

P2

Evaluation Diagnostique et Dépistage

Nathalie GAULT
nathalie.gault@bch.aphp.fr

Département d'Epidémiologie, de Biostatistiques et de
Recherche Clinique
Hôpital Bichat – Claude Bernard
Unité INSERM U738 : « Modèles et méthodes de l'évaluation
thérapeutique »

Plan

Objectif

Les tests

Performances diagnostiques - Courbe ROC

Valeurs prédictives

Rapport de vraisemblance

La prévention

Objectif

- Etude d'évaluation diagnostique = étude évaluant des **performances diagnostiques** d'un test de dépistage ou d'un test diagnostique
 - Par rapport à un **test de référence** (gold standard)
- Performance diagnostique = capacité à bien classer les sujets malades et non malades

Les tests

Test diagnostique

- **Un signe clinique** subjectif (céphalée) ou objectif (éruption)
- **Un examen de laboratoire** : qualitatif (hémoculture) ou quantitatif (glycémie).
- **Un examen radiographique**
- **Un acte invasif** : biopsie, intervention chirurgicale, autopsie
- Etc...

Test diagnostique

2 situations selon le résultat du test

↙

Variable qualitative

Signe présent oui / non

Ex: Hémoculture positive

Abolition d'un réflexe

↘

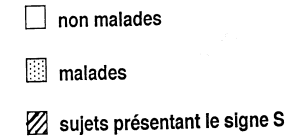
Variable quantitative

Ex: Glycémie

Cholestérolémie

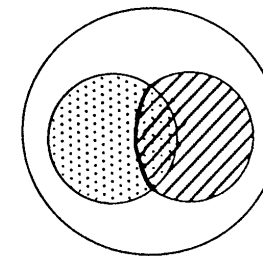
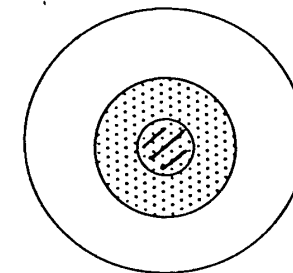
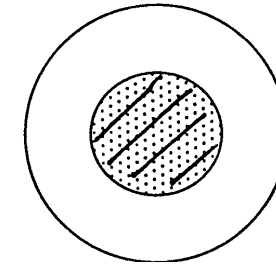
Pression artérielle

Test diagnostique



- **Variable qualitative**

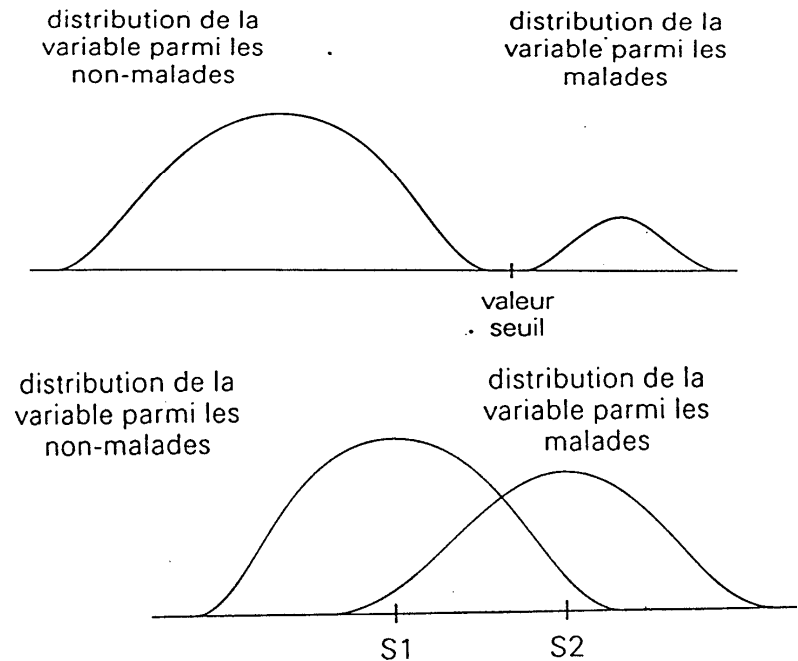
- Test idéal : le test est positif si et seulement si le sujet est malade. Permet de classer sans erreur en Malade ou Non Malade
- Signe pathognomonique : Seuls les sujets malades ont le signe mais tous les sujets malades ne l'ont pas. Permet de classer sans erreur les sujets malades
 - ex : signe de Köplick dans la rougeole
- 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



Test diagnostique

- **Variable quantitative**

- Situation idéale : les distributions de la valeur résultat du test sont bien distinctes. Permet de classer sans erreur en Malade ou Non Malade
- Situation habituelle : Les valeurs résultat du test s'échelonnent et se superposent entre les malades et les non-malades. Ne permet pas de classer sans erreur en Malade ou Non-Malade



- Astuce : choix d'une **valeur seuil** pour transformer en variable qualitative
 - Ex : Malade si glycémie > 1,40 g/l

Test de référence (gold standard)

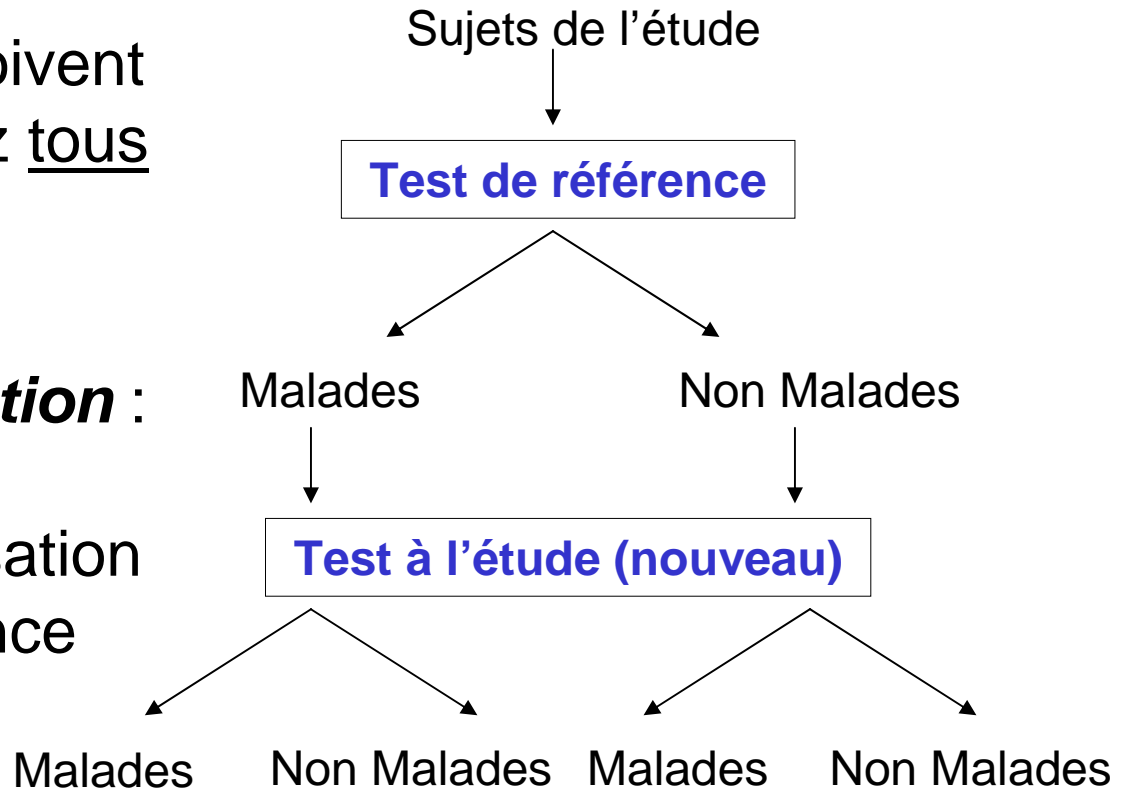
- Test déjà existant, validé, qui permet d'affirmer que la maladie est présente ou absente
 - Idéalement avec critère indiscutable
 - *Ex* : présence de BK dans le LCR, cellules cancéreuses dans une tumeur
 - Souvent plus problématique
 - Subjectif : donnée d'interrogatoire (angine de poitrine)
 - Cher
 - Risqué ou difficile à obtenir (biopsies, exploration chirurgicale, autopsie, ...)
 - Test inexistant : évolution naturelle (maladie à prions, maladie à évolution lente...)
- Si inexistant : classer les sujets avec un faisceau d'arguments dont la validité doit avoir été étudiée
- Gold-standard = meilleur test disponible (même si imparfait)

Performances d'un test

Courbe ROC

Performance d'un test

- Les deux tests doivent être réalisés chez tous les sujets
- **Biais de vérification** : si le test à l'étude influence la réalisation du test de référence



Performance d'un test

- Obtention des valeurs :
 - Vrai Positif (VP), Vrai Négatif (VN), Faux Positif (FP), Faux Négatif (FN)

		Test de référence	
		Malade (M+)	Non Malade (M-)
Test à l'étude	+	VP	FP
	-	FN	VN

Performance d'un test

- Performance diagnostique mesurée par
 - **Sensibilité** : probabilité (%) d'avoir un test positif quand on est malade

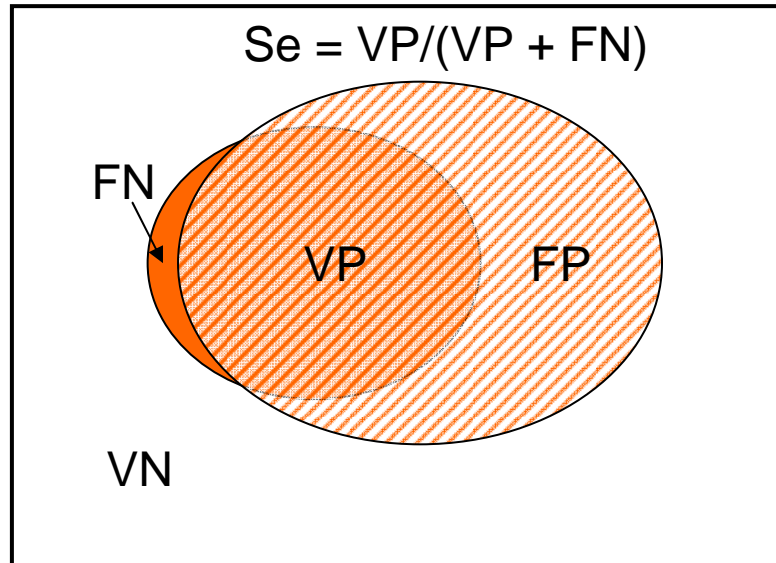
$$Se = VP/(VP + FN)$$

- **Spécificité** : probabilité (%) d'avoir un test négatif quand on est non-malade

$$Sp = VN/(VN + FP)$$

- Propriétés prises en compte quand on **décide de prescrire ou non un test** (avant le résultat)

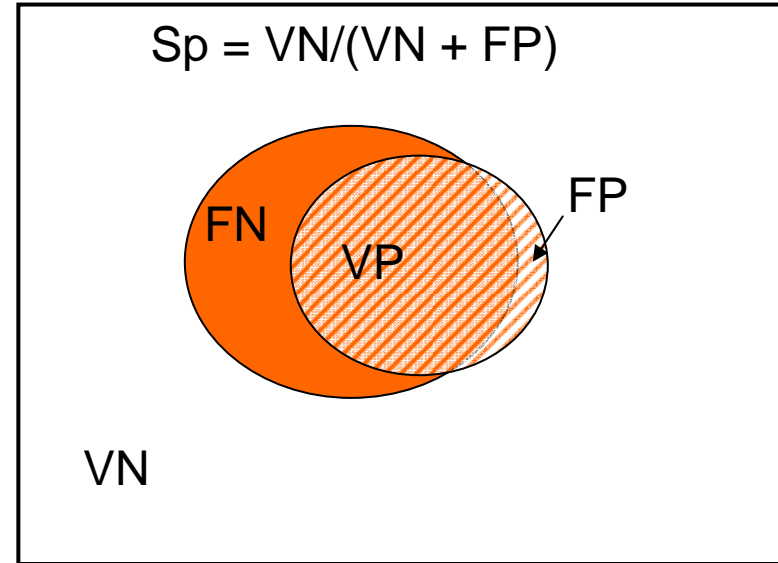
Performance d'un test



Test sensible

La **sensibilité** est l'indice qui évalue la capacité d'une mesure à **bien classer les malades** (ne manquera pas les malades)

$Se = 1$: le signe est présent chez tous les malades (pas de FN).



Test spécifique

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

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

Performance d'un test

- En pratique : usage des **tests sensibles**
 - maladie très rare
 - maladie curable
 - diagnostic précoce d'une maladie grave en améliore le pronostic : pénalité à manquer la maladie
 - *Ex* : tuberculose, maladie de Hodgkin, ...
 - stade précoce d'une action diagnostique pour éliminer des hypothèses (ex : sérologie HIV devant un infiltrat pulmonaire)
 - Utile quand négatif : permet d'exclure la maladie si négatif

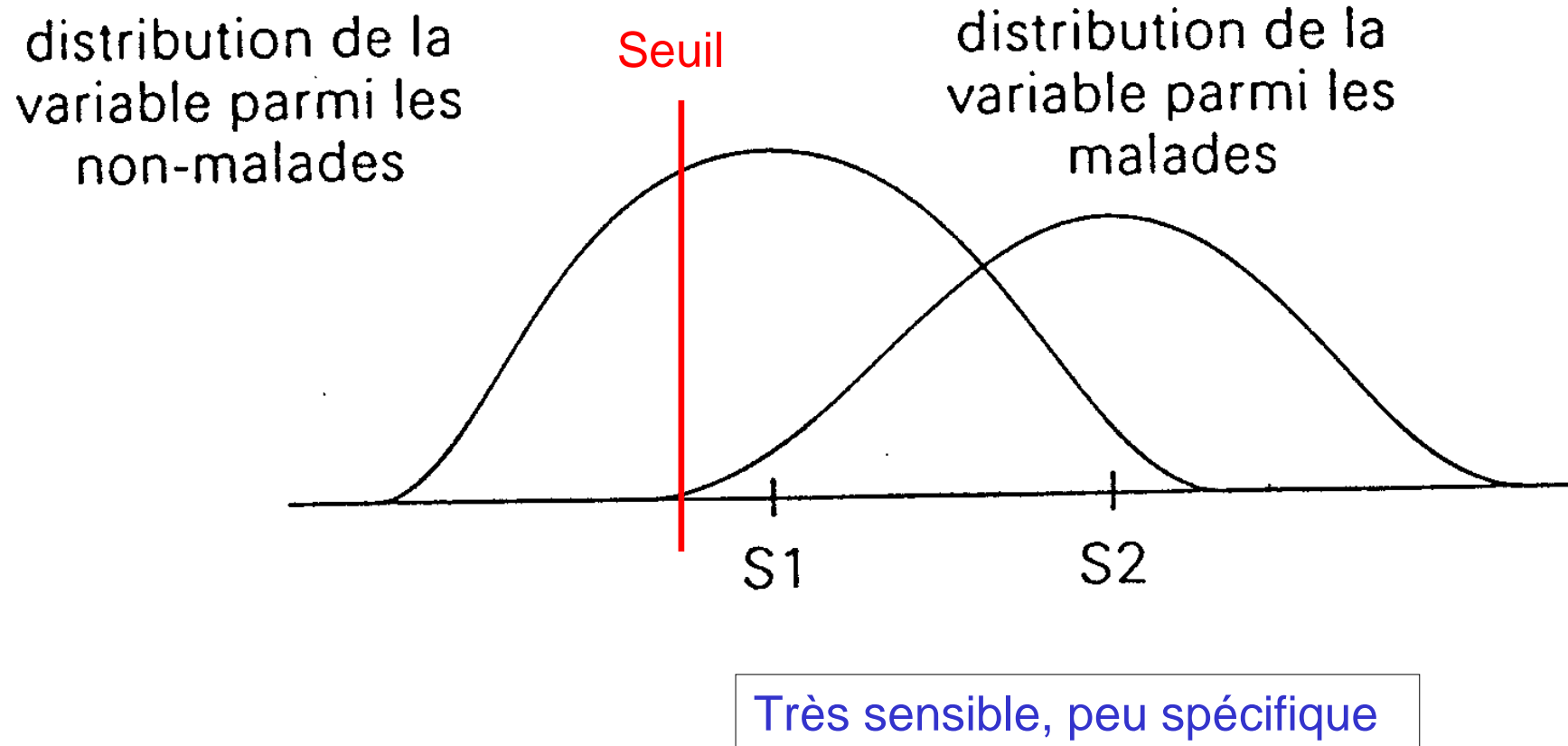
Performance d'un test

- En pratique : usage des **tests spécifiques**
 - confirmer un diagnostic suggéré par d'autres résultats
 - si les résultats faussement positifs peuvent être préjudiciables au patients (ex cancer et preuve histologique)
 - Utile quand positif : il sera rarement positif en l'absence de maladie

Performance d'un test

- Idéal : test à la fois très sensible et très spécifique
- Idéal \neq réalité : **compromis** entre la sensibilité et la spécificité
- Pour les tests quantitatifs
 - **choix de la valeur seuil** délimitant Malades et Non Malades est **arbitraire**
 - en fonction du seuil choisi, quand la sensibilité augmente, la spécificité diminue et inversement

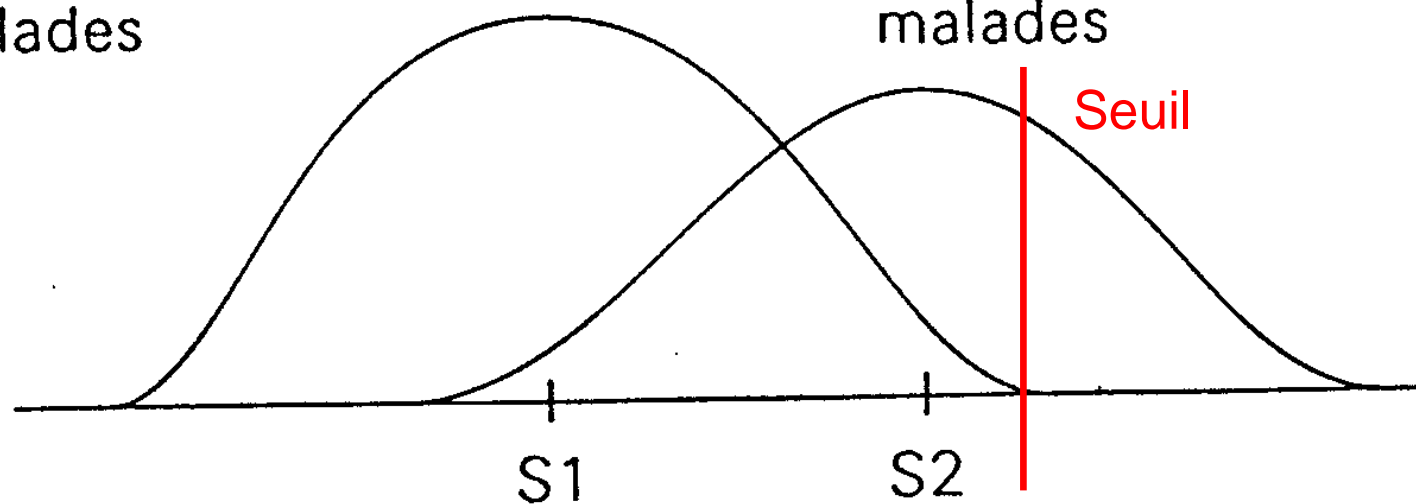
Performance d'un test



Performance d'un test

distribution de la variable parmi les non-malades

distribution de la variable parmi les malades



Peu sensible, très spécifique

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

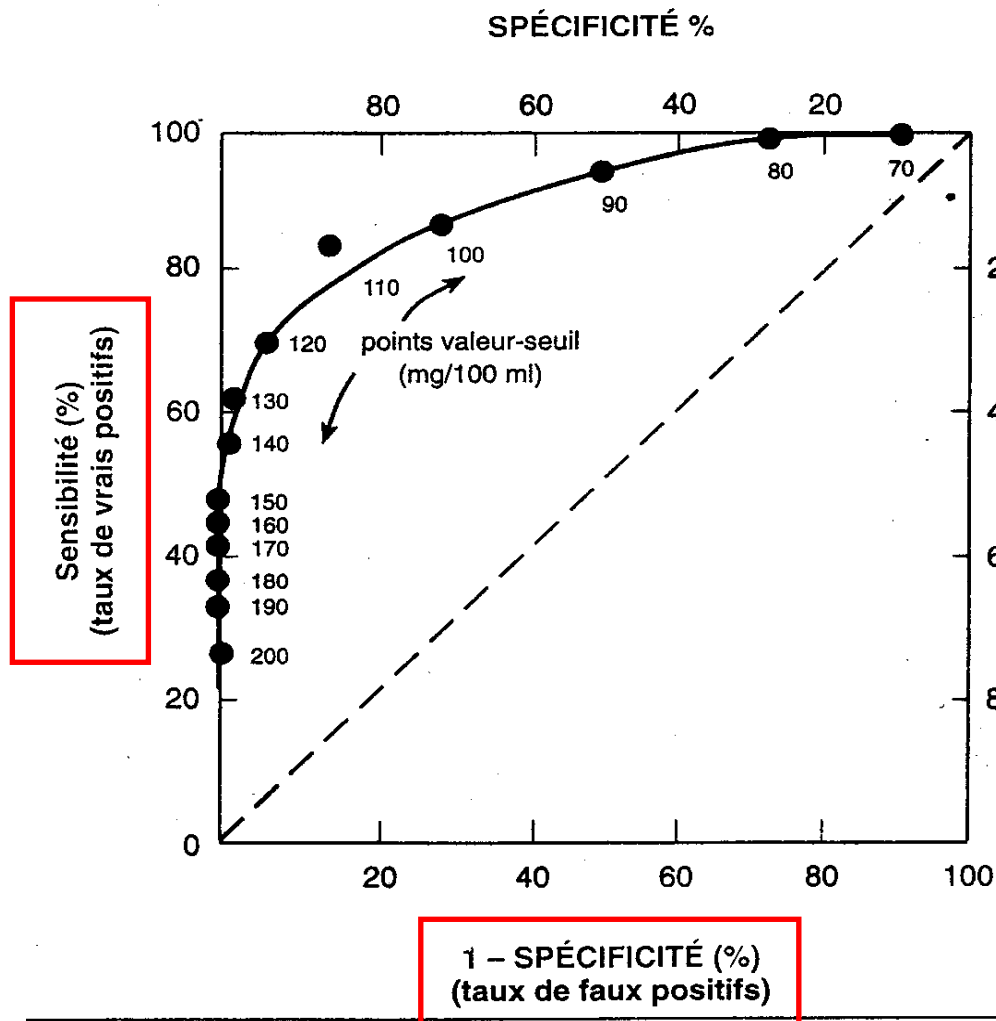
Glycémie postprandiale t + 2 h (mg/100ml)	Sensibilité (%)	Spécificité (%)
70	98,6	8,8
80	97,1	25,5
90	94,3	47,6
100	88,6	69,8
110	85,7	84,1
120	71,4	92,5
130	64,3	96,9
140	57,1	99,4
150	50	99,6
160	47,1	99,8
170	42,9	100
180	38,6	100
190	34,3	100
200	27,1	100

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

Courbe ROC

- Courbe ROC = Receiver Operating Characteristic (courbe caractéristique de la performance d'un test)
- Façon d'exprimer le compromis entre Se et Sp
- But : ***choix du seuil*** diagnostique d'un test en recherchant le meilleur rapport entre sensibilité et spécificité (uniquement pour les tests dont le résultat est quantitatif)

Courbe ROC



- Les **tests discriminants** sont dans le coin supérieur gauche. Les **tests peu performants** se rapprochent de la diagonale (pile ou face).
- La valeur seuil permettant de minimiser à la fois les FP et les FN (maximiser à la fois la spécificité et la sensibilité) est au **point d'inflexion** de la courbe ROC
- L'exactitude globale du test est **l'aire sous la courbe ROC**. Plus l'aire est grande, meilleur est le test.

Une courbe ROC. L'exactitude d'une glycémie postprandiale à t + 2h comme tes du diabète. (D'après *Public Health Service. Diabetes program guide. Publication n° 161* Washington, DC : U.S. Government Printing Office, 1960)

Valeurs prédictives

Valeurs prédictives

- Obtention des valeurs VPP et VPN

		Test de référence	
		Malade (M+)	Non Malade (M-)
Test à l'étude	+	VP	FP
	-	FN	VN

Valeurs prédictives

- Valeurs prédictives mesurées par
 - **Valeur prédictive positive (VPP)** : Probabilité (%) d'être malade si on a un test positif

$$VPP = VP / (VP + FP) \quad VPP = (Se * p) / [(Se * p) + (1 - Sp) * (1 - p)]$$

- **Valeur prédictive négative (VPN)** : Probabilité (%) de ne pas être malade si on a un test négatif

$$VPN = VN / (VN + FN) \quad VPN = Sp * (1 - p) / [(1 - Se) * p + Sp * (1 - p)]$$

- Propriétés prises en compte quand les résultats du test sont connus pour **savoir si le patient est malade ou non** (probabilité a posteriori ou post-test).

Valeurs prédictives

- 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 **dépendent de la prévalence**, elles ne sont pas indépendantes du cadre dans lequel le test est appliqué

Valeurs prédictives

- *Exemple* : 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

Prévalence = $104/195 = 53\%$

Se = $55/104 = 53\%$

Sp = $84/91 = 92\%$

VPP = $55/62 = 89\%$

VPN = $84/133 = 63\%$

Valeurs prédictives

- *Exemple* : Valeurs prédictives de L'ECG d'effort en fonction de la prévalence des sténoses coronaires

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

Prévalence = $104/624 = 17\%$

Se = $55/104 = 53\%$

Sp = $478/520 = 92\%$

VPP = $55/97 = 57\%$

VPN = $478/527 = 91\%$

Propriétés des tests diagnostiques

		Test de référence	
		(M+)	(M-)
Test à l'étude	+	VP	FP
	-	FN	VN

VPP

VPN

Propriétés extrinsèques

(performance en situation réelle) dépendent de:

- qualités intrinsèques du test
- la prévalence de la maladie en situation réelle

Se

Sp

Propriétés intrinsèques
(performances expérimentales)

Rapport de vraisemblance

Rapport de vraisemblance (RV) "Likelihood ratios"

- Façon alternative de décrire les performances d'un test
- Pré-requis : le concept d'**odds** (rapport de 2 probabilités)

$$\text{Odds} = \text{Proba}(\text{événement}) / 1 - \text{Proba}(\text{événement})$$

- *Ex* : « un odds à 4:1 pour qu'un cheval gagne la course », équivaut à « ce cheval a 80% de chances de gagner ».

Rapport de vraisemblance (RV) "Likelihood ratios"

- **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 obtenu "RV-" fois plus souvent chez les sujets malades que chez les sujets non malades »

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

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

Rapport de vraisemblance (RV) "Likelihood ratios"

- $RV = 1$: 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, plus le test présente de l'intérêt

– $RV+ > 10$ ou $RV- < 0,1$	⇒	– Le test a un <i>très fort gain diagnostic</i>	} Confiance dans le diagnostic
– $RV+$ entre 5 et 10 ou $RV-$ entre 0,1 et 0,2	⇒	– Le test a un fort gain diagnostic	
– $RV+$ entre 2 et 5 ou $RV-$ entre 0,2 et 0,5	⇒	– Le test a un gain diagnostic modéré	} Infirmerie du diagnostic
– $RV+ < 2$ ou $RV- > 0,5$	⇒	– Le test a un gain diagnostic faible	

Rapport de vraisemblance (RV) "Likelihood ratios"

Exemple :

CPK	IDM+ N (%)	IDM- N (%)	RV+ P(T+/M+) / P(T+/M-)
≥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 / 0.200 = 0.3
1-39	2 (0.009)	88 (0.677)	0.009 / 0.677 = 0.01
Total	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. » Indication à poursuivre les examens

« 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. » On peut **exclure** la maladie avec une grande certitude.

Rapport de vraisemblance (RV) "Likelihood ratios"

$$\text{Odds pré-test} * RV = \text{Odds post-test}$$



proba pré-test
= prévalence



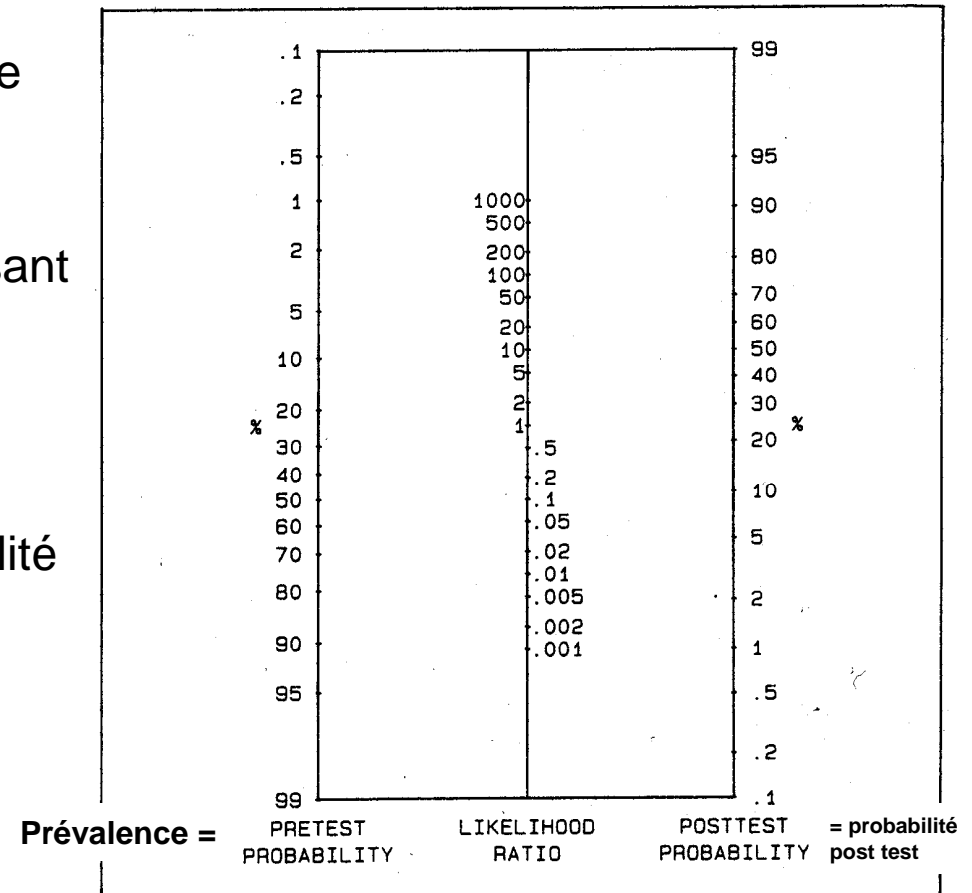
proba post-test
= VP

Pour mémoire : Odds = Proba(événement) / 1-Proba(événement)

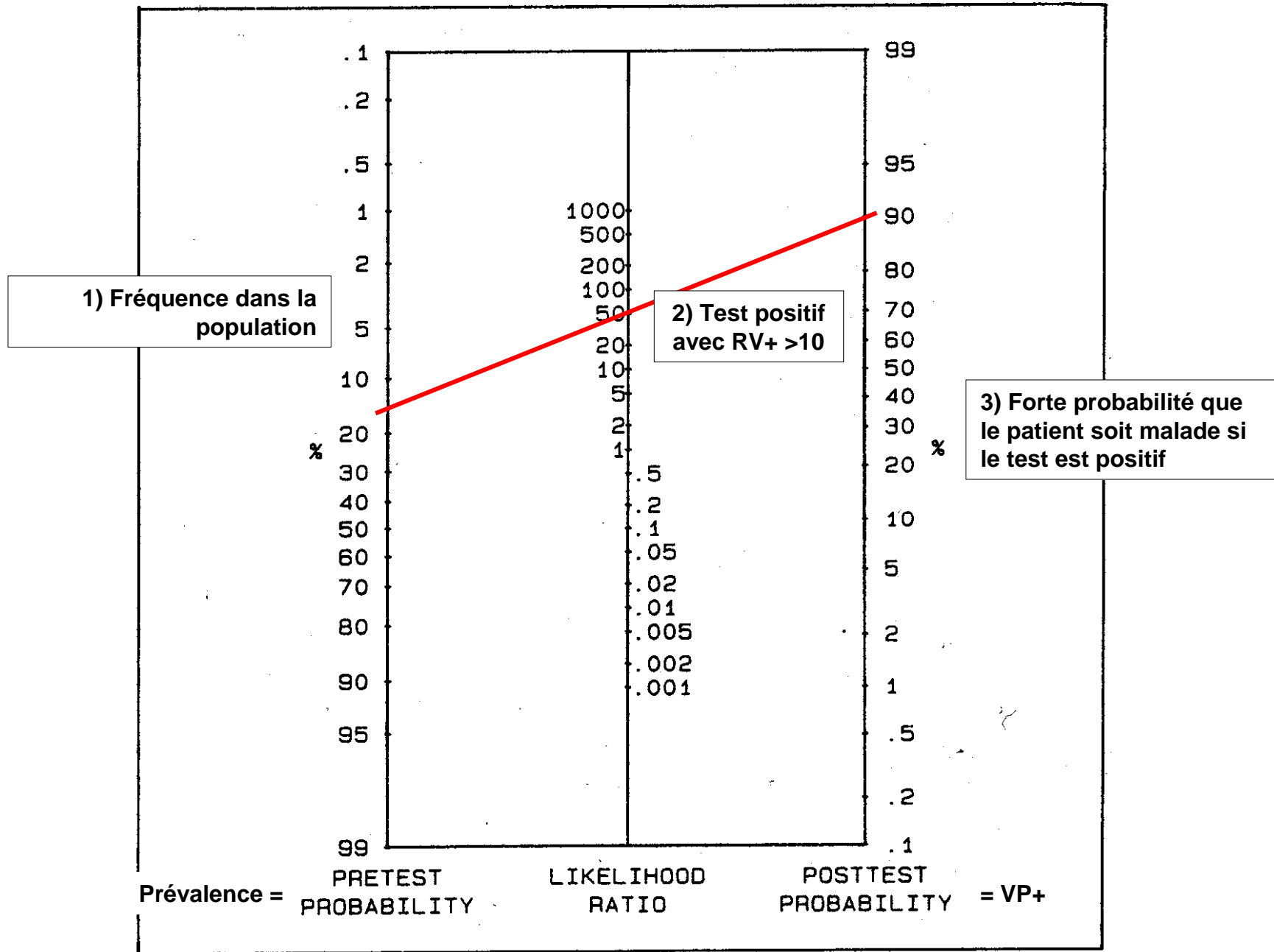
Rapport de vraisemblance (RV) "Likelihood ratios"

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

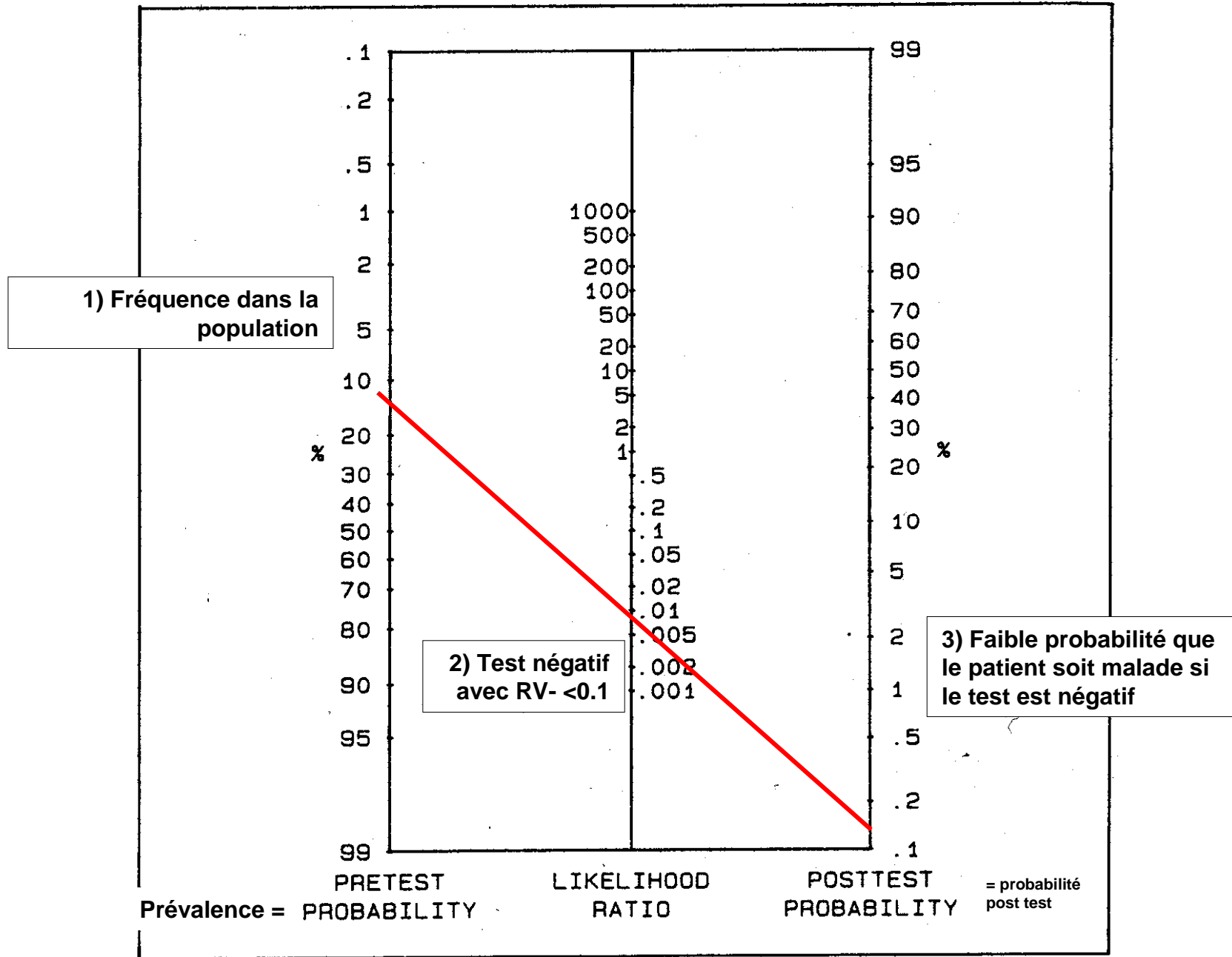
Sur ce diagramme, tracer la droite passant par la prévalence de la maladie dans le groupe étudié et par le rapport de vraisemblance du test utilisé. Le point d'intersection de cette droite avec l'axe "valeur prédictive" détermine la probabilité post test pour le patient considéré.



A nomogram for applying likelihood ratios. (Adapted from T. J. Fagan. Nomogram for Bayes' theorem. *N. Engl. J. Med.* [Letter] 293:257, 1975.)



A nomogram for applying likelihood ratios. (Adapted from T. J. Fagan. Nomogram for Bayes' theorem. *N. Engl. J. Med.* [Letter] 293:257, 1975.)

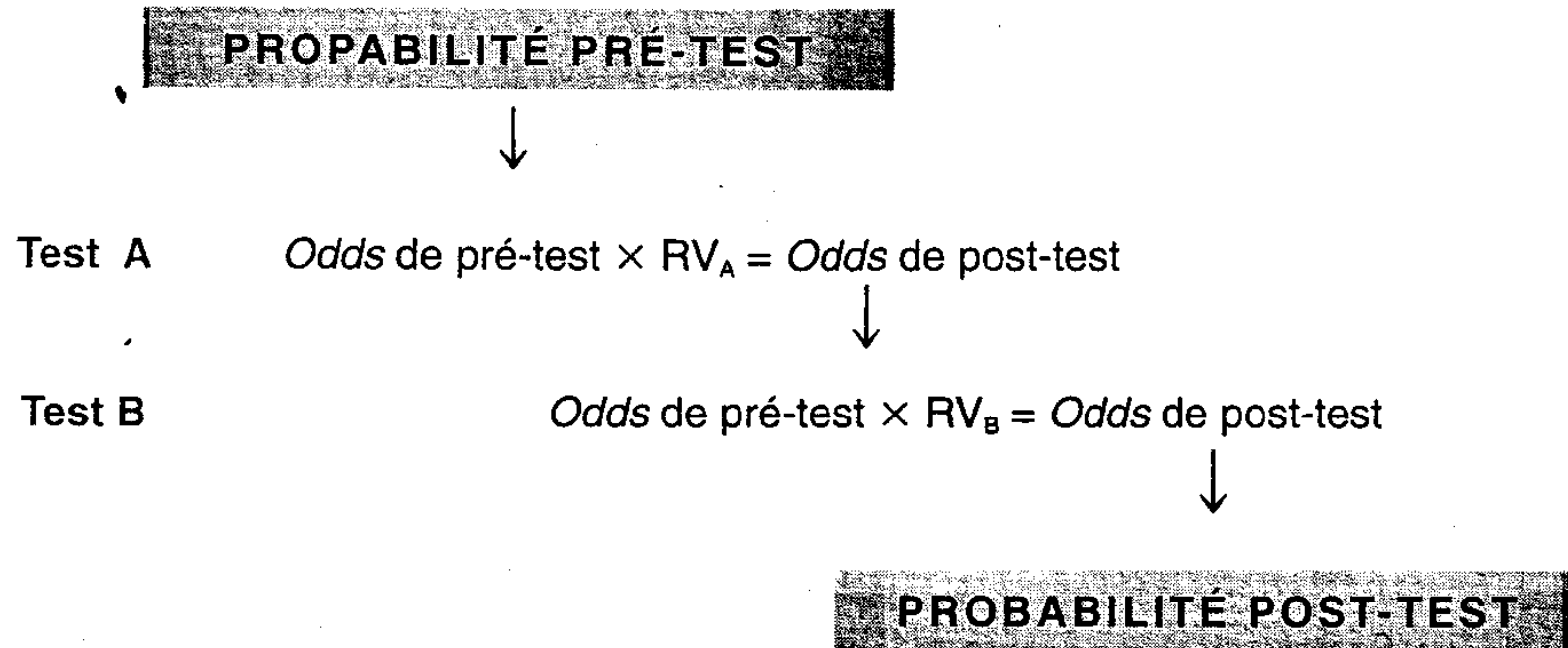


A nomogram for applying likelihood ratios. (Adapted from T. J. Fagan. Nomogram for Bayes' theorem. *N. Engl. J. Med.* [Letter] 293:257, 1975.)

Rapport de vraisemblance (RV) "Likelihood ratios"

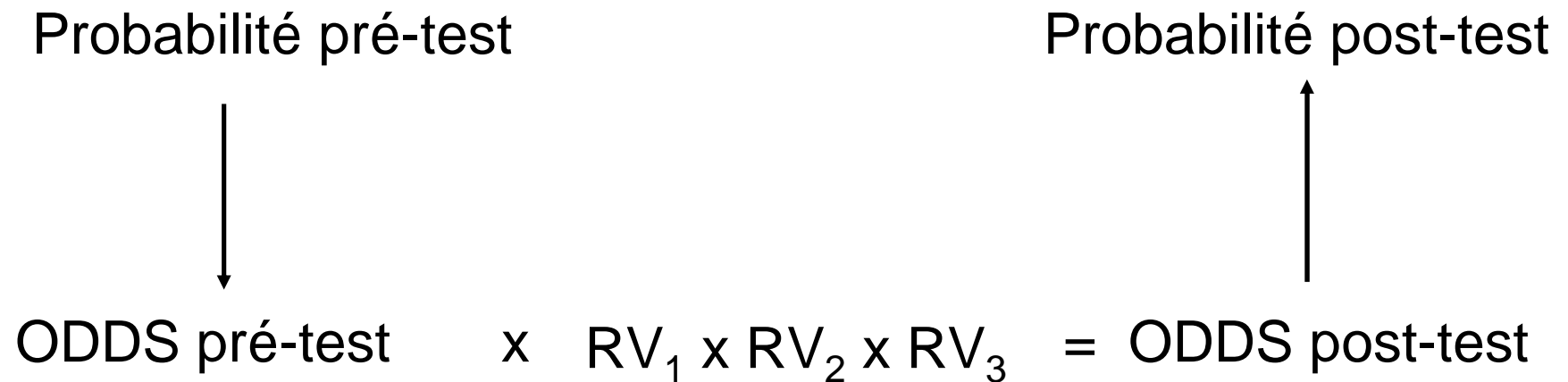
- **Tests multiples**

- Quand une série de test est utilisée, une probabilité globale peut être calculée, utilisant les RV de chaque résultat.



Rapport de vraisemblance (RV) "Likelihood ratios"

- **Tests multiples**



Rapport de vraisemblance (RV) "Likelihood ratios"

- **Avantages** :
 - Indépendant de la prévalence de la maladie (probabilité pré-test). Bon reflet de la valeur du test quel que soit le groupe auquel il est appliqué.
 - Résume l'information en un seul paramètre : à quel point la maladie est vraisemblable?
 - Permet de calculer la probabilité de maladie après une série de tests
- **Inconvénient** :
 - Utilisation des Odds peu aisée

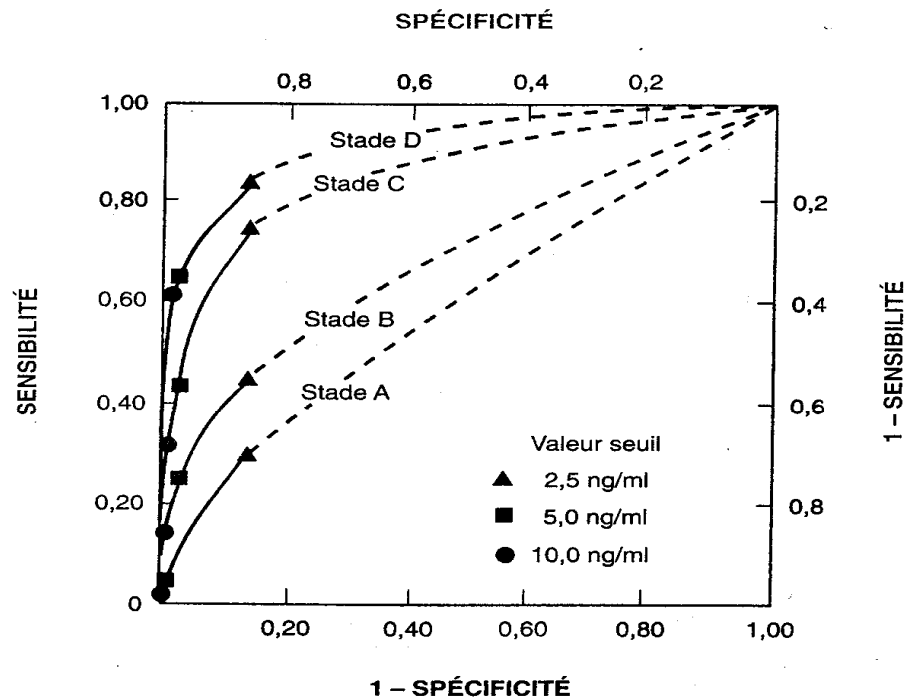
Validité interne

- Validité méthodologique de l'étude:
 - Comparaison en aveugle, indépendante avec un test de référence
 - Biais de vérification? Tous les patients doivent avoir les 2 tests
 - Description de la méthodologie du test

Validité externe

- **Population cible** : sur quels patients le test a-t-il été validé?

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



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

Les **patients** utilisés pour décrire les propriétés du test doivent être **identiques** à ceux pour lesquels le test s'applique en pratique clinique

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

Validité externe

- Reproductibilité du test, qualification des évaluateurs?
 - Certains tests diagnostiques nécessitent une **interprétation** (imagerie, examen clinique...) : part de **subjectivité**
 - Faible intérêt si tests non reproductibles
 - entre différents patients => mesurer cette reproductibilité
 - entre différents évaluateurs => Importance de la **description** précise des tests diagnostiques utilisés et de leur interprétation (qualification de l'évaluateur, courbe d'apprentissage)

La prévention

Dépistage et Diagnostic

- Objectif commun : identifier une maladie pour pouvoir la traiter
- Contexte différent :
 - **Diagnostic** : le patient se plaint d'un symptôme et veut en être soulagé
 - Trouver la maladie expliquant les symptômes du patient
 - **Dépistage** : le patient est en bonne santé apparente
 - Trouver les malades parmi les sujets asymptomatiques

La prévention

- **Définition** : acte d'empêcher la survenue d'un événement négatif
- 3 niveaux de prévention selon le moment de l'évolution de la maladie où l'intervention est faite (selon l'OMS)

Niveaux de prévention

- ***Prévention primaire*** (OMS) : ensemble des moyens mis en œuvre pour empêcher l'apparition d'un trouble, d'une pathologie ou d'un symptôme (en éliminant ses causes)
- Cadre du système de soins (en ville) ou hors cadre du système de soins (média)
- *Exemples*
 - Information de la population
 - Vaccination contre tétanos
 - Campagne anti-tabac
 - Utilisation du préservatif
 - « *Pour votre santé, évitez de manger trop gras, trop sucré, trop salé* »

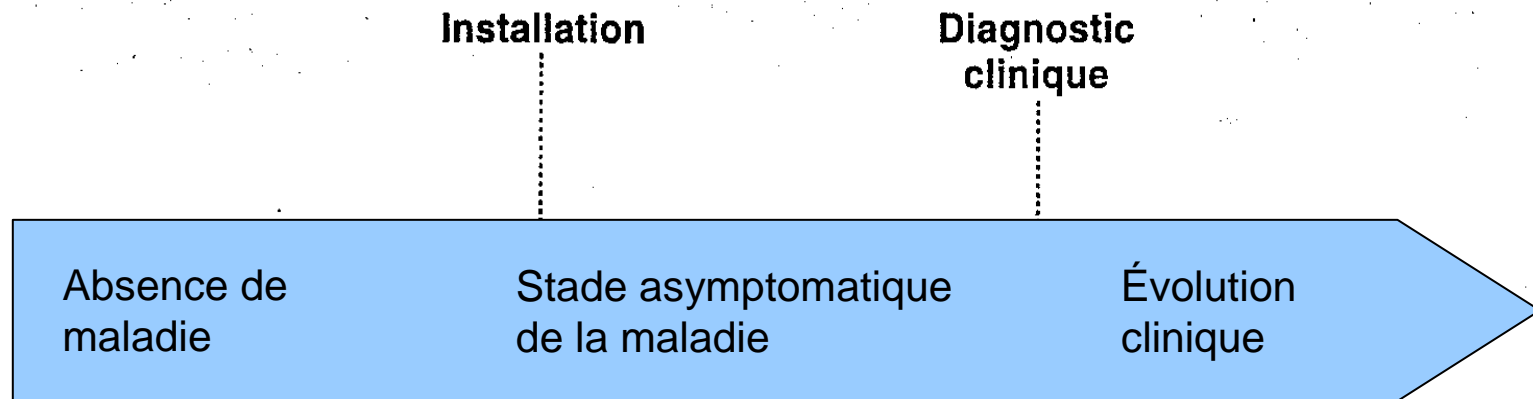
Niveaux de prévention

- **Prévention secondaire** (OMS) : vise la détection précoce des maladies, dans le but de les découvrir à un stade où elles peuvent être traitées (stade asymptomatique)
- Pré requis = **dépistage**
- Cadre du système de soins (en ville)
- Réalisé au **niveau individuel** avec éventuelle incitation **collective**
- *Exemples :*
 - Frottis cervicaux et dépistage du cancer du col de l'utérus
 - Mammographie et dépistage du cancer du sein
 - Test de Guthrie et dépistage de la phénylcétonurie

Niveaux de prévention

- **Prévention tertiaire** (OMS) : tend à éviter les complications dans les maladies déjà présentes (stade symptomatique)
- Pré requis = **diagnostic**
- Cadre du système de soins (hôpital)
- Réalisé au **niveau individuel**
- *Exemples :*
 - Détection rétinopathie chez des patients diabétiques
 - Béta-bloquants post IDM

Dépistage et Diagnostic



Prévention :

PRIMAIRE
Supprime les facteurs de risque



Empêche la maladie de survenir

SECONDAIRE
Détection et traitement précoce



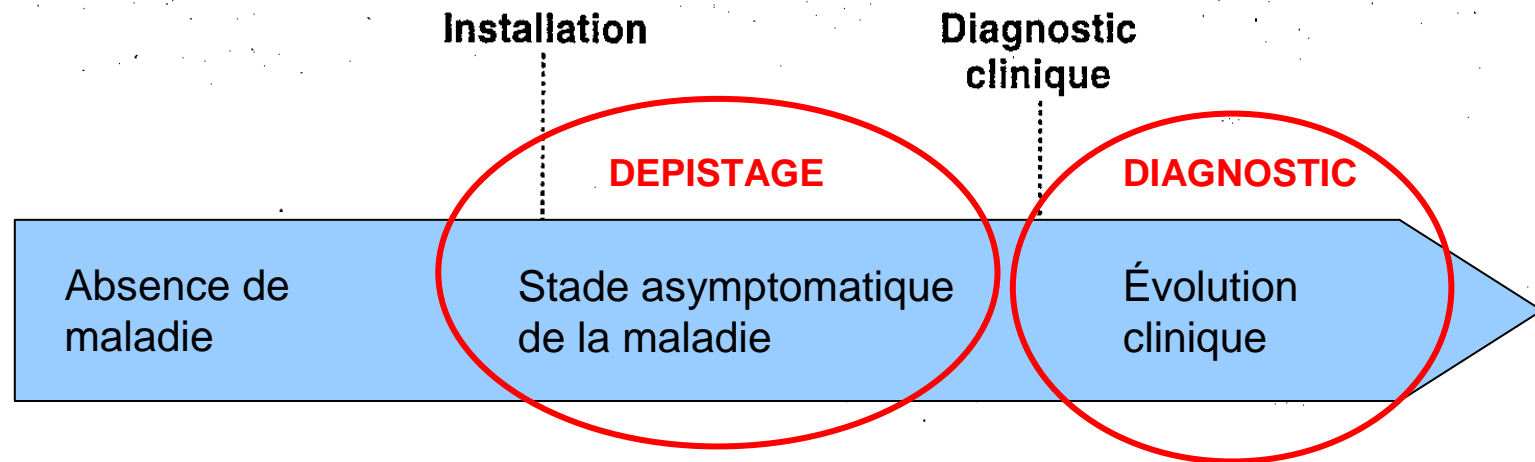
Détecte la maladie précocement lorsqu'elle est asymptomatique et qu'un traitement peut arrêter sa progression

TERTIAIRE
Réduit les complications



Évite la détérioration et les complications

Dépistage et Diagnostic



Prévention :

PRIMAIRE

Supprime les facteurs de risque



Empêche la maladie de survenir

SECONDAIRE

Détection et traitement précoce



Détecte la maladie précocement lorsqu'elle est asymptomatique et qu'un traitement peut arrêter sa progression

TERTIAIRE

Réduit les complications



Évite la détérioration et les complications

Point critique

- On appelle ***point critique*** dans l'histoire naturelle d'une maladie, un point avant lequel un traitement est soit plus efficace, soit plus facile à appliquer qu'après ce point
 - *Exemple* : l'apparition de métastases, point critique d'un certain nombre de cancers.
 - *Exemple* : certaines maladies n'ont pas de point critique (connu à ce jour), comme le cancer du pancréas
- La localisation de ces points critiques est cruciale pour la valeur d'un diagnostic précoce. Le dépistage est utile s'il permet de poser un diagnostic avant le point critique

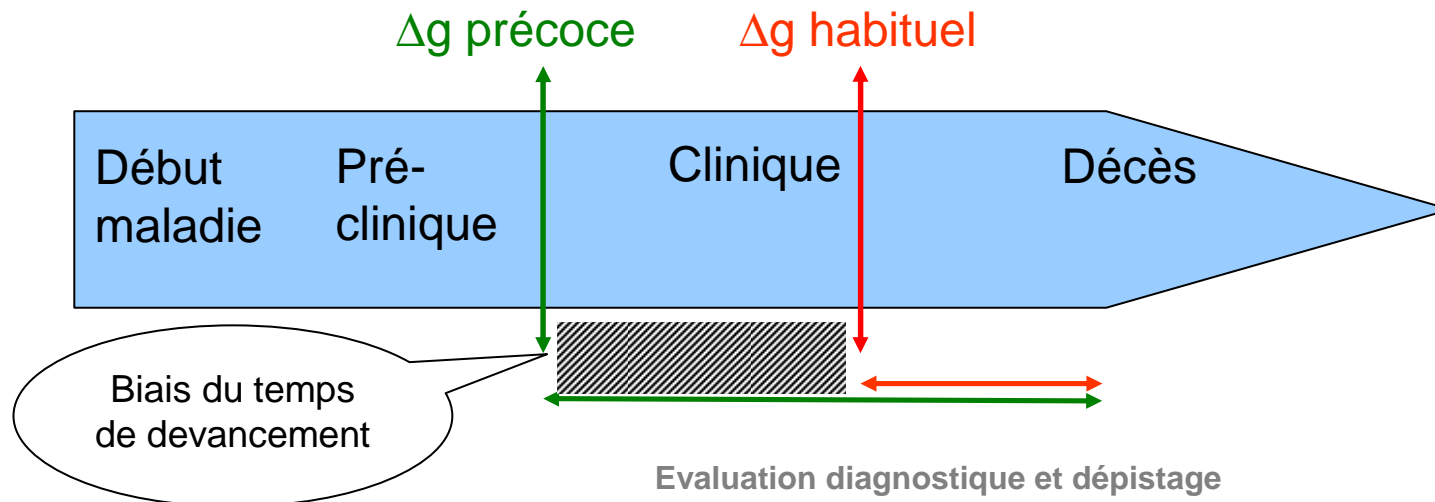
Évaluation d'un programme de dépistage

- Evaluation à 2 niveaux :
 - Performance propre du test de dépistage
 - Sensibilité
 - Spécificité
 - Efficacité d'un programme de dépistage
 - Utilité du dépistage précoce : allonger la durée de vie et la qualité de vie des sujets atteints
 - Essai d'intervention, souvent randomisé (~ essai thérapeutique)
 - Coût du dépistage
- *Exemple : dépistage généralisé du cancer du sein*
 - Performance diagnostique de la mammographie
 - Efficacité du dépistage sur la mortalité par cancer du sein

Évaluation d'un programme de dépistage

Biais du temps de devancement

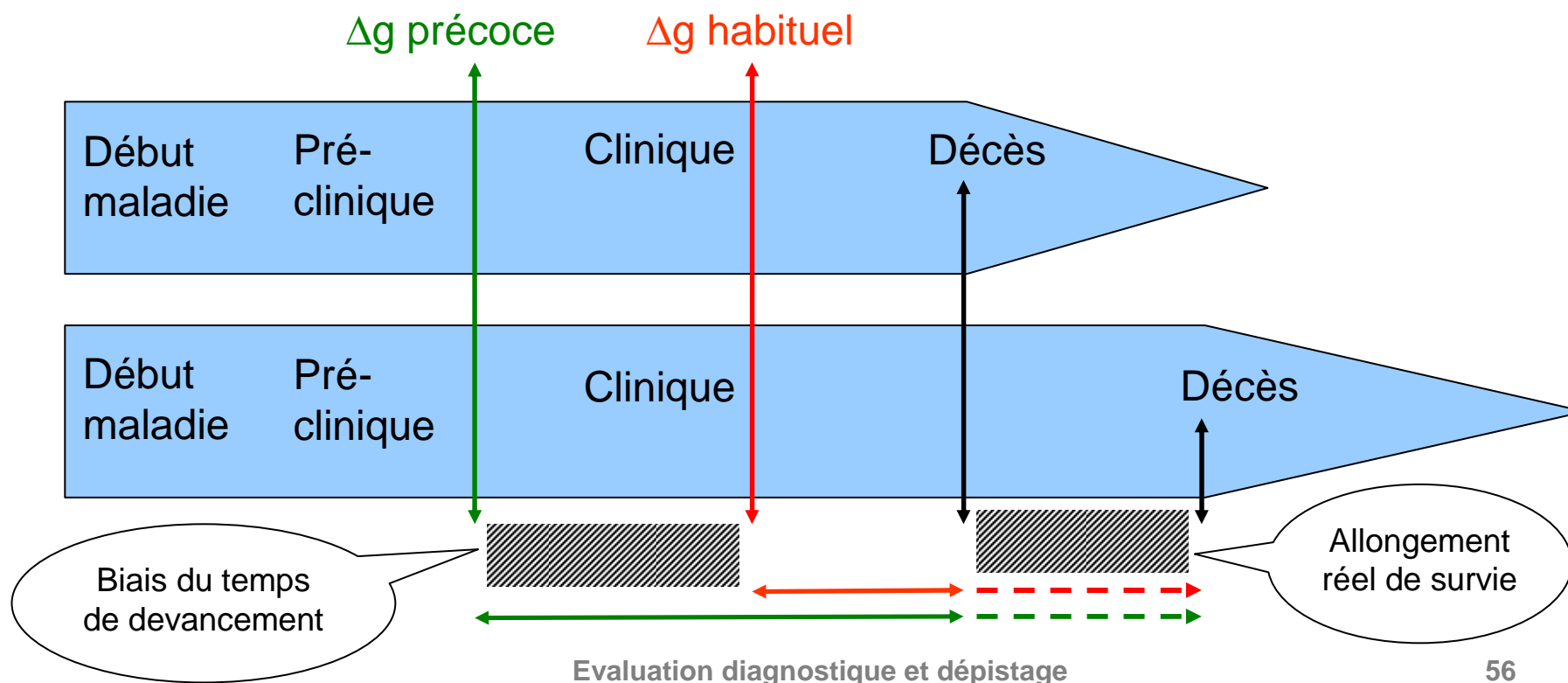
- Biais du diagnostic précoce : déplacement du temps zéro (zero-time shift ou lead-time bias). Période de temps entre la détection d'une maladie par dépistage et le moment où elle aurait été diagnostiquée habituellement après apparition des symptômes.
- **Augmentation artificielle de la durée de survie** par allongement de la durée de maladie (mais ne retarde pas la date de la mort)



Évaluation d'un programme de dépistage

Biais du temps de devancement

- Objectif idéal d'un dépistage : allonger la durée de survie en retardant la date de décès



Évaluation d'un programme de dépistage

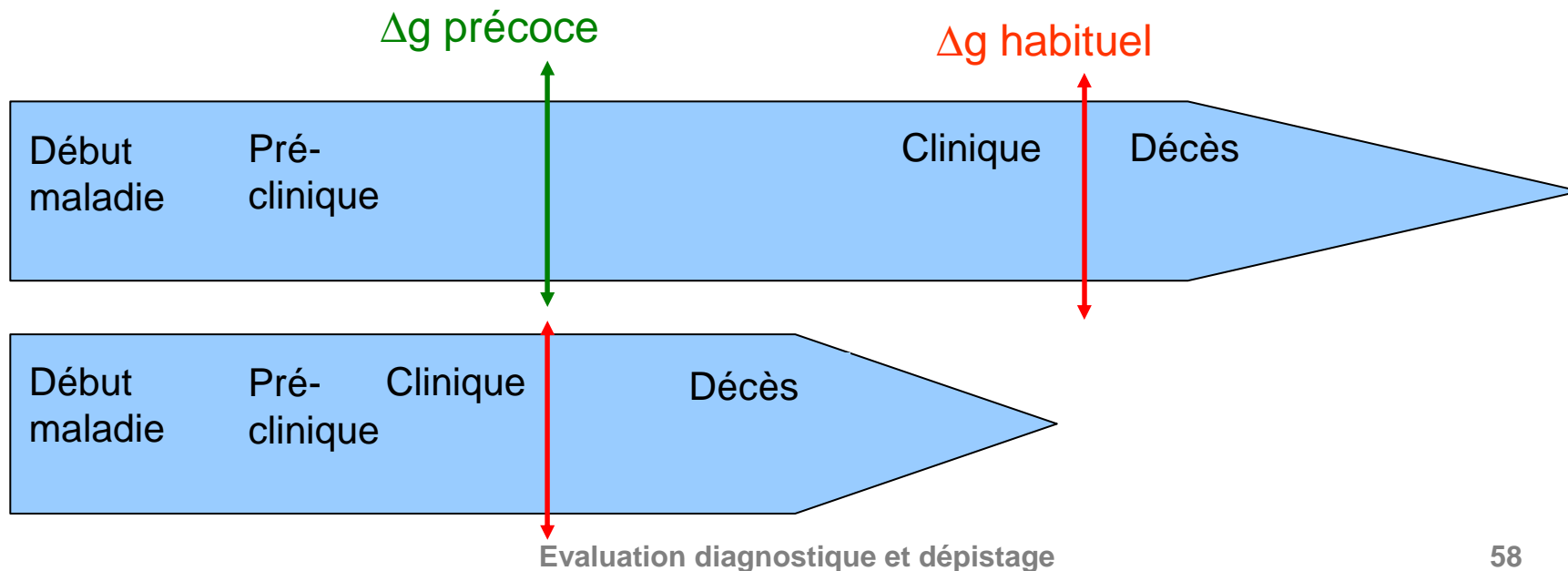
Biais du temps de devancement

- Prise en compte de ce biais : évaluer les taux de **mortalité par classe d'âge** plutôt que des taux de survie depuis le diagnostic
- *Exemples* : SIDA

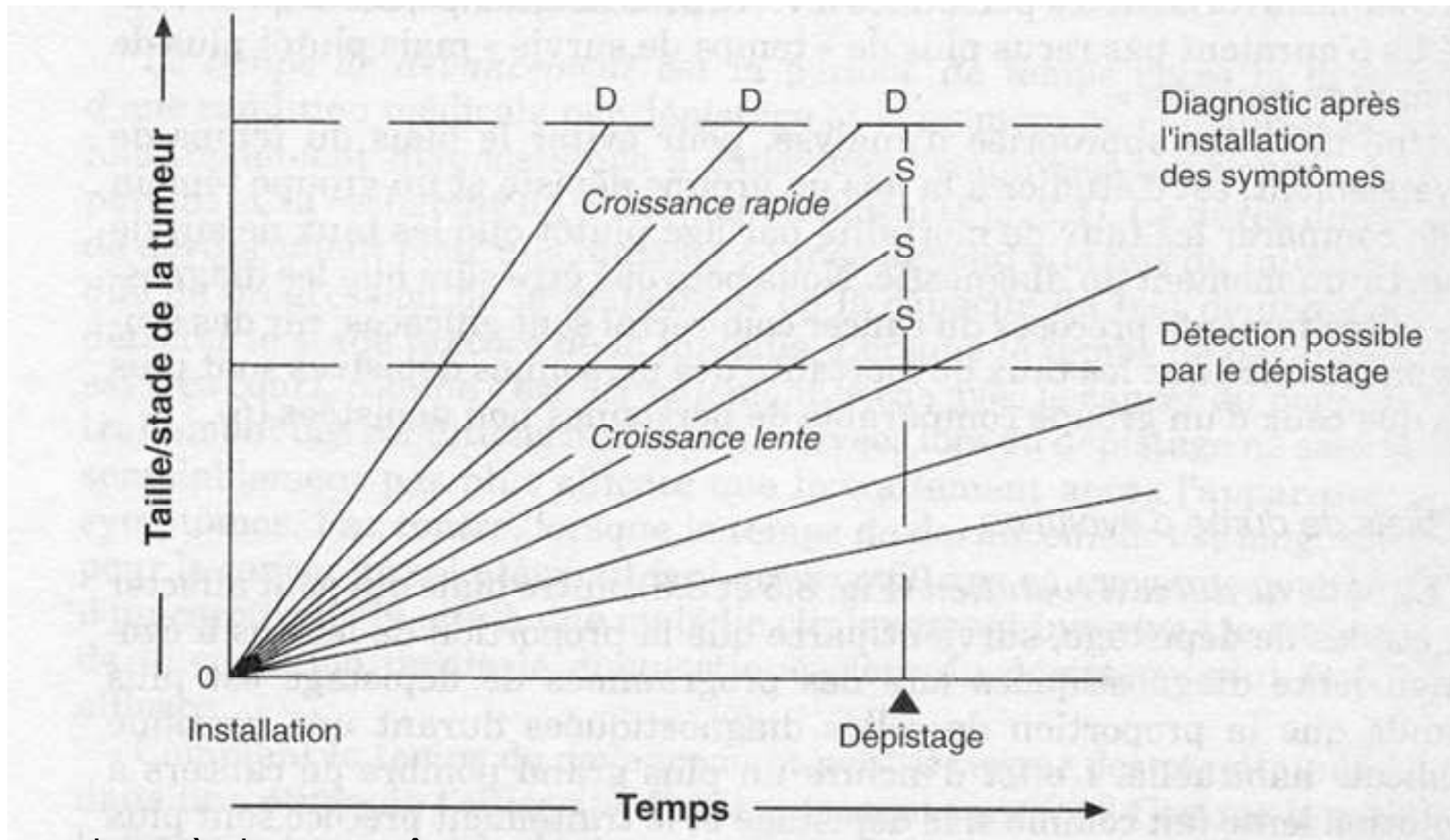
Évaluation d'un programme de dépistage

Biais de durée d'évolution

- On dépiste plus de maladie à évolution lente (phase préclinique longue) qu'à évolution rapide
 - Exemple : cancer de la prostate, cancer du pancréas
- Le dépistage favorise les cas les plus favorables, c'est à dire ceux dont l'histoire naturelle a l'évolution la plus lente



Évaluation d'un programme de dépistage



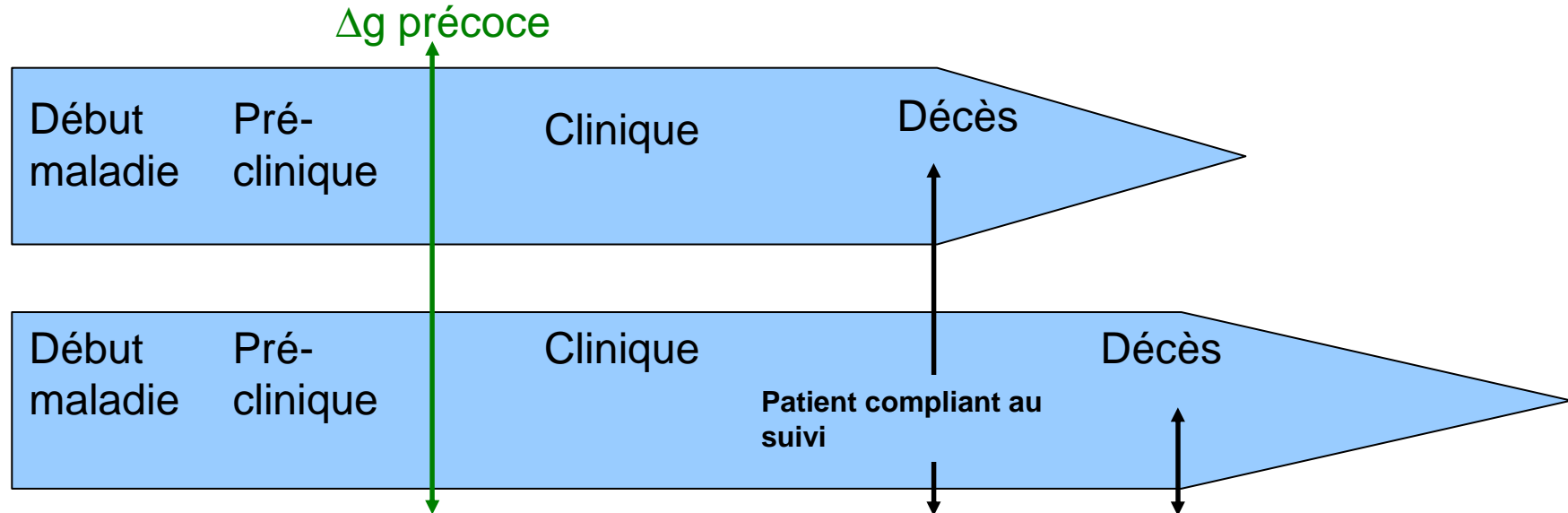
D = diagnostic après les symptômes

S = diagnostic après le dépistage

Évaluation d'un programme de dépistage

Biais d'observance ou de compliance

- Les patients compliants tendent à avoir 1 meilleur pronostic sans influence du dépistage
 - Rôle de facteurs liés à l'observance
 - Patients également plus compliants aux traitements



Évaluation d'un programme de dépistage

Biais de durée d'évolution et biais d'observance ou de compliance

- Prise en compte de ces biais : **essais contrôlés** avec **randomisation** du dépistage
 - Groupes comparables en termes de tumeurs à évolution lente ou rapide, et en termes de niveaux d'observance
- Suivi comparable des patients dans les 2 groupes
- Analyse de l'efficacité sur la population en **Intention de dépister** (équivalent de l'intention de traiter pour les essais thérapeutiques)

Assurer la comparabilité initiale des groupes

Maintenir la comparabilité initiale des groupes

Limites du dépistage

- Impact psychologique
 - Diminution de la qualité de vie du patient : inquiétude, sujet bien portant devient malade
 - Si pas de traitement disponible
 - Si faux positif
 - Augmentation des examens invasifs de « vérification »
- Augmentation de l'incidence de la maladie
 - Identification de cas qui n'auraient pas été diagnostiqués en l'absence de dépistage
 - Régression spontanée
 - Non progression
 - Durée d'évolution supérieure à l'espérance de vie

Conditions pour la mise en place d'un programme de dépistage de masse

- Maladie fréquente dans le groupe dépisté
- Maladie grave
- Existence d'une phase préclinique
- Amélioration attendue du pronostic si prise en charge précoce
- Traitement disponible
- Pas d'autres méthodes diagnostiques applicables
- Bonne performances diagnostiques (sensibilité, spécificité)
- Test fiable, reproductible
- Test peu coûteux
- Test non traumatisant, peu invasif : acceptabilité
- Test applicable à la population ciblée, simple

Récapitulatif

- Tests diagnostiques
 - Gold standard + test à étudier, biais de vérification
 - Performances diagnostiques = Se, Sp
 - Choix de valeur seuil (critère quantitatif), courbe ROC
 - Valeurs prédictives, RV, nomogramme de Fagan
 - Validité interne, validité externe d'une étude diagnostique
- Dépistage / prévention
 - Niveaux de prévention
 - Point critique
 - Evaluation d'un programme de dépistage, leur biais et moyens de les prendre en compte
 - Conditions de mise en place d'un programme de dépistage

Références

- Epidémiologie Clinique, R. FLETCHER, ed. Pradel
- U. Glenck, *Evidence-based medicine : comment juger une étude sur un test diagnostique?* Forum Med Suisse, N°9, 28 février 2001, http://www.medicalforum.ch/f/set_archiv.html