

# Application de la médecine factuelle aux tests diagnostiques

Ch. Mélot, MD, PhD, MSciBiostat  
Service des Urgences  
Hôpital Universitaire Erasme

[cmelot@ulb.ac.be](mailto:cmelot@ulb.ac.be)

Certificat en médecine factuelle  
23 & 24 février 2017

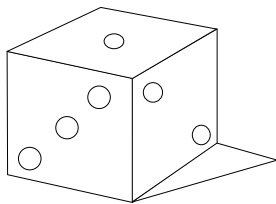
## BUTS DU TRAITEMENT STATISTIQUE DES DONNEES EN MEDECINE CLINIQUE

- Le traitement statistique:
  - essaie de répondre à la séparation entre le normal et le pathologique pour un patient:
    - intervalle de prédiction à 95% ( $m \pm 2 SD$ )
    - valeurs prédictives des tests & scores
  - identifie une différence qualifiée de significative entre deux ou plusieurs groupes de patients:
    - intervalle de confiance à 95% ( $m \pm 2 SE$ )
    - $SE = SD/\sqrt{n}$
    - probabilité de l'efficacité d'un traitement

## Probabilité

- La probabilité est un nombre entre 0 (0 % de chance) et 1 (100 % de chance)
- Probabilité subjective (approche Bayésienne)
- Probabilité statistique (approche fréquentiste)

## Odds (cote) et Probabilité



$$\text{Probabilité} = \frac{1}{6} = 0.166$$

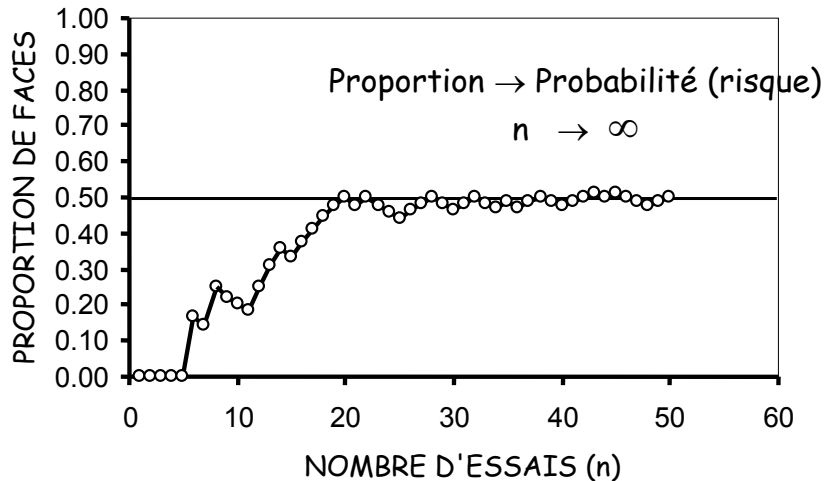
$$\text{Odds en faveur} = \frac{\frac{1}{6}}{\frac{5}{6}} = 0.20$$

$$\text{Odds contre} = \frac{\frac{5}{6}}{\frac{1}{6}} = 5 \text{ contre } 1$$



# Probabilité statistique

## JET D'UNE PIECE



## Test diagnostique et décision médicale

- Par exemple: biomarqueurs et diagnostic
- Il y a 2 approches statistiques pour l'évaluation d'un biomarqueur:
  - Modélisation du risque (disease outcome) par régression logistique (ou régression de Cox)  
→ le biomarqueur est utile s'il a un lien avec le risque ( $p < 0.05$ )

Approche fréquentiste

- Performance dans la classification malade/non malades (sensibilité, spécificité, valeur prédictive, LR et courbe ROC).  
→ le biomarqueur est utile si il a une bonne performance de classification (discriminant)

Approche bayésienne

## Approche fréquentiste.

### Dans l'approche fréquentiste, la valeur de p (p-value) est...

- la probabilité d'obtenir un résultat égal ou plus grand que le résultat observé, dans la situation où l'hypothèse nulle est vraie (hasard).
- n'est pas la probabilité que l'hypothèse nulle soit vraie car elle (p-value) est calculée, en assumant que l'hypothèse nulle est vraie (100%).

## Approche bayésienne.

The reverend Thomas Bayes 1702-1761



*Essay towards solving a problem in the doctrine of chances*  
published in the *Philosophical Transactions of the Royal Society of London* in  
1764.

## Approche fréquentiste versus approche bayésienne.

	Approche fréquentiste Pr(résultats observés Ho)	Approche Bayésienne Pr (Ha   résultats observés)
Information	p - valeur	Valeur prédictive

## Théorème de Bayes

- Le théorème de Bayes permet d'inverser les probabilités. C'est-à-dire que si l'on connaît les conséquences (effets) d'une cause, l'observation des effets (signes cliniques, anomalie biologique, image radiologique, ...) permet de remonter aux causes (maladies).
- Il permet de calculer la probabilité des causes.

$$\Pr (M_1|S) = \frac{\Pr (S|M_1) * \Pr (M_1)}{\sum_{i=1}^k \Pr (S|M_i) * \Pr (M_i)}$$

## Exemple d'inférence bayésienne: d'où vient ce biscuit au chocolat?



40

Boîte A: 30 biscuits au chocolat ( $p = 75\%$ ) et 10 ordinaires.

$p = 0.0377$  (approche fréquentiste)



40

Boîte B: 20 biscuits au chocolat ( $p = 50\%$ ) et 20 ordinaires

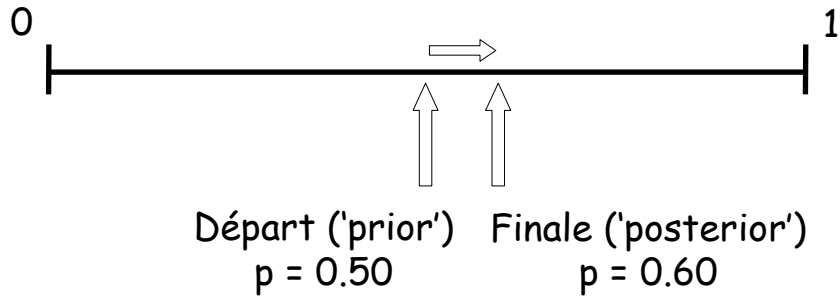
## Exemple d'inférence bayésienne: d'où vient ce biscuit au chocolat?

- La formule de Bayes nous donne:

$$\begin{aligned} \Pr(\text{Boîte}_A | BC) &= \frac{\Pr(BC | \text{Boîte}_A) * \Pr(\text{Boîte}_A)}{\Pr(BC | \text{Boîte}_A) * \Pr(\text{Boîte}_A) + \Pr(BC | \text{Boîte}_B) * \Pr(\text{Boîte}_B)} \\ &= \frac{0.75 * 0.5}{0.75 * 0.5 + 0.5 * 0.5} = 0.60 \end{aligned}$$

- Avant de regarder le biscuit, notre probabilité d'avoir choisi la boîte A était  $P(\text{Boîte}_A)$ , soit 0.50
- Après l'avoir regardé, nous révisons cette probabilité à  $P(\text{Boîte}_A | BC)$ , qui est 0.60

## Approche Bayésienne



### Exemple d'inférence Bayésienne:

Tirage	Séquence des résultats	Outcome	Probabilité finale (Boîte A)
1	Biscuit au chocolat (BC)	Succès	0.600
2	BC	Succès	0.692
3	Biscuit ordinaire (BO)	Echec	0.529
4	BC	Succès	0.628
5	BC	Succès	0.717
6	BC	Succès	0.792
7	BO	Echec	0.655
8	BC	Succès	0.740
9	BO	Echec	0.587
10	BC	Succès	0.681
11	BC	Succès	0.762
12	BC	Succès	0.828
13	BO	Echec	0.706
14	BC	Succès	0.783
15	BC	Succès	0.844
16	BC	Succès	0.890
17	BC	Succès	0.924
18	BC	Succès	0.948
19	BC	Succès	0.965
20	BO	Echec	0.932
21	BC	Succès	0.954



### Exemple d'inférence Bayésienne:

Tirage	Séquence des résultats	Outcome	Probabilité finale (Boite A)
22	BC	Succès	0.969
23	BO	Echec	0.939
24	BC	Succès	0.958
25	BC	Succès	0.972
26	BC	Succès	0.981
27	BO	Echec	0.963
28	BC	Succès	0.975
29	BC	Succès	0.983
30	BO	Echec	0.967
31	BC	Succès	0.978
32	BC	Succès	0.985
33	BC	Succès	0.990
34	BC	Succès	0.993
35	BO	Echec	0.987
36	BC	Succès	0.991

$27/36 = 75\%$  succès  $\rightarrow$  0.991 probabilité finale

Je suis certain à 99.1% que la boite A contient 75 % de biscuits au chocolat

### En résumé...

- A la fin de l'essai randomisé, je suis sûr avec un niveau de confiance de 99.1 % ( $\neq$  intervalle de confiance), que le nouveau traitement est bénéfique chez 75 % des patients.
- L'intervalle de confiance à 95 % des 75 % de bonnes réponses est  $\pm 14\%$  (61 % à 89 %) pour  $n = 36$

## Bayesian approach

Prior + Evidence = Beliefs in how the world is



Beliefs + Utilities/Costs = Decisions about I should do



Decisions + [Stuff] = Behavior

Recent research in behavioral economics says much that our behavior does not follow traditional decision theory

Probabilité a priori + Evidence = Croyance  
L'évidence seule n'influence pas la croyance



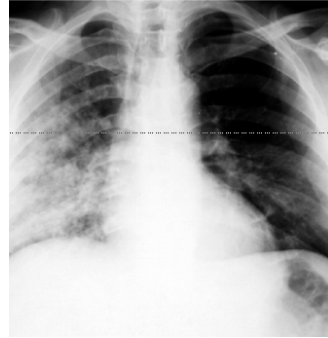
Probabilité a priori fortement négative en faveur des OVNIIs  
+ Evidence faible

---

Très faible croyance aux OVNIIs

## La probabilité a priori influence aussi notre croyance à un diagnostic

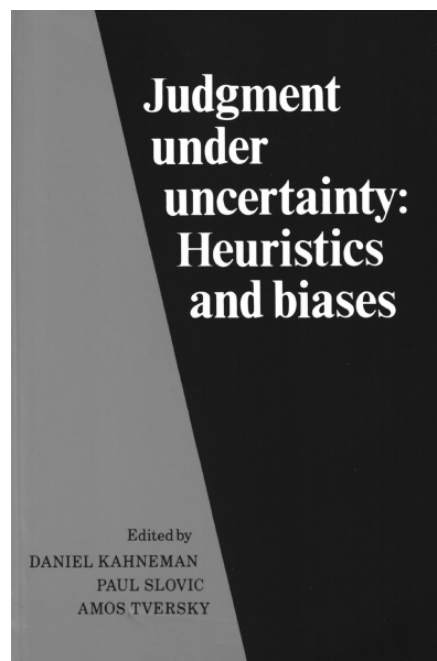
- H 68 ans, fumeur.  
Depuis 5 jours: fièvre,  
dyspnée et expectorations  
jaunes
- AngioCT montre « 2  
emboles subsegmentaires »



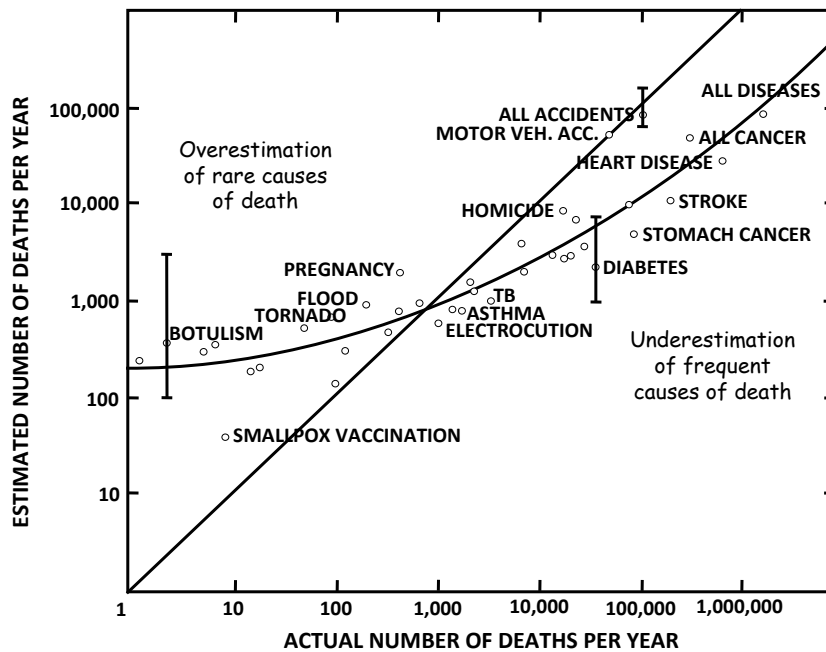
Probabilité a priori fortement négative en faveur de l'EP  
+ Evidence faible

---

Très faible croyance au diagnostic d'EP



## OBJECTIVE AND SUBJECTIVE RISK ESTIMATION



Intensive Care Med  
DOI 10.1007/s00134-015-3814-1

EDITORIAL



Christian Mélot

### To score or not to score during triage in the emergency department?

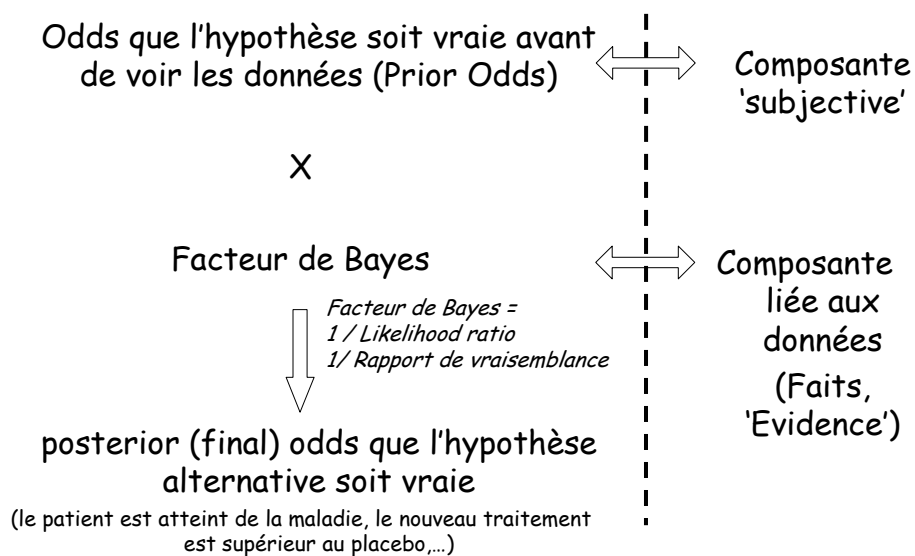
Received: 6 April 2015  
Accepted: 7 April 2015  
© Springer-Verlag Berlin Heidelberg and ESICM 2015

C. Mélot (✉)  
Head of the Emergency Medicine Department, Erasme University Hospital, Lennik Road 808, 1070 Brussels, Belgium  
e-mail: cmelot@ulb.ac.be  
Tel.: +32-2-555.33.82

On the other hand, the Early Warning Scoring System (EWSS), first introduced in 1997 at the James Paget University Hospital in Great Yarmouth (UK), was developed to detect deteriorating physiological parameters in ward patients [9]. More recently, the National Early Warning Score (NEWS), developed in conjunction with the Royal College of Physicians of London, has been more rigorously tested and performs better than any of the 33 published systems commonly in use [10]. It has a good ability to discriminate at risk of cardiac arrest death or

# Facteur de Bayes (Bayes factor)

## Théorème de Bayes



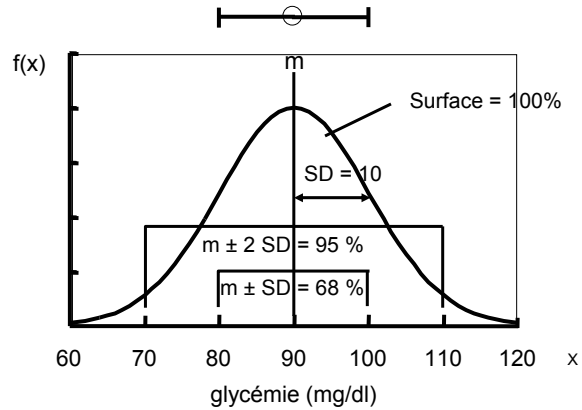
## Calibrage des rapports de vraisemblance (RVs)

Force (poids) des faits (evidence)	RV (LR)	Probabilité finale pour une probabilité a priori égale à		
		25%	50%	75%
Zero	1	0.25	0.50	0.75
Faible	5	0.62	0.83	0.94
Modérée	10	0.77	0.91	0.97
Forte	20	0.83	0.95	0.98
Très forte	40	0.93	0.975	0.99
Très très forte	80	0.96	0.99	0.996

LIMITES DE LA NORMALE POUR UN TEST

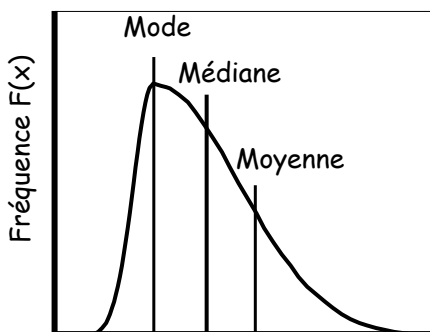
## Distribution gaussienne

Glycémie (mg/100ml):  $90 \pm 10$  (moyenne  $\pm$  SD)

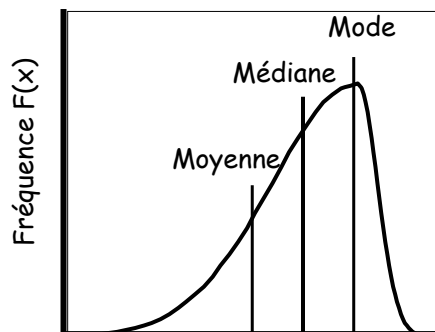


Limites pour la glycémie: 70 à 110 mg/dl  
(Intervalle de prédiction à 95 %)

## Distributions asymétriques (non gaussiennes)



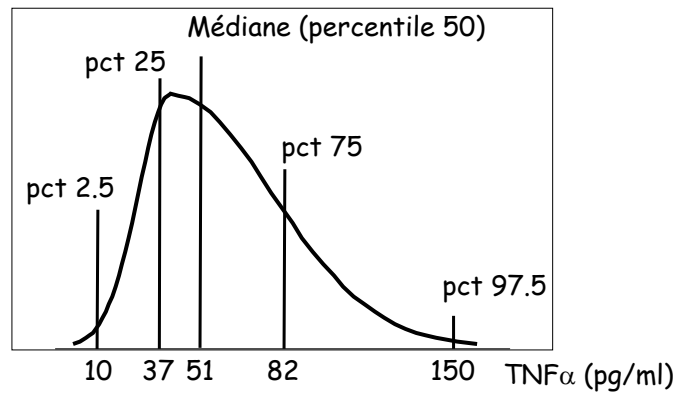
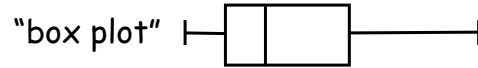
Skewness > 0.5  
(Coefficient d'asymétrie)



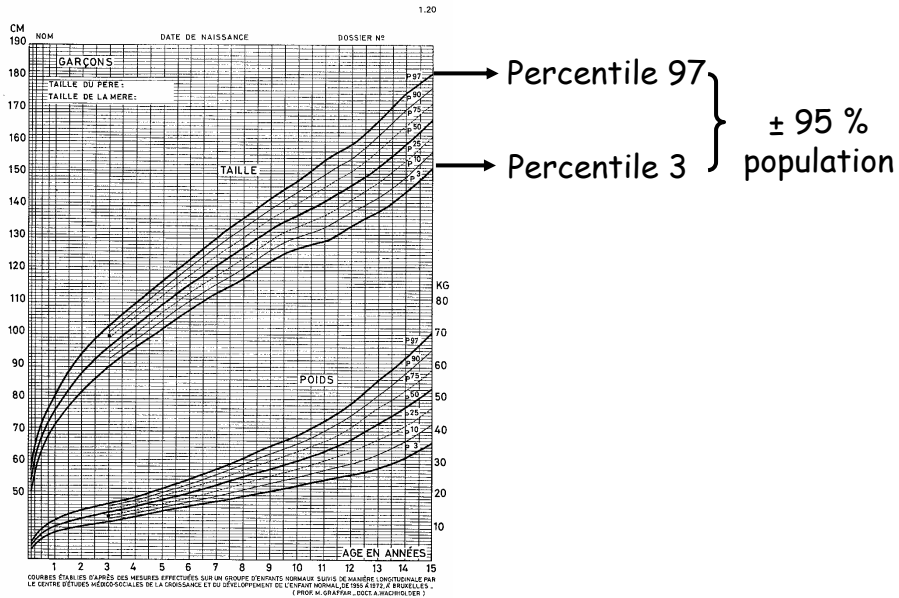
Skewness < - 0.5

## Distribution asymétrique: percentiles

TNF $\alpha$  (pg/ml) : 51 (37 - 82) médiane (espace interquartile)



Limites pour le TNF $\alpha$ : 10 à 150 pg/dl  
(Intervalle de prédiction à 95 %)





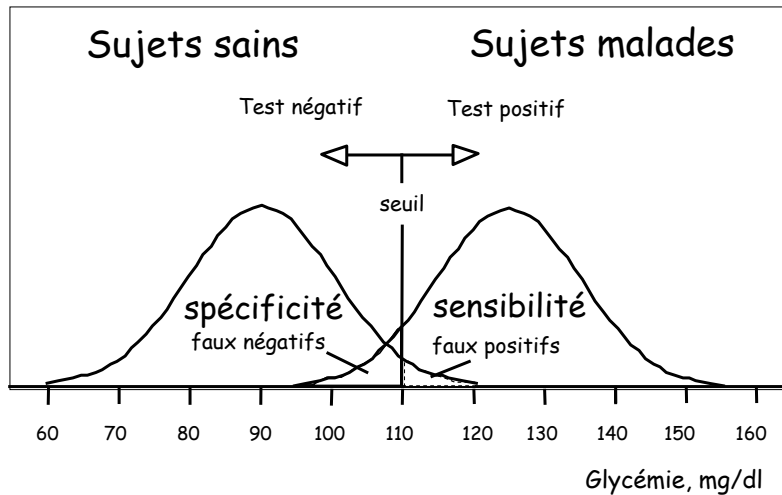
## TESTS DIAGNOSTIQUES

### PRINCIPES D'UN TEST DIAGNOSTIQUE

- Comme médecin, nous prenons des décisions dans un climat d'incertitude. Nous estimons la vraisemblance d'une maladie et nous décidons de la nécessité de tests complémentaires pour confirmer un diagnostic.
- L'information que nous utilisons pour modifier notre hypothèse diagnostique est l'anamnèse, l'examen clinique et les résultats de tests diagnostiques.

## TEST DIAGNOSTIQUE

Spécificité + Faux Positifs = 1    Sensibilité + Faux Négatifs = 1



### Test diagnostique: sensibilité et spécificité Gold standard (par ex.: histologie)

		Tumeur	
		Maligne	Bénigne
Mammographie	Positive	<b>297</b> (Sens: 0.85)	<b>108</b> (FP: 0.30)
	Négative	<b>51</b> (FN: 0.15)	<b>252</b> (Spéc: 0.70)
		348	360

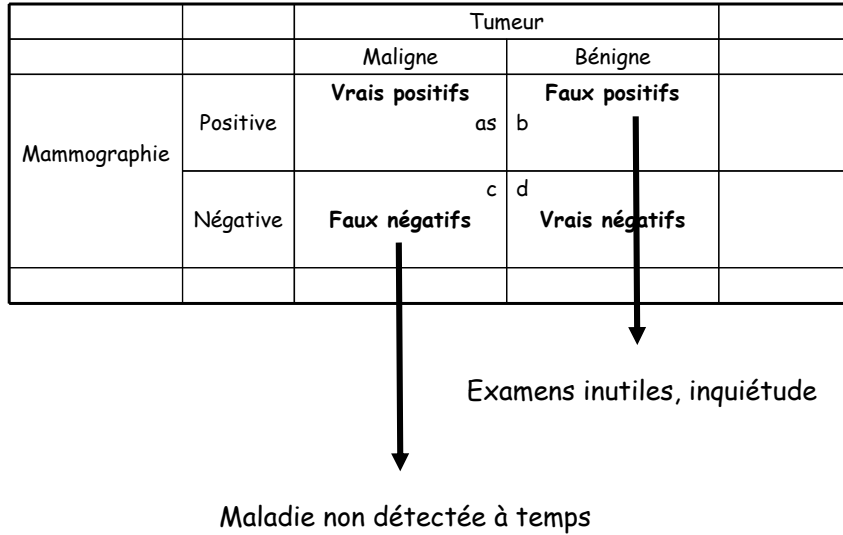
Sensibilité =  $297/348 = 0.85$  (85 %)

Faux négatifs =  $51/348 = 0.15$  (15 %)

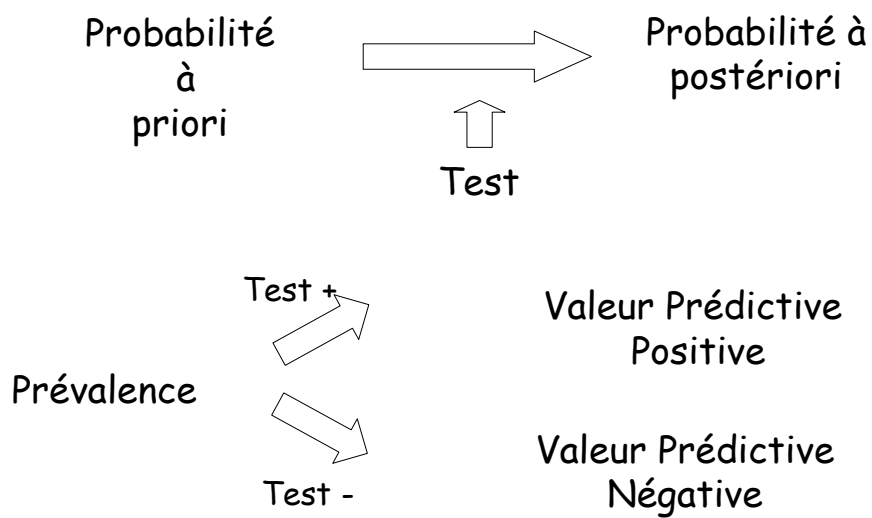
Spécificité =  $252/360 = 0.70$  (70 %)

Faux positifs =  $108/360 = 0.30$  (30 %)

Valeur diagnostic des tests diagnostiques  
Gold standard (par ex.: histologie)



Probabilités a priori et a posteriori



**Test diagnostique: valeurs prédictives (VP)**  
Gold standard (par ex.: histologie)

		Tumeur		
		Maligne	Bénigne	
Mammographie	Positive	<b>85</b> (Sens: 0.85)	<b>270</b> (FP: 0.30)	355
	Négative	<b>15</b> (FN: 0.15)	<b>630</b> (Spéc: 0.70)	645
		100	900	1000

Prévalence = 100/1000 = 0.10 (10 %)

Valeur prédictive positive (VPP) :  
VPP = 85/355 = 0.24 (24 %)

Valeur prédictive négative (VPN) :  
VPN = 630/645 = 0.98 (98 %)

**Test diagnostique: valeurs prédictives (VP)**  
Gold standard (par ex.: histologie)

		Tumeur		
		Maligne	Bénigne	
Mammographie	Positive	<b>127</b> (Sens: 0.85)	<b>255</b> (FP: 0.30)	382
	Négative	<b>23</b> (FN: 0.15)	<b>595</b> (Spéc: 0.70)	618
		150	850	1000

Prévalence = 150/1000 = 0.15 (15 %)

Valeur prédictive positive (VPP) :  
VPP = 127/382 = 0.33 (33 %)

Valeur prédictive négative (VPN) :  
VPN = 595/618 = 0.96 (96 %)

Les VPP et VPN  
changent avec  
la prévalence

Rapport de vraisemblance  
(Likelihood Ratio: LR)

		Tumeur		
		Maligne	Bénigne	
Mammographie	Positive	<b>85</b> (Sens: 0.85)	<b>270</b> (FP: 0.30)	355
	Négative	<b>15</b> (FN: 0.15)	<b>630</b> (Spéc: 0.70)	645
		100	900	1000

Prévalence = 100/1000 = 0.10 (10 %)

Rapport de Vraisemblance Test + (LR+):

$$p_a / p_b = 0.85/0.30 = 2.83$$

Rapport de Vraisemblance Test - (LR-):

$$p_c / p_d = 0.15/0.70 = 0.21$$

Rapport de vraisemblance  
(Likelihood Ratio: LR)

		Tumeur		
		Maligne	Bénigne	
Mammographie	Positive	<b>127</b> (Sens: 0.85)	<b>255</b> (FP: 0.30)	382
	Négative	<b>23</b> (FN: 0.15)	<b>595</b> (Spéc: 0.70)	618
		150	850	1000

Prévalence = 150/1000 = 0.15 (15 %)

Rapport de Vraisemblance Test + :

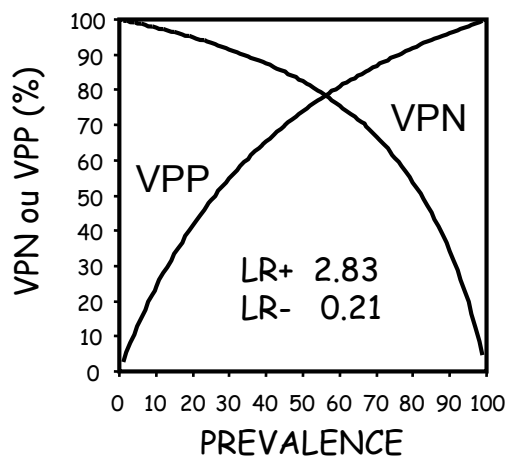
$$p_a / p_b = 0.85/0.30 = 2.83$$

Rapport de Vraisemblance Test - :

$$p_c / p_d = 0.15/0.70 = 0.21$$

Les LR+ et LR-  
ne changent pas  
avec la  
prévalence

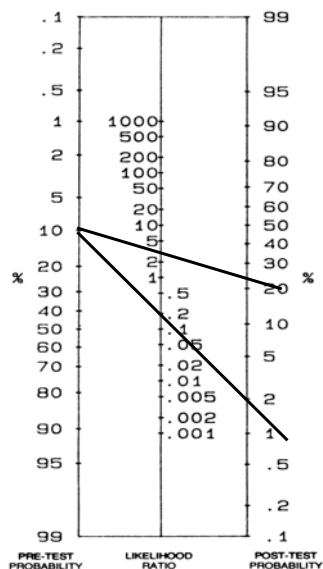
## Valeurs prédictives d'un test en fonction de la prévalence



Sensibilité 85 %  
Spécificité 70 %  
(mammographie)

### Rapport de Vraisemblance (Likelihood Ratio LR)

- Probabilité pré-test:
  - Prévalence = 0.10
  
- Mammographie:
  - $LR+ = 2.83$
  
  - 24 % de chance d'avoir une tumeur maligne (VPP = 24 %)
  
  - $LR- = 0.21$
  
  - 2 % de chance d'avoir une tumeur maligne (VPN = 98 %)



*Nomogram for Bayes's theorem, Fagan TJ, NEJM 1975;293:257*

## Multiplicativité du rapport de vraisemblance (LR)

### Probabilité post-test d'être malade

- Calcul du odds post-test

$$\text{Odds}_{\text{post-test}} = \text{LR+} \times \text{Odds}_{\text{pré-test}}$$

(LR+  $\gg$  1 en faveur de la maladie  
 $\approx$  0 en faveur de l'absence de la maladie  
= 1 pas d'information au sujet de la maladie)

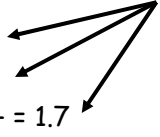
- Transformation en probabilité du odds post-test:

$$\text{Prob (maladie)} = \frac{\text{Odds}_{\text{post-test}}}{1 + \text{Odds}_{\text{post-test}}}$$

Gurney JW. Radiology 1993;186:405-413

## Exemple

- Homme 81 ans, tabagique (60 paquets-année), présente un nodule de 2.5 cm au CT-scan thoracique.
- Quelle est la probabilité qu'il s'agisse d'un cancer?

- Prévalence dans le milieu clinique: 0.40
  - Odds<sub>pré-test</sub> =  $0.4 / (1-0.4) = 0.67$
  - LR+ sont: 70-83 ans → LR+ = 5.7  
Tabagisme → LR+ = 3.9  
Nodule 2.3-3.2 cm → LR+ = 1.7
  - Odds<sub>post-test</sub> =  $0.67 \times 5.7 \times 3.9 \times 1.7 = 25.3$
  - Probabilité d'avoir un cancer:  $p = 25.3 / (1+25.3) = 0.96$
- Facteurs de risques indépendants (≠ Trop & CK<sub>MB</sub>)
- 

Gurney JW. Radiology 1993;186:405-413

## Exemple

- Homme 81 ans, tabagique (60 paquets-année), présente un nodule de 2.5 cm au CT-scan thoracique.
  - Quelle est la probabilité qu'il s'agisse d'un cancer?
- Si TEP positive pour le nodule → LR+ = 4
  - → Odds<sub>post-test</sub> =  $25.3 \times 4 = 101.2$
  - →  $p = 101.2 / (1+101.2) = 0.99$

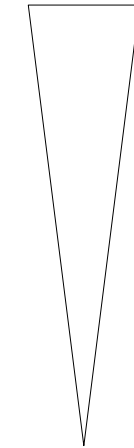
Gurney JW. Radiology 1993;186:405-413



## Propriétés d'un test

- Propriétés intrinsèques:
  - Sensibilité
  - Spécificité
- Propriétés extrinsèques:
  - Statiques:
    - Valeur prédictive positive
    - Valeur prédictive négative
  - Dynamiques:
    - Rapport de vraisemblance d'un test positif
    - Rapport de vraisemblance d'un test négatif

Population



Individu

## Pourquoi dit-on d'un test qu'il est très spécifique de la maladie ou très sensible pour exclure la maladie?

- Des valeurs très élevées de sensibilité et de spécificité sont utiles mais pas pour les raisons que l'on peut invoquer dans une approche simpliste qui consisterait à dire:
  - Une spécificité élevée est utile pour exclure le diagnostic de la maladie
  - Une sensibilité élevée est utile pour affirmer le diagnostic.
- En fait,
  - Plus le test est sensible, moins il comportera de faux négatifs (Faux Négatifs = 1 - Sensibilité).  
Dès lors, un test négatif permettra d'exclure la maladie.  
Les auteurs anglo-saxons ont un moyen mnémotechnique: « SnNout » (when a test or a sign has a high Sensitivity, a Negative result rules Out the diagnosis).  
Le test est dit « très sensible pour exclure la maladie ».

## Pourquoi dit-on d'un test qu'il est très spécifique de la maladie ou très sensible pour exclure la maladie?

- En fait,
  - Plus le test est spécifique, moins il comportera de faux positifs (Faux Positifs = 1 - Spécificité).  
Dès lors, un test positif permettra de confirmer la maladie.  
Les auteurs anglo-saxons ont un moyen mnémotechnique: « SpPin » (when a test or a sign has a high Specificity, a Positive result rules In the diagnosis).  
Le test est dit « très spécifique de la maladie ».

*Référence: DL Sackett et al. Evidence-Based Medicine. 2nd edition, Churchill Livingstone, 2000 pp74-75.*

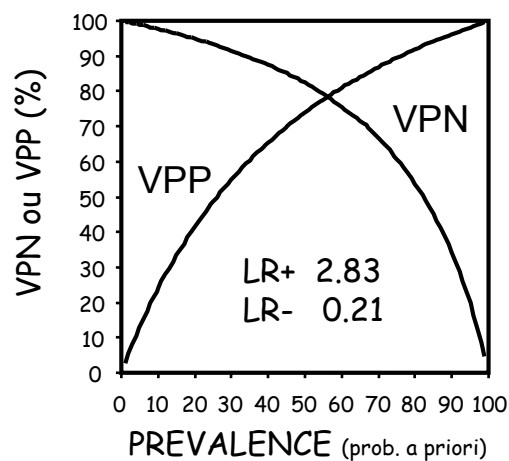
## Exemple: utilisation en pratique d'un test diagnostique: dosage des D-dimères pour exclure le diagnostic de maladie thrombo-embolique

- DIAGNOSTIC POSITIF (faible spécificité et faible VPP)
  - Le dosage des D-Dimères est inutile dans le diagnostic positif d'embolie pulmonaire ou de TVP.
  - La présence de D-Dimères ne signifie pas obligatoirement présence de thrombose. En effet, on peut observer une élévation des D-Dimères en post-opératoire, au cours de la résorption d'un hématome, au cours de la grossesse, dans les phénomènes inflammatoires qui entraînent une augmentation du taux de fibrinogène. Le taux augmente aussi avec l'âge.
- DIAGNOSTIC D'EXCLUSION (forte sensibilité et forte VPN)
  - A défaut de poser le diagnostic positif, en raison de leur excellente valeur prédictive négative les D-Dimères peuvent être utilisés pour identifier avec le plus de certitude possible, les patients qui ne présentent pas de thrombose.

Exemple: utilisation en pratique d'un test diagnostique: dosage des D-dimères pour exclure le diagnostic de maladie thrombo-embolique

- Cependant il faut établir une distinction entre les patients hospitalisés et les patients ambulatoires.
- En effet, une grande majorité de patients hospitalisés ont des valeurs élevées de D-Dimères, en l'absence de maladie thrombo-embolique.
- La valeur prédictive négative est donc meilleure pour les patients en ambulatoire.

Résumé



Sensibilité 85 %  
Spécificité 70 %  
(mammographie)

- Stratégie d'application répétée du même test:
  - Stratégie en série: contrôle des tests +
  - Stratégie en parallèle: contrôle des tests -
- Combinaison de plusieurs tests

- Stratégie en série

### Répétition du même test: stratégie en série

	Malades (M+)	Non malades (M-)		
Test +	<b>90</b> (Sens: 0.90)	<b>90</b>	180	VP+ = 50 %
	a	b		
Test -	<b>10</b>	<b>810</b> (Spéc: 0.90)	900	VP- = 98.78 %
	c	d		
Prévalence 10 %	100	900	1000	

### Répétition du test chez les positifs

	Malades (M+)	Non malades (M-)		
Test +	<b>81</b> (Sens: 0.90)	<b>9</b>	90	VP+ = 90 %
	a	b		
Test -	<b>9</b>	<b>81</b> (Spéc: 0.90)	90	VP- = 90 %
	c	d		
Prévalence 50 %	90	90	180	

### Répétition du même test: stratégie en série

	Malades (M+)	Non malades (M-)		
Test +	<b>73</b> (Sens: 0.90)	<b>1</b>	74	VP+ = 98.78 %
	a	b		
Test -	<b>8</b>	<b>8</b> (Spéc: 0.90)	16	VP- = 50 %
	c	d		
Prévalence 90 %	81	9	90	

### Résultat final: test réalisé 3 fois en série

	Malades (M+)	Non malades (M-)		
Tous Test +	<b>73</b> (Sens finale: 0.73)	<b>1</b> (minimise les FP)	74	VP+ = 98.78 %
	a	b		
1 des Test -	<b>27</b>	<b>899</b> (Spéc finale:0.999)	926	VP- = 97 %
	c	d		
Prévalence 10 %	100	900	1000	

- Stratégie en parallèle

Répétition du même test: stratégie en parallèle

	Malades (M+)	Non malades (M-)		
Test +	<b>810</b> (Sens: 0.90)	<b>10</b>	820	VP+ = 98.78 %
	a	b		
Test -	<b>90</b>	<b>90</b> (Spéc: 0.90)	180	VP- = 50 %
	c	d		
Prévalence 90 %	900	100	1000	

Répétition du test chez les négatifs

	Malades (M+)	Non malades (M-)		
Test +	<b>81</b> (Sens: 0.90)	<b>9</b>	90	VP+ = 90 %
	a	b		
Test -	<b>9</b>	<b>81</b> (Spéc: 0.90)	90	VP- = 90 %
	c	d		
Prévalence 50 %	90	90	180	

### Répétition du même test: stratégie en parallèle

	Malades (M+)	Non malades (M-)		
Test +	<b>8</b> (Sens: 0.90)	<b>8</b>	16	VP+ = 50 %
		a b		
Test -	<b>1</b>	<b>73</b> (Spéc: 0.90)	74	VP- = 98.78 %
		c d		
Prévalence 10 %	9	81	90	

### Résultat final: test réalisé 3 fois en parallèle

	Malades (M+)	Non malades (M-)		
1 des Test +	<b>899</b> (Sens finale: 0.999)	<b>27</b>	926	VP+ = 97 %
		a b		
Tous Test -	<b>1</b> (minimise les FN)	<b>73</b> (Spéc finale:0.73)	74	VP- = 99.9 %
		c d		
Prévalence 90 %	900	100	1000	

- Stratégie d'application répétée du même test:
  - Stratégie en série: contrôle des tests +
  - Stratégie en parallèle: contrôle des tests -
  
- Combinaison de plusieurs tests

Combinaison de plusieurs tests: appliquer le plus sensible en premier

HIV	Malades (M+)	Non malades (M-)		
Test + <b>ELISA</b>	<b>291</b> (Sens: 0.97)	<b>1999</b>	2290	VP+ = 12.7 %
	a	b		
Test -	<b>9</b>	<b>997,701</b> (Spéc: 0.998)	997,710	VP- = 99.99 %
	c	d		
Prévalence 0.03 %	300	999,700	1,000,000	

Répétition du test chez les positifs (stratégie en série)

HIV	Malades (M+)	Non malades (M-)		
Test + <b>Western Blot</b>	<b>276</b> (Sens: 0.95)	<b>0</b>	276	VP+ = 99.93 %
	a	b		
Test -	<b>15</b>	<b>1999</b> (Spéc: 0.999)	2014	VP- = 99.28 %
	c	d		
	291	1999	2290	

Combinaison de plusieurs tests

HIV	Malades (M+)	Non malades (M-)		
<b>ELISA et WB +</b>	<b>276</b> (Sens finale: 0.921)	<b>0</b> (minimise les FP)	276	VP+ = 99.93 %
	a	b		
<b>ELISA ou WB -</b>	<b>24</b>	<b>999,700</b> (Spéc finale:0.999)	997,724	VP- = 99.99 %
	c	d		
Prévalence 0.03 %	300	999,700	1,000,000	

Exemple: donneurs de sang



Combinaison de plusieurs tests: appliquer le plus sensible en premier

HIV	Malades (M+)	Non malades (M-)		
Test + <b>ELISA</b>	<b>582</b> (Sens: 0.97)	<b>1</b>	583	VP+ = 99.86 %
	a	b		
Test -	<b>18</b>	<b>399</b> (Spéc: 0.998)	417	VP- = 95.68 %
	c	d		
Prévalence 60 %	600	400	1000	

Répétition du test chez les négatifs (stratégie en parallèle)

HIV	Malades (M+)	Non malades (M-)		
Test + <b>Western Blot</b>	<b>17</b> (Sens: 0.95)	<b>0</b>	17	VP+ = 99.76 %
	a	b		
Test -	<b>1</b>	<b>399</b> (Spéc: 0.999)	400	VP- = 99.77 %
	c	d		
	18	399	417	

Combinaison de plusieurs tests

HIV	Malades (M+)	Non malades (M-)		
<b>ELISA ou WB +</b>	<b>599</b> (Sens finale: 0.998)	<b>1</b>	600	VP+ = 99.85 %
	a	b		
<b>ELISA et WB -</b>	<b>1</b> (minimise les FN)	<b>399</b> (Spéc finale:0.998)	400	VP- = 99.77 %
	c	d		
Prévalence 60 %	600	400	1000	

Exemple: toxicomanes

## Diagnostic de cancer: nombre minimum de biopsies nécessaires

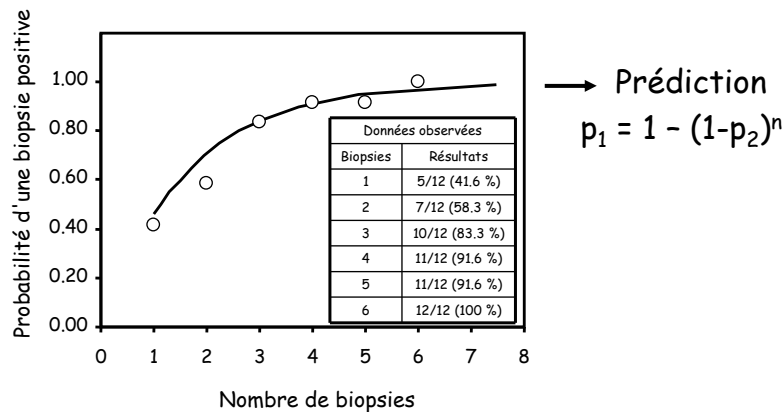
$$p_1 = 1 - \underbrace{(1-p_2)^n}$$

probabilité de n'avoir aucune biopsie positive sur n prélèvements

- $p_2$  = probabilité d'avoir un résultat positif par biopsie (par ex.:  $p_2 = 32/70 = 0.46$ )
- $P_1$  = probabilité d'avoir au moins un résultat positif sur n biopsies (e.g.:  $p_1 = 0.99$ )  
 $n = \log(1-p_1)/\log(1-p_2)$   
 $n = \log(1-0.99)/\log(1-0.46) = 7.47$   
 $n = 8$  biopsies

*Gilman MJ, Wang KP Am. Rev. Respir. Dis. 1980;122:721-724*

## Diagnostic de cancer: nombre minimum de biopsies nécessaires



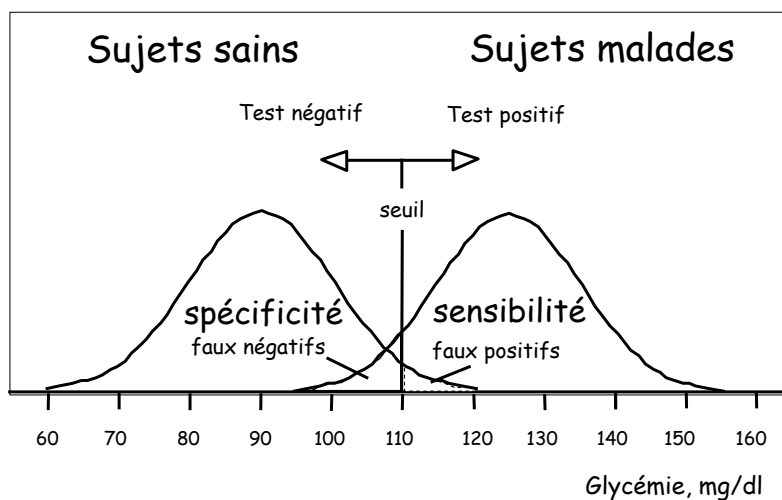
*Gilman MJ, Wang KP Am. Rev. Respir. Dis. 1980;122:721-724*

L'intervalle de prédiction à 95% donne-t-il le meilleur seuil pour un test?

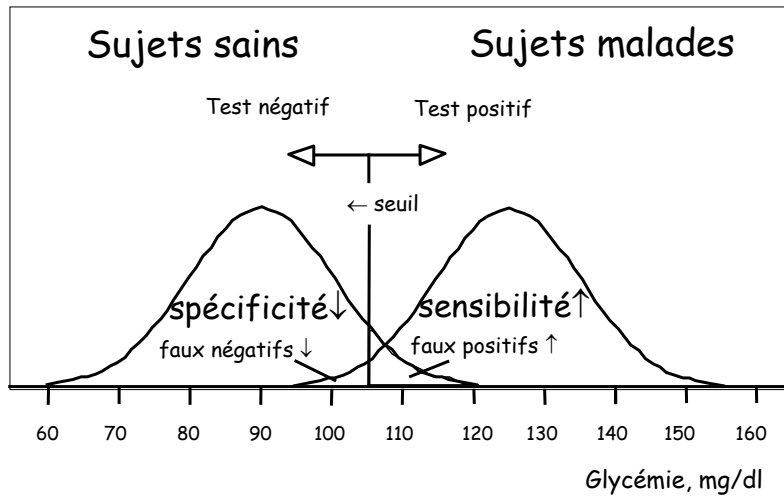
- Courbe ROC (Receiver Operating Characteristic)

### TEST DIAGNOSTIQUE

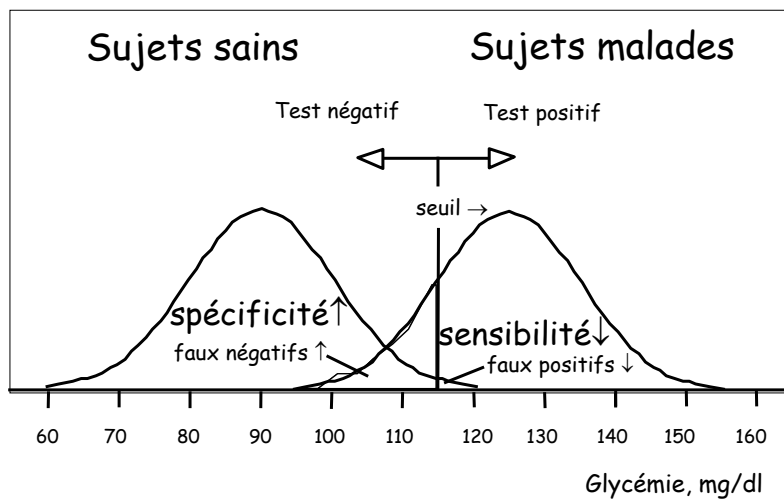
Spécificité + Faux Positifs = 1    Sensibilité + Faux Négatifs = 1



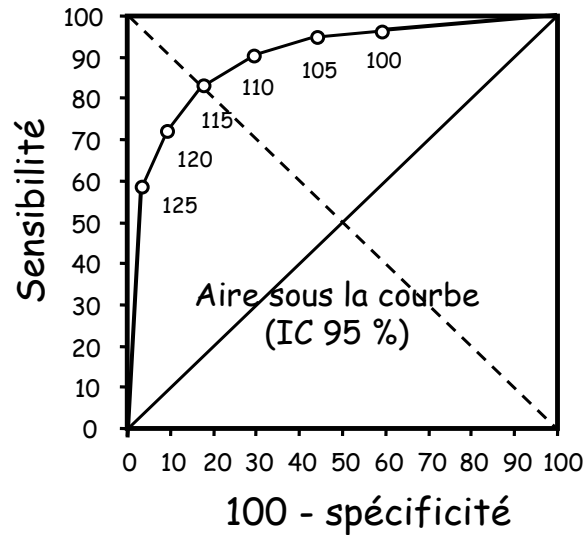
## TEST DIAGNOSTIQUE: CHANGEMENT DU SEUIL



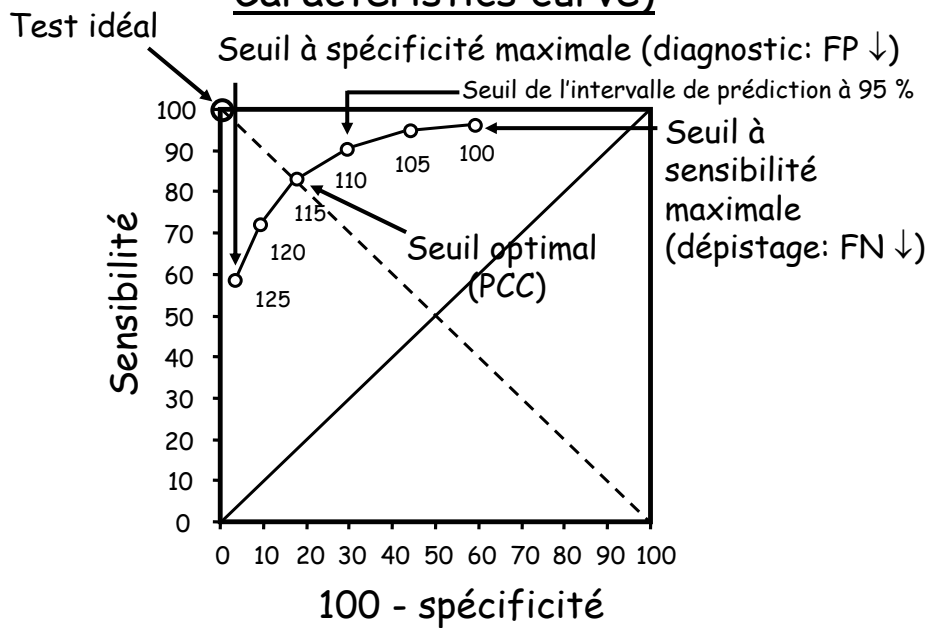
## TEST DIAGNOSTIQUE: CHANGEMENT DU SEUIL



## COURBE ROC (Receiver Operating Characteristics curve)



## COURBE ROC (Receiver Operating Characteristics curve)



Test diagnostique: pourcentage de classification correcte (PCC)

Gold standard (par ex.: histologie)

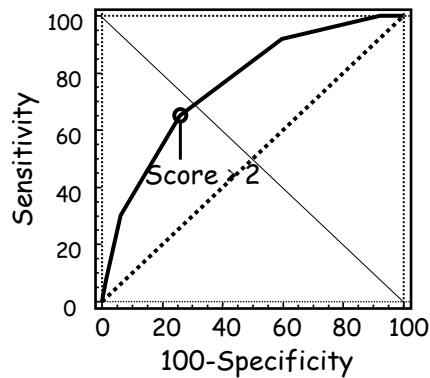
		Tumeur		
		Maligne	Bénigne	
Mammographie	Positive	<b>85</b> (Sens: 0.85) a	<b>270</b> (FP: 0.30) b	355
	Négative	<b>15</b> (FN: 0.15) c	<b>630</b> (Spéc: 0.70) d	645
		100	900	1000

Pourcentage de classification correcte:

$$(85 + 630)/1000 = 0.715 (71.5 \%)$$

Diagnostic ou Dépistage?

	Diagnostic	Dépistage
But	Identifier avec certitude la maladie chez le patient	Eviter de manquer des cas de maladie dans la population
Coût unitaire	Elevé	Bas
Acceptabilité	Faible	Forte (rapide et peu invasif)
Risque	Eventuellement	Nul ou très faible
Propriété du test	Minimiser les FP Maximiser la VPP Spécificité élevée	Minimiser les FN Maximiser la VPN Sensibilité élevée



## Embolie pulmonaire (Score de Genève)

Prevalence 19.2 %

Area: 0.759 (0.709 - 0.804)

Criterion	Sens. (95% CI)	Spec. (95% CI)	LR+	LR-	PV+	PV-	
≥ 0	100 (94.4 - 100)	0 (0 - 1.4)	1.00		19.2		100
> 0	100 (94.4 - 100)	7.7 (4.8 - 11.5)	1.08	0.00	20.4	100	108
> 1	92.3 (82.9 - 97.4)	39.8 (33.9 - 45.8)	1.53	0.19	26.7	95.6	132
> 2*	66.2 (53.4 - 77.4)	73.7 (68.1 - 78.8)	2.52	0.46	37.4	90.2	140
> 3	29.2 (18.6 - 41.8)	93.4 (89.8 - 96.1)	4.45	0.76	51.4	84.8	123
> 4	7.7 (2.6 - 17.1)	98.5 (96.3 - 99.6)	5.27	0.94	55.6	81.8	106
> 5	1.5 (0 - 5.6)	99.6 (98.0 - 99.9)	4.22	0.99	50.0	81.0	101
> 6	0 (0 - 5.6)	100 (98.6 - 100)	∞	1.00		80.8	100

## Test diagnostique et décision médicale

- Il y a 2 approches statistiques pour l'évaluation des biomarqueurs:
  - Modélisation du risque (disease outcome) par régression logistique (ou régression de Cox)
    - le biomarqueur est utile si il a un lien avec le risque ( $p < 0.05$ )
  - Performance dans la classification malade/non malades (sensibilité, spécificité, valeur prédictive, LR et courbe ROC).
    - le biomarqueur est utile si il a une bonne performance de classification (discriminant)

Approche fréquentiste

Approche bayésienne

## Modélisation du risque: prédictibilité (régression logistique)

- Exemple: cancer de la prostate

Variable	OR	P
Log (PSA)	3.67	< 0.001
Age (yrs)	1.03	0.02
Digital Rectal Examination (DRE) positive	2.69	< 0.001
Prior biopsy (PB)	0.0026	0.04

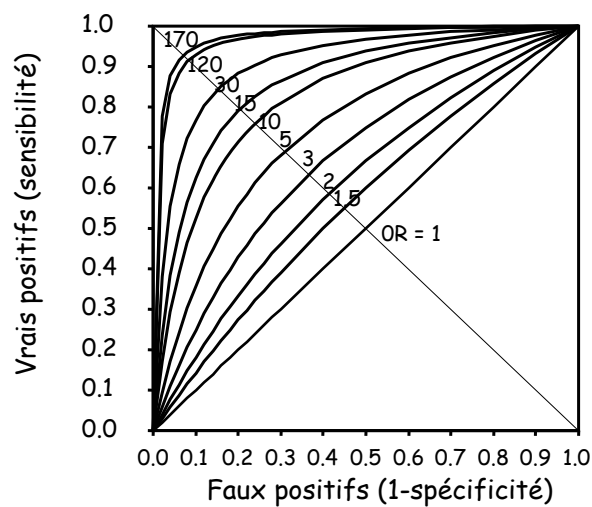
$$y = \frac{e^{-5.94 + 1.30 \text{ Log(PSA)} + 0.03 \text{ Age} + 0.99 \text{ DRE} - 0.037 \text{ PB}}}{1 + e^{-5.94 + 1.30 \text{ Log(PSA)} + 0.03 \text{ Age} + 0.99 \text{ DRE} - 0.037 \text{ PB}}}$$

Cancer = 1

No cancer = 0

Pepe MS et al. Am J Epidemiol 2008;167:362-368

## Courbe ROC et Odds Ratio



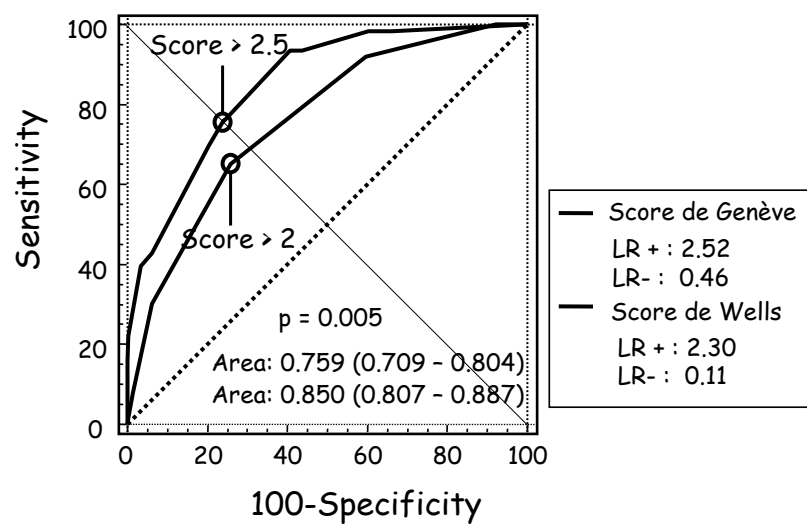
Pepe MS et al. Am J Epidemiol 2004;159:882-890



- Comparaison du pouvoir discriminant de deux tests

### Diagnostic d'embolie pulmonaire aux urgences

333 patients (63 embolies pulmonaires)

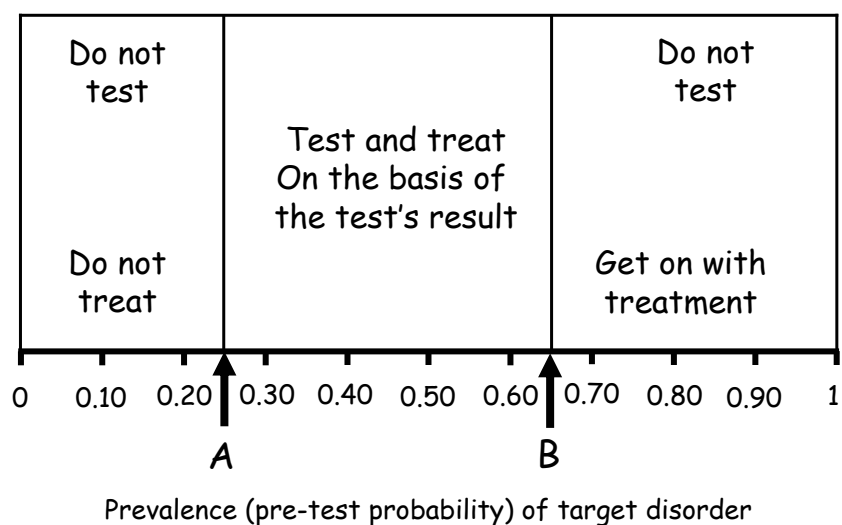


## Tests diagnostiques

- Y a-t-il une comparaison aveugle avec un test standard pour le diagnostic (« gold standard ») (discrimination)
- Le test a-t-il été évalué pour un spectre approprié de patients? (calibration)
- Le test a-t-il été validé dans un second group de patients? (building set; validation set)
- Le test standard de référence a-t-il été appliqué indépendamment du résultat du test? (ex: test négatif)

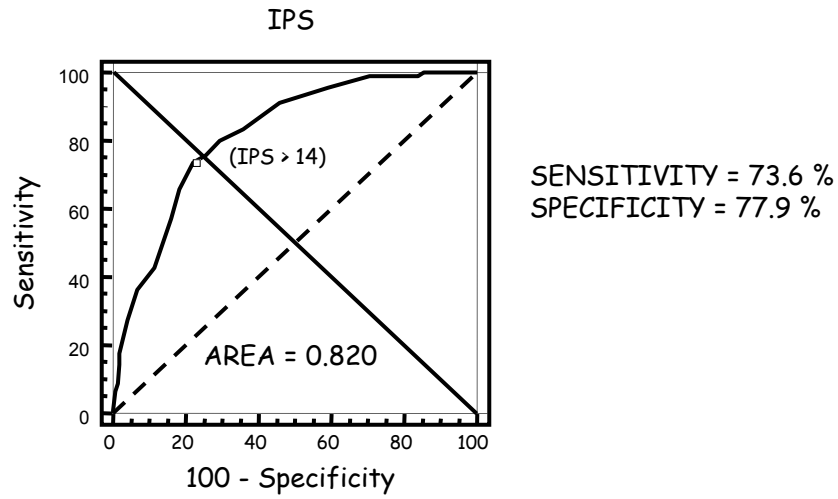
Evidenced-based Medicine : how to practice and to teach it, 4th Ed, 2011 p 84

## Tests diagnostiques



Evidenced-based Medicine : how to practice and to teach it, 4th Ed, 2011 p 84

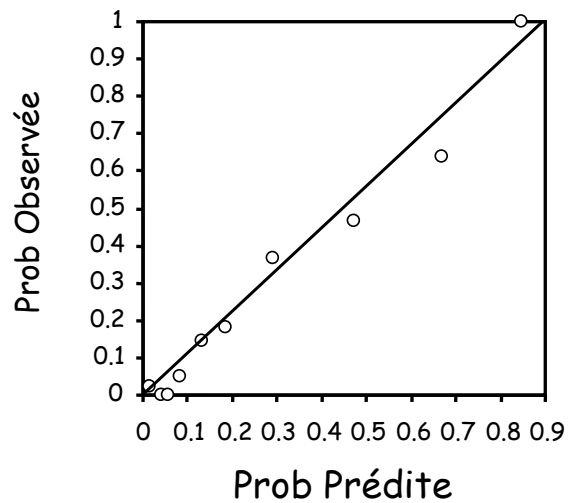
## ROC CURVE ON THE BUILDING DATA SET



PREVALENCE OF INFECTION: 91 / 353 = 25.8 %  
POSITIVE PREDICTIVE VALUE (IPS > 14) = 53.6 %  
NEGATIVE PREDICTIVE VALUE (IPS ≤ 14) = 89.5 %

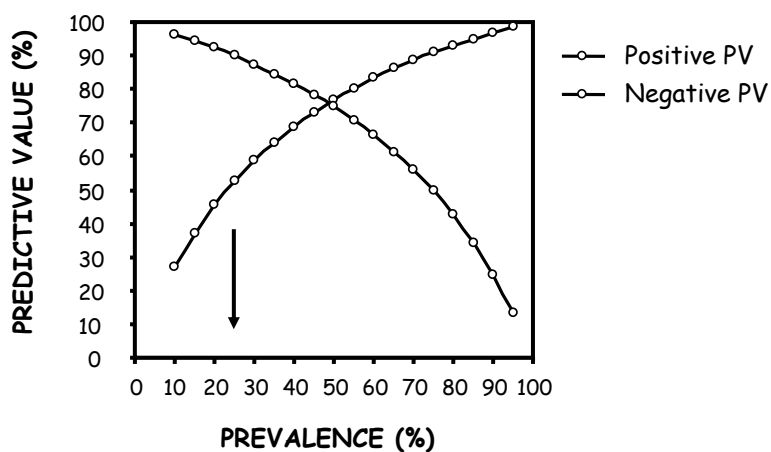
Peres Bota D, Mélot C, Lopes Ferreira F, Vincent JL. Crit Care Med. 2003;31:2579-84.

## CALIBRATION



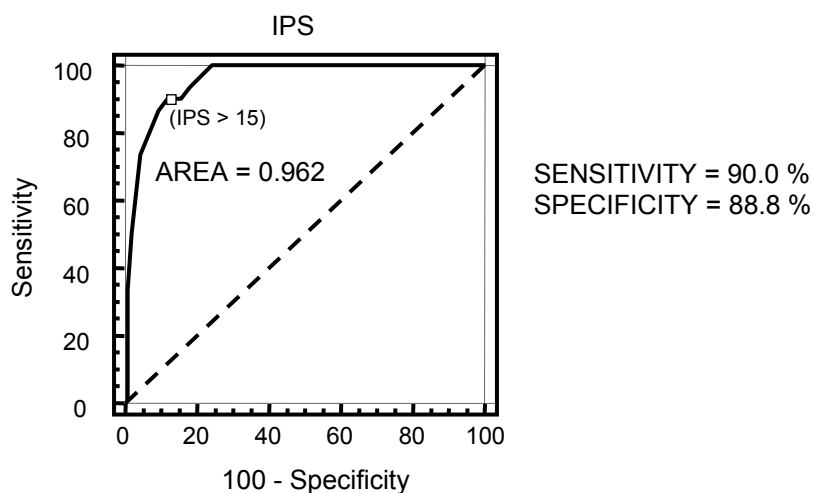
Peres Bota D, Mélot C, Lopes Ferreira F, Vincent JL. Crit Care Med. 2003;31:2579-84.

## PREDICTIVE VALUES OF IPS



Peres Bota D, Mélot C, Lopes Ferreira F, Vincent JL. Crit Care Med. 2003;31:2579-84.

## ROC CURVE ON THE VALIDATION DATA SET



PREVALENCE OF INFECTION: 30 / 110 = 27.3 %  
 POSITIVE PREDICTIVE VALUE (IPS > 15) = 75.0 %  
 NEGATIVE PREDICTIVE VALUE (IPS ≤ 15) = 95.9 %

Peres Bota D, Mélot C, Lopes Ferreira F, Vincent JL. Crit Care Med. 2003;31:2579-84.