

# SDA Biostatistiques

VALEUR INFORMATIONNELLE D'UN SIGNE: SENSIBILITE,  
SPECIFICITE ET VALEURS PREDICTIVES



# Rappel : Les probabilités conditionnelles

- ▶  $P(A/B) = P(A \cap B) / P(B)$
- ▶ Formule des probabilités conditionnelles

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{et} \quad P(B/A) = \frac{P(A \cap B)}{P(A)}$$

- ▶ D'où le théorème de la multiplication :

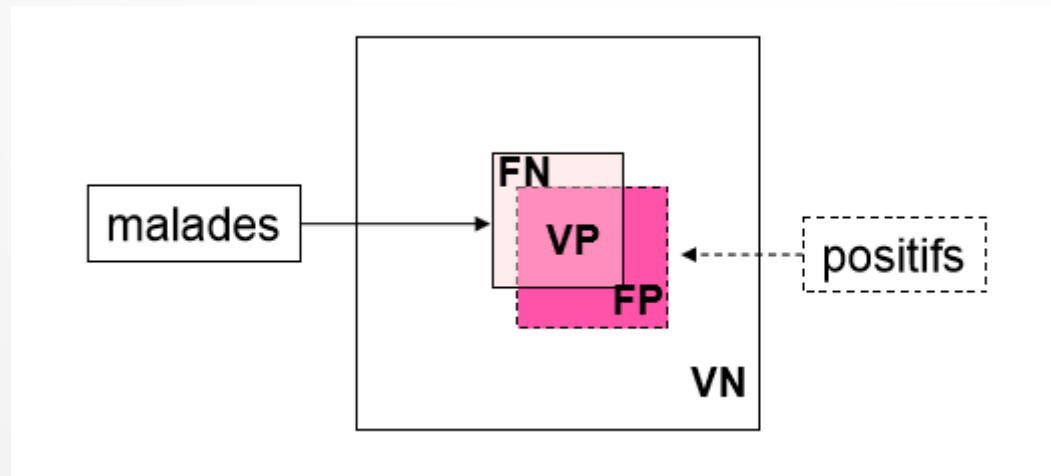
$$P(A \cap B) = P(A/B) \times P(B) = P(B/A) \times P(A)$$

- ▶ Formule de Bayes :

$$P(B/A) = \frac{P(A/B) \times P(B)}{P(A)}$$

# I. Introduction

- ▶ Importance de la démarche diagnostique :
  - Accueil du patient et de sa plainte
  - Probabilité à priori de plusieurs maladies
  - Recueil d'informations utiles (observation du patient, interrogation,...)
  - Probabilité à postériori
  - Diagnostique : probabilité post-test



# Les sources d'informations utiles

- ▶ Le « Gold Standard » : permet de déterminer avec certitude par un moyen incontestable le statut du patient : « malade » ou « non malade »
  - Moyen incontestable
  - Test de référence

Mais le gold standard n'est pas toujours applicable (examens pénibles, coûteux, ex: biopsie, parfois uniquement utilisable après la mort,...).

- ▶ Les tests diagnostiques :
  - Test « imparfait », substitue le GS
  - Laisse la place à un degré d'incertitude

Il est donc nécessaire de les évaluer

**Test diagnostique** : moyen d'obtenir une information utile dans la démarche diagnostique face au patient

# Le test diagnostique

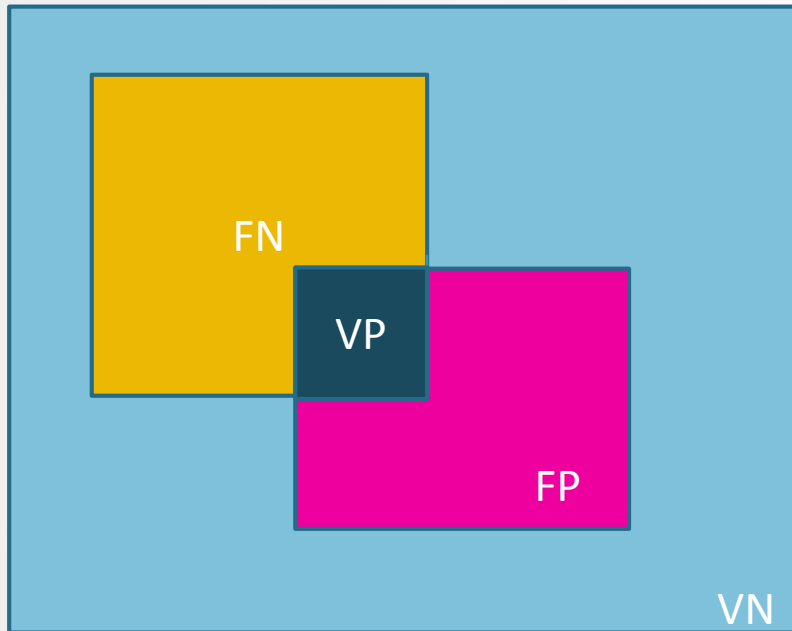
- ▶ L'objectif du test diagnostique est d'identifier les malades
- ▶ Mais : il pourra être positif (T+) alors que le sujet n'est pas malade (NM)
- ▶ Il pourra être négatif (T-) alors que le sujet est réellement malade (M)
- ▶ Alors qu'avec le Gold Standard il n'y aura aucune erreur

# Pourquoi évaluer un test diagnostique ?

- ▶ En pratique on utilise donc le test diagnostique qui présente un risque d'erreur : Il faut donc qu'au préalable, les performances du test soient évaluées pour savoir si les résultats seront fiables
- ▶ On recherche **la valeur informationnelle** du test diagnostique
- ▶ Calcul des paramètres : Sensibilité, Spécificité, Valeurs prédictives

*Valeur informationnelle : capacité d'un test à nous donner les bonnes informations*

# I. Paramètres mesurant la valeur informationnelle d'un test diagnostique



→ **Vrai positifs (VP)** : Patients à la fois Malades (M) ET Positifs au test (T+) le diagnostic est posé avec succès par le test

→ **Vrai Négatifs (VN)** : Patients à la fois Non Malades (NM) ET Négatifs au test (T-) le diagnostic est écarté avec succès par le test

→ **Faux Positif (FP)** : Patients à la fois Non Malades (NM) ET Positifs au test (T+) : déclarés malades à tort = **Fausse alerte !**

→ **Faux Négatifs (FN)** : Patients à la fois Malades (M) ET Négatifs au test (T-) : déclarés sains à tort = **Diagnostic manqué !**

		Information sur la maladie (résultat du gold standard)	
		M	NM
Résultat du test	T+	VP	FP
	T-	FN	VN

# I. Paramètres mesurant la valeur informationnelle du test diagnostique

## A) Sensibilité et Spécificité

- ▶ La Sensibilité se définit dans le groupe des malades (M)
- ▶ La Spécificité se définit dans le groupe des non malades (NM)
- ▶ Se et Sp permettent d'évaluer la valeur diagnostique d'un test avant son utilisation
- ▶ Se et Sp sont propres au test
- ▶ Indépendantes de la prévalence : **intrinsèques** du test diagnostique

# La Sensibilité

- ▶ Probabilité qu'un sujet soit positif au test, sachant qu'il est réellement malade (PT+/M)
- ▶ Capacité d'un test à détecter les cas d'une maladie (à identifier correctement comme malades ceux qui ont la maladie)

- ▶  $Se = P(T+/M) = \frac{P(T+\cap M)}{P(M)}$

- ▶ Estimé par :  $Se = \frac{VP}{VP+FN}$

		Information sur la maladie (résultat du gold standard)	
		M	NM
Résultat du test	T+	VP	FP
	T-	FN	VN

# La Sensibilité

- ▶ Plus la sensibilité est élevée ( $\approx 1$ ) plus le risque de manquer le bon diagnostic ( $P(T- / M)$ ) est faible ( $\approx 0$ ) car  $P(T- / M) = 1 - P(T+ / M)$
- ▶ Si  $Se = 100\%$  alors il n'y aura donc aucun Faux Négatifs (FN)
- ▶ On privilégie la sensibilité lorsque les erreurs par défaut (FN) sont plus graves que les erreurs par excès (FP)

*Exemple: Mieux vaut détruire une poche de sang non contaminée par le VIH (FP) qu'en transfuser une contaminée (FN) !*

# La Spécificité

- ▶ Probabilité qu'un sujet soit négatif au test, sachant qu'il est réellement non malade  $P(T-/NM)$
- ▶ Capacité d'un test à ne pas alerter faussement les non malades Plus la spécificité est élevée, plus le déclenchement de l'alerte est spécifique de la maladie

- ▶  $Sp = P(T-/NM) = \frac{P(T- \cap NM)}{P(NM)}$

- ▶  $Sp = \frac{VN}{VN+FP}$

		Information sur la maladie (résultat du gold standard)	
		M	NM
Résultat du test	T+	VP	FP
	T-	FN	VN

# La Spécificité

- ▶ Plus la spécificité est élevée ( $\approx 1$ ) plus la probabilité de faire une fausse alerte ( $P(T+/NM)$ ) est faible ( $\approx 0$ ) car  $P(T+/NM) = 1 - P(T- / NM)$
- ▶ Si  $Sp = 100\%$  : Aucun Faux Positifs (FP)
- ▶ On privilégie la Spécificité lorsque les erreurs par excès (FP) sont plus graves que les erreurs par défaut (FN)

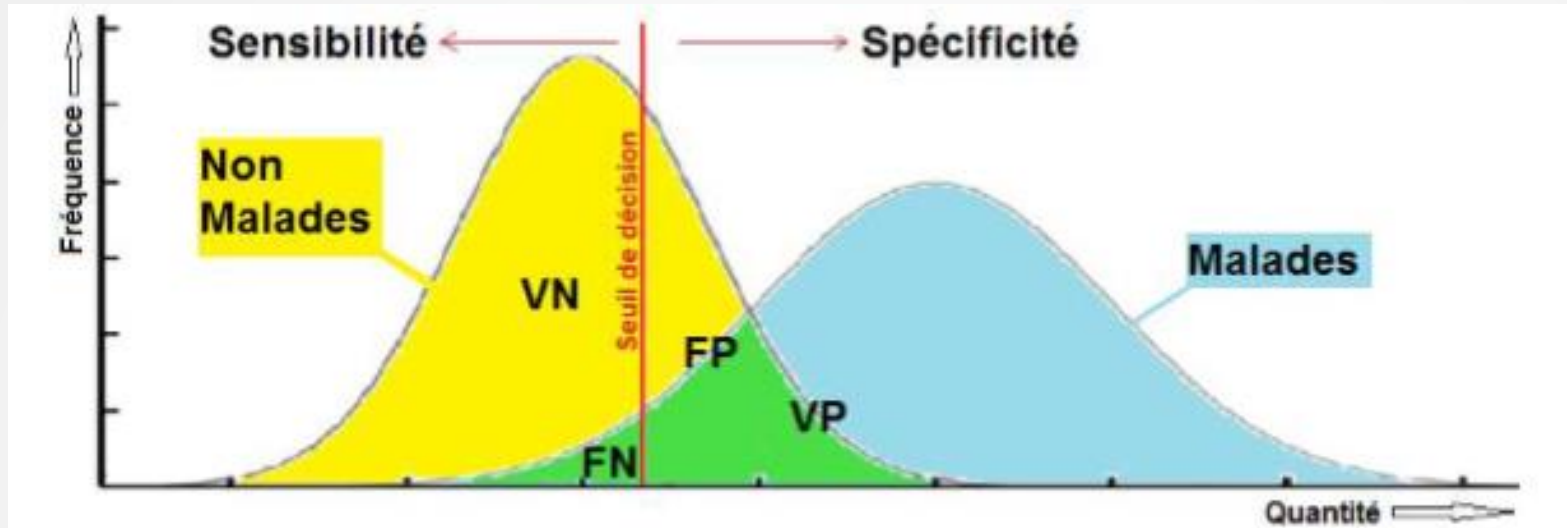
*Exemple : Mieux vaut manquer dans un premier temps le diagnostic anténatal d'une spina bifida (FN), que de diagnostiquer à tort cette malformation (FP) et risquer une interruption volontaire de grossesse !*



- ▶ Si un test a une sensibilité à 100% : tous les malades sont correctement identifiés, il n'y a aucun faux négatif
- ▶ Si un test a une spécificité de 100% : tous les non malades sont correctement identifiés, il n'y a aucun faux positif
- ▶ Spécificité et sensibilité sont comprises entre 0 et 1, exprimées en %
- ▶ Sensibilité et spécificité sont propres au test, elles sont indépendantes de la prévalence de la maladie

# Choix d'un seuil

- ▶ Répercussion sur la Sensibilité et la Spécificité : elles varient l'une au dépend de l'autre
- ▶ Selon le seuil choisi pour le test, un plus grand nombre de Faux Négatifs (FN) ou de Faux Positif (FP) est engendré
- ▶ Nécessité de fixer un seuil correspondant à la sensibilité ou la spécificité souhaitée :
  - Se privilégie pour éviter un trop grand nombre de Faux Négatif (FN)
  - Sp privilégiée pour éviter un trop grand nombre de Faux Positif (FP)



- Si on **diminue** le seuil (on le déplace vers la gauche) → Sensibilité **augmente** → Nombre de Vrai Positif (VP) **augmente** → mais le nombre de Faux Positifs (FP) **augmente** aussi → Spécificité **diminue**
- Si on **augmente** le seuil (on le déplace vers la droite) → Spécificité **augmente** → Nombre de Vrai Négatif (VN) **augmente** → mais le nombre de Faux Négatif (FN) **augmente** aussi → Sensibilité **diminue**

# I. Paramètres mesurant la valeur informationnelle du test diagnostique (en situation réelle)

## A) Valeurs prédictives

- ▶ La Valeur Prédictive Positive (VPP) est définie parmi le groupe des sujets positifs (T+)
- ▶ La Valeur Prédictive Négative (VPN) est définie parmi le groupe des sujets négatifs (T-)

# La Valeur Prédictive Positive

- ▶ C'est la probabilité d'être malade lorsque le résultat est positif :  $P(M/T+)$
- ▶ Capacité d'un test à prédire la présence de la maladie
- ▶ Lorsque VPP = 100% tous les résultats positifs correspondent à des malades : il n'y a aucun Faux Positifs (FP)
- ▶  $VPP = P(M/T+) = \frac{P(T+ \cap M)}{P(T+)}$
- ▶ On trouve donc :  $VPP = \frac{VP}{VP+FP}$

# La Valeur Prédictive Négative

- ▶ C'est la probabilité d'être indemne de la maladie quand le test est négatif :  $P(NM/T-)$
- ▶ Lorsque VPN = 100% tous les résultats négatifs correspondent à des non malades : il n'y a aucun Faux Négatifs (FN)
- ▶  $VPN = P(NM/T-) = \frac{P(NM \cap T-)}{P(T-)}$
- ▶ On trouve donc :  $VPN = \frac{VN}{VN+FN}$

# Influence de la prévalence

- ▶ La Sensibilité et la Spécificité sont indépendantes de la prévalence (intrinsèque)
- ▶ La VPP et la VPN dépendent quand à elle de la prévalence : qualité extrinsèque du test

L'estimation des VPP et VPN à partir des VP, VN, FP, FN, ne peut se faire que si la proportion de malade dans l'échantillon d'étude correspond à la prévalence de la maladie dans la population cible !

- ▶ Si la prévalence dans l'échantillon ne correspond pas à celle dans la population cible on ne pourra pas estimer les VPP et VPN directement avec les VP, VN, FP, FN tirés de l'étude !!
- ▶ Dans ce cas là :  $VPP \neq \frac{VP}{VP+FP}$   $VPN \neq \frac{VN}{VN+FN}$
- ▶ On pourra calculer les valeurs prédictives en utilisant la Se, la Sp mais aussi la **prévalence P** de la population cible !!!

# Détermination de la VPP avec le Théorème de Bayes

- ▶  $VPP = P(M/T+) = \frac{P(M \cap T+)}{P(T+)} = \frac{P(M \cap T+)}{P(T+ \cap M) + P(T+ \cap NM)}$
- ▶ D'après le Théorème de Bayes :  $VPP = \frac{P(T+/M) \times P(M)}{P(T+/M) \times P(M) + P(T+/NM) \times P(NM)}$
- ▶ Sachant que  $Se = P(T+/M)$  ,  $Sp = P(T-/NM) = 1 - P(T+/NM)$  et que  $P(M)$  représente la prévalence on obtient :

$$VPP = \frac{SeP}{(SeP) + (1 - Sp) \times (1 - P)}$$

Si **VPP = 100%** : aucun **Faux Positifs** (FP)

# Détermination de la VPN avec le Théorème de Bayes

- ▶  $VPN = P(NM/T-) = \frac{P(NM \cap T-)}{P(T-)} = \frac{P(NM \cap T-)}{P(T- \cap NM) + P(T- \cap M)}$
- ▶ D'après le Théorème de Bayes :  $VPN = \frac{P(T-/NM) \times P(NM)}{P(T-/NM) \times P(NM) + P(T-/M) \times P(M)}$
- ▶ Sachant que  $Se = P(T+/M) = 1 - P(T-/M)$ ,  $Sp = P(T-/NM)$  et que  $P(M)$  représente la prévalence :

$$VPN = \frac{Sp(1-P)}{Sp(1-P) + (1-Se)P}$$

Si **VPN = 100%** : **Aucun Faux Négatifs (FN)**



- ▶ Si  $Se = 100\% \rightarrow VPN = 100\%$ . En effet une  $Se$  de  $100\%$  signifie qu'il n'y a aucun Faux Négatifs

Sachant que  $VPN = \frac{VN}{VN+FN}$ , si  $FN=0$  alors  $VPN=100\%$

- ▶ Si  $Sp = 100\%$  alors  $VPP = 100\%$  puisqu'il n'y a aucun FP

$VPP = \frac{VP}{VP+FP}$  donc si  $FP=0$   $VPP=100\%$

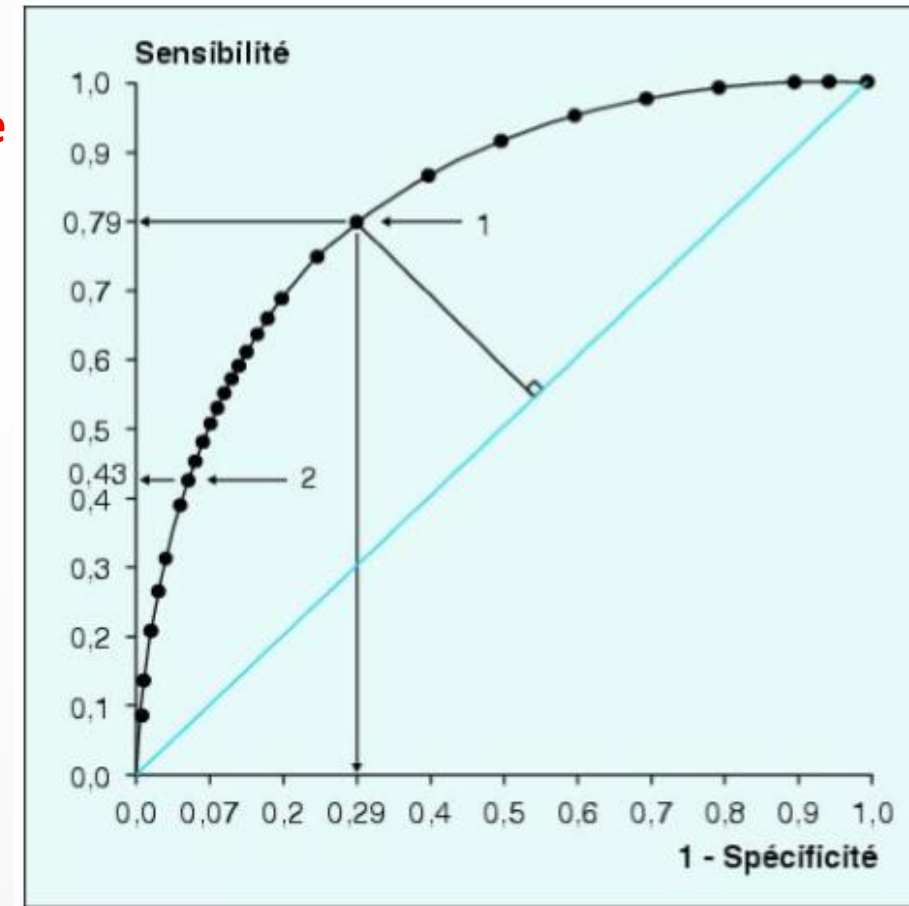
- ▶  $VPP$  et  $VPN$  permettent d'estimer la probabilité **post-test** de la maladie **après** le résultat du test (T+/T-)



# La Courbe ROC

- ▶ La courbe ROC est un outil qui permet de représenter l'évolution **conjointe** de la Sensibilité et de la Spécificité, en fonction de la valeur seuil choisie
- ▶ L'objectif des courbes ROC est de **déterminer la valeur seuil optimale**
- ▶ Le seuil optimum est celui qui correspond au point le plus près du **coin supérieur gauche**

▲ 1-Spécificité !!!



# Interprétation de la courbe ROC

- ▶ Un test avec un fort pouvoir discriminatoire occupera la partie supérieure gauche du graphique
- ▶ La valeur seuil optimale : Point le plus proche du Point parfait ( $Se = Sp = 1$ )
- ▶ Le test est d'autant plus performant que la courbe est proche du point Parfait et s'éloigne de la diagonale
- ▶ Un test dont la courbe ROC est située sur la diagonale n'apporte aucune information diagnostique

# Super point méthode

		Information sur la maladie (résultat du gold standard)	
		<b>M</b>	<b>NM</b>
Résultat du test	<b>T+</b>	VP	FP
	<b>T-</b>	FN	VN

# QCMS

- ▶ A propos des définitions donnez la(les) vraie(s) :
- A) On appelle "Vrais Positifs" les patients qui sont à la fois malade et négatifs au test
- B) Avec les Faux Négatifs la maladie est diagnostiquée plus tardivement
- C) Avec les Faux Positifs on induit des examens inutiles et anxiogènes
- D) On appelle "Faux Négatifs" les patients non malades et positifs au test
- E) Toutes les propositions sont fausses

# QCMS

- ▶ A propos des définitions donnez la(les) vraie(s) :
- A) On appelle "Vrais Positifs" les patients qui sont à la fois malade et négatifs au test
- B) Avec les Faux Négatifs la maladie est diagnostiquée plus tardivement
- C) Avec les Faux Positifs on induit des examens inutiles et anxiogènes
- D) On appelle "Faux Négatifs" les patients non malades et positifs au test
- E) Toutes les propositions sont fausses

# QCM 2

► Donnez la(les) proposition(s) vraie(s) :

A)  $Se = P(T+/M)$

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

C)  $Se = P(T+ \cap M)/P(M)$

D)  $Sp = P(T-/M)$

E) Toutes les propositions sont fausses

		Information sur la maladie (résultat du gold standard)	
		<b>M</b>	<b>NM</b>
Résultat du test	<b>T+</b>	VP	FP
	<b>T-</b>	FN	VN

# QCM 2

► Donnez la(les) proposition(s) vraie(s) :

A)  $Se = P(T+/M)$

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

C)  $Se = P(T+ \cap M)/P(M)$

D)  $Sp = P(T-/M)$

E) Toutes les propositions sont fausses

		Information sur la maladie (résultat du gold standard)	
		M	NM
Résultat du test	T+	VP	FP
	T-	FN	VN

# QCM 3

- On met en place un test pour dépister une hyperthyroïdie sur 1000 sujets. On a trouvé 75 sujets positifs, parmi lesquels 35 étaient réellement malades. On trouve également 5 Faux Négatifs. Quelle(s) est(sont) la(les) proposition(s) exacte(s) ?
- A) Le nombre de patients ayant un test négatif est de 925
  - B) La sensibilité du test est égale à  $35/40$
  - C) La Valeur Prédictive Négative du test est égale à  $920/960$
  - D) Il y a 40 Faux Positifs
  - E) Toutes les propositions sont fausses

# QCM 3

- On met en place un test pour dépister une hyperthyroïdie sur 1000 sujets. On a trouvé 75 sujets positifs, parmi lesquels 35 étaient réellement malades. On trouve également 5 Faux Négatifs. Quelle(s) est(sont) la(les) proposition(s) exacte(s) ?
- A) Le nombre de patients ayant un test négatif est de 925
  - B) La sensibilité du test est égale à  $35/40$
  - C) La Valeur Prédictive Négative du test est égale à  $920/960$
  - D) Il y a 40 Faux Positifs
  - E) Toutes les propositions sont fausses

# QCM 4

- ▶ Donnez la(les) réponse(s) vraie(s) :
- A) Si un test à une Sensibilité de 100% il n'y a aucun faux positifs
- B) Si un test à une Spécificité de 100% il n'y aucun faux négatifs
- C) Se et Sp dépendent de la prévalence de la maladie
- D) Les Valeurs Prédicatives sont indépendantes de la prévalence de la maladie
- E) Toutes les propositions sont fausses

# QCM 4

- ▶ Donnez la(les) réponse(s) vraie(s) :
- A) Si un test à une Sensibilité de 100% il n'y a aucun faux positifs
- B) Si un test à une Spécificité de 100% il n'y a aucun faux négatifs
- C) Se et Sp dépendent de la prévalence de la maladie
- D) Les Valeurs Prédicatives sont indépendantes de la prévalence de la maladie
- E) Toutes les propositions sont fausses

# QCM 5

- ▶ Concernant les valeurs prédictives positives (VPP) et négatives (VPN), donnez la(les) proposition(s) vraie(s) :
- A) La VPP correspond à la capacité d'un test à prédire la présence de la maladie
- B) La VPN correspond à la capacité d'un test à prédire l'absence de la maladie
- C) La VPP et VPN dépendent de la prévalence de la maladie donc sont des qualités extrinsèques d'un test
- D) VPP et VPN sont indépendants de la sensibilité et de la spécificité du test diagnostique
- E) Toutes les propositions sont fausses

# QCM 5

- ▶ Concernant les valeurs prédictives positives (VPP) et négatives (VPN), donnez la(les) proposition(s) vraie(s) :
- A) La VPP correspond à la capacité d'un test à prédire la présence de la maladie
- B) La VPN correspond à la capacité d'un test à prédire l'absence de la maladie
- C) La VPP et VPN dépendent de la prévalence de la maladie donc sont des qualités extrinsèques d'un test
- D) VPP et VPN sont indépendants de la sensibilité et de la spécificité du test diagnostique
- E) Toutes les propositions sont fausses

# QCM 6

- ▶ Selon le choix du seuil de décision d'un test, la sensibilité et la spécificité varient. En considérant que l'on diminue le seuil, donner la ou les proposition(s) vraie(s)
  - A) La spécificité augmente
  - B) La sensibilité diminue
  - C) Le nombre de Vrais Positifs (VP) diminue
  - D) Le nombre de Faux Positifs (FP) diminue
  - E) Toutes les propositions sont fausses

# QCM 6

- ▶ Selon le choix du seuil de décision d'un test, la sensibilité et la spécificité varient. En considérant que l'on diminue le seuil, donner la ou les proposition(s) vraie(s)
  - A) La spécificité augmente
  - B) La sensibilité diminue
  - C) Le nombre de Vrais Positifs (VP) diminue
  - D) Le nombre de Faux Positifs (FP) diminue
  - E) Toutes les propositions sont fausses

# QCM 7

- ▶ Le tracé d'une courbe ROC (Receiver Operating Characteristics) : /!\ QCM tiré du livre du Pr STACCINI /!\
- A) Est constitué de couples de valeurs (sensibilité, spécificité)
- B) Permet de catégoriser le niveau de performance d'un test diagnostique
- C) Indique une performance maximale du test lorsqu'il est proche de la diagonale
- D) Comporte deux axes dont les valeurs vont de  $-1$  à  $+1$
- E) Toutes les propositions sont fausses

# QCM 7

- ▶ Le tracé d'une courbe ROC (Receiver Operating Characteristics) : /!\ QCM tiré du livre du Pr STACCINI /!\
- A) Est constitué de couples de valeurs (sensibilité, spécificité)
- B) Permet de catégoriser le niveau de performance d'un test diagnostique
- C) Indique une performance maximale du test lorsqu'il est proche de la diagonale
- D) Comporte deux axes dont les valeurs vont de  $-1$  à  $+1$
- E) Toutes les propositions sont fausses

FIN



N'oubliez pas, la biostats vous aime ♡♡