse

Rapports de vraisemblance

Vraisemblance du test

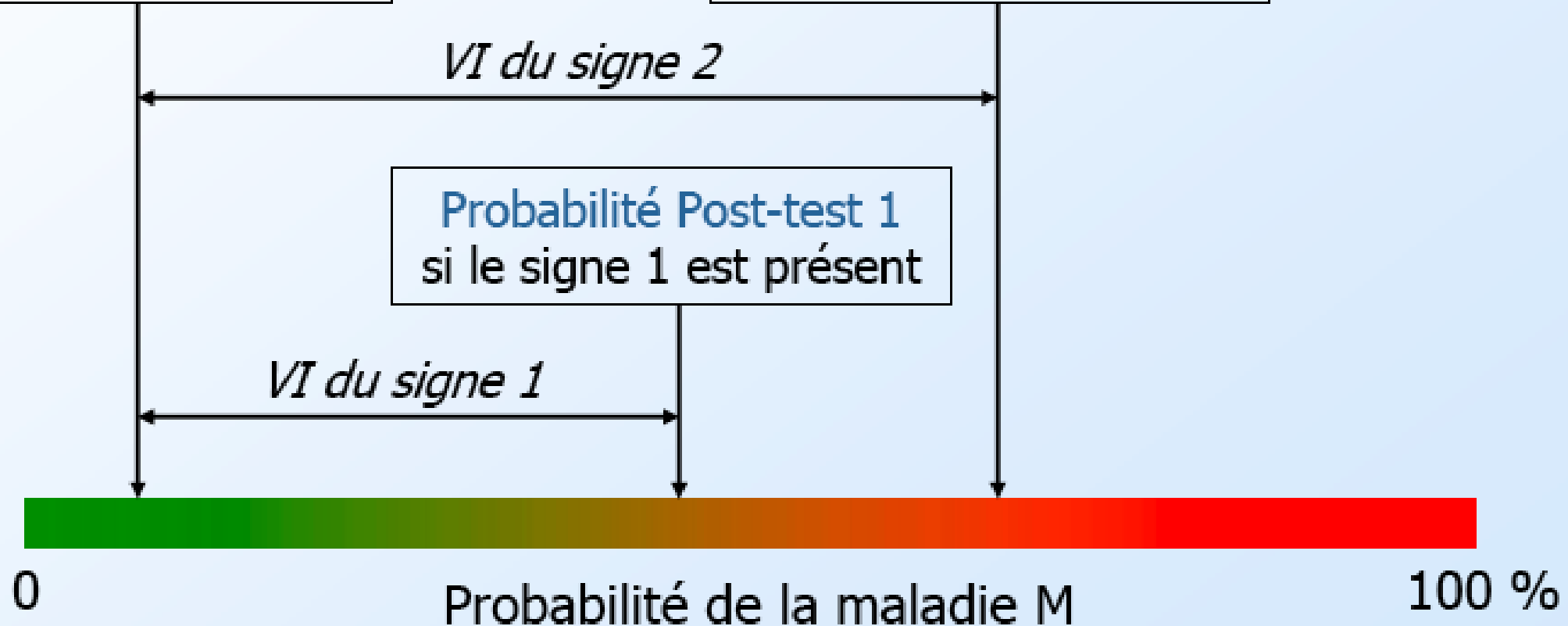
- ▶ Les rapports de vraisemblance (“likelihood ratios”) sont des indices qui facilitent le calcul de la probabilité post test de maladie.
- ▶ Ils correspondent, pour un résultat d'examen donné, au rapport de la probabilité d'être malade sur la probabilité de ne pas l'être.
- ▶ Il varie donc selon que le résultat de l'examen diagnostique est positif ou négatif.

Probabilités pré et post-test

Probabilité Pré-test
(prévalence de M
pour une population
donnée)

Probabilité Post-test 2
si le signe 2 est présent

Probabilité Post-test 1
si le signe 1 est présent



Vraisemblance du test

- ▶ 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.

Vraisemblance positive

- ▶ Rapport de vraisemblance positif
 - ▶ crédibilité d'un test positif
 - ▶ $L+ = P(T+/M+) / P(T+/M-) = Se / (1-Sp)$
- ▶ Varie de 0 ($Se=0$) à l'infini (Sp tend vers 1)
- ▶ Plus $L+$ s'écarte de la valeur 1, plus le rapport $L+$ est élevé, plus il nous permet de confirmer la maladie

Vraisemblance négative

- ▶ Rapport de vraisemblance négatif
 - ▶ crédibilité d'un test négatif
 - ▶ $L^- = P(T-/M+) / P(T-/M-) = (1-Se) / Sp$
- ▶ Rapport nul quand $Se = 1$
- ▶ Test totalement spécifique (pathognomonique), L^- est égal à $1-Se$ ($Sp = 1$)
- ▶ Plus L^- est petit (tend vers 0), plus il permet d'exclure la maladie.

Vraisemblance

- ▶ si LR+ ou LR- = 1 alors test sans intérêt, n'apporte rien

LR+	LR-	changement entre prob pré-test et post-test
≥ 10	$\leq 0,1$	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

Nomogramme de Bayes

examens d'imagerie pour la surveillance d'un cancer traité

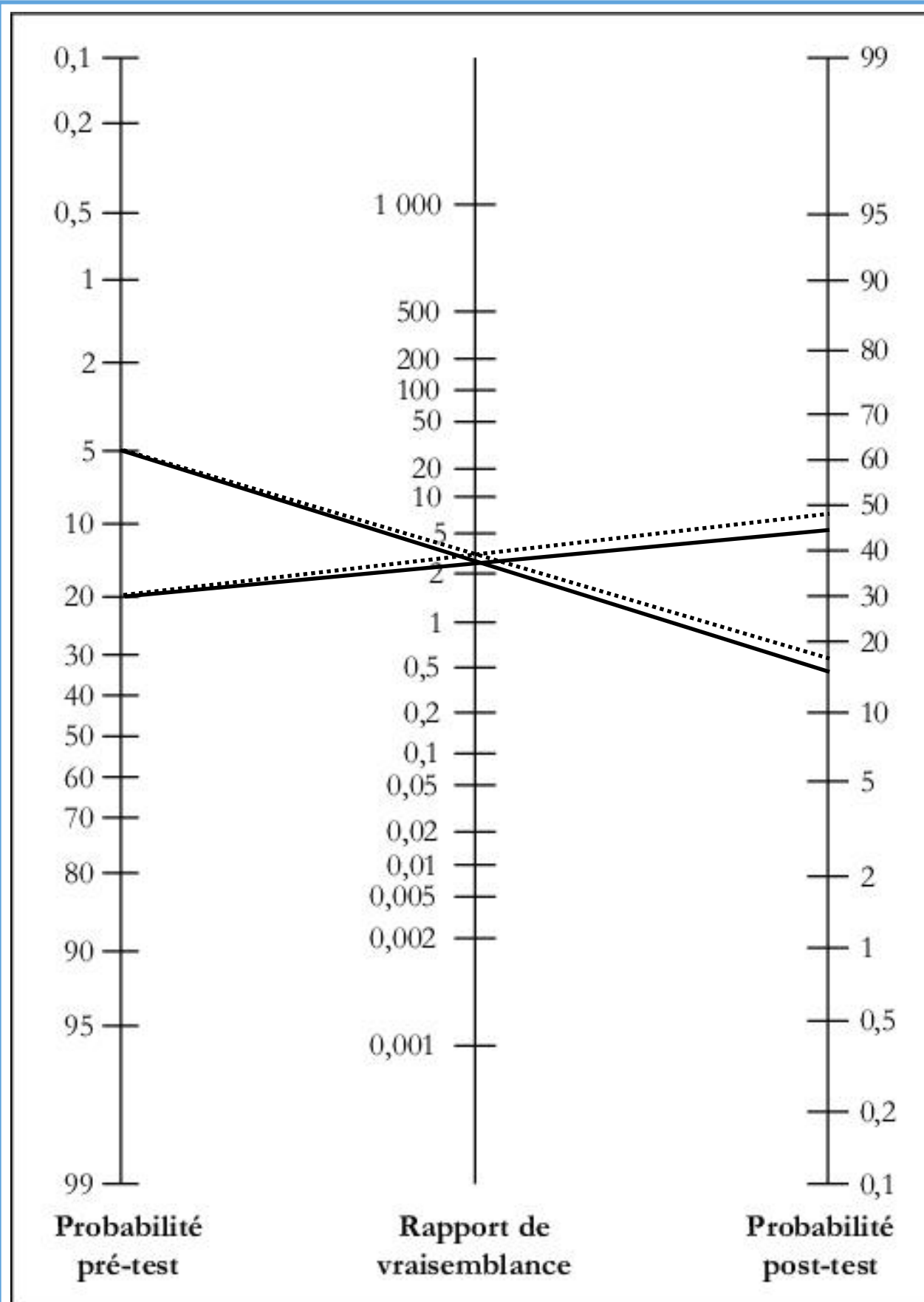
LR + Radiographie thoracique : 3 (ligne pleine) ;

LR+ scanner thoracique : 3,3 (ligne pointillée)

deux situations :

patient à faible risque :
prévalence 5 %,

patient à haut risque :
prévalence 20 %,



Pour résumer

- ▶ 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 $LR+$ est plus élevé
 - ▶ L'apport diagnostique d'un résultat négatif d'autant plus grand que $LR-$ est petit et proche de 0