

Tests Diagnostiques

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

Introduction

On appelle raisonnement clinique la suite d'étapes (= démarche clinique) aboutissant à l'élaboration d'une décision.

La Démarche clinique est la suite d'observations et d'évaluations (interprétations)

Exemple du professeur : mesure de la flèche hépatique

Raisonnement clinique : 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.

Démarche clinique : 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) S'initier à l'analyse critique d'un article d'évaluation d'un test diagnostique

L'interprétation clinique d'une mesure est assez simple dans le cas de la flèche hépatique : on compare la valeur obtenue à une valeur de référence qu'on a obtenue par analyse statistique (courbe de gauss). Cependant c'est plus difficile pour poser un diagnostic et décider d'une thérapeutique dans un contexte donné puisqu'il y a de nombreux signes cliniques, des résultats d'examens complémentaires...

L'objectif va donc être de choisir le ou les signe(s) qui m'apportent le plus selon leur **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, il faut observer, mesurer et/ou interpréter un signe direct ou indirect (ou une combinaison de signes) qui évoque, oriente ou prouve la présence d'une pathologie sous-jacente sans recourir à des examens plus invasifs, plus coûteux, à risque pour le patient...

Validité interne d'un test diagnostique

Comment évaluer un test

Le test doit mesurer de manière effective ce qu'on veut mesurer, il faut donc qu'il soit capable d'identifier (= différencier = discriminer) le mieux possible les malades et les sujets sains :

On vérifie d'abord que :

- ⇒ si un sujet est Malade → Test positif
- ⇒ si un sujet est Sain → Test négatif

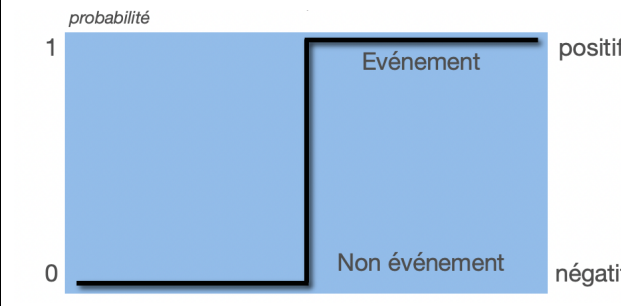
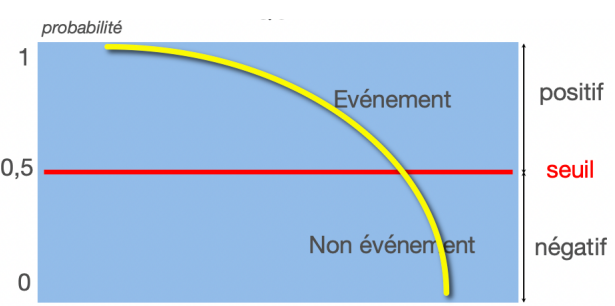
Pour pouvoir dire ensuite :

- ⇒ si le Test positif → le sujet est Malade
- ⇒ si le Test négatif → le sujet est Sain

Pour l'évaluer de cette façon nous avons besoin 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 | Variable Quantitative |
|--|---|
|  <p>On a une réponse binaire, le patient peut :</p> <ul style="list-style-type: none"> - Avoir le signe = positif - Ne pas avoir le signe = négatif |  <p>On a une réponse sous forme d'une valeur numérique, il faut donner un seuil pour savoir si le patient est positif ou non.</p> |

Définition des patients

On appelle population **cible**, la population susceptible de bénéficier de cet examen.

Expression des résultats

L'Expression des résultats dépend de la nature de celui-ci :

- Résultat qualitatif
 - ⇒ Ordinal → absent / peu important / très important
 - ⇒ Binaire → absent / présent ou positif / négatif
- Résultat quantitatif (mesure, valeur continue)
 - X ml de ... / X g.l⁻¹ ...

Test de référence

L'examen de référence aussi appelé anatomo-pathologique ou consensus doit être indépendant de l'examen évalué et le résultat non connu (lorsqu'on analyse le résultat du test à évaluer). Son résultat est **incontestable** et toujours vrai.

Le résultat de l'examen ou du test évalué est comparé avec le diagnostic réel

Pour cela, on soumet les sujets au test de référence (diagnostic béton) et au test à évaluer puis on met en place un tableau de contingence.

Ajout de ma part qui n'est pas dit dans le cours cette année : Le Gold standard est peu utilisé en routine car trop cher, faisable qu'en post-mortem, peu acceptable ou examens pénibles (biopsie de prostate pour un taux élevé de PSA)

Qualités du test :

| | |
|-----------------------------|---|
| Sensibilité (Se) | <p>On peut la définir comme la :</p> <p>⇒ Capacité à détecter tous les malades</p> <p>ou</p> <p>⇒ Probabilité d'avoir un test positif sachant que je suis malade</p> <p>ou</p> <p>⇒ Probabilité de tests positifs chez les malades</p> <div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin: 10px auto; width: fit-content;"> <p><u>FORMULE :</u></p> $Se = VP / (VP + FN)$ </div> <p>⇒ C'est un nombre compris entre 0 et 1, exprimé en %</p> <p style="text-align: center; color: #c08040;">Qualité recherchée pour le dépistage</p> |
| Spécificité (Sp) | <p>On peut la définir comme la :</p> <p>⇒ Capacité à ne reconnaître que les non malades (=capacité à écarter les sujets sains)</p> <p>ou</p> <p>⇒ Probabilité d'avoir un test négatif sachant que je suis sain</p> <p>ou</p> <p>⇒ Proportion de test négatifs chez les sujets sains</p> <div style="border: 1px solid black; border-radius: 10px; padding: 5px; margin: 10px auto; width: fit-content;"> <p><u>FORMULE :</u></p> $Sp = VN / (VN + FP)$ </div> <p>⇒ C'est un nombre compris entre 0 et 1, exprimé en %</p> <p style="text-align: center; color: #c08040;">Qualité recherchée avant de débuter un traitement invasif</p> |

On parle de qualités **intrinsèques** car elles sont propres au test et indépendantes de la prévalence de la maladie et du contexte d'application :

La sensibilité et la spécificité d'un examen pour une maladie reposent sur la définition des "malades", et sont donc caractéristiques d'UNE maladie et d'UN signe. Elles ne sont pas susceptibles de varier d'un centre à l'autre (*d'un service hospitalier spécialisé à une consultation de médecin généraliste, par exemple*).

Test parfait

Un test est dit parfait lorsqu'il ne se trompe jamais et donc quand on a 0 FN et 0 FP.

| | | MALADIE | |
|--------|---------|---------|-----|
| | | OUI | NON |
| SIGNES | présent | 140 | 0 |
| | absent | 0 | 160 |

Signe pathognomonique :

- Sensibilité : 100% => détecte tous les malades (aucun faux négatif)
- Spécificité : 100% => détecte tous les sains (aucun faux positif)

Cas d'un test qualitatif

Exemple : Prédiction du cancer colorectal par un questionnaire et un système de notation

- 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 à la suite de la coloscopie

| | | CANCER COLORECTAL (Diagnostic histologique) | | |
|-------------------|---------|--|----------------------------|-------|
| | | OUI | NON | TOTAL |
| SAIGNEMENT RECTAL | Présent | 82 (VP) Vrais positifs | 1505 (FP) Faux positifs | 1587 |
| | Absent | 13 (FN) Faux négatifs | 668 (VN) Vrais négatifs | 681 |
| | Total | 95 | 2173 | 2268 |

Questions à se poser :

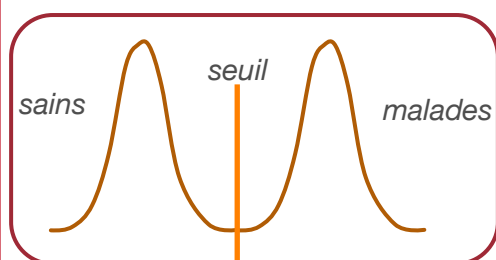
1. Quels sont les patients bien classés ? Quels sont ceux mal classés ?
2. Quelle est la fréquence de la maladie dans cette étude ?
3. Quelle est la proportion de patients malades qui ont un test positif ?
4. Quelle est la proportion de patients non malades qui ont un test négatif ?

Réponses

1. Patients bien classés : 82 (VP) et 668 (VN) (diagonale principale du tableau = diagonale verte)
Patients mal classés : 13 (FN) et 1505 (FP) (diagonale secondaire = diagonale rouge)
Proportion de patients bien classés : $(82+668) / (82+668+13+1505) = 33\% \rightarrow$ Accuraton (= Exactitude) = 33%
2. Prévalence de la maladie = $95 / 2268 = 4,2\%$
3. Proportion de patients malades qui ont un test positif = $82 / 95 = 86,3\% = Se$ (on utilise les cases orange)
4. Proportion de patients non malades avec un test négatif = $668 / 2173 = 30,7\% = Sp$ (on utilise les cases rose)

Cas d'un test quantitatif

Les tests quantitatifs sont problématiques puisque les résultats sont exprimés sous forme d'une valeur numérique, les valeurs sont différentes d'un sujet à l'autre et ils nécessitent de poser un seuil pour classer les malades et les sujets sains.

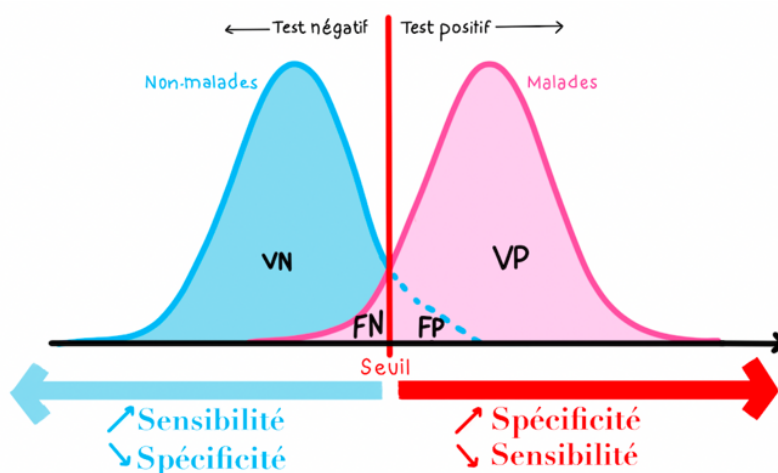


Test discriminant :

Quand notre test est discriminant, la distribution dans le groupe des cas bien séparée de distribution des valeurs dans un groupe de sujets sains. La valeur seuil est donc très simple à déterminer.

Malheureusement la situation est rarement observée...

Le plus souvent nos courbes se chevauchent et on obtient ça :



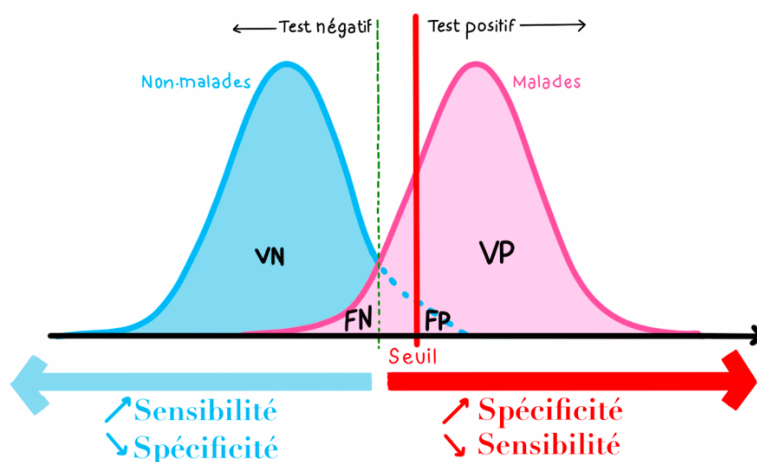
Variation de seuil :

⇒ Cas n°1 :

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.

Pour cela il faut donc hausser le seuil de positivité.

Exemple : dépistage anténatal de l'anencéphalie → un FP aurait des conséquences lourdes puisque cela nécessite une ITG
→ un FN aurait des conséquences peu importantes puisque rattrapable ultérieur par échographie

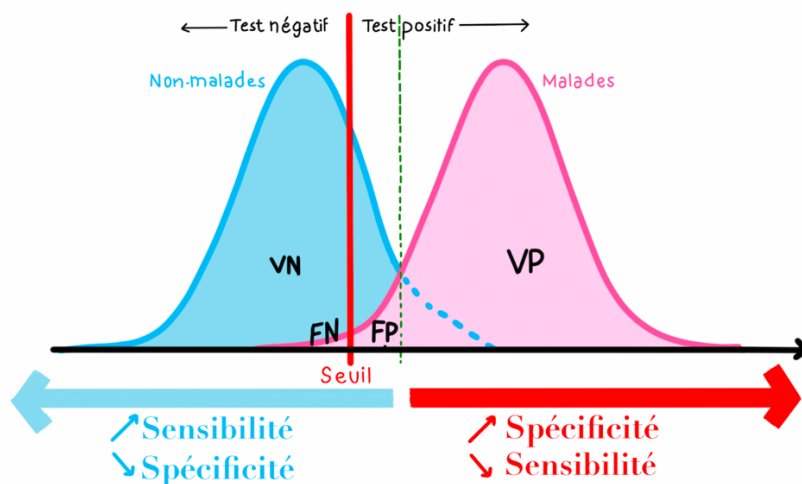


⇒ Cas n°2 :

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.

Pour cela il faut donc baisser le seuil de positivité.

Exemple : Dépistage de la phénylcétonurie à la naissance → un FN aurait des conséquences lourdes avec le développement de la maladie
→ un FP aurait des conséquences peu importantes avec une 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 ce qui nous donne donc une liste de couples Se-Sp.

On peut ainsi dessiner un graphe appelé courbe ROC (*Receiver Operating Caractéristiques*), avec en ordonnée la sensibilité de chaque seuil (Se : vrais positifs) et en abscisse le pourcentage de faux positifs (1-Sp).

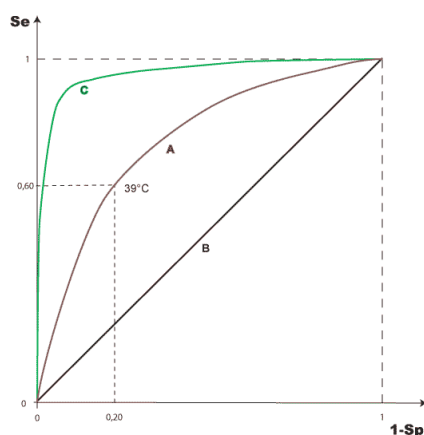
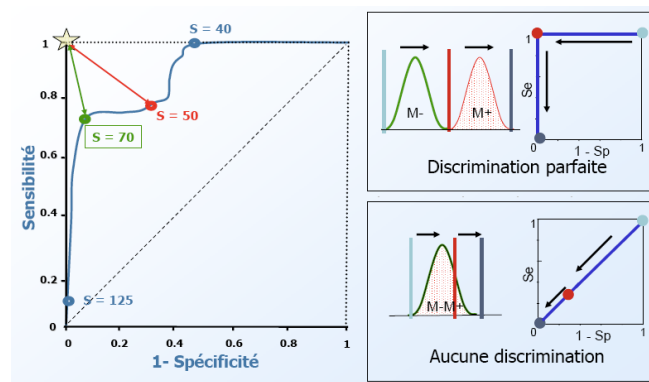
L'aire sous la courbe ROC (ou AUC : *Area Under the Curve*) représente un indice global de la performance discriminante du test, quel que soit le seuil considéré

⇒ $AUC = 1 \rightarrow$ Discrimination parfaite

⇒ $AUC = 0,5 \rightarrow$ Pas de discrimination

La courbe ROC possède deux grands avantages : Elle permet une **analyse globale** des VP et VN et n'effectue pas d'hypothèse sur le seuil

Plus un test est discriminant plus il sera sur la partie **supérieure gauche**, ainsi un test peu discriminant, donc de qualité médiocre, s'aplatira en diagonale.



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.

→ Dans notre graphique à gauche, la technique C est la plus discriminante (donc de meilleure qualité) puisque c'est la courbe la plus convexe vers le point en haut à gauche

La Stratégie à utiliser dépend de l'utilisation de l'examen :

⇒ Faut-il privilégier la sensibilité ? => Cas du dépistage

⇒ Faut-il privilégier la spécificité => Cas du diagnostic de certitude

Pour cela on s'interroge sur quelle est notre priorité ?

⇒ Dépister les patients malades ? → Sensibilité

⇒ Exclure les patients non malades ? → Certitude diagnostique = Spécificité

⇒ Ne pas inquiéter à tort les Faux Positifs (maladie grave) → Privilégier Spécificité

⇒ Exactitude de la mesure → que tout le monde soit bien classés

Apport décisionnel d'un test

La Sensibilité et la Spécificité ne sont pas utilisables directement pour le patient puisqu'elles définissent la probabilité d'avoir un test positif ou négatif selon si on est malade ou non.

Pour notre patient la question est donc inverse : si le test est positif, quelle est la probabilité qu'il soit vraiment malade ?

On va donc introduire la notion de **valeurs prédictives**.

Les valeurs prédictives dépendent de la Se, de la Sp et de la prévalence (p) de la maladie. Ce sont des qualités extrinsèques.

Les 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.

| <u>Valeur Prédictive Positive (VPP)</u> | <u>Valeur Prédictive Négative (VPN)</u> |
|--|--|
| → Probabilité d'être malade sachant que le test est positif | → Probabilité d'être sain sachant que le test est négatif |
| <p><u>FORMULE :</u></p> $VPP = \frac{VP}{(VP + FP)}$ $VPP = \frac{Se \cdot P}{(Se \cdot P) + (1 - Sp)(1 - p)}$ | <p><u>FORMULE :</u></p> $VPN = \frac{VN}{(VN + FN)}$ $VPN = \frac{Sp \times (1 - P)}{Sp \times (1 - P) + (1 - Se) \times P}$ |

La Prévalence (p)

La prévalence correspond au nombre de cas dans une population à un instant donné, c'est une probabilité pré-test.

Elle peut être modifiée entre le début et la fin d'un test, dans ce cas l'échantillon n'est plus représentatif et on utilisera pour la VPP et la VPN les formules qui impliquent directement la prévalence.

| | Cancer | Sains | |
|--|------------|------------|----------------------------|
| N=2268 Prévalence = 4,2% | | | |
| Saignement rectal | + | 82 | 1505 |
| | - | 13 | 668 |
| | Se = 86,3% | Sp = 30,7% | |
| | | | VPP = 5,2% VPN = 98,1% |
| N=2268 Pop. à risque Prévalence = 29,7% | | | |
| Saignement rectal | + | 581 | 1105 |
| | - | 92 | 490 |
| | Se = 86,3% | Sp = 30,7% | |
| | | | VPP = 34,5% VPN = 84,2% |

Indice de synthèse

EXACTITUDE

L'exactitude correspond à la proportion de sujets bien classés, c'est le % de résultats exacts soit les VP et VN par rapport à la cohorte

Elle varie de 0 à 1 et :

- ♥ 1 correspond à une absence de FP et de FN
- ♥ un tirage au sort correspond à 0,5

FORMULE :

$$\text{Exactitude (\%)} = \frac{VP + VN}{VP + VN + FP + FN}$$

INDICE DE YODEN

L'indice de Youden c'est l'addition des 2 qualités du test : Se et Sp
Il varie de -1 à 1 et lorsqu'il vaut 0 on n'a pas d'orientation diagnostique

Il permet de trouver le meilleur seuil de discrimination d'une courbe ROC

⇒ **Cet indice est peu utilisé**

FORMULE :

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

RAPPORTS DE VRAISEMBLANCE

Les rapports de vraisemblance, aussi appelés “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.

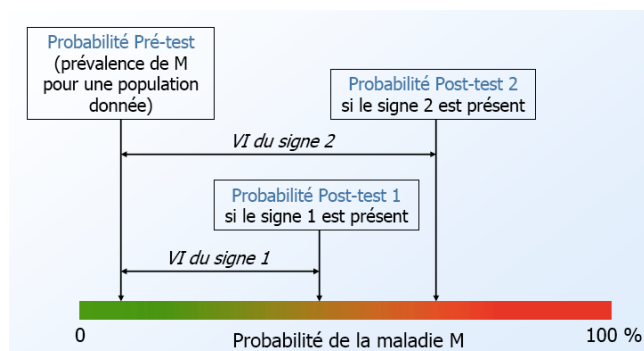
Ce rapport **varie** donc selon que le résultat de l'examen diagnostique est positif ou négatif.

C'est 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.

Pour qu'un test soit jugé utile cliniquement il faut :

- ♥ Que le rapport de vraisemblance d'un test positif soit le plus grand possible
- ♥ Que le rapport de vraisemblance d'un test négatif soit le plus petit possible

Probabilités pré et post-test



Vraisemblance positive

Le rapport de vraisemblance **positif** exprime la crédibilité d'un test **positif**

Il varie de 0 (quand Se = 0) à l'infini (quand 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

FORMULE :

$$L+ = \frac{P(T+/M+)}{P(T+/M-)} = \frac{Se}{(1-Sp)}$$

Vraisemblance négative

Le rapport de vraisemblance **négatif** exprime la crédibilité d'un test **négatif**

- ♥ Le rapport est nul quand Se = 1
- ♥ Le test est totalement spécifique (=pathognomonique) quand L- = 1-Se (si Sp = 1)

⇒ Plus L- est petit (tend vers 0), plus il permet d'exclure la maladie.

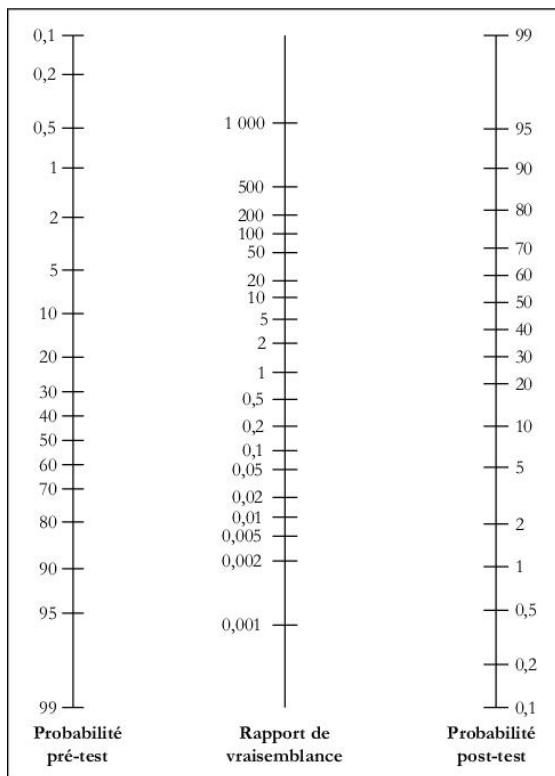
FORMULE :

$$L- = \frac{P(T-/M+)}{P(T-/M-)} = \frac{(1-Se)}{Sp}$$

Vraisemblance

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

| LR+ | LR- | Changement entre prob pré-test et post-test |
|------|---------|---|
| ≥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 |

Exemple du Nomogramme de Bayes :

À gauche nous avons un examen 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 grand (LR+ plus élevé)
- ♥ L'apport diagnostique d'un résultat négatif est grand (LR- est petit et proche de 0)

- ♥ Dédi à ma maman quand même parce que sans elle bon...
- ♥ Dédi à Valentine comme d'hab parce qu'elle y a toujours plus cru que moi que j'y arriverais et que je ne la remercierais jamais assez de m'avoir supporté H24 😊
- ♥ Dédi à Aymeric parce que même un message lambda devient drôle avec lui
- ♥ Dédi à Carla Campagne, une tutrice incroyable
- ♥ Dédi à Bryan et Elly (mais beurk la bioch dsl)
- ♥ Dédi à Mathilde et Grégory (joyeux anniversaire avec 1h30 d'avance)
- ♥ Dédi à tous ceux qui envoient des messages gentils sur discord <3
- ♥ Pas dédi à ceux qui n'aiment pas la biostat 😊
- ♥ Dédis aux CTs (et à notre CT réf 😊) parce qu'en vrai ils sont super investit pour vous à côté de leur D1 et on est pas facile avec eux ptdrrr