

## Cours n°2 : Evaluation des performances diagnostiques/Evaluation d'un programme de dépistage

*La prof commence son cours en précisant que cette année, ce qui est important, ce n'est pas de se concentrer sur les aspects purement statistiques, mais d'insister sur la méthodologie, autrement dit comment on construit une étude. Les points importants de ce cours sont : les propriétés intrinsèques d'un test ( $Se$ ,  $Sp$ ) ainsi que le rapport de vraisemblance ( $RV$ ), les propriétés extrinsèques (valeurs prédictives) et le biais de vérification.*

*Tout le long du cours, elle a développé un exemple de cas clinique qui n'est pas à apprendre, je l'indiquerai parfois juste pour la compréhension du cours mais ce n'est pas à retenir.*

*Elle relira la ronéo si elle a le temps.*

**Plan du cours :**

**I. Généralités**

1. *Tests diagnostiques*
2. *Test de référence*
3. *Biais de vérification*

**II. Caractéristiques intrinsèques d'un test diagnostique**

1. *Sensibilité*
2. *Spécificité*
3. *Courbes ROC*
4. *Rapport de vraisemblance*

**III. Caractéristiques extrinsèques d'un test diagnostique**

1. *Valeur prédictive positive*
2. *Valeur prédictive négative*

**IV. Validité interne / Validité externe**

1. *Validité interne*
2. *Validité externe*

## I. Généralités

**Objectif** : L'objectif de l'évaluation diagnostique c'est d'évaluer des performances diagnostiques d'un test diagnostique ou d'un test de dépistage par rapport à un **test de référence**, qu'on appelle « **gold standard** ».

C'est important de parler **des** performances, et pas d'une performance car il y a plusieurs estimateurs. De plus, il faut bien différencier un test diagnostique d'un test de dépistage, bien que la méthode reste la même en cherchant à déterminer plusieurs indices, on ne va pas privilégier le même indice que l'on soit dans un contexte de diagnostic ou de dépistage.

Déterminer la **performance d'une stratégie diagnostique** c'est mesurer la **capacité du test** à bien classer les **sujets malades** et les **sujets non malades**.

### 1. Tests diagnostiques

Un test peut être un test biologique (ex : CRP pour les infections), un signe clinique, un examen de laboratoire, un examen radiographique ou encore un acte invasif (ex : biopsie pour les cancers).

**Résultat d'un test** : Le résultat d'un test peut être **continu** (quantitatif) comme la glycémie par exemple, ou d'ordre **qualitatif**, c'est à dire binaire (signe présent : oui/non) par exemple l'abolition d'un réflexe.

#### a) Variable qualitative

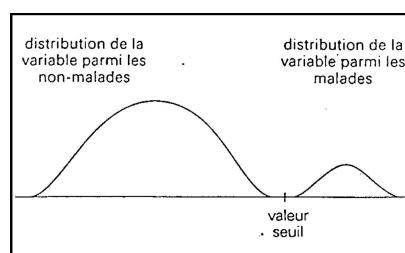
**Test idéal** : Un test idéal serait **positif si et seulement si le sujet est malade**. Il permettrait ainsi de classer sans erreur les résultats en malade ou en non malade.

Un **signe pathognomonique** est présent chez les malades et **jamais chez les non malades**, mais tous les malades ne l'ont pas forcément, il permet ainsi de classer sans erreur les sujets non-malades (ex : signe de Köplick dans la rougeole).

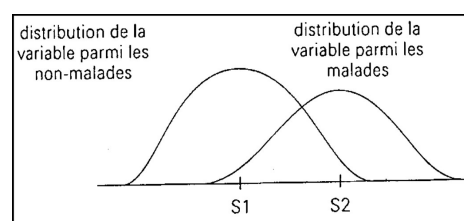
Mais dans la vraie vie, ce n'est pas aussi simple, certains malades ont le signe et d'autres non, certains non-malades ne l'ont pas et d'autres l'ont. Ainsi, on ne peut pas classer sans erreur les sujets, et il y aura un certain nombre de faux positifs et de faux négatifs.

#### b) Variable quantitative

Situation idéale : Dans une situation idéale, les distributions du résultat du test sont bien **distinctes**, et cela permet ainsi de classer sans erreur les sujets en malade ou non malade.



Mais dans la vraie vie, les valeurs résultat du test s'échelonnent et se superposent entre les malades et les non-malades, il y aura ainsi des faux positifs et des faux négatifs.



## 2. Test de référence

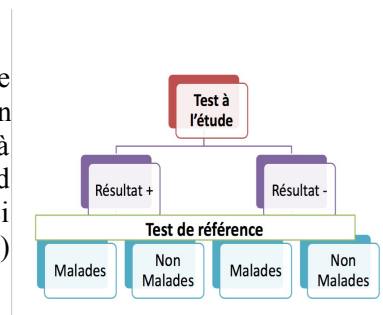
Le test de référence, aussi appelé gold standard, est un test **déjà existant, validé** et qui **permet d'affirmer que la maladie est présente ou absente** (ex : dans un cancer le test de référence est la biopsie). Idéalement le test de référence présente un critère indiscutable. Mais souvent c'est plus problématique : parfois le test de référence n'existe pas, il est parfois subjectif (ex : une donnée d'interrogatoire), il peut être cher ou difficile à obtenir (ex : on ne peut pas faire subir des biopsies à toute une population).

Attention à ne pas dire qu'on compare les performances d'un test avec les performances de u gold standard, **on se sert du gold standard pour déterminer les performances d'un test.**

Si le test de référence n'existe pas, on va classer les sujets avec un **faisceau d'arguments** dont la validité doit avoir été étudiée.

Finalement, le gold standard est en fait le **meilleur test disponible**.

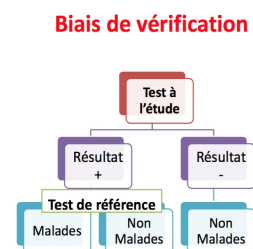
Pour l'étude d'un test diagnostique, le test de référence et le test à l'étude doivent **tous deux** être réalisés chez **tous les sujets**. Ensuite, l'interprétation des résultats de l'un se fait en insu des résultats de l'autre. Ainsi, le test à l'étude va donner des résultats positifs et des résultats négatifs, le Gold Standard va ensuite classer ces résultats en malades et en non malades. Ceci va nous permettre d'établir un tableau de contingence (qu'on verra plus loin) et de calculer nos estimateurs.



Si le test à l'étude influence la réalisation du test de référence, il y a un **biais de vérification**.

## 3. Biais de vérification

Le **biais de vérification** c'est quand le test de référence n'est **pas disponible pour tout le monde** (invasif, trop cher...). La prof a précisé qu'un biais de vérification pouvait avoir un impact important en augmentant la sensibilité et en diminuant la spécificité (pas à retenir). Pour nous, ce qui est important c'est de reconnaître qu'il y a un biais de vérification et qu'il peut avoir un impact sur la surestimation d'un indice ou la sous-estimation d'un autre.



*Exemple : La prof a donné l'exemple d'une étude qu'elle a menée, et qu'elle a ensuite développé dans tous le cours. L'étude portait sur le dosage de la CRP après une chirurgie pariétale, pour évaluer si le patient allait développer des fistules gastriques. Ainsi si le taux de la CRP était trop élevé on garderait le malade 6-7jours comme dans la procédure habituelle, mais si on estimait qu'il était assez bas indiquant qu'il ne développerait pas de complications, on ferait directement sortir le patient.*

*Le gold standard pour détecter les fistules gastriques est le scanner. Ici, le gold standard (donc le scanner) n'était disponible que chez les patients qui présentaient des signes inquiétants, en outre une CRP élevée. Donc il y a un risque de biais de vérification.*

## II. Caractéristiques intrinsèques d'un test diagnostique

Les **caractéristiques intrinsèques** du test sont des estimateurs qui **ne dépendent pas de la prévalence de la maladie dans la population**.

### 1. Sensibilité

La **sensibilité** d'un test c'est la **capacité du test à détecter les cas d'une maladie**.

$Se = P(T+/M+)$   $Se = VP/(VP+FN)$  C'est la probabilité que le test soit positif si le patient est malade.

Si le test a une **sensibilité de 100%**, cela signifie que tous les malades sont correctement identifiés et donc qu'il n'y a **aucun faux négatif (FN)**.

C'est d'autant plus important d'avoir un **test sensible** dans le cas d'un **dépistage**.

L'utilité des test sensibles est surtout pour les **maladies très rares** qui sont **curables**, et quand le **diagnostic précoce d'une maladie grave en améliore le pronostic**, ainsi il y a une pénalité à manquer la maladie (ex: tuberculose).

Un test sensible est utile au **stade précoce d'une action diagnostique pour éliminer des hypothèses**, le diagnostic sera ensuite confirmé par d'autres tests plus spécifiques.

Ainsi s'il est très sensible il est **utile quand négatif car il permet d'exclure la maladie**.

*Exemple : le taux de complications septiques après les chirurgies pariétales est très rare (2-3%) donc on a besoin d'un test très sensible. Ainsi avec la CRP, si le taux était élevé il y avait un risque de complications et il fallait faire des tests plus spécifiques. Par contre si le taux était faible, on était sûr que le patient n'était pas à risque de développer des complications.*

### 2. Spécificité

La **spécificité** d'un test c'est la **capacité du test à détecter les sujets qui ne sont pas malades**.

$Sp = P(T-/M-)$   $Sp = VN/(VN+FP)$  C'est la probabilité que le test soit négatif si le patient n'est pas malade.

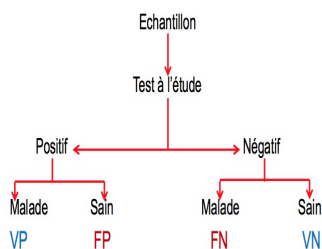
Si le test a une **spécificité de 100%**, cela signifie que tous les non-malades sont correctement identifiés et donc qu'il n'y a **aucun faux positif (FP)**.

On veut éviter les cas de faux positifs quand il y a un impact thérapeutique (ex : faire subir une chimiothérapie à un patient qui n'a en fait pas de cancer).

Un test spécifique est utile à un **stade plus avancé d'une action diagnostique pour confirmer des hypothèses ou un diagnostic suggéré par d'autres tests plus sensibles**, particulièrement dans le cas où les résultats faussement positifs peuvent être préjudiciables au patient.

Ainsi s'il est très spécifique il est **utile quand positif car il marque la certitude d'être malade**.

*Exemple : on peut avoir une CRP élevée due à une infection urinaire par exemple, un test plus spécifique permettait de démontrer si les complications étaient chirurgicales ou non.*



	Maladie présente	Maladie absente
Test +	VP	FP
Test -	FN	VN

Tableau de contingence

$$Se = VP / (VP+FN) \quad Sp = VN/(VN+FP)$$

$$Se = VP/(VP+FN) \quad Sp = VN/(VN+FP)$$

### 3. Courbes ROC

Dans l'idéal, un test est à la fois très sensible et à la fois très spécifique. On considère qu'entre 0,75 et 0,80 on peut dire qu'on a une sensibilité ou une spécificité intéressante. Plus un test est subjectif (par exemple l'anatomopathologiste qui étudie des cellules cancéreuses), moins le test est performant.

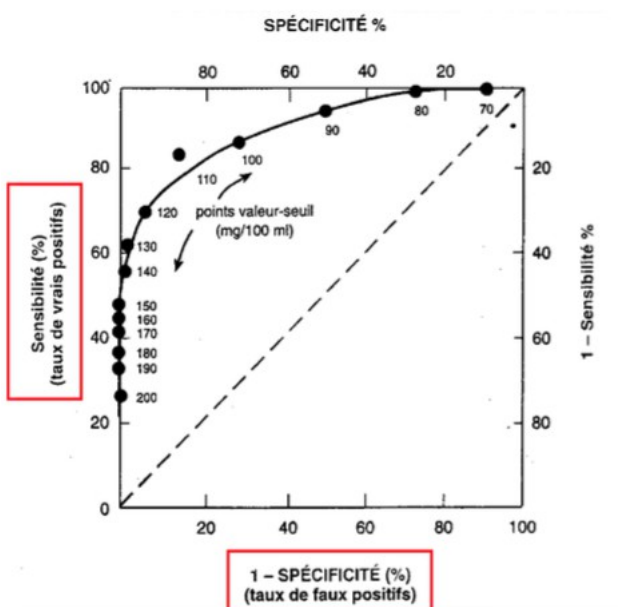
En réalité, on fait un **compromis** entre la sensibilité et la spécificité, qu'on évalue **en fonction des situations**, pour diagnostiquer ou pour dépister.

Pour les tests quantitatifs, il faut définir une **valeur seuil** qui délimite si le test est positif ou négatif. Il y a différentes façons de déterminer les seuils, la façon la plus courante est de déterminer le seuil qui maximise à la fois la spécificité et la sensibilité, mais on peut aussi choisir de maximiser la sensibilité ou la spécificité en fonction des situations.

En fonction du seuil choisi, **quand la sensibilité augmente la spécificité diminue et inversement**.

La courbe ROC est la courbe qui va résumer les performances du test. Elle a pour objectif de déterminer la **valeur seuil optimale** de la sensibilité et de la spécificité du test. Plus l'aire sous la courbe est proche de 1, plus le test est intéressant. Un test qui a une aire sous la courbe de 0,5 (correspond à la diagonale) ne fait pas mieux que le hasard.

Elle permet ainsi de visualiser les performances du test en fonction des seuils choisis ainsi que d'étudier les variations de la sensibilité et de la spécificité pour différentes valeurs seuils.



On a en abscisse le taux de faux positifs et en ordonnée le taux de faux négatifs.

Plus le seuil se rapproche de l'angle du carré (en haut à gauche, donc du 100%), plus le test est performant.

### 4. Rapport de vraisemblance

La prof a traité cette partie à part, dans son diapo elle figurait après les caractéristiques extrinsèques du test. Cependant, attention, c'est bien une caractéristique **intrinsèque**, car le rapport de vraisemblance ne dépend pas de la prévalence de la maladie.

Le rapport de vraisemblance ou « likelihood ratios » est un **autre mode d'expression, un résumé, des caractéristiques intrinsèques d'un test (Se, Sp)**.

Il estime le **rapport entre la probabilité d'avoir un test positif (ou négatif) chez les sujets malades et celle d'avoir un test positif (ou négatif) chez les sujets sains**.

Le rapport de vraisemblance est calculé **à partir de la sensibilité et de la spécificité**, il est donc **indépendant de la prévalence de la maladie dans la population**.

Il y a un estimateur pour un test positif, et un estimateur pour un test négatif.

- **Rapport de vraisemblance positif (RV+)**

C'est le **rapport** entre la probabilité de présenter un **test positif quand la personne est malade** et la probabilité de présenter un **test positif quand la personne n'est pas malade**.

$$RV+ = P(T+/M+) / P(T+/M-) \quad RV+ = Se / 1-Sp$$

Il quantifie le gain diagnostique d'un test positif, **un individu malade ayant RV+ fois plus de chance d'avoir un test positif qu'un individu sain**.

*Ex : dans l'étude avec la CRP, le rapport de vraisemblance positif était de 13, ainsi un individu ayant une complication septique avait 13 fois plus de chance d'avoir un test positif qu'un sujet sain.*

- **Rapport de vraisemblance négatif (RV-)**

C'est le **rapport** entre la probabilité de présenter un **test négatif quand la personne est malade** et la probabilité de présenter un **test négatif quand la personne n'est pas malade**.

$$RV- = P(T-/M+) / P(T-/M-) \quad RV- = 1-Se / Sp$$

Il quantifie le gain diagnostique d'un test négatif, **un individu malade ayant RV- fois plus de chance d'avoir un test négatif qu'un individu sain**.

*Ex : si on détermine le seuil de la CRP à 40 pour déterminer que le test est positif, le rapport de vraisemblance négatif représente à quel point on peut être sûr qu'un patient ayant une CRP en dessous de 40 ne développera pas de complications.*

*De plus, si le rapport de vraisemblance négatif est de 0,25, alors il y aura 4 fois plus de chance (1/0,25) de présenter un test négatif lorsque la personne n'est pas malade que lorsque la personne est malade.*

**Plus le rapport de vraisemblance positif est élevé et plus le rapport de vraisemblance négatif est faible, plus le gain diagnostique du test est important.**

RV+	RV-	Apport diagnostique
>10	<0,1	Très fort
5-10	0,1-0,2	Fort
2-5	0,2-0,5	Modéré
1-2	0,5-1	Faible
1	1	Nul

Le rapport de vraisemblance permet directement de voir si un test est utile en clinique ou pas. On veut le RV+ le plus élevé possible et le RV- le plus faible possible.

Les rapports de vraisemblance permettent de calculer la **probabilité post-test de la maladie** par application du théorème de Bayes, donc de répondre à la question du clinicien : **quelle est la probabilité que le patient soit malade si le test est positif ou sain si le test est négatif ?**

*Exemple : ainsi le test avec la CRP permettait d'avoir une anticipation sur les complications, et ainsi de les prévenir avant que les signes cliniques n'apparaissent, et de ne pas laisser des patients à risque sortir de l'hôpital.*

Les rapports de vraisemblance étant **indépendants de la prévalence de la maladie** dans la population, ce calcul est **transposable d'une population à une autre**.

Le rapport de vraisemblance peut être considéré comme un « **odd-ratio** » (il ne faut pas l'écrire formellement mais on peut le comprendre comme ça), c'est-à-dire le **risque d'avoir la maladie si on est exposé à tel facteur de risque**.

Il est important de calculer la probabilité post-test, à partir de la probabilité pré-test qui correspond à la prévalence et du rapport de vraisemblance. Cette probabilité post-test ne nous met plus à l'échelle théorique d'une population, mais on est face à un patient donné, qui a eu un résultat par rapport au test.

$$\text{Odd post-test} = \text{odd pré-test} \times \text{rapport de vraisemblance}$$

$\text{Odd} = P / 1-P$

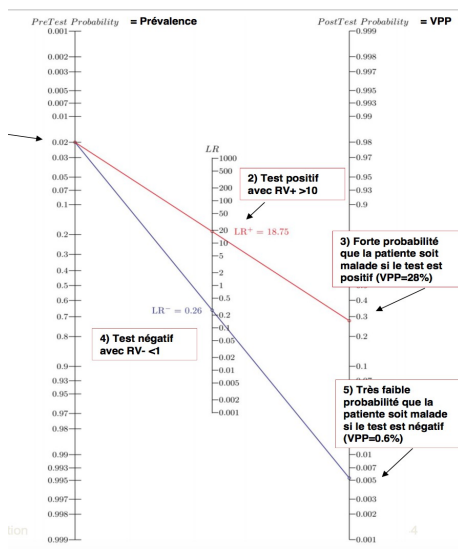
Exemple : un patient lambda a 2-3% de risques d'avoir des complications, tandis que si l'on se met face à un patient qui a eu un test positif à la CRP (valeur au dessus du seuil choisi), son risque passe de 2% (probabilité pré-test) à  $2 \times 13 = 26\%$  (probabilité post-test). Ce risque ne peut pas être ignoré et il faut donc avoir recours à des examens complémentaires (scanner...).

**Attention : la formule indique bien « odd » et non pas « probabilité », les deux sont à différencier.**

L'odd est une façon d'exprimer une probabilité, comme aux champs de course. L'odd exprime au numérateur la probabilité qu'un événement se produise, et au dénominateur la probabilité qu'il ne se produise pas (P/1-P).

Exemple : Si l'odd de souffrir d'une maladie est de 4, cela signifie que le risque de présenter la maladie est 4 fois plus élevé que de ne pas la présenter donc pour calculer la probabilité on retourne la formule et **probabilité = odd/(1+odd)=4/5=80%**.

En pratique, on utilise le **nomogramme de Fagan**. Il permet de déterminer la probabilité post-test en traçant une droite qui passe par la probabilité pré-test (donc la prévalence) et le rapport de vraisemblance, positif si le test est positif, et négatif si le test est négatif.



La prévalence de la maladie est de 2%. On cherche à déterminer la probabilité post-test que la patiente soit malade quand son test est positif : ici son test est positif et le RV+ est de 18,75. On trace la droite formée par les deux points et on obtient la probabilité post-test : la patiente a 28% de chance d'être malade si le test est positif.

On cherche ensuite à déterminer la probabilité post-test qu'une patiente soit malade si son test est négatif avec un RV- de 0,26. La droite formée par les deux points nous indique que la probabilité post-test est de 0,6% : la patiente a une probabilité de 0,6% (très faible) d'être malade si le test est négatif.

Dans la vraie vie on ne fait pas un seul test : on va faire des tests en séries (on attend d'avoir le résultat d'un test avant d'en faire un autre), ou des tests en parallèle.

Ainsi **l'odd post-test va dépendre des rapports de vraisemblance de chaque test** :

$$\text{Odd post-test} = \text{Odd pré-test} \times \text{RV test 1} \times \text{RV test 2} \times \text{RV test 3} \dots$$

**Avantages des rapports de vraisemblance :**

- 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.

Cependant, ils présentent comme inconvénient l'utilisation peut aisée des odds.



### III. Caractéristiques extrinsèques d'un test diagnostique

Les **caractéristiques extrinsèques** d'un test **dépendent de la prévalence de la maladie dans la population**, qui est un paramètre autre, non propre au test. On s'intéresse à la probabilité que le sujet soit malade en fonction des résultats du test, ainsi on ne s'intéresse plus seulement aux performances du test mais au sujet.

#### 1. Valeur prédictive positive

La valeur prédictive positive est la **probabilité d'être malade lorsque le résultat est positif**.

**VPP=P(M+T+)** Une stratégie diagnostique ayant une **VPP de 100%** signifie que tous les résultats positifs correspondent à des malades (**aucun FP**).

*Exemple : la prévalence des complications était de 2-3% et ils ont déterminé un seuil au dessus duquel il le patient avait 30% de risques d'avoir des complications. Ainsi, quand le test était positif, le patient avait dix fois plus de risque d'apparition de complications septiques.*

#### 2. Valeur prédictive négative

La valeur prédictive négative est la **probabilité d'être indemne de la maladie lorsque le résultat est négatif**.

**VPN= P(M-/T-)** Une stratégie diagnostique ayant une **VPN de 100%** signifie que tous les résultats négatifs correspondent à des non-malades (**aucun FN**).

*Exemple : ils n'ont pas réussi à identifier un seuil en dessous duquel ils pouvaient être sûrs que le patient n'allait absolument pas développer de complications.*

Deux exemples qui ne sont pas à retenir :

On étudie les valeurs prédictives de l'ECG d'effort en fonction de la prévalence des sténoses coronaires. Le test à l'étude est ainsi l'ECG d'effort, et le test de référence est la coronographie.

	Sténose +	Sténose -	Total
ECG +	55	7	62
ECG-	49	84	133
Total	104	91	195

Ici, la prévalence correspond au nombre de patients présentant une sténose par rapport au nombre total de patient, soit  $104/195 = 53\%$ . La prévalence est élevée.  
 $Se=55/104=53\%$                        $VPP=55/62=89\%$   
 $Sp=84/91=92\%$                        $VPN =84/133=63\%$

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

Si on diminue la prévalence, ici  $104/624=17\%$ .  
 $Se=55/104=53\%$                        $VPP=55/97=57\%$   
 $Sp=478/520=92\%$                        $VPN=478/527=91\%$

On voit que la sensibilité et la spécificité n'ont pas changé car ce sont des caractéristiques intrinsèques, tandis que la prévalence influence la VPP et la VPN.

Ainsi, les propriétés extrinsèques du test (VPP et VPN) dépendent de la prévalence de la maladie dans la population dans laquelle la stratégie diagnostique est destinée, ainsi que de la sensibilité et de la spécificité du test.

#### **IV. Validité interne / Validité externe**

Il n'y a pas de lien entre la validité interne et la validité externe.

##### ***1. Validité interne***

La validité interne prend en compte tout ce qui permet d'**éviter les biais**, c'est-à-dire **tout ce qui pourrait compromettre l'étude**.

Elle correspond à la **validité méthodologique** de l'étude, il faut donc :

- Une **comparaison en aveugle, indépendante avec un test de référence**.
- Savoir évaluer le **biais de vérification**. Tous les patients doivent avoir les deux tests.
- Une **description de la méthodologie du test**.

##### ***2. Validité externe***

La validité externe correspond à la **généralisabilité** du test, c'est à dire à quel point les résultats du test peuvent être **appliqués à d'autres populations**. Elle correspond donc à la **reproductibilité du test**, et interroge donc la **qualification des évaluateurs**. En effet, certains tests diagnostiques nécessitent une interprétation (imagerie, examen clinique...) : la part de subjectivité rend donc le test moins reproductible.

Le test a un faible intérêt s'il est non reproductible :

- **entre différents patients** : il faut 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).

Elle prend donc en compte la population cible c'est-à-dire sur quels patients le test a été validé.

*Dédicaces :*

- *A Julie, ma ronéoficheuse, avec qui j'ai tout traversé depuis le lycée !*
- *A Pierre Huz, celui qui chopait plus vite que son ombre, mais qui a enfin réussi à se caser*
- *Au Med'Sing, parce que vous êtes tous beaucoup trop géniaux, et que j'ai adoré cette année avec vous*
- *Mention spéciale à Sacha et Anna, parce que je vous aime bien*
- *A Nesrine, meilleure VP, c'est sûrement toi ma bonne étoile au final !*
- *A mes co-stagiaires de l'ambiance : Alex, Hermine et Antoine, je changerais de stage pour rien au monde !*
- *A mes petites Alex et Vivi, courage les filles, l'année prochaine c'est la P2*
- *Au BDA, meilleure asso, à la Barge du Crous et au DGT*
- *A Medistarf*
- *Mention spéciale à Dbh, parce que toi aussi je t'aime bien*
- *Et le meilleur pour la fin : à mon Caporal, merci de me supporter 💙*