

## LCA: Essai diagnostique

Plan		
<ul style="list-style-type: none"> <li>- Types de tests diagnostiques</li> <li>- Paramètres opérationnels</li> <li>- Analyse d'un test quantitatif</li> <li>- Analyse d'un test qualitatif</li> <li>- Critères de validité</li> </ul>		
Objectifs CNCI		
- cf liste des <a href="#">objectifs en LCA</a> du CNCI		
Recommandations	Mots-clés / Tiroirs	NPO / PMZ
<ul style="list-style-type: none"> <li>- Grades <a href="#">ANAES 00</a></li> <li>- <a href="#">STARD checklist</a></li> </ul>	<ul style="list-style-type: none"> <li>- Tests de dépistage / de confirmation</li> <li>- Se-Sp / RVP-RVN / VVP-VPN</li> <li>- Bayes / Fagan / VVP = <math>p \times RVP</math></li> <li>- ROC / Pearson / Bland-Altman</li> <li>- Reproductibilité / concordance</li> <li>- Gold standard validé / description ++</li> </ul>	<ul style="list-style-type: none"> <li>- Tous les patients ont les 2 tests</li> <li>- Présence des IC95</li> </ul>

### - Types de test diagnostiques

#### Tests qualitatifs et quantitatifs

- Test qualitatif
  - Résultat:
    - binaire (dichotomique): + ou - (ex: crachats BK)
    - nominal: échelle sans hiérarchie (ex: groupes sg: A/B/AB/O)
    - ordinal: échelle hiérarchique (ex: ABG: S/I/R)
  - → évaluation de la reproductibilité par coefficient kappa ( $\kappa$ )
- Test quantitatif
  - Résultat
    - toujours une valeur numérique continue (ex: glycémie)
    - problème = valeur seuil normal / pathologique
  - → détermination de la valeur seuil par courbe ROC (+/- Pearson / Bland-Altman)

#### Tests de dépistage et de confirmation

	Test diagnostique de dépistage	Test diagnostique de confirmation
--	--------------------------------	-----------------------------------

Objectif	<ul style="list-style-type: none"> <li>Dépistage d'une maladie latente</li> <li>Diminution de la mortalité</li> </ul>	<ul style="list-style-type: none"> <li>Diagnostic de certitude</li> <li>Pour décision thérapeutique</li> </ul>
Contexte	<ul style="list-style-type: none"> <li>Problème de santé publique</li> <li>Prévention secondaire</li> <li>Histoire naturelle connue</li> </ul>	<ul style="list-style-type: none"> <li>A la suite d'un test de dépistage</li> <li>Si doute diagnostique</li> </ul>
Sujets	<ul style="list-style-type: none"> <li>sur une population</li> <li>dépistage de masse</li> </ul>	<ul style="list-style-type: none"> <li>sur un individu</li> <li>confirmation ciblée</li> </ul>
Se/Sp	<ul style="list-style-type: none"> <li><b>Se</b> +++ / Sp ↓</li> <li>VPN maximale</li> </ul>	<ul style="list-style-type: none"> <li><b>Sp</b> +++ / Se ↓</li> <li>VPP maximale</li> </ul>
Avantages	<ul style="list-style-type: none"> <li>Coût faible / rapide</li> <li>Bonne faisabilité</li> <li>Bonne acceptabilité</li> </ul>	<ul style="list-style-type: none"> <li>Validité élevée</li> <li>Bonne reproductibilité</li> <li>Peu de biais</li> </ul>
Inconvénients	<ul style="list-style-type: none"> <li>Faible validité: biais souvent nombreux</li> </ul>	<ul style="list-style-type: none"> <li>Coût plus élevé</li> <li>Plus lourd, souvent invasif</li> </ul>

### Conditions pour le dépistage de masse (OMS) +++

- **Maladie**
  - fréquente / grave / problème de santé publique
  - curable si diagnostic précoce / ↓ mortalité
  - latent: évolution naturelle connue et lente
- **Population cible**
  - **connue**: données démographiques (ex: femmes entre 50 et 75ans)
  - accessible / mobilisable / dépistage ressenti comme un besoin..
- **Tests**
  - test de dépistage: validé / simple / acceptable / peu coûteux / reproductible
  - test de confirmation: existant / validé / acceptable

### - Paramètres opérationnels d'un test diagnostique

#### Quelque soit le paramètre évalué

- **Tableau de contingence** +++

	M+	M-
--	----	----

T+	VP	FP
T-	FN	VN

- !! Pour que l'article soit valable, vérifier
  - La présence d'une définition précise et valable pour FP/FN/VP/VN +++
  - La représentativité de l'échantillon: critères de sélection / prévalence

### Sensibilité (Se) / Spécificité (Sp)

- !! Paramètres intrinsèques = ne dépendent pas de la prévalence
  - Mais dépendent du seuil si quantitatif (cf courbe ROC: infra)
  - Déterminent (avec RV) la validité intrinsèque du test
- Sensibilité (Se)
  - Se = probabilité d'avoir un test positif si patient malade
  - **Se =  $p(T+/M+) = p(T+ \cap M+) / p(M+) = VP / (VP+FN)$**
- Spécificité (Sp)
  - Sp = probabilité d'avoir un test négatif si patient sain
  - **Sp =  $p(T-/M-) = p(T- \cap M-) / p(M-) = VN / (VN+FP)$**
- Remarque: indice de Youden
  - $Y = Se + Sp - 1$  ; permet un aperçu global

### Valeur prédictive positive (VPP) / négative (VPN)

- !! Paramètre extrinsèques = dépendent de la prévalence
  - Déterminent la validité extrinsèque du test
  - On a **VPP = RVP x prévalence**
- Valeur prédictive positive (VPP)
  - VPP = probabilité d'être malade si le test est positif
  - **VPP =  $p(M+/T+) = p(M+ \cap T+) / p(T+) = VP / (VP+FP)$**
- Valeur prédictive négative
  - VPN = probabilité d'être sain si le test est négatif
  - **VPN =  $p(M-/T-) = p(M- \cap T-) / p(T-) = VN / (VN+FN)$**

### Rapport de vraisemblance positif (RVP) / négatif (RVN)

- RV = LR: Likelihood Ratios
- !! Paramètres intrinsèques: ne dépendent pas de la prévalence
  - Déterminent (avec Se/Sp) la validité intrinsèque du test
- RVP (= RV d'un test positif)
  - RVP = combien de fois plus vraisemblable d'être M+ sachant qu'on a T+
  - **RVP =  $p(T+/M+) / p(T+/M-) = (VP/M+) / (FP/M-) = Se / (1-Sp)$**
  - Exemple: RVP = 5 → le patient a 5x plus de chance d'être malade si T+
- RVN (= RV d'un test négatif)
  - RVN = combien de fois plus vraisemblable d'être M- sachant qu'on a T-
  - **RVN =  $p(T-/M+) / p(T-/M-) = (FN/M+) / (VN/M-) = (1-Se) / Sp$**
  - Exemple: RVN = 0.2 → le patient a 1/0.2 (= 5x) plus de chance d'être sain si T-
- Théorème de Bayes / Nommogramme de Fagan
  - **p (post-test) = p (pré-test) x RV** → on a **VPP = prévalence x RVP**
  - Plus le RVP est > 1, plus il augmente la probabilité post-test d'être M+

- Plus le RVN est  $< 1$ , plus il augmente la probabilité post-test d'être M-
- → Nomogramme de **Fagan** associe  $p(\text{post-test})$  à chaque  $p(\text{pré-test})$  et RV
- !! Remarque
  - Si on veut une VPN ↑ il faut une  $p(\text{pré-test})$  assez basse
    - exemple = test diagnostique de confirmation
    - → D-dimères inutile si la probabilité clinique d'EP élevée
  - Si on veut une VPP ↑ il faut une  $p(\text{pré-test})$  assez élevée
    - exemple = test diagnostique de dépistage
    - → pas de dépistage du cancer de l'ovaire par échographie
- Bonnes valeurs pour les RV
  - Fort gain diagnostique si **RVP > 10** et **RVN < 0.1**
  - Le test est moyen si  $RVP = 5-10$  et  $RVN = 0.1-0.2$
  - Le test est mauvais si  $RVP < 2$  et  $RVN > 0.5$
- Intérêts du rapport de vraisemblance
  - Indépendant de la prévalence (intrinsèque): ne dépend pas du groupe de patient
  - Intérêt individuel en fonction de la probabilité pré-test d'un patient donné
  - Valeur différente selon le niveau quantitatif du test

#### Paramètres faisant varier les indicateurs de performances

- Pour les caractéristiques intrinsèques (Se-Sp)
  - seuil de détection du test
    - si seuil bas:  $Se \uparrow - Sp \downarrow / VPP \downarrow - VPN \uparrow / RVP \downarrow - RVN \uparrow$
    - si seuil élevé: tout est inversé
  - stade de la maladie / quantité de l'inoculum
    - si maladie est avancée ou inoculum important
    - $Se \uparrow - Sp \downarrow / VPP \downarrow - VPN \uparrow / RVP \downarrow - RVN \uparrow$
- Pour les caractéristiques extrinsèques (VPP-VPN)
  - prévalence de la maladie
    - si  $p \uparrow$  alors  $VPP \uparrow$  et  $VPN \downarrow$
    - si  $p \downarrow$  alors  $VPP \downarrow$  et  $VPN \uparrow$
  - représentativité de l'échantillon
    - conséquence directe de ce qui précède

#### - Analyse d'un test diagnostique qualitatif

##### Paramètres opérationnels du test

- Détermination Se-Sp / VPP-VPN / RVP-RVN pour tout test qualitatif
- Pas de calcul de significativité ( $p$ ) mais valeurs de l'**IC95** indispensables +++

##### Fiabilité (ou reproductibilité)

- Principe
  - Pour évaluation tests diagnostiques **qualitatifs** seulement
  - Reproductible: à travers le temps / à travers les expérimentateurs
- Reproductibilité mesurée par coefficient de concordance kappa ( $\kappa$ ):
  - bonne si  $\kappa > 0.6$
  - intermédiaire si  $\kappa = 0.3-0.6$
  - faible si  $\kappa < 0.3$

## - Analyse d'un test diagnostique quantitatif

### Courbe ROC (Receiver Operating Characteristic)

- Principe
  - ROC = Se (ordonnées) en fonction de 1-Sp (en abscisse)
  - Pour évaluation des tests diagnostiques **quantitatifs** seulement
  - si ROC médiane ayant 0 pour origine: test sans intérêt
- Identification de la valeur seuil optimale
  - = point **d'inflexion** de la courbe: le plus en haut / à gauche
  - Point d'intersection entre la courbe et la diagonale [Se = 1 / 1-Sp = 1]
  - Valeur seuil pour laquelle Se et Sp sont optimales
  - Remarque: seuil optimal selon le type de test
    - **Dépistage**: on veut Se ~ 100% avec la meilleure Sp possible
    - **Confirmation**: on veut Sp ~ 100% avec la meilleure Se possible
- Evaluation de la capacité discriminante globale du test
  - = calcul de l'aire sous la courbe (+/- comparaison statistique avec un autre test)
  - Test d'autant plus discriminant que l'aire sous la courbe est proche de 1 (= 100%)
  - En pratique la valeur de l'aire sous la courbe doit être  $\geq 70-75\%$  pour être valable
- !! Validation ultérieure du seuil
  - La validation de l'utilité du test avec la valeur seuil optimale déterminée en ROC devra se faire au cours d'un **autre** essai avec ce seuil uniquement +++

### Corrélation de Pearson

- Principe
  - Pour chaque patient, mesure avec chacun des deux tests en abscisse / ordonnée
  - On trace (en fait régression linéaire) la droite passant en moyenne entre les points
- → calcul du coefficient de corrélation (r)
  - Si R = 1: corrélation parfaite dans le sens positif de l'association
  - Si R = -1: corrélation parfaite dans le sens contraire
  - En pratique: on met au carré (R<sup>2</sup>: élimine les signes)
- → calcul de la significativité de la corrélation
  - Obtention d'un p : significatif si  $p < 0.05$
  - En pratique: R<sup>2</sup> > 0.6 (bon) ou > 0.8 (excellent)

### Diagramme de Bland-Altman

- Principe
  - différence entre les 2 mesures pour chaque patient (ordonnée)
  - en fonction de la valeur moyenne de la variable mesurée (abscisse)
  - On rajoute: bornes de l'IC95 / le biais / (0 = concordance parfaite)
- Intérêt
  - Donne la **dispersion** des variables pour l'ensemble du spectre des moyennes
  - Donne le **biais**: différence moyenne entre les mesures par les 2 tests
  - Permet de repérer si le test est plus performant pour une valeur particulière

## - Validité d'un essai diagnostique

Critères indispensables à la validité interne de l'essai (10) +++ (STARD checklist)

- Inclusion **prospective** / étude multicentrique / calcul du NSN
- Sélection et **représentativité** de l'échantillon (cf paramètres extrinsèques)
- Examen de référence (Gold standard): le test doit être **validé** / Se-Sp fournies
- Définition de la maladie par le gold standard / description **précise** du test
- Tous les sujets on eu **à la fois** le test à l'étude **et** le test de référence (**PMZ**)
- Réalisation et interprétation des 2 tests **en aveugle** / comité d'adjudication
- **Reproductibilité** du test / étude de la variabilité inter et intra-observateurs
- Ensemble des paramètres opérationnels fournis (3×2) avec **IC95** pour chacun (**PMZ**)
- Choix du seuil argumenté = courbe ROC fournie
- Validité externe: applicabilité / risque / coût-bénéfice

### Biais fréquents dans les études diagnostiques

- Biais de sélection
  - Comme pour tout essai: étude monocentrique / patients volontaires / restriction..
  - Si la sélection des patients sur le résultat du test à l'essai et non sur la référence +++
- Biais de classement
  - Biais d'interprétation: peu d'évaluateurs / biais d'apprentissage (machine neuve..)
  - Biais d'information: pas d'aveugle: on connait le résultat de l'un pour interpréter l'autre
  - **Biais de vérification**: le gold standard n'est pas le même pour tous les patients
- Biais de confusion
  - Si le gold standard est déterminé par le résultat du test à l'essai +++
  - ex: histologie (GS) que si suspicion de tumeur à la TEP (T) et surveillance sinon
- **!! Remarque: « lead-time bias » dans les études de dépistage**
  - Biais lié au fait que la maladie est dépistée plus tôt sans pour autant améliorer la survie
  - Exemple
    - SLA dépistée par génétique à 10ans → 50ans de survie « apparente »
    - SLA diagnostiquée sur symptômes à 40ans → seulement 10ans de survie
    - Pour autant le dépistage de la SLA n'améliore pas la survie du patient...

### Pour tout test diagnostique, l'intérêt dépendra de:

- Test reproductif / simple / fiable / acceptable
- Impact diagnostique: paramètres de performance
- Impact thérapeutique: modifie la prise en charge
- Impact de santé publique: coût / résistances, etc.

### - Classification des essais diagnostiques

#### Essai de phase I

- On fait subir le test à l'essai à une population saine et une population malade
- On recherche une différence significative entre les deux populations
- Ex: faire un BNP à des patients sains et des patients en décompensation cardiaque

#### Essai de phase II

- On fait subir le test à l'essai à une population indifférenciée de sains et malades
- On regarde si le test permet de bien classer les patients en sains / malades
- Ex: comparer les résultats d'un BNP avec ceux de l'ETT / étude **rétrospective**

#### Essai de phase III (+++)

- On fait subir le test à l'essai à une population de patients suspects d'être malades

- On regarde si le test permet de différencier les patients sains et malades
- Ex: voir si le BNP est prédictif du résultat de l'ETT / étude **prospective**

#### Essai de phase IV

- On pose une hypothèse avec un critère de jugement principal / on randomise
- On compare les résultats sur les critères d'évaluation avec et sans le test à l'essai
- Ex: voir si le dosage du BNP aux urgences améliore la survie à 6 mois

#### - Résumé

##### Objectif de l'étude

- Evaluer la **performance diagnostique** du test à l'étude (« TDR »)
- **!! NPO de préciser**
  - **par rapport** à quel test de référence (« culture »)
  - **dans** quelle maladie (« angines à SGA »)
  - dans quel contexte (aux urgences)

##### Critère principal d'évaluation

- = performance diagnostique du test
- en pratique, ce sont les **paramètres opérationnels** (Se-Sp / VPP-VPN / RVP-RVN)
- **!! Dans le cadre d'un test quantitatif**
  - critère principal = les paramètres opérationnels pour un seuil prédéfini
  - critères secondaires: le seuil optimal (ROC) / le Pearson / le Bland-Altman

##### Résultats dans un essai diagnostique quantitatif: donner

- Paramètres opérationnels: Se-Sp ; VPP-VPN ; RVP-RVN pour **le** seuil de définition
- Courbe ROC: valeur seuil optimale / aire sous la courbe +++
- Coefficient de corrélation de Pearson (r et p)
- Bland-Altman: biais systématique