

Évaluation Diagnostique

Cours P2
Ph. Ravaud

Test diagnostic

Ce peut être :

- **Un signe clinique** subjectif (caractéristiques cliniques d'une céphalée) ou objectif (aspect clinique d'une éruption).
- **Un test de laboratoire** : qualitatif (hémoculture) ou quantitatif (glycémie).
- **Un examen radiographique.**
- **Un acte invasif** : biopsie, intervention chirurgicale.

Évaluation d'une procédure diagnostique

But : classer les sujets malades et non malades

2 situations

```
graph TD; A[2 situations] --> B[Variable qualitative]; A --> C[Variable quantitative];
```

Variable qualitative

Signe présent oui / non

Hémoculture positive

Abolition d'un réflexe

Oui / non

Variable quantitative

Glycémie

Cholestérolémie

Pression artérielle

Variable Qualitative

- **Test idéal** : le test est positif si et seulement si le sujet est malade.

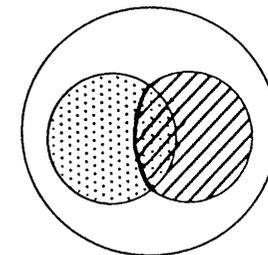
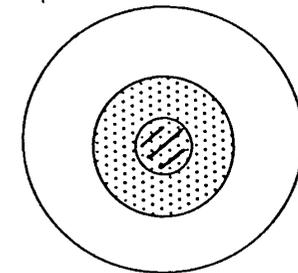
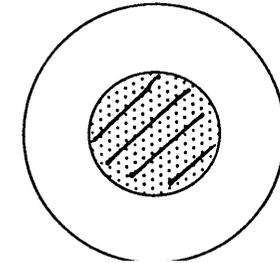
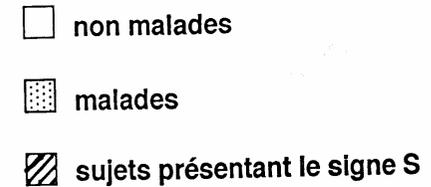
Permet de classer sans erreur en Malade ou Non Malade.

- **Signe pathognomonique** : Seuls les sujets malades ont le signe mais tous les sujets malades ne l'ont pas (ex : s. de Köplick dans la rougeole).

Permet de classer sans erreur les sujets malades.

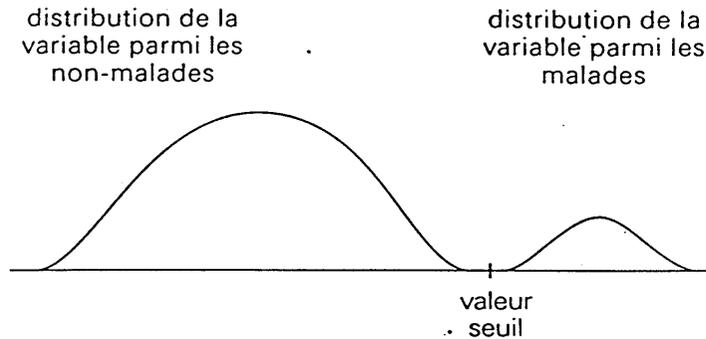
- **Situation habituelle** : Certains malades ont le signe, d'autres non; certains non-malades ne l'ont pas, d'autres l'ont.

Ne permet pas de classer sans erreur en Malade ou Non-Malade.



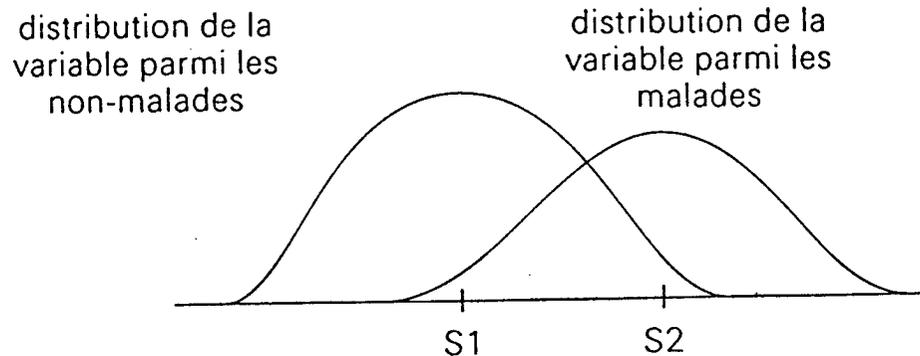
Variable Quantitative

- Situation idéale



Permet de classer sans erreur en Malade ou Non Malade.

- Situation habituelle



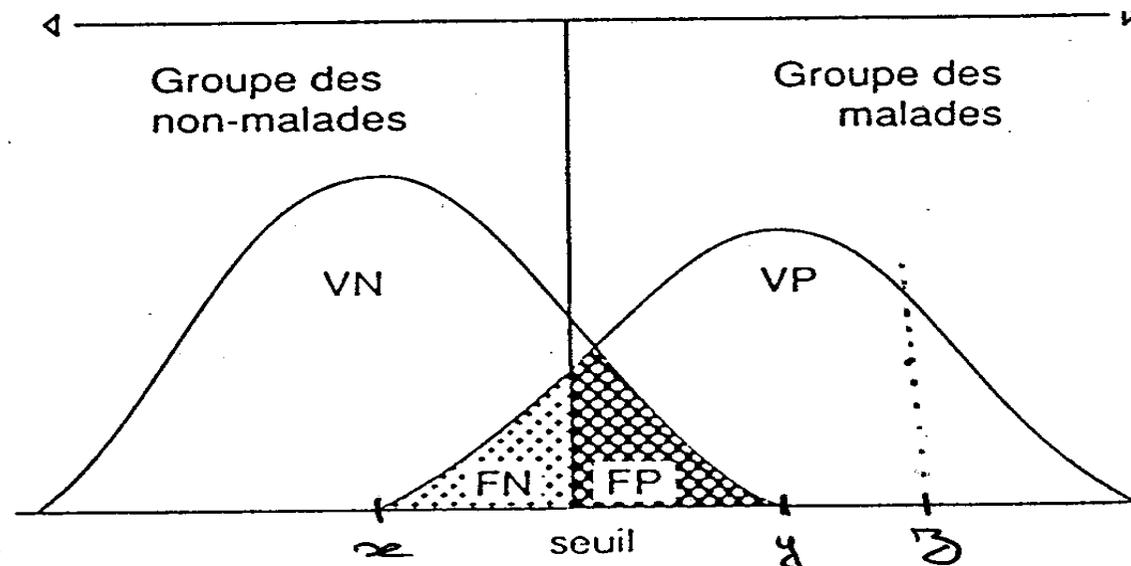
Le choix d'une valeur seuil permet de transformer une variable quantitative (glycémie) en variable qualitative dichotomique (glycémie $> 1,40$ g/l : oui/non).

Malade / Non Malades

Définition thérapeutique :

Choisir comme seuil le point (z) au delà duquel on a démontré que les traitements spécifiques font plus de bien que de mal. Évite de qualifier des sujets comme étant malades s'ils ne doivent pas être traités.

Ex : Pression artérielle



Pour évaluer un test diagnostic

- Définir les termes Malades et Non Malades
- Comparer le test à l'étude à un test de référence

"GOLD STANDARD"

Pour évaluer les performances d'un test, il faut un **test de référence.**

⇒ Critère permettant d'affirmer que la maladie M est présente ou absente.

- Parfois critère indiscutable (présence de BK dans le LCR).
- Souvent plus problématique :
 - Subjectif : donnée d'interrogatoire (angine de poitrine)
 - Cher
 - Risqué ou difficile à obtenir (biopsies, exploration chirurgicale, autopsie, ...).
 - Parfois c'est l'évolution (cancers, maladie à prions, ...).
- L'absence de test de référence nécessite de définir le diagnostic recherché avec un faisceau d'arguments dont la validité doit avoir été étudiée.

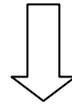
Conséquence des standards imparfaits

Le test de référence peut-être imparfait

(choisi car d'usage plus ancien ou considéré comme le meilleur par des experts).

Si un nouveau test est comparé avec un test standard ancien (mais inexact), **le nouveau test pourra sembler plus mauvais alors qu'il est en réalité meilleur...**

si le nouveau test est plus sensible que l'ancien
(cad s'il détecte mieux les sujets malades)



les patients supplémentaires identifiés par le nouveau test
seront considérés comme de faux positifs
(alors qu'il s'agissait de faux négatifs du test de référence)

Ex : comparaison de l'échographie et de la cholécystographie orale pour la détection des calculs vésiculaires.

Performance d'un test

Le but de toutes les études diagnostiques est d'obtenir a, b, c, et d.

		Maladie M (méthode de référence)	
		Présente	Absente
Test diagnostique à l'étude	Positif	Vrais positifs a	b Faux positifs
	Négatif	Faux négatifs c	d Vrais négatifs

- Manque d'informations sur les tests négatifs (il n'est pas toujours éthique de pratiquer le test de référence si le test à l'étude est négatif : PSA et biopsie de prostate).
- Manque d'informations sur les résultats des tests chez les non malades (il n'est pas toujours éthique (ou trop coûteux de pratiquer le test de référence si le sujet n'a pas de symptôme : IRM lombaire et lombalgie).

Magnetic resonance imaging of the lumbar spine in people without back pain.

Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. *N Engl J Med* 1994 Jul 14;331(2):69-73.

- Contexte : Chez les patients lombalgiques, on visualise fréquemment en IRM (plus de 2 fs/3) un disque intervertébral saillant ou protru, qui sert souvent à justifier la douleur et à guider le traitement.
- Des IRM lombaires ont été réalisées chez 98 volontaires asymptomatiques.
- Des disques saillants ou protrus étaient notés chez presque 2/3 des sujets asymptomatiques, un peu moins seulement que la fréquence des anomalies similaires chez des patients atteints de lombalgies. Les auteurs concluent que de tels résultats peuvent être dus à une coïncidence.
- Conclusion : L'IRM lombaire aurait été considérée comme un bon examen pour le diagnostic de lombalgie si l'on n'avait eu les résultats dans le groupe des non malades. Risque de surestimation de l'intérêt du test.

Performance d'un test

		Maladie M (méthode de référence)	
		Présente	Absente
Test diagnostique à l'étude	Positif	Vrais positifs a	b Faux positifs
	Négatif	Faux négatifs c	d Vrais négatifs

$Se = VP/(VP + FN)$ Probabilité d'avoir un signe positif quand on est malade.

$Sp = VN/(VN + FP)$ Probabilité d'avoir un signe négatif quand on est non-malade.

Sensibilité et Spécificité

- La **sensibilité** est l'indice qui évalue la capacité d'une mesure à **bien classer les malades**.

$$Se = \frac{VP}{(VP + FN)}$$

Se = 1 Le signe est présent chez 100% des malades (pas de FN).

- La **spécificité** est l'indice qui évalue la capacité d'une mesure à **bien classer les non-malades**.

$$Sp = \frac{VN}{(VN + FP)}$$

Sp = 1 Le signe est absent chez tous les non-malades (pas de FP).

- La sensibilité et la spécificité sont les propriétés prises en compte quand on décide de **prescrire ou non un test diagnostic**.

Usage des tests sensibles

(habituellement positif en cas de maladie, peu de FN)

- Quand la maladie est très rare.
- Quand la maladie est curable.
- Quand le diagnostic précoce d'une maladie grave en améliore le pronostic (tuberculose, maladie de Hodgkin, ...).
- Au stade précoce d'une action diagnostique pour éliminer des hypothèses (ex : sérologie HIV devant un infiltrat pulmonaire).
- Un test sensible est surtout utile au clinicien lorsqu'il est négatif.

Usage des tests spécifiques

(rarement positif en l'absence de maladie, peu de FP)

- Confirmer un diagnostic suggéré par d'autres résultats.
- Quand des résultats faussement positifs peuvent être préjudiciables au patients (ex cancer et preuve histologique).
- Quand la maladie est incurable
- Un test spécifique est surtout utile au clinicien lorsqu'il est positif.

Compromis entre sensibilité et spécificité

- Il est évidemment souhaitable d'avoir un test à la fois très sensible et très spécifique.
- C'est rarement le cas, il faut donc faire un **compromis entre la sensibilité et la spécificité** du test.
- **Le choix de la valeur seuil** délimitant Malades et Non Malades **est arbitraire.**
- Pour un même test, en fonction du seuil choisi, quand la sensibilité augmente, la spécificité diminue et inversement.

Compromis entre la sensibilité et la spécificité lors d'un test diagnostique du diabète ^a

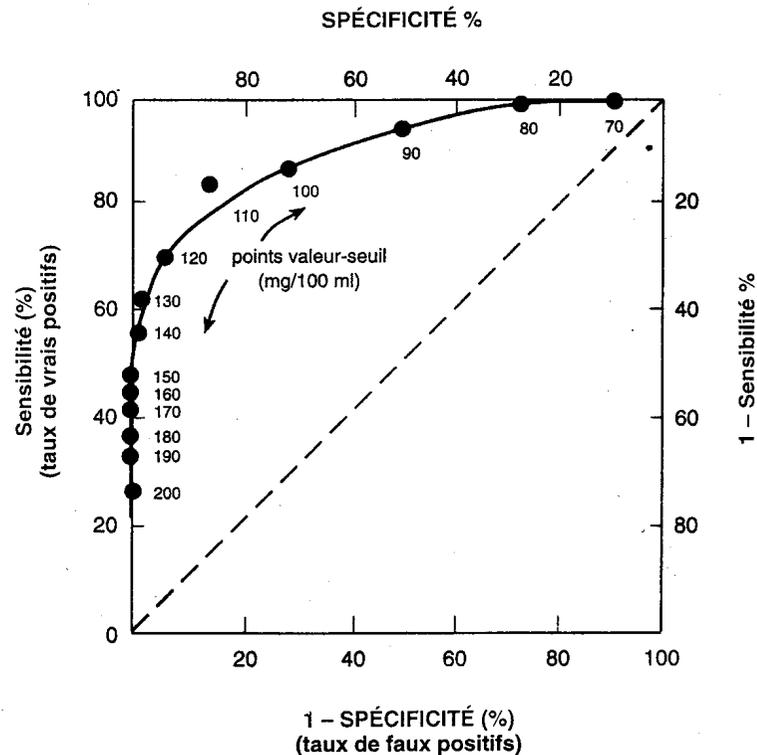
Glycémie postprandiale t + 2 h (mg/100ml)	Sensibilité (%)	Spécificité (%)
70	98,6	8,8
80	97,1	25,5
90	94,3	47,6
100	88,6	69,8
110	85,7	84,1
120	71,4	92,5
130	64,3	96,9
140	57,1	99,4
150	50	99,6
160	47,1	99,8
170	42,9	100
180	38,6	100
190	34,3	100
200	27,1	100

^a D'après *Public Health Service*. Diabete program guide. Publication n° 506. Washington, DC : U.S. Government Printing Office, 1960.

Choix du seuil

Courbes ROC (Receiver Operating Characteristic ou *courbe caractéristique de la performance d'un test*)

But : Choisir le seuil diagnostique d'un test en recherchant le meilleur rapport entre sensibilité et spécificité.

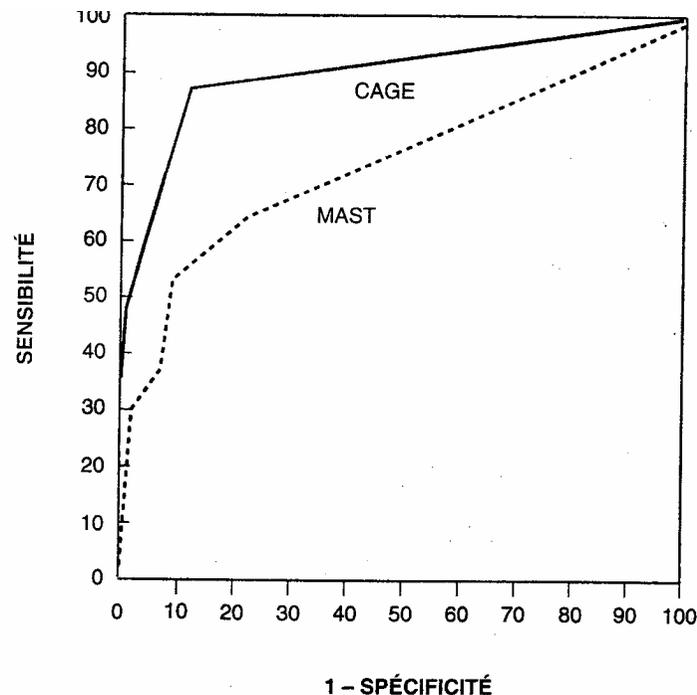


- Confronte le taux de VP au taux de FP pour une gamme de valeurs seuil.
- Les tests bien discriminants sont dans le coin supérieur gauche. Les tests peu performants se rapprochent de la diagonale.
- La valeur seuil permettant de minimiser à la fois les FP et les FN est au point d'inflexion de la courbe ROC.

Une courbe ROC. L'exactitude d'une glycémie postprandiale à t + 2h comme test diagnostique du diabète. (D'après *Public Health Service. Diabetes program guide. Publication n° 506.* Washington, DC : U.S. Government Printing Office, 1960)

Alcoholism screening questionnaires : are they valid in elderly medical outpatients?

Jones TV, Lindsey BA, Yount P, Soltys R, Farani-Enayat B. *J Gen Intern Med* 1993;8:674-678

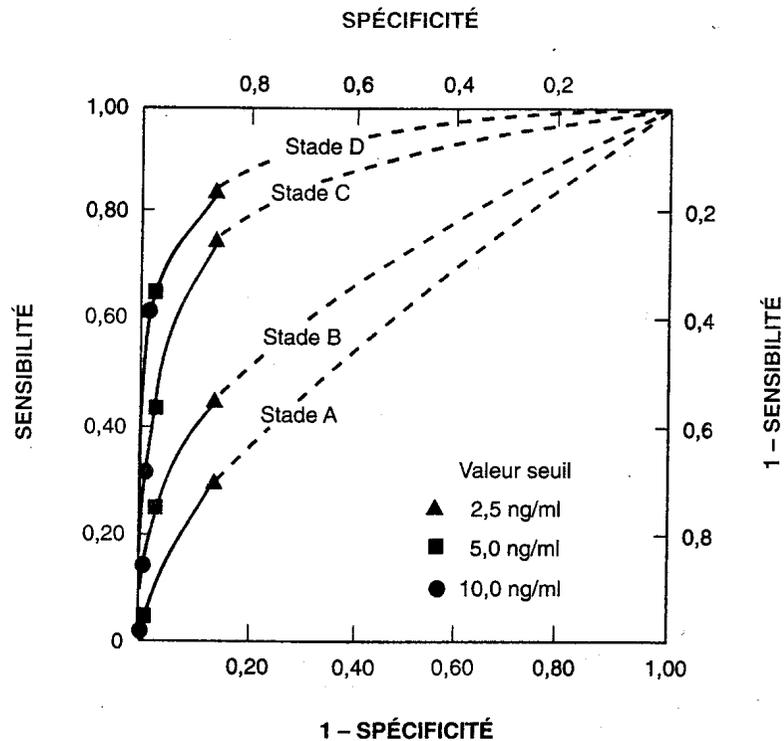


- Les courbes ROC permettent de comparer des tests alternatifs pour un même diagnostic.
- L'exactitude globale du test est l'aire sous la courbe ROC. Plus l'aire est grande, meilleur est le test.

Les courbes ROC des questionnaires CAGE et MAST chez des personnes âgées alcooliques et non alcooliques. (D'après Jones TV, Lindsey BA, Yount P, Soltys R, Farani-Enayat B. Alcoholism screening questionnaires : are they valid in elderly medical outpatients ? *J Gen Intern Med* 1993 ; 8 : 674-678)

Carcinoembryonic antigen.

Fletcher RH. *Ann Intern Med* 1986;104:66-73)



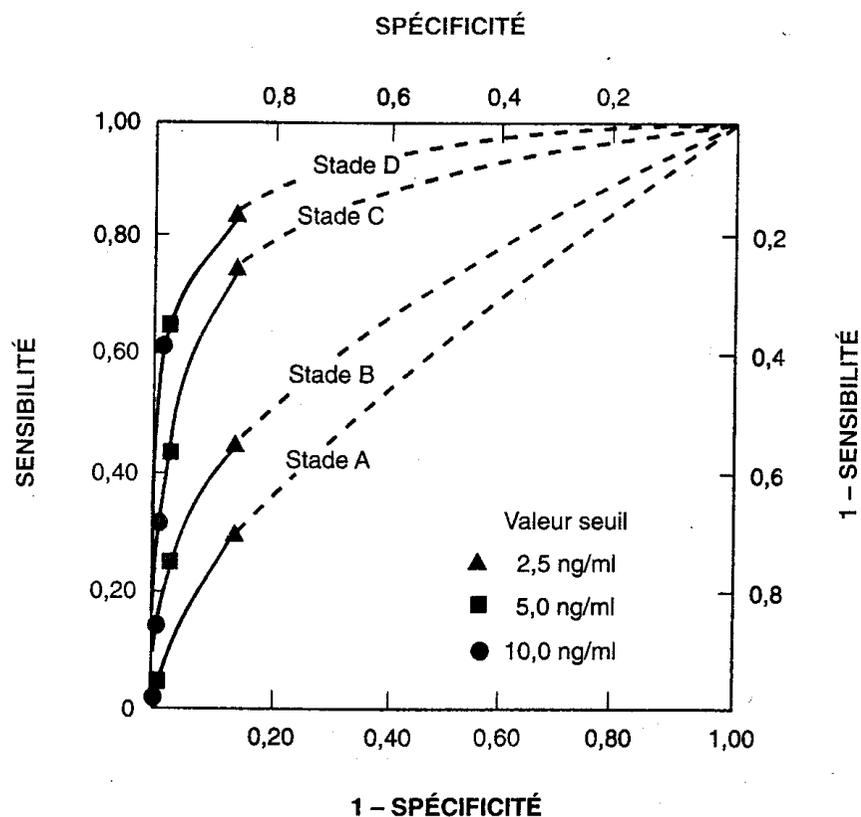
La courbe ROC de l'ACE comme test diagnostique du cancer colo-rectal, selon le stade de la maladie. La sensibilité et la spécificité d'un test varient selon le stade de la maladie. (D'après Fletcher RH. Carcinoembryonic antigen. *Ann Intern Med* 1986 ; 104 : 66-73)

La sensibilité et la spécificité varient en fonction du stade ou de la sévérité de la maladie.

Problème :

Quand les patients utilisés pour décrire les propriétés du test sont différents de ceux pour lesquels le test s'applique en pratique clinique.

Variation de la performance du dosage de l'ACE selon le score de cancer colo-rectal.



La courbe ROC de l'ACE comme test diagnostique du cancer colo-rectal, selon le stade de la maladie. La sensibilité et la spécificité d'un test varient selon le stade de la maladie. (D'après Fletcher RH. Carcinoembryonic antigen. *Ann Intern Med* 1986 ; 104 : 66-73)

- Évaluation de marqueurs tumoraux : comparaison maladies métastatiques / étudiants en médecine! Ne correspond pas à la pratique clinique.

- But du test : pouvoir discriminer sujets atteints de cancers / sujets atteints d'autres pathologies pouvant faire évoquer un cancer.

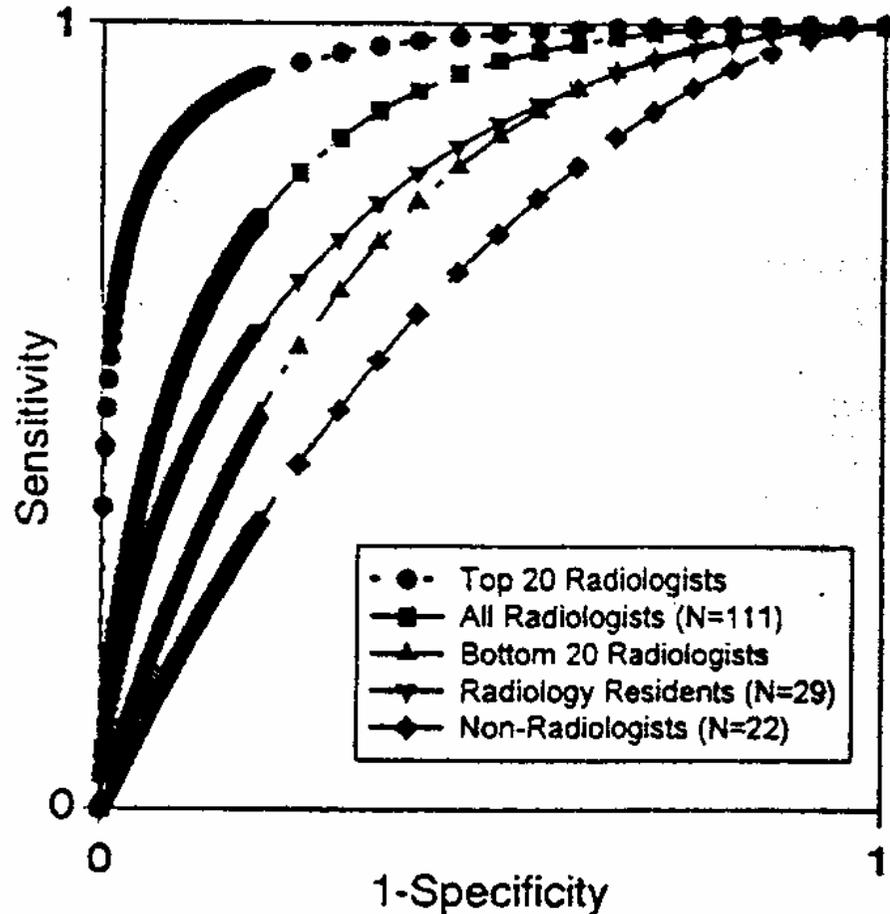
- L'ACE reflète bien une maladie métastatique et moins bien un cancer localisé.

- Des sujets indemnes de la maladie peuvent avoir d'autres conditions positivant le test. Ex : L'ACE est augmenté en cas de colite ulcéreuse et de cirrhose.

- Seule indication de l'ACE : suivi des patients ayant un cancer colo-rectal.

Measuring performance in chest radiography.

Potchen EJ, Cooper TG, Sierra AE, Aben GR, Potchen MJ, Potter MG, Siebert JE.
Radiology 2000;217:456-459



Variabilité inter observateurs.

Problème pour généraliser les résultats.

Valeurs prédictives

- Se, Sp \Rightarrow décision de prescrire ou non un test
- VPP, VPN Les résultats du test sont connus \Rightarrow savoir si le patient est malade ou non

$$VPP = \frac{VP}{(VP + FP)}$$

Probabilité d'être malade si on a un test positif.

$$VPN = \frac{VN}{(VN + FN)}$$

Probabilité de ne pas être malade si on a un test négatif.

$$\text{Prévalence} = \frac{(a + c)}{(a + b + c + d)}$$

	M +	M -	Total
T +	VP a	b FP	a + b
T -	FN c	d VN	c + d
Total	a + c	b + d	a + b + c + d

Les valeurs prédictives positive et négative dépendent des sensibilité et spécificité du test et de la prévalence de la maladie dans la population étudiée.

Valeurs prédictives

- $VPP = \frac{Se \cdot prévalence}{(Se \cdot prévalence) + (1 - Sp) \cdot (1 - prévalence)}$
- Plus un test est sensible, meilleure est sa VPN (plus le clinicien est sûr qu'un résultat négatif élimine la maladie suspectée).
- Plus un test est spécifique, meilleure est sa VPP (plus le clinicien est sûr qu'un résultat positif confirme le diagnostic suspecté).
- Les VPP et VPN dépendant de la prévalence, elles ne sont pas indépendantes du cadre dans lequel le test est appliqué.

Valeurs prédictives de L'ECG d'effort en fonction de la prévalence des sténoses coronaires

	sténose +	sténose -	Total
ECG+	55	7	62
ECG -	49	84	133
Total	104	91	195

$$Se = 55/104 = 53\%$$

$$Spe = 84/91 = 92\%$$

$$VPP = 55/62 = 89\%$$

$$VPN = 84/133 = 63\%$$

$$\text{Prévalence} = 104/195 = 53\%$$

	sténose +	sténose -	Total
ECG+	55	42	97
ECG -	49	478	527
Total	104	520	624

$$Se = 55/104 = 53\%$$

$$Spe = 478/520 = 92\%$$

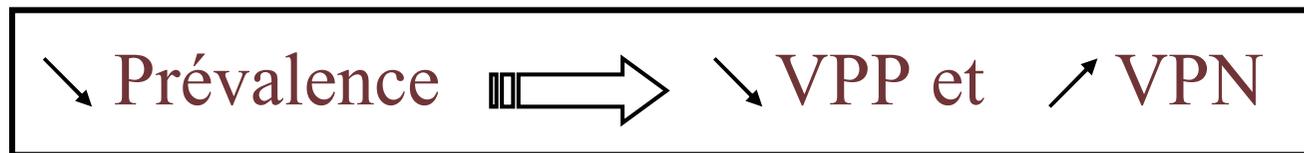
$$VPP = 55/97 = 57\%$$

$$VPN = 478/527 = 91\%$$

$$\text{Prévalence} = 104/624 = 17\%$$

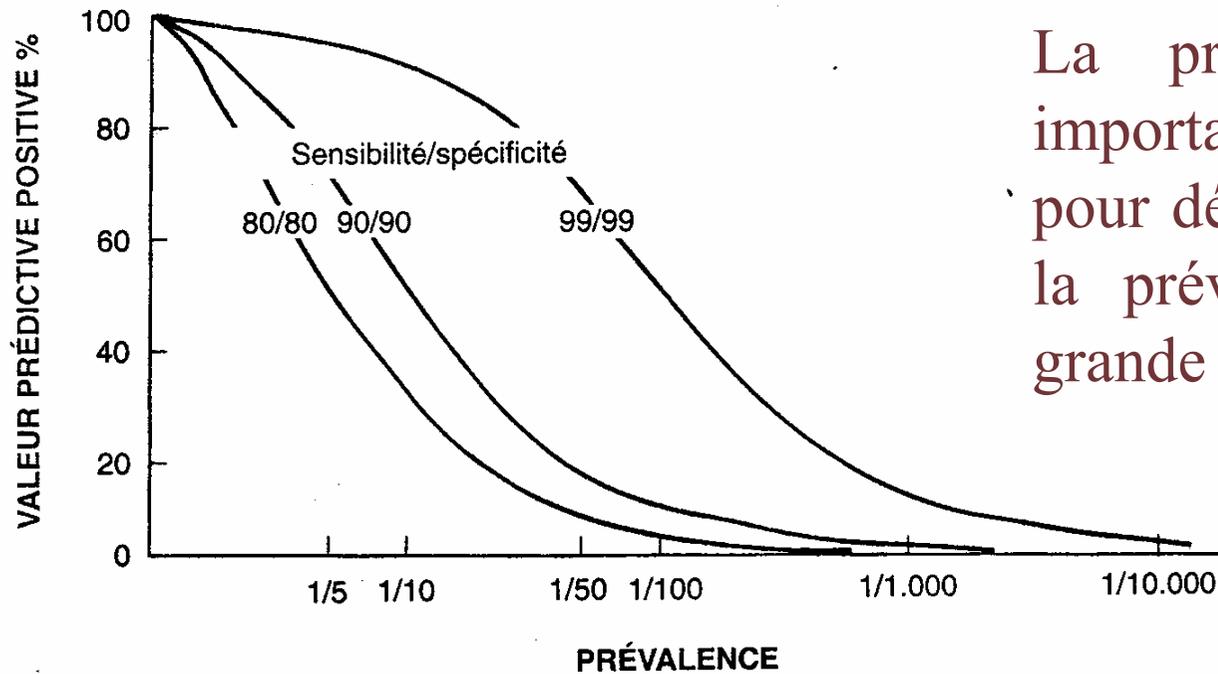
Valeurs prédictives et prévalence

- D'une manière générale, la valeur prédictive positive chute et la valeur prédictive négative augmente quand le test diagnostic initialement développé dans un groupe de patients où la prévalence de la maladie est élevée, est appliqué ensuite à un groupe de patients dans lequel la prévalence est plus faible.



- Les résultats positifs d'un test même très spécifique dans un groupe où la prévalence est faible seront largement faux positifs.
- Les résultats négatifs d'un test même très sensible dans un groupe où la prévalence est forte seront largement faux négatifs.

Valeurs prédictives et prévalence



La prévalence est plus importante que Se et Spe pour déterminer les VP, car la prévalence a une plus grande gamme de variation.

La valeur prédictive positive selon la sensibilité, la spécificité et la prévalence de la maladie.

Quand la prévalence de la maladie est élevée, le test a de bonnes performances. Pour des prévalences basses, la VPP est proche de 0 et le test pratiquement inutile pour diagnostiquer la maladie.

Efficacité diagnostique

L'efficacité diagnostique d'un test correspond au pourcentage de sujets bien classés par le test.

$$\text{Eff. Diag.} = (VP + VN) / (\text{Malades} + \text{Non Malades})$$

Performance d'un test

- Propriétés intrinsèques

- Sensibilité
- Spécificité

- Propriétés extrinsèques :

- VPP, *probabilité a posteriori ou post-test*
- VPN

dépendantes - des qualités intrinsèques du test

- de la prévalence de la maladie*

*Prévalence : proportion de personnes dans une population définie, à un temps donné, ayant la maladie en question. *Probabilité a priori ou pré-test.*

Rapports de vraisemblance

"Likelihood ratios"

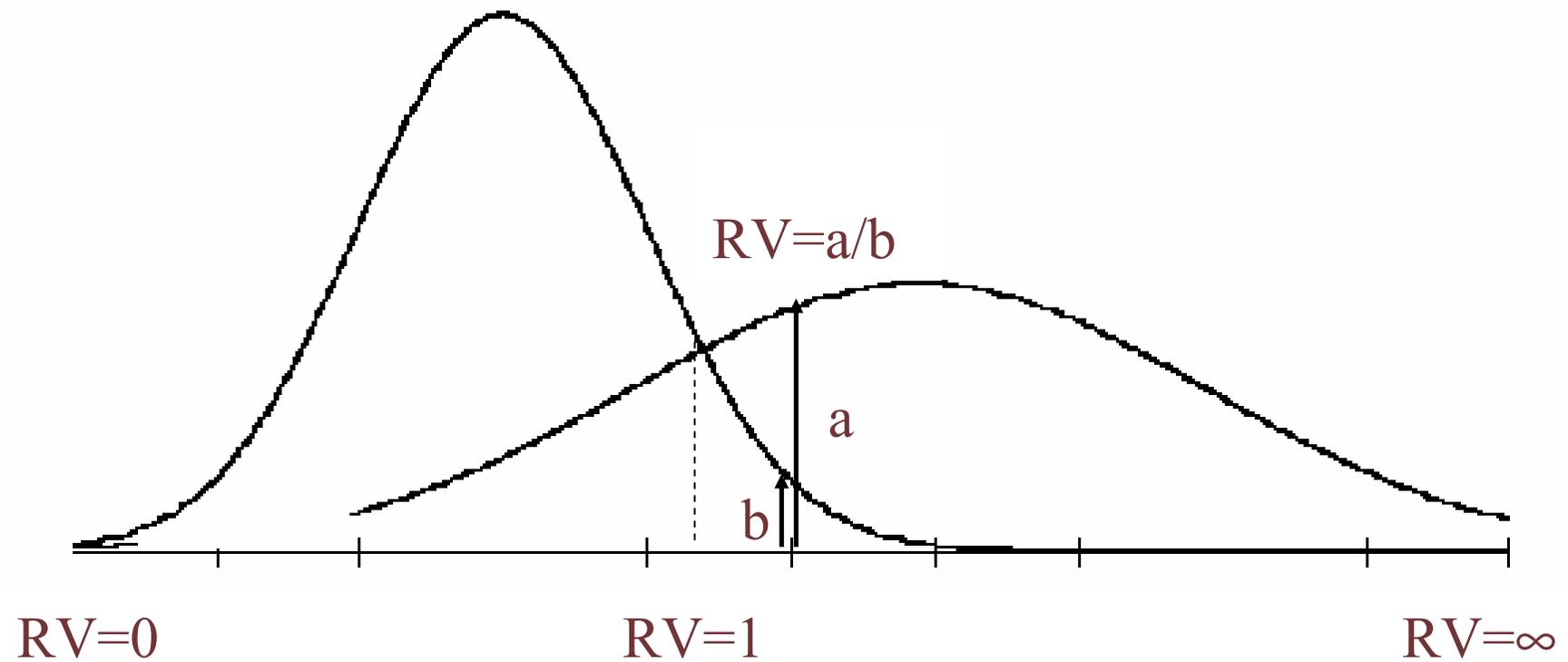
Les RV sont une façon alternative de décrire les performances d'un test.

Le RV compare la probabilité du résultat d'un test chez les malades et chez les non malades. Un test positif est obtenu RV+ fois plus souvent chez les sujets malades que chez les sujets non malades ; un test négatif est RV- fois moins fréquent chez les malades que chez les non malades.

$$\text{RV positif} = \frac{P(T+/M+)}{P(T+/M-)} = \frac{Se}{(1 - Sp)} = \frac{(VP / malades)}{(FP / non - malades)}$$

$$\text{RV négatif} = \frac{P(T-/M+)}{P(T-/M-)} = \frac{(1 - Se)}{Sp} = \frac{(FN / malades)}{(VN / non - malades)}$$

Évaluation diagnostique : rapport de vraisemblance

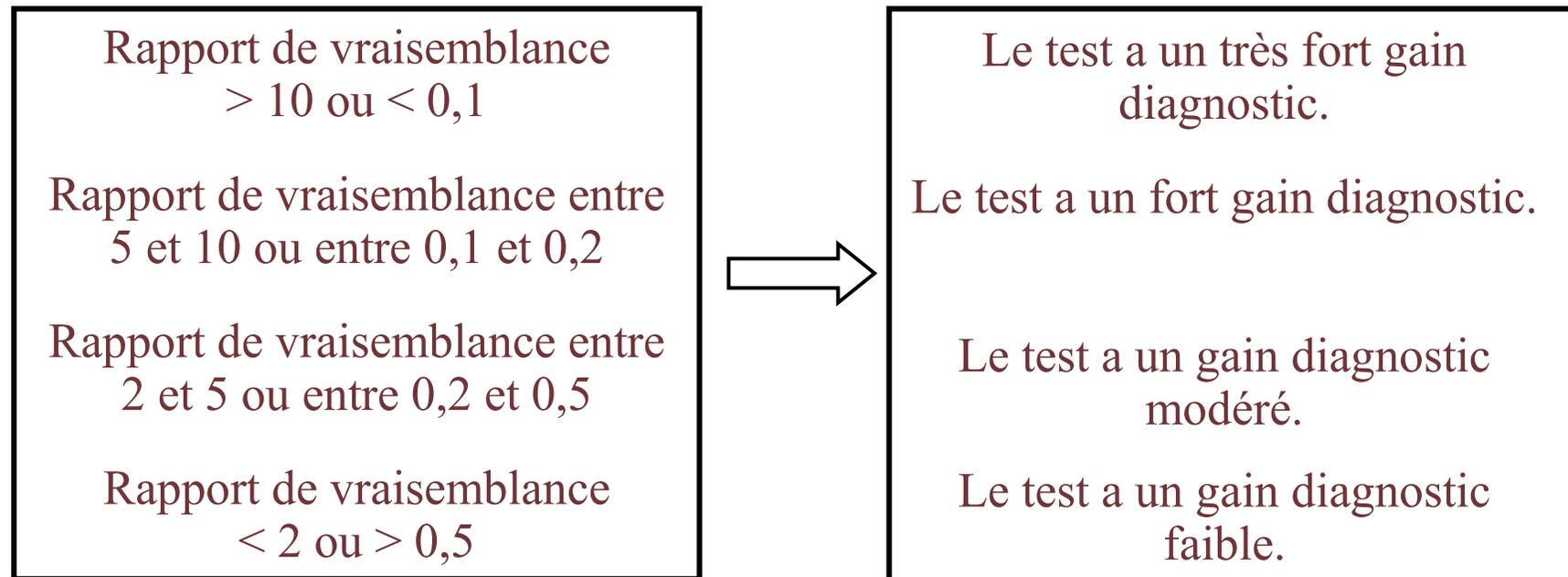


Rapport de vraisemblance

Si le RV est égal à 1, la probabilité du diagnostic est la même avant et après le test. Le test est donc inutile.

Plus les valeurs du RV s'éloignent de 1 et plus le test présentera de l'intérêt.

Les valeurs du RV+ qui sont supérieures à 1 montrent une augmentation de la confiance dans le diagnostic, alors que les valeurs inférieures à 1 reflètent une infirmation du diagnostic.



Rapport de vraisemblance

CPK	IDM+	IDM-	RV+
≥280	97 (0.422)	1 (0.008)	0.422/0.008 = 55
80-279	118 (0.513)	15 (0.115)	0.513/0.115 = 4.4
40-79	13 (0.057)	26 (0.200)	0.057/0.2 = 0.3
1-39	2 (0.009)	88 (0.677)	0.009/0.677 = 0.01
	230	130	

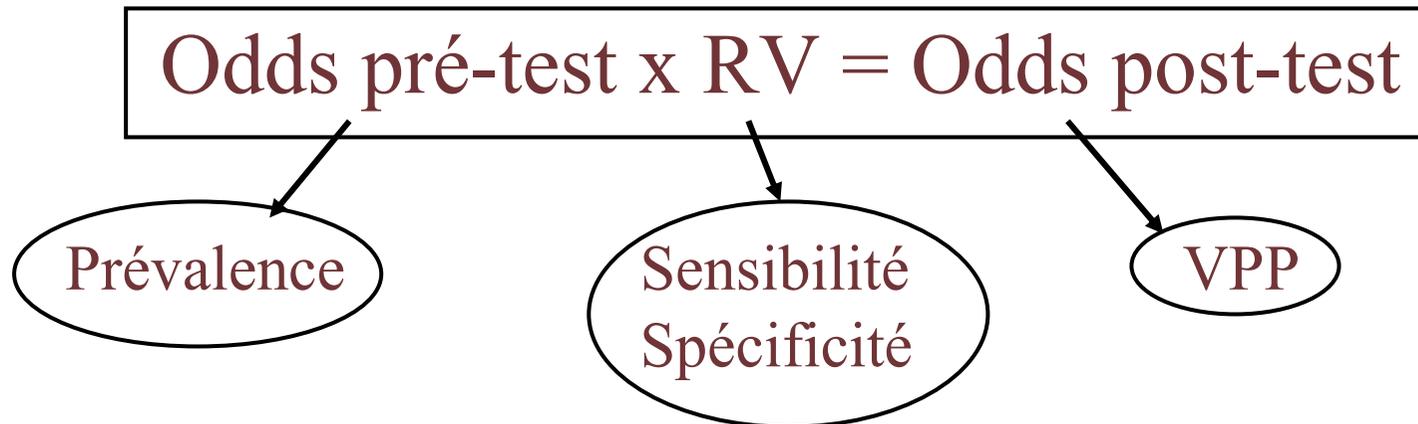
- Un taux de CPK ≥280 a 55 fois plus de chances de provenir d'un patient faisant un infarctus du myocarde que d'un patient non malade.
- Un taux de CPK compris entre 1 et 39 a 0.01 fois plus de chances (soit 100 fois moins de chances) de provenir d'un patient faisant un infarctus du myocarde que d'un patient non malade.

Les Odds

- La probabilité P c'est la proportion de personnes chez qui un caractère donné est présent (fréquence de ce caractère).
- L'Odds et la probabilité expriment la même information différemment : l'Odds est le rapport de deux probabilités.

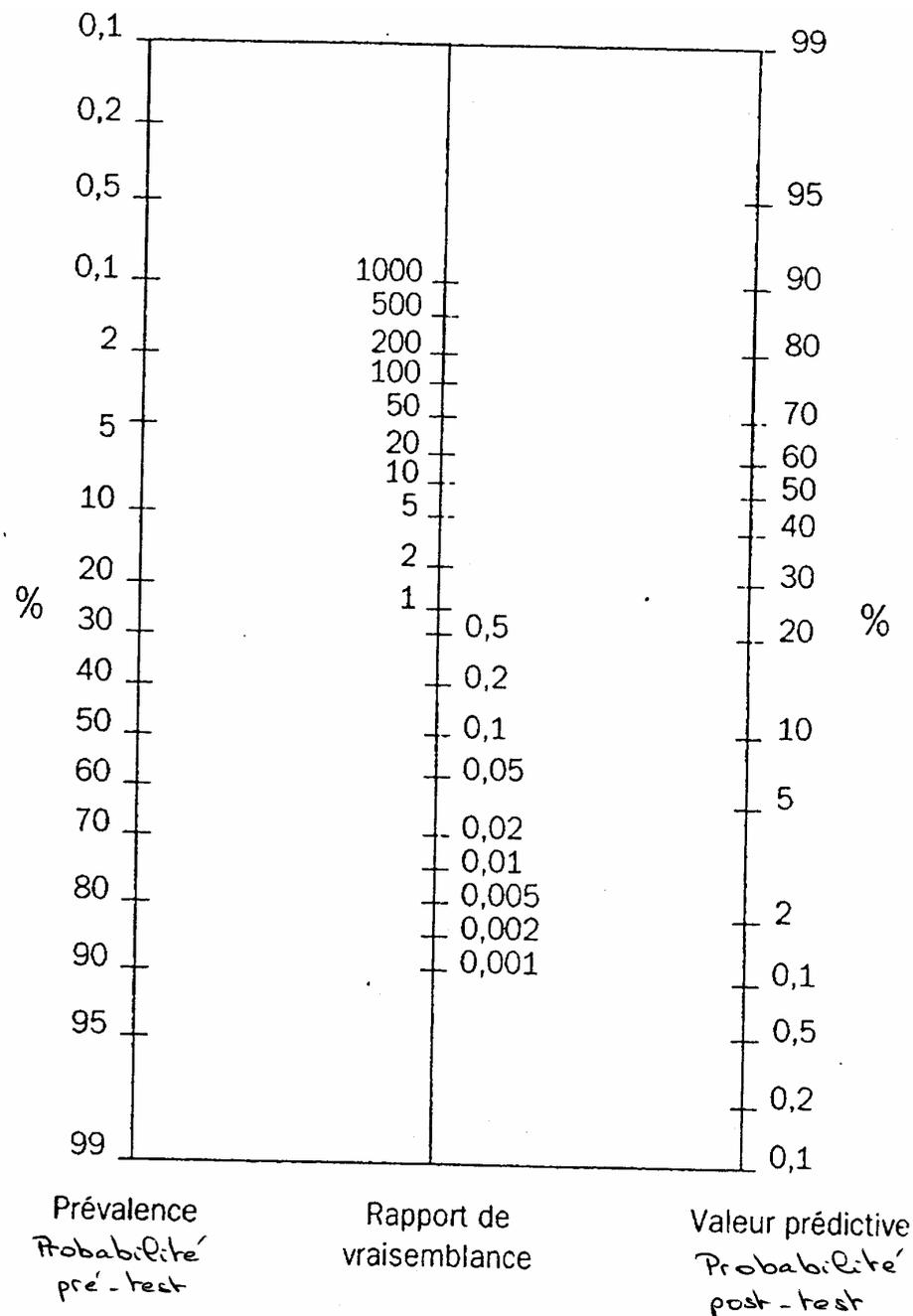
$$\text{Odds} = \frac{P}{(1 - P)} \quad \text{d'où} \quad P = \frac{\text{Odds}}{(\text{Odds} + 1)}$$

Ex : l'Odds que la France batte la Bulgarie est de 4:1, c'est à dire que la France a 80% de chances de gagner.



L'utilisation du rapport de vraisemblance est facilité par l'utilisation du **nomogramme de Fagan**.

Sur ce diagramme, tracer la droite passant par la prévalence de la maladie dans le groupe étudié et par le rapport de vraisemblance du test utilisé. Le point d'intersection de cette droite avec l'axe "valeur prédictive" détermine la probabilité post test pour le patient considéré.



Rapport de vraisemblance

Avantages

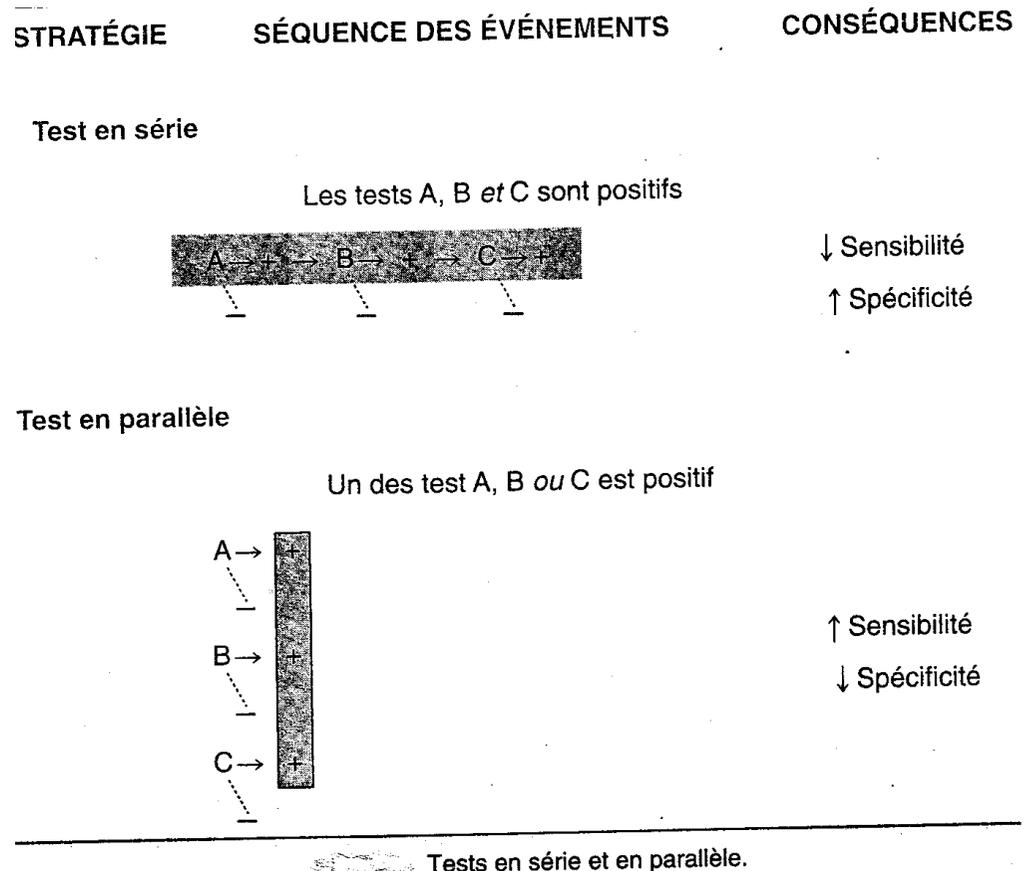
- Est indépendant de la prévalence de la maladie. Bon reflet de la valeur du test quel que soit le groupe auquel il est appliqué.
- Résume l'information en un seul paramètre.
- Utilisable pour différents niveaux de résultats d'un test. Pour chaque niveau il procure une information différente permettant d'interpréter au mieux les résultats du test.
- Il permet de calculer de manière individuelle l'intérêt de réaliser le test à partir de la probabilité initiale de maladie du patient.

Les tests multiples

- Les cliniciens utilisent habituellement des tests imparfaits (Se et $Sp < 100\%$, RV intermédiaires), un test unique menant à une probabilité de maladie entre 10 et 90%. Pour aller plus loin dans le diagnostic, le médecin va donc prescrire plusieurs tests.

Tests en série : consécutivement, basés sur les résultats du test précédent. Tous les tests doivent être positifs pour faire le diagnostic car le processus diagnostique est arrêté quand un résultat est négatif. Stratégie quand une évaluation rapide est nécessaire.

Tests en parallèle : tous au même moment. Un seul résultat positif suffit au diagnostic.



Tests multiples

Une stratégie diagnostique utilisant deux tests en série sera plus efficace si le test ayant la plus haute spécificité est utilisé en premier.

Ainsi moins de patients auront à subir les deux tests, et un nombre égal de malades seront diagnostiqués.

Cependant si l'un des tests est moins coûteux ou moins risqué, il est plus prudent de l'utiliser en premier.

Tests multiples

Limites de la normalité : 95%
(courbes de Gausse, risque $\alpha = 5\%$)
 \Rightarrow **la probabilité d'être normal s'amenuise avec la multiplication des tests diagnostiques.**

La probabilité qu'un patient soit appelé "normal" est égale à $(1 - \alpha)^n$
n : nombre de tests diagnostiques indépendants appliqués.

Ainsi un patient qui a subi 100 tests a seulement 6 chances sur 1000 d'être appelé "normal" à la fin de son bilan.

Nombre de tests indépendants	Risque (%) qu'au moins un résultat soit interprété à tort comme anormal (FP) $1 - (1 - \alpha)^n$
1	5
2	10
4	19
6	26
10	40
20	64
50	92
30	99

Évaluation Diagnostique : Hiérarchisation

Niveau 1 : Performance technique

L'IRM donne-t-elle des images de qualité représentative de l'anatomie ?

Niveau 2 : Performance diagnostique

Les images produites par l'IRM permettent-elles de faire un diagnostic précis ?

Niveau 3 : Impact diagnostique

Les résultats de l'IRM modifient-ils la confiance que l'on a dans le diagnostic (différence entre les probabilités pré et post test) et évitent-ils des examens ?

Niveau 4 : Impact thérapeutique

Les résultats de l'IRM contribuent-ils aux prises de décision thérapeutiques ?

Niveau 5 : Impact sur les résultats

Les résultats de l'IRM contribuent-ils à améliorer l'état de santé du patient ?

Niveau 6 : Impact de santé publique

Analyse coût-bénéfice et analyse coût efficacité du point de vue de la société.³⁹