

LCA des études diagnostiques

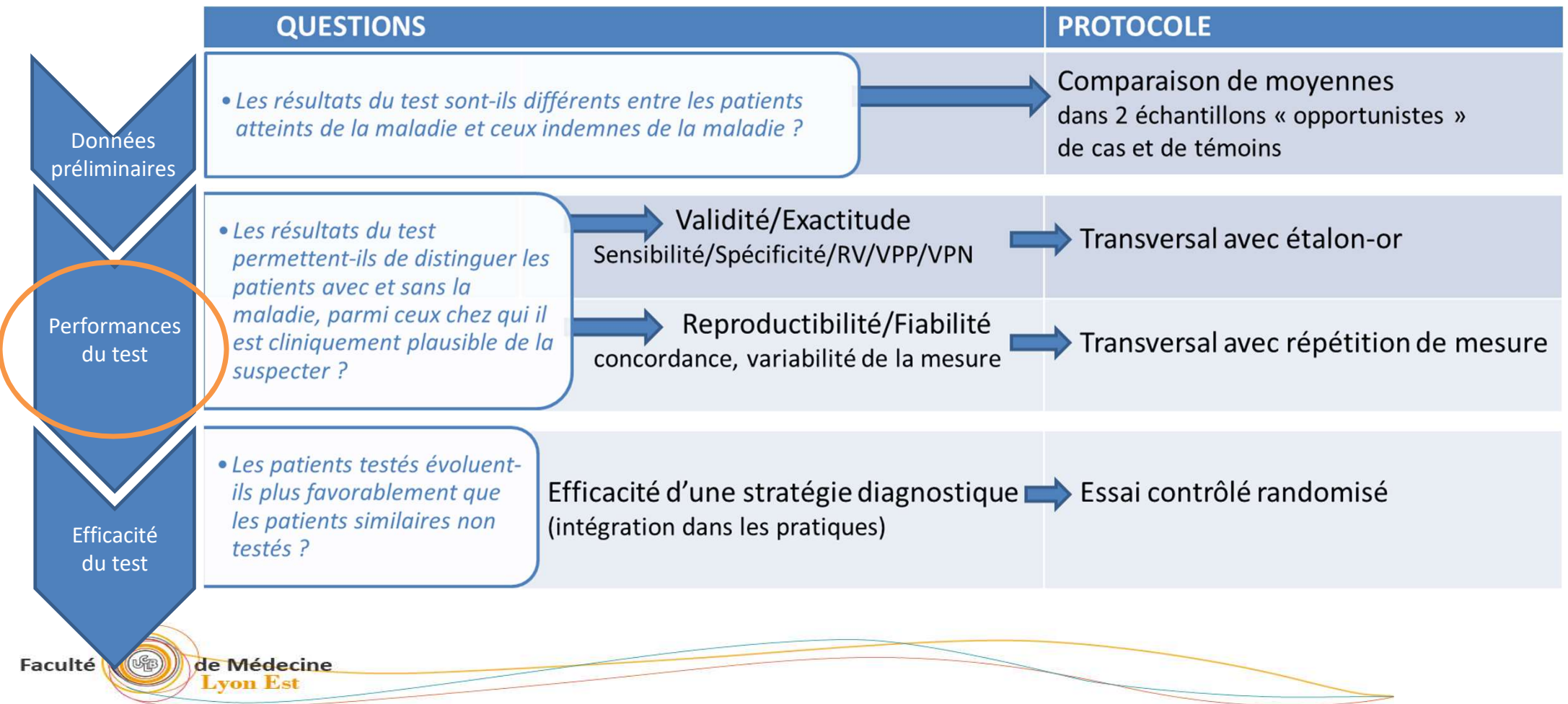


DFASM 1
Lyon Est 2025-2026

Laure Huot
laure.huot@chu-lyon.fr

ETAPES D'UNE DEMARCHE D'EVALUATION D'UN TEST DIAGNOSTIQUE

Quelle est la **question de recherche** et la **méthode** possible associée pour y répondre ?

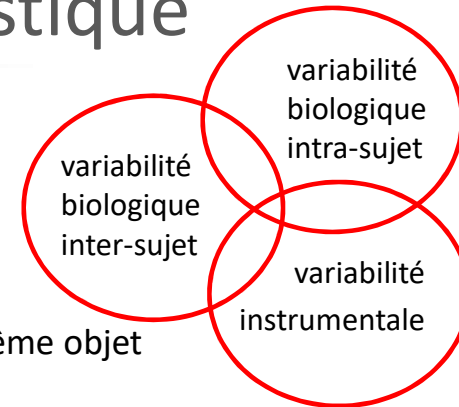


Evaluation de la performance d'un test diagnostique

VALIDATION SCIENTIFIQUE D'UN NOUVEAU TEST DIAGNOSTIQUE

❖ Fiabilité de la mesure

- Etude de la fiabilité = étude de la variabilité totale d'une méthode de mesure
- **Reproductibilité** : résultat similaire de deux ou plusieurs mesures de même nature se rapportant au même objet
 - Variabilité instrumentale (opérateur) → Estimation de la concordance (% Kappa)



❖ Validité du test

- Capacité du test à mesurer effectivement ce qu'il est censé mesurer
- Echantillon de n sujets recrutés et soumis à la fois
 - **Test diagnostique évalué**
 - **Test de référence (gold standard)** = Examen de référence / diagnostic de certitude
- Classement des individus (**tableau de contingences**)
 - Performances : **Sensibilité, Spécificité, Valeur Prédictive Positive** et **Valeur Prédictive Négative**
 - Pouvoir discriminant d'un test diagnostique : **Rapport de Vraisemblance**
 - Notion de seuil : **Courbe ROC**

Evaluation de la performance d'un test diagnostique

TABEAU DE CONTINGENCE

	Test de référence		
	MALADE (M+)	NON MALADE (M-)	
TEST POSITIF (T+)	VP	FP	$N1 = VP + FP$
TEST NEGATIF (T-)	FN	VN	$N0 = VN + FN$
	$M1 = VP + FN$	$M0 = VN + FP$	N

NB : Prévalence = $m1 / n$

Evaluation de la performance d'un test diagnostique

SENSIBILITÉ

	Test de référence		
	MALADE (M+)	NON MALADE (M-)	
TEST POSITIF (T+)	VP	FP	N1 = VP + FP
TEST NEGATIF (T-)	FN	VN	N0 = VN + FN
	M1 = VP + FN	M0 = VN + FP	N

Valeur intrinsèque

$$0 \leq Se \leq 1$$

$$Se = P(T+ / M+) = \frac{P(T+ \text{ et } M+)}{P(M+)} = \frac{VP}{(VP + FN)}$$

S N OUT

SN = Sensitivity (sensibilité)

N = Negative (négatif)

OUT = to rule Out (exclure)

Evaluation de la performance d'un test diagnostique

SPECIFICITÉ

	Test de référence		
	MALADE (M+)	NON MALADE (M-)	
TEST POSITIF (T+)	VP	FP	N1 = VP + FP
TEST NEGATIF (T-)	FN	VN	N0 = VN + FN
	M1 = VP + FN	M0 = VN + FP	N

Valeur intrinsèque

$$0 \leq Sp \leq 1$$

$$Sp = P(T- / M-) = \frac{P(T- \text{ et } M-)}{P(M-)} = \frac{VN}{(VN + FP)}$$

S P IN

SP = Specificity (spécificité)

P = Positive (positif)

IN = to rule In (confirmer)

Evaluation de la performance d'un test diagnostique

Exemple : Ultrasound for diagnosing acute salpingitis: a prospective observational diagnostic study

Table IV Diagnoses suggested by the ultrasound examiner on the basis of subjective evaluation of ultrasound findings.

Ultrasound diagnosis	True diagnosis		
	Salpingitis (n = 17)	Endometritis (n = 9)	Others (n = 26)
Salpingitis	14	4	4 ^a
Appendicitis	0	0	4 ^b
Normal ultrasound findings	2	3	10 ^c
No diagnosis suggested	1	2	5 ^d
Ovarian cyst	0	0	2 ^e
Adnexal mass of unclear origin	0	0	1 ^f

Vrais Positifs

Faux Négatifs

The **sensitivity** with regard to acute salpingitis of subjective interpretation of the ultrasound findings by the ultrasound examiner was **82% (14/17, 95% CI 57 – 96%)**
the **specificity** was **77% (27/35, 95% CI 60 – 90%)**

Faux Positifs

Vrais Négatifs

Evaluation de la performance d'un test diagnostique

VALEURS PREDICTIVES

	Test de référence		
	MALADE (M+)	NON MALADE (M-)	
TEST POSITIF (T+)	VP	FP	N1 = VP + FP VPP
TEST NEGATIF (T-)	FN	VN	
	M1 = VP + FN	M0 = VN + FP	N

Valeurs extrinsèques (dépendent de la prévalence de la maladie)

- Valeur prédictive positive = $VPP = P(M+/T+) = \frac{P(M+ \text{ et } T+)}{P(T+)} = \frac{VP}{VP+FP}$
- Valeur prédictive négative = $VPN = P(M-/T-) = \frac{P(M- \text{ et } T-)}{P(T-)} = \frac{VN}{VN+FP}$

Evaluation de la performance d'un test diagnostique

RAPPORTS DE VRAISEMBLANCE = pouvoir discriminant

❖ **Rapport de Vraisemblance Positif (RV+ ou L) :**
$$L = \frac{P(T+ / M+)}{P(T+ / M-)} = \frac{Se}{1 - Sp}$$
 $1 \leq L \leq +\infty$

- Ex : Salpingite $RV+=L = 0,82/(1-0,77)=3,6$
→ Le test est 3,6 fois plus souvent positif chez les malades que chez les non-malades
→ Le patient a 3,6 fois plus de chance d'être malade si le test est positif
- Pour éviter les FP, privilégier les tests avec un rapport de vraisemblance positif le plus élevé possible (VPP↑)

❖ **Rapport de Vraisemblance Négatif (RV- ou λ) :**
$$\lambda = \frac{P(T- / M+)}{P(T- / M-)} = \frac{1 - Se}{Sp}$$
 $0 \leq \lambda \leq 1$

- Ex : Salpingite $\lambda = 0,18/0,77=0,23$
→ Le test est 4 à 5 fois (1/0,23) plus souvent négatif chez les non malades que chez les malades
- Pour éviter les FN, privilégier les tests avec un rapport de vraisemblance négatif le plus faible possible (VPN↑)

Evaluation de la performance d'un test diagnostique

Rapport de vraisemblance d'un test : exemples de grandeurs

	Likelihood ratio	Change in likelihood of disease after test
RV+/ LR+	>10	Large increase
	5 - 10	Moderate increase
	2 - 5	Small increase
	1 - 2	Minimal increase
	1	No change
RV- /LR-	0.5 - 1.0	Minimal decrease
	0.2 - 0.5	Small decrease
	0.1 - 0.2	Moderate decrease
	<0.1	Large decrease

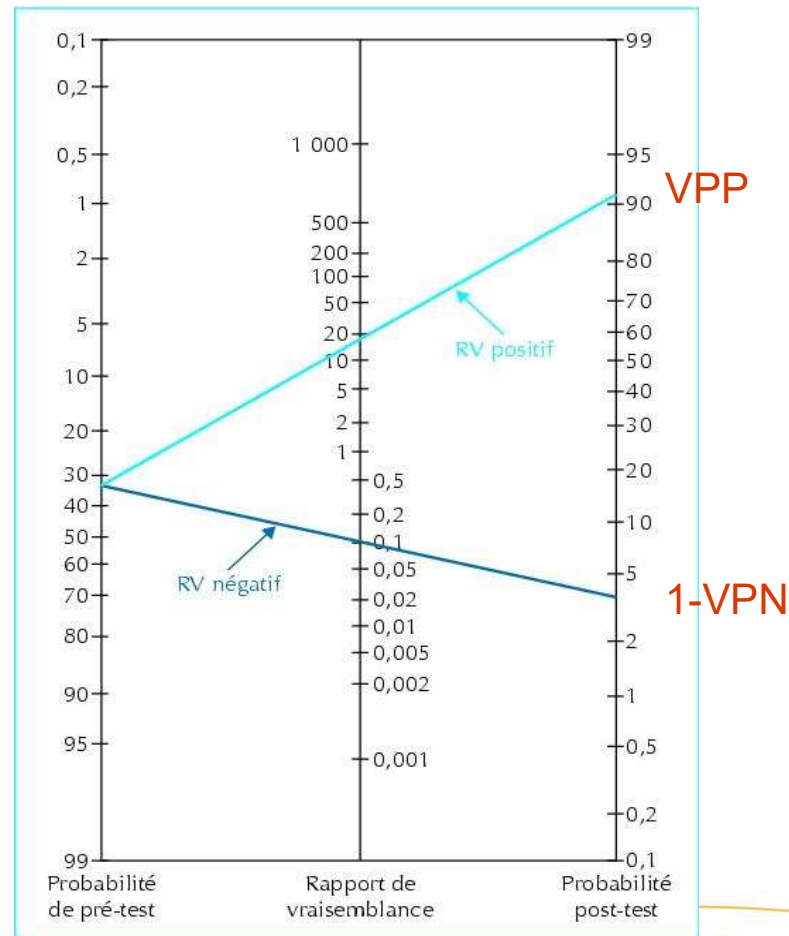
Evaluation de la performance d'un test diagnostique

NOMOGRAMME DE FAGAN

Probabilité pré-test de la maladie :

Probabilité que le sujet ait la maladie avant de connaître le résultat du test diagnostique

(Prévalence)



Probabilité post-test de la maladie :

Probabilité révisée que le sujet ait la maladie connaissant le résultat (positif ou négatif) du test diagnostique

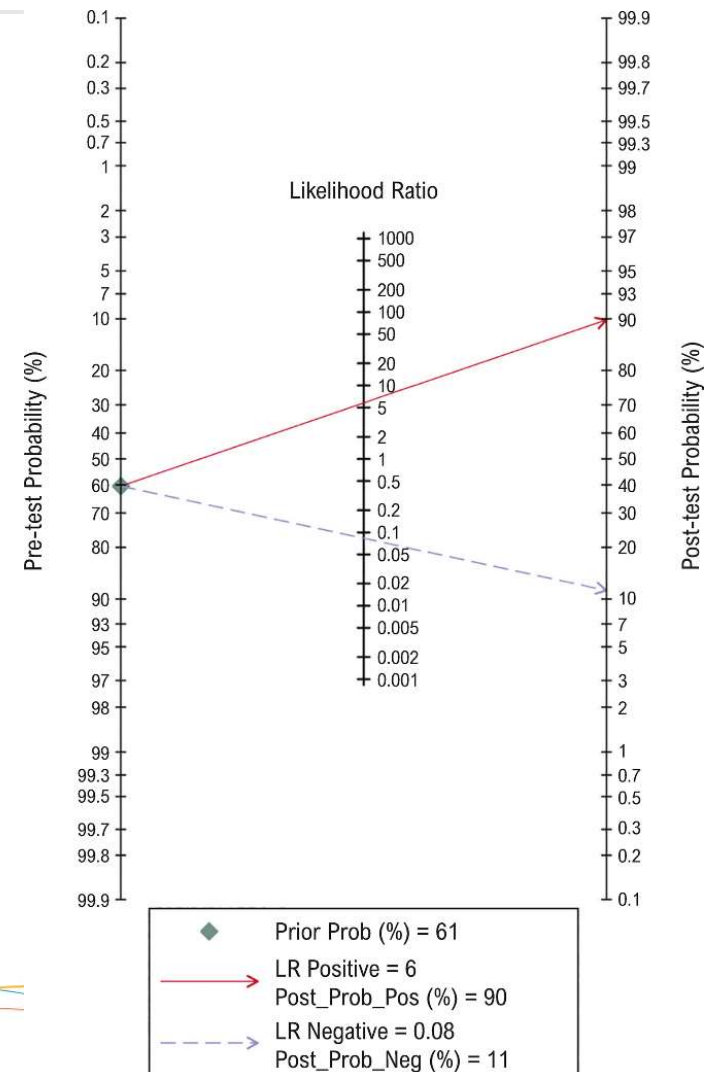
(Valeur prédictive)

Evaluation de la performance d'un test diagnostique

Exemple : Diagnostic Role of NT-proBNP in Patients with Cardiac Amyloidosis Involvement

“Fig: The clinical utility of NT-proBNP.

The Fagan nomogram plot showed the pre-test probability, PLR, NLR, and post-test probability of NT-proBNP for the diagnosis of heart injury.”



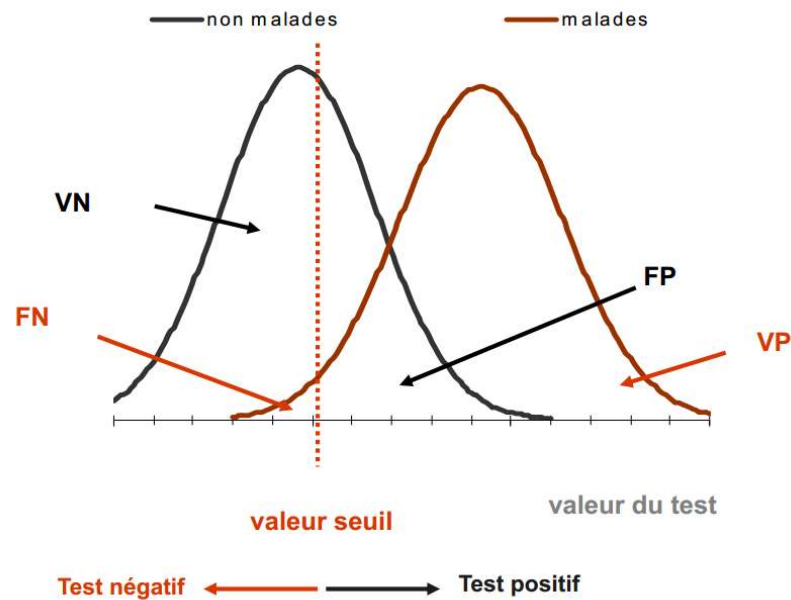
Zhang Y et al. Arq Bras Cardiol. 2022;119(2):212-222.
PMID: 35544851

Evaluation de la performance d'un test diagnostique

COURBE ROC

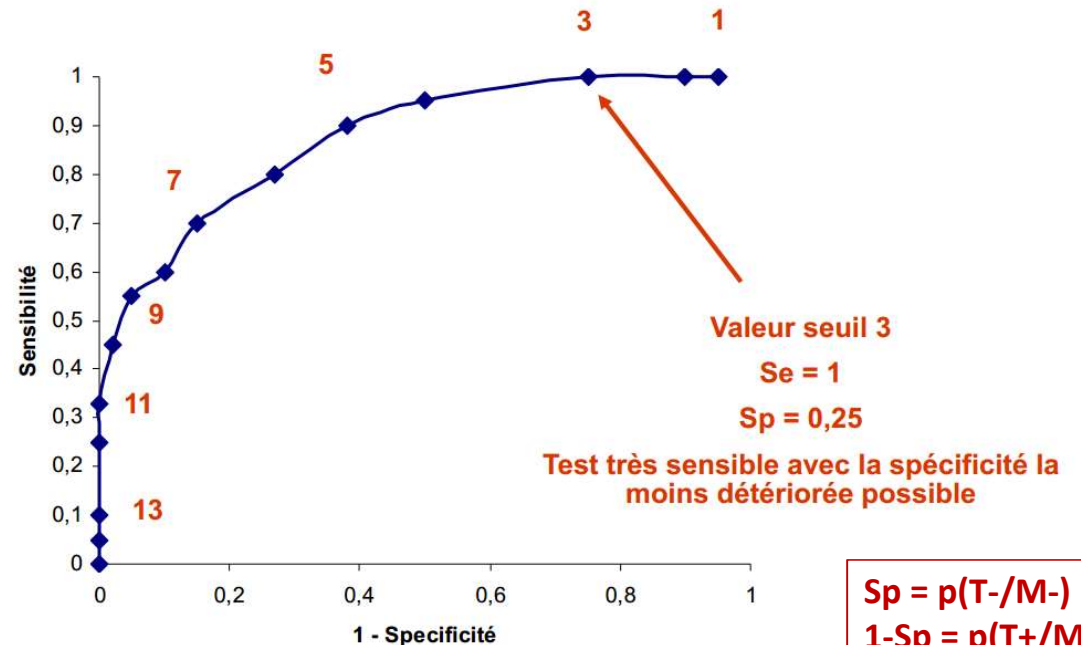
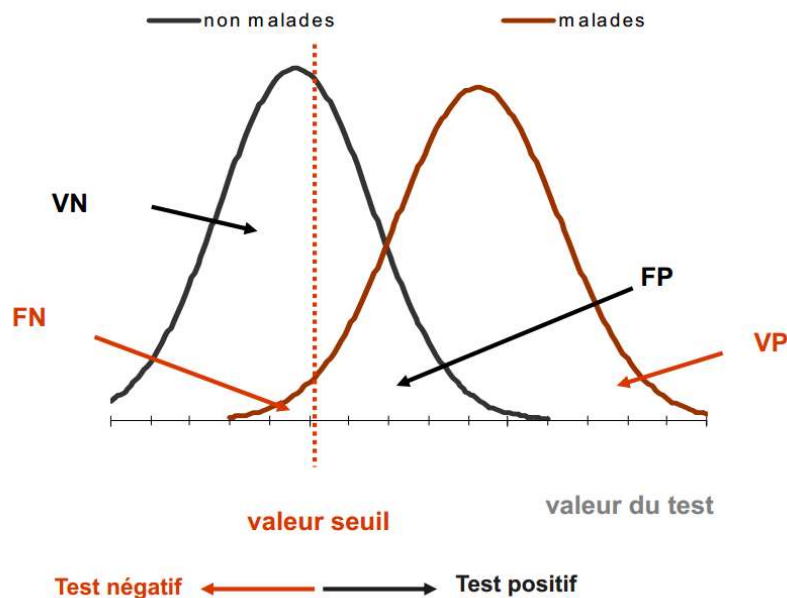
❖ **Test diagnostique quantitatif** = résultat sous forme de valeur numérique continue (ex: glycémie, PSA...)

- Choix d'un seuil ? En fonction de la valeur retenue, il y aura plus de FP ou FN -> conséquences sur performances



Evaluation de la performance d'un test diagnostique

COURBE ROC = Représentation graphique des caractéristiques intrinsèques pour l'ensemble des seuils envisageables -> dichotomisation



ROC : Receiver Operating Characteristics

Evaluation de la performance d'un test diagnostique

COURBE ROC

❖ Aire sous la courbe ROC :

- Estimation globale des caractéristiques intrinsèques (Se, Sp) du test pour ses différentes valeurs
- Comprise entre 0,5 et 1
- Comparaison possible des performances globales de 2 ou plusieurs tests quantitatifs

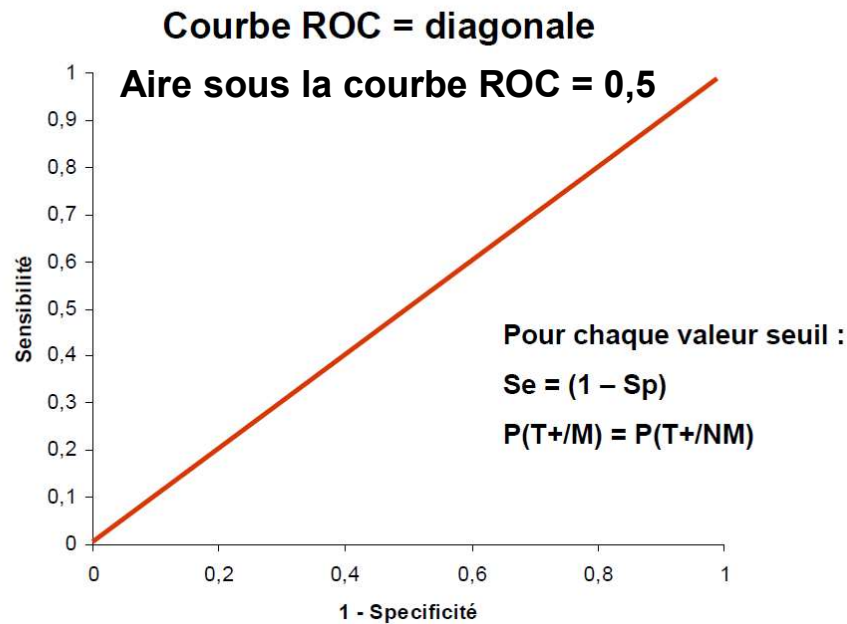
AUC : Area Under Curve - AUROC : Area Under the Receiver Operating Characteristics



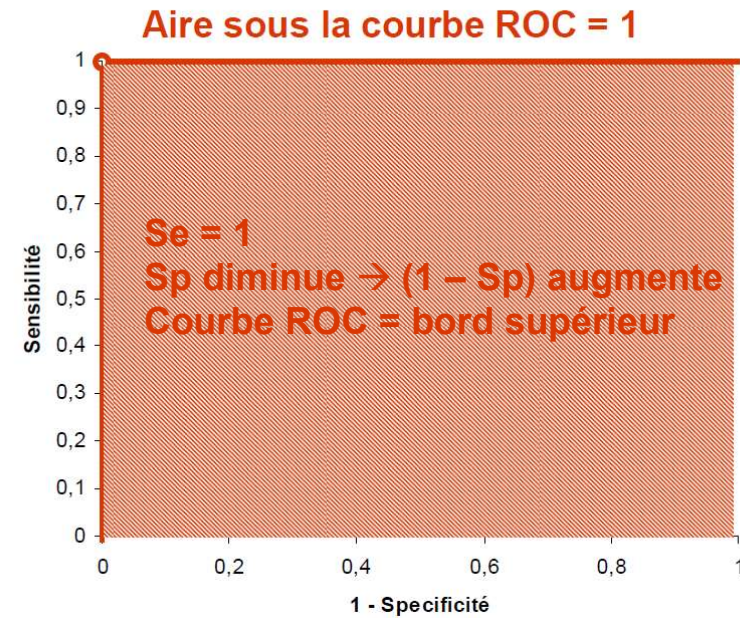
Evaluation de la performance d'un test diagnostique

COURBE ROC

Test quantitatif qui ne fait pas mieux que le hasard

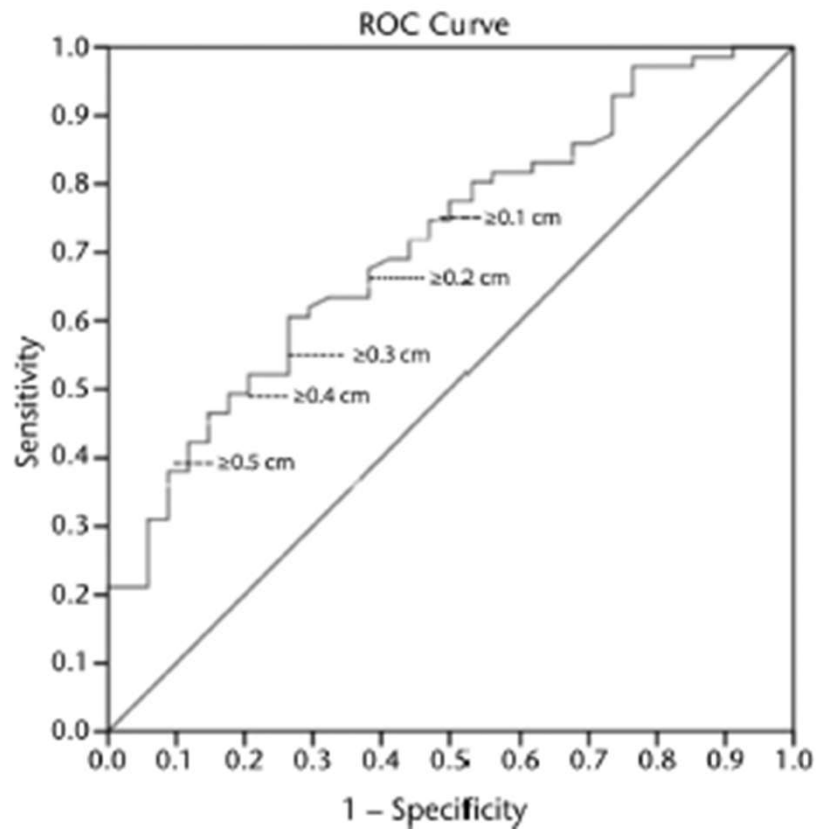


Test quantitatif parfaitement discriminant



Evaluation de la performance d'un test diagnostique

Exemple ECN 2019 : Échographie versus palpation digitale pour le diagnostic de luxation glénohumérale post-AVC



Receiver operating characteristic (ROC) curve with area under the receiver operating characteristic (**AUROC**) curve value of **0.73** (95% confidence interval=0.63–0.83).

Curved line shows sensitivity and specificity varied around different cutoff points.

Straight line indicates AUROC curve 0.5 (test not useful). Cutoff points 0.1 to 0.5 cm correspond to sensitivity and specificity.
Optimal cutoff point of 0.2 cm indicates 68% sensitivity and 62% specificity.

Evaluation de la performance d'un test diagnostique

Indice	Probabilité	Calcul	Valeurs	Varie en fonction de la prévalence
Se	$P(T+/M+)$	$VP / (VP + FN)$	$0 \rightarrow 1$	Non
Sp	$P(T-/M-)$	$VN / (VN + FP)$	$0 \rightarrow 1$	
RV+	$P(T+/M+) / P(T+/M-)$	$Se / (1 - Sp)$	$1 \rightarrow +\infty$	
RV-	$P(T-/M+) / P(T-/M-)$	$(1 - Se) / Sp$	$0 \rightarrow 1$	
VPP	$P(M+/T+)$	$VP / (VP + FP)$	$0 \rightarrow 1$	Oui
VPN	$P(M-/T-)$	$VN / (VN + FN)$	$0 \rightarrow 1$	

	Examen de référence		
	MALADE (M+)	NON MALADE (M-)	
TEST POSITIF (T+)	VP	FP	N1 = VP + FP
TEST NEGATIF (T-)	FN	VN	N0 = VN + FN
	M1 = VP + FN	M0 = VN + FP	

Biais des études de performance de tests diagnostiques

BIAIS À RECHERCHER

▪ Biais de sélection ?

= Echantillon non représentatif :

- Caractéristiques $\neq$ de la population cible
- Population insuffisamment variée
- Prévalence $M \neq$

L'absence de maladie est plus facile à étudier chez les patients sans comorbidités.

La présence de maladie est plus facile à diagnostiquer si la maladie est sévère.

Biais des études de performance de tests diagnostiques

BIAIS À RECHERCHER

▪ Biais de sélection ?

○ Biais de spectre : spectre des patients insuffisant

- Profil clinique des sujets d'étude différent de celui de la population cible : le plus souvent, stades précoces (ex : asymptomatique) sous-représentés -> Se surestimée
- Recrutement de type 'cas témoins' + à risque : inclusion de cas sévère et de contrôles en bonne santé -> Se / Sp surestimées (- de FN et - de FP)

→ Mauvaise estimation de la performance

→ Extrapolation limitée

Using a test in a population who **obviously do not have a disease** will artificially improve the test's **sensitivity**.

Using a test in a population who **obviously do have a disease** will artificially improve the test's **specificity**.

! Pour l'éviter : recrutement consécutif sans connaître le statut malade ou non malade (approche 'cohorte')

Biais des études de performance de tests diagnostiques

BIAIS À RECHERCHER

▪ Biais de classement ?

○ Observation ou évaluation investigateur dépendant

- Selon l'examen réalisé, le type et le nombre d'observateurs
- Variabilité inter-observateurs et intra-observateur : lié à l'expérience ou à la pratique de l'investigateur

! Pour l'éviter / le réduire :

- Standardiser le recueil
- Interpréter les 2 tests indépendamment et en insu l'un de l'autre

○ Biais d'interprétation : interprétation non indépendante des deux tests

- Connaissance du résultat du test de référence (absence d'insu)

○ Biais d'incorporation : le résultat du test évalué est pris en compte pour déterminer le statut réel des sujets (cf. exemple)

Biais des études de performance de tests diagnostiques

Exemple : Diagnostic de bactériémie chez le nourrisson fébrile

- Test de référence = avis clinicien + hémoculture
- Test évalué = CRP

Problématique = pour orienter le diagnostic final, les cliniciens se basent aussi sur des biomarqueurs – dont taux de CRP !

-> **Biais d'incorporation = surestimation Se et Sp de la CRP**

Car classement M+ / M- basé en partie sur le test évalué

Biais des études de performance de tests diagnostiques

BIAIS À RECHERCHER

▪ Biais de classement ?

- **Biais de vérification / Work-up bias** : lorsque le test de référence n'est pas réalisé chez tous les participants de l'étude (cf. exemples)

- Gold standard imparfait
- Gold standard invasif / coûteux

➤ Classement des patients :

- Statut vérifié = malade
- Statut vérifié = non malade
- Statut non vérifié -> patient considéré comme non malades (alors qu'il n'a pas eu de test de référence) ou exclu de l'étude

	Vérifié			Non-vérifié
	Malade	Non-malade	Total	
Positif	90	10	100	5
Négatif	25	75	100	20

Se = 78%

-> 75,9%

Sp = 88%

-> 89,6%

! Surestimation Se (peu de FN) et sous-estimation de Sp (plus de FP)

Biais des études de performance de tests diagnostiques

Exemple : Diagnostic de bactériémie chez le nourrisson fébrile

- Test évalué = PCT
- Test de référence = hémoculture

Problématique = hémoculture est un test invasif pour le nourrisson -> le clinicien ne souhaite le réaliser que s'il a une forte présomption

-> **Biais de vérification**, car le résultat du test évalué ne sera pris en compte que chez les nourrissons ayant eu le test de référence -> ont un risque plus élevé d'avoir une bactériémie

! Le test de référence doit être réalisé chez tous les patients éligibles / de l'étude

Biais des études de performance de tests diagnostiques

Exemple : Diagnostic de l'embolie pulmonaire

- Test évalué = D-Dimères
- Test de référence =
 - 1) scintigraphie de ventilation-perfusion (si D-Dimères +)
 - 2) suivi clinique de routine (si D-Dimères -)

Problématique = 2 gold standard, plus ou moins invasif, dont le choix est basé sur le résultat du test évalué !

-> **Biais de vérification:** patients avec embolie pulmonaire asymptomatique, mais résultats négatifs pour les D-dimères, peuvent ne pas avoir été diagnostiqués par le suivi de routine / les symptômes peuvent avoir disparu dans l'intervalle (FN)

Evaluation de la performance d'un test diagnostique

BIAIS À RECHERCHER

▪ Biais de classement ?

- **Biais d'ordre séquentiel** : lorsque le délai de réalisation entre test et gold standard n'est pas suffisamment court
 - Risque d'évolution de la maladie entre les 2 tests
 - NB Autre nom = biais d'évolutivité

Evaluation de la performance d'un test diagnostique

BIAIS A RECHERCHER / CHECK-LIST !

- Comparaison avec un test de référence (gold standard) unique, indépendant, en insu ?
- Le spectre des patients correspond-t-il à celui auquel on souhaiterait proposer le test en pratique ?
- Est-ce que le gold standard était appliqué quel que soit le résultat du test évalué ?

Evaluation de la performance d'un test diagnostique



A RETENIR

❖ Validité de l'étude :

- Utilisation d'un étalon-or ou « gold standard » fiable et validé ?
 - Test de référence pratiqué chez tous les patients inclus dans l'étude ?
 - Test de référence et test étudié interprétés indépendamment ?
- Échantillons Malades et Non Malades représentatifs de la population à laquelle le test sera appliquée ?

❖ Performances du test

- Choix du cut-off point sous le contrôle des investigateurs ?
- Sensibilité/Spécificité/Rapport vraisemblance + et –
 - Intervalle de confiance ?

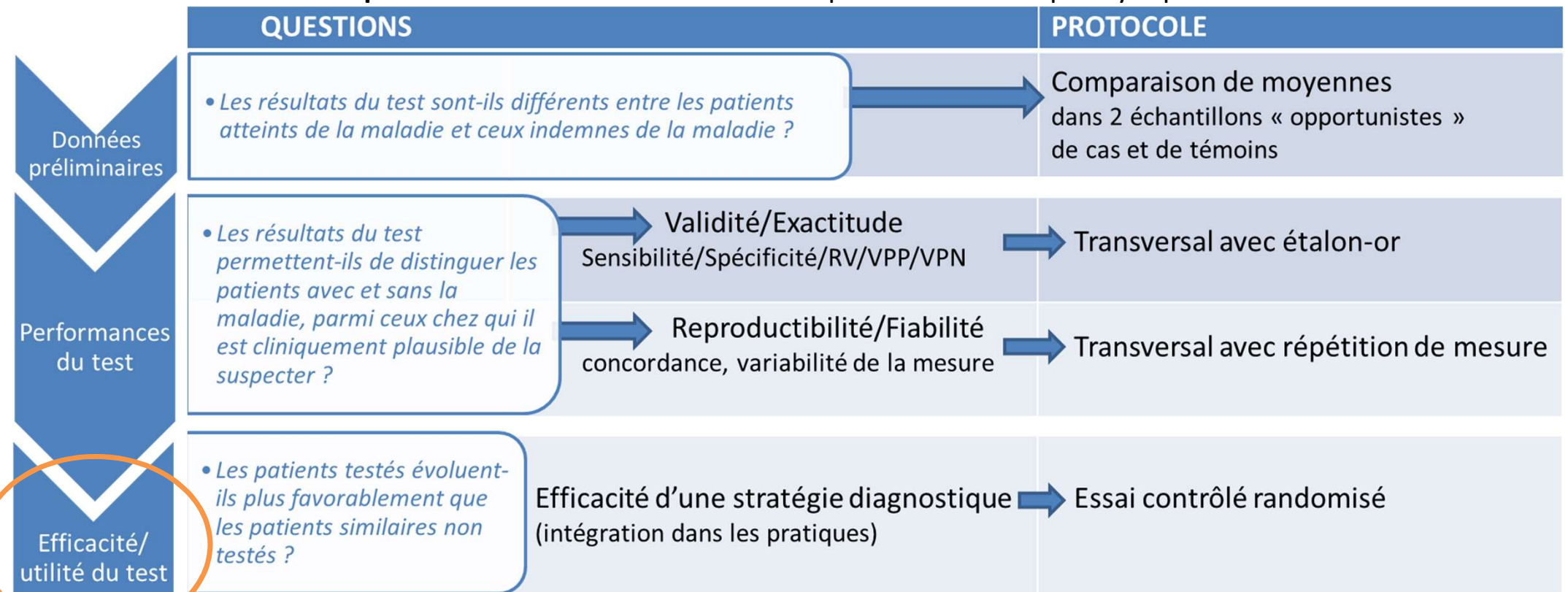
❖ Applicabilité des résultats :

- Reproductibilité évaluée ?
- L'utilisation du test améliore l'état de santé → étude stratégie diagnostique



ETAPES D'UNE DEMARCHE D'EVALUATION D'UN TEST DIAGNOSTIQUE

Quelle est la **question de recherche** et la **méthode** possible associée pour y répondre ?



Evaluation de L'EFFICACITE d'une STRATEGIE diagnostique/de dépistage

L'ÉVALUATION DIAGNOSTIQUE DÉPEND À LA FOIS :

- ❖ De l'exactitude et de la fiabilité du test → études évaluant la PERFORMANCE d'un test diagnostique
- ❖ De la prise en charge entreprise à la suite du résultat du test : **IMPACT et UTILITÉ** du test
 - Etude de référence : **ESSAI CLINIQUE randomisé** évaluant l'efficacité d'une **STRATÉGIE diagnostique/de dépistage**
 - Idem efficacité d'une intervention : nouvelle stratégie diagnostique vs. ancienne/PEC habituelle
 - Formulation de l'objectif reprend le PICO :
 - **P** = Population à risque de la maladie
 - **I** = Procédure de diagnostic/dépistage
 - **C** = Absence de stratégie de diagnostic/dépistage
 - **O** = Mortalité ou guérison

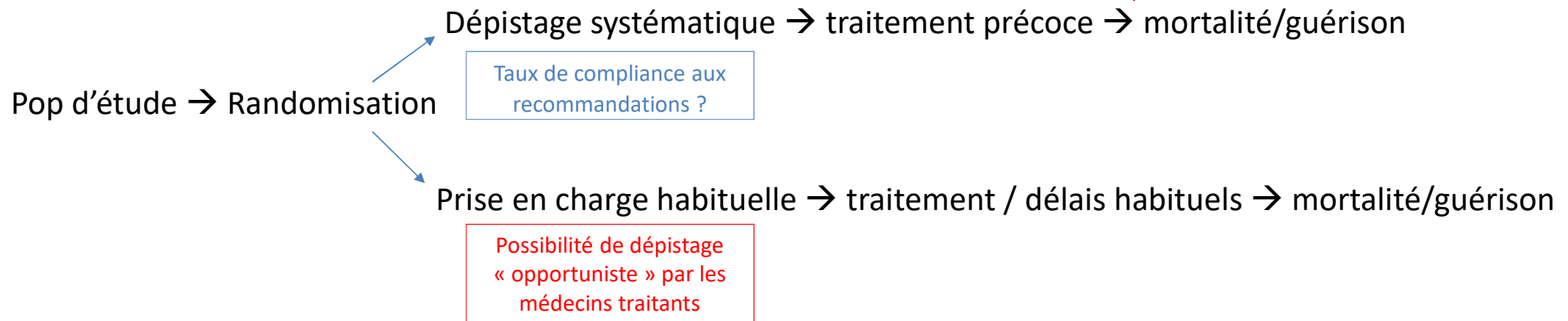
Exemple: évaluer l'efficacité d'un dépistage organisé par frottis cervico-vaginal (I) par rapport à une absence de dépistage organisée (C) en termes de mortalité (O) chez les femmes de 25 à 65 ans (P)

Evaluation de L'EFFICACITE d'une STRATEGIE diagnostique/de dépistage

Type d'étude : Essai randomisé contrôlé généralement à deux bras parallèles

Exemple d'étude visant à évaluer une stratégie de dépistage systématique

Intervention contrôlée par l'EC | PEC non contrôlée par l'EC | Evaluation de l'intervention



Evaluation de L'EFFICACITE d'une STRATEGIE diagnostique/de dépistage

Application des notions des essais cliniques

❖ Qualité de la randomisation +++

❖ Critère de jugement

- Pertinence clinique ?
- Subjectivité ?
- Mode d'évaluation ?

❖ Analyse en intention de dépister = idem ITT

- Analyse de tous les patients randomisés dans leur groupe **quel que soit leur suivi ou leur compliance**
- Reflet de l'impact de la procédure de dépistage dans la vraie vie

❖ Interprétation des résultats, considérer :

- La réduction absolue de risque
- Le nombre de sujets à dépister pour éviter un évènement

➡ 4 biais identifiables dans les EC :

- biais de sélection : différentiel/non différentiel
- biais d'attrition
- biais de suivi
- biais de mesure



Evaluation de L'EFFICACITE d'une STRATEGIE diagnostique/de dépistage

- ❖ Histoire de la maladie bien connue
- ❖ Maladie détectable en phase asymptomatique
 - Test de dépistage avec forte Sp et forte Se (étude de performance reportée dans l'article ?)
 - Simple à utiliser
 - Bonne acceptabilité des sujets
 - Compliance ?
- ❖ Moyens diagnostiques et traitement disponible
- ❖ Existence d'un traitement efficace
- ❖ Coût du dépistage/diagnostic/traitement non disproportionné pour la collectivité
- ❖ Possibilité de réaliser le programme sur le long terme de façon continue (implémentation et maintenance)
 - coût ?
 - diminution du suivi des recommandations de la population à risque au cours du temps ?

- ❖ Preuve de son efficacité / EC randomisé

A retenir

- Les modes de calcul des indicateurs de performance
- SPIN SNOOUT
- Reproductibilité et fiabilité
- Caractéristiques du gold standard
- Les grands types de biais
- Evaluation d'efficacité

In English

- Diagnosis accuracy study
- Reproducibility
- Reliability
- Sensitivity / Specificity
- Likelihood ratio
- Area Under the Curve