

# Evaluation d'une méthode diagnostique (II)

# Plan du cours

- Évaluation d'une méthode/test diagnostique quantitatif (non-dichotomique)
  - I: validité, fiabilité
  - II: intrinsèque: Se, Sp, LR+, LR-
    - **Courbe ROC (Receiver Operator Characteristic)**
    - **L'aire de la courbe ROC (AUC ou AUROC)**
  - III: extrinsèque: VPP, VPN
  - IV: cout – bénéfice
- Exemples des études
- Exemples des exercices

# Évaluation d'une méthode diagnostique

- **Définition:** des études qui permet l'évaluation
  - des **avantages** (les **qualités diagnostiques**: Se, Sp, ...) et **inconvénients** offerts par
  - un nouveau test de diagnostic, ou plusieurs
    - par rapport aux méthodes standard de diagnostic d'une maladie
- **Utilité:**
  - réaliser le **diagnostic d'une maladie**
  - identifier la gravité/stade d'un maladie afin de **choisir le meilleur traitement**
- **Exemples** des méthodes diagnostiques:
  - imagerie médicale (radiographie du crane, tomodensitométrie, radiographie dentaire),
  - examen anatomopathologique,
  - biochimie du sang/ urine... (glycémie, glycosurie, ...)

# Évaluation d'une méthode diagnostique

- On doit préciser dans le titre, fin d'Introduction (but/objective), résumé, chapitre méthodes de l'article/thèse qu'on étudie les qualités diagnostiques.
- Recherché **analytique** – on compare deux tests diagnostiques
- le nouveau test diagnostique peut être un test **qualitatif** (résultat négatif/positif) – voir le cours 05
- ou un **test quantitatif** (résultats quantitatifs=variable quantitative ou qualitative ordinale avec au moins 5 classes/catégories/modalités):
- Exemples des tests/méthodes quantitatifs:
- taux d'hormone dans le sang,
- taille d'une tumeur,
- les scores d'un questionnaire de santé bucco-dentaire

# Pour des tests diagnostiques **quantitatifs**: la **courbe ROC**

- l'objectif de l'utilisation de la **courbe ROC** (Receiver Operator Characteristic) pour un **test diagnostique quantitatif** est:  
de savoir si les valeurs « **élevées** » du test correspondent à des **individus malades** et si les valeurs « **faibles** » du test correspondent à des **individus sains** (dans la majorité des études cliniques)
- les notions de « **élevées** » et « **faibles** » se réfèrent au fait que les valeurs obtenues par le test diagnostique seront supérieures ou inférieures à une **valeur seuil**

# Pour des tests diagnostiques quantitatifs.....

- On utilise les **courbes ROC (Receiver Operator Characteristic)**
  - Pour trouver la valeur seuil (en anglais « cut-off value ») pour la meilleur choix des **Se** et **Sp**
  - Pour évaluer la capacité diagnostique globale d'un test (avec AUROC ou AUC)
  - Pour comparer différents tests diagnostiques entre eux (qui est le meilleur, s'ils ont des capacités diagnostiques statistiquement différentes ou pas)
- **L'aire de la courbe ROC** (*en anglais – Area Under the Receiver Operator Characteristic*) : **AUROC ou AUC** = c'est la surface en dessous de la courbe ROC, qui donne une valeur diagnostique globale du test diagnostique quantitatif.

## Pour des tests diagnostiques **quantitatifs**

- **La méthode de construction de la courbe ROC**

- **chaque valeur** numérique du test diagnostique sera **considère** comme une **valeur seuil**, pour laquelle les valeur plus grands sont positives, et les valeurs plus petits sont négatives.
- après on fait **un tableau de contingence** -> on **calcule Se, Sp**, et on **met un point** sur le graphique ROC.
- on **répète l'opération** pour toutes les valeurs observées du test.
- au final on va **joindre toutes les points** pour **obtenir la courbe ROC**.

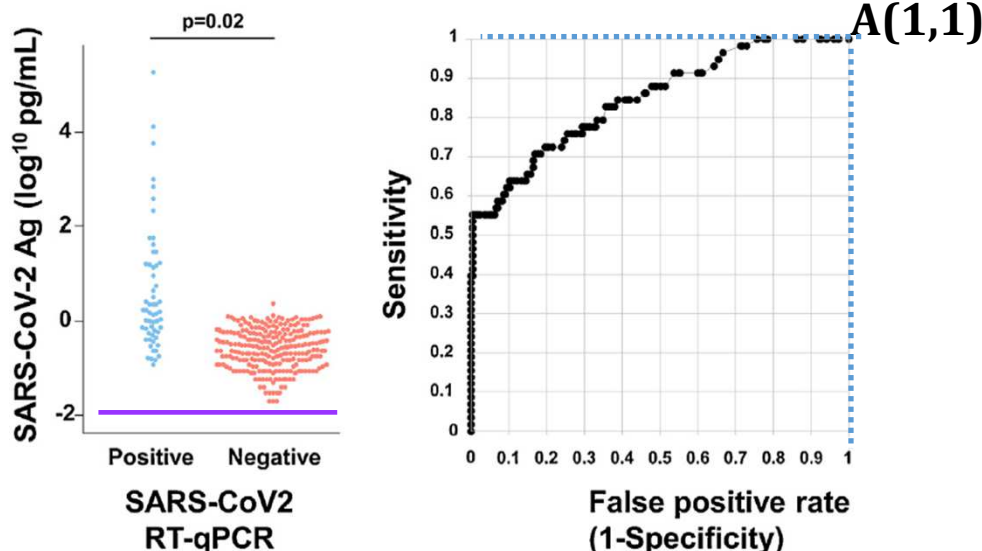
Exemple: on va évaluer la concentration des antigènes anti SARS-COV-2 mesurées par un test rapide antigénique, en le comparant avec un test RT-PCR

# Pour des tests diagnostiques quantitatifs

## • La méthode de construction de la courbe ROC

### Valeur seuil du test antigène = $-2$ (0,01 pg/mL):

Toutes les sujets ayant le test RT-PCR positif – **en bleu** – ont le test Ag positif (en haut de la ligne). 0 sujets avec le test RT-PCR positif – **en bleu** – ont le test Ag négatif (en bas de la ligne). Toutes les sujets avec le test RT-PCR négatif – **en rouge** – ont le test Ag positif (en haut de la ligne). 0 sujets avec le test RT-PCR négatif – **en rouge** – ont le test Ag négatif (en bas de la ligne).



RT-PCR vs. Test Ag	SARS- Cov2+	SARS- Cov2-	Total
Test Ag + ( $\geq -2$ )	58	255	313
Test Ag - ( $< -2$ )	0	0	0
Total	58	255	313

$M+ = \text{SARS-Cov2+}; M- = \text{SARS-Cov2-}$

$\text{Se} = \text{Pr}(\text{Test+}/M+) = 58 / (58 + 0) = 58 / 58 = 1$  (100%)

$\text{Sp} = \text{Pr}(\text{Test-}/M-) = 0 / (0 + 255) = 0 / 255 = 0$  (0%)

$1-\text{Sp} = 1$  (100%)  $\Rightarrow$  le point A(1,1) de la courbe

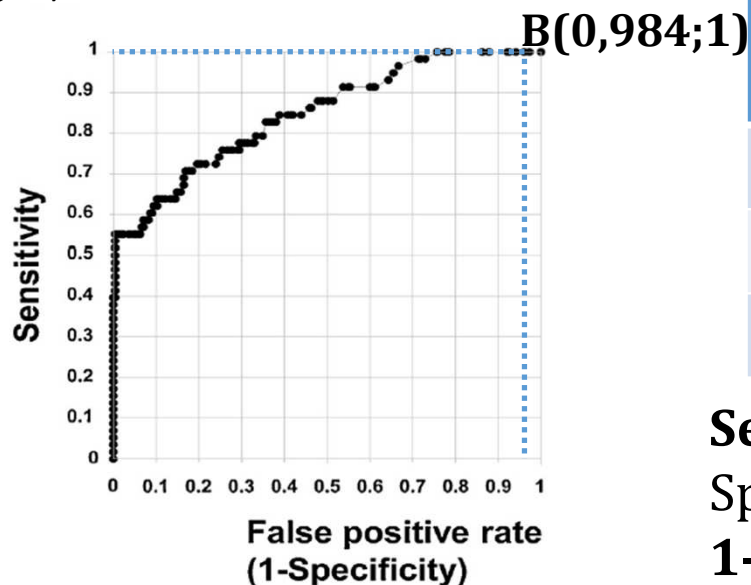
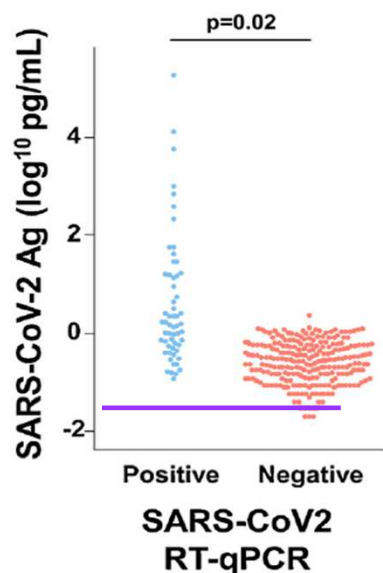


# Pour des tests diagnostiques quantitatifs

- La méthode de construction de la courbe ROC

Valeur seuil du test antigène =  $-1,7$  ( $0,02$  pg/mL):

Toutes les sujets avec le test RT-PCR positif – **en bleu** - ont le test positif (en haut de la ligne). 0 sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag négatif (en bas de la ligne). La majorité des sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag positif (en haut de la ligne) mais quelque (4) sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag négatif (en bas de la ligne).



RT-PCR vs. Test Ag	SARS- Cov-2 +	SARS- Cov-2-	Total
Test Ag +	58	251	309
Test Ag -	0	4	4
Total	58	255	313

$$Se = 58 / (58 + 0) = 58 / 58 = 1 \text{ (100\%)}$$

$$Sp = 4 / (4 + 251) = 4 / 255 = 0,016 \text{ (1,6\%)}$$

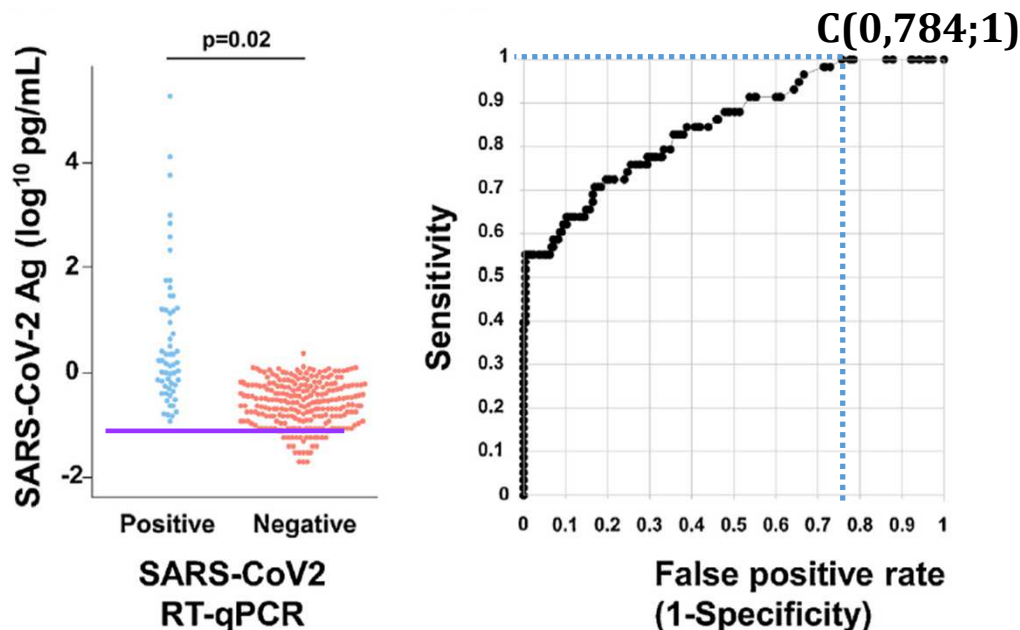
$$1-Sp=0,984$$

# Pour des tests diagnostiques quantitatifs

- La méthode de construction de la courbe ROC

Valeur seuil du test antigène =  $-1$  (0,1 pg/mL):

Toutes les sujets avec le test RT-PCR positif – en bleu – ont le test Ag positif (en haut de la ligne). 0 sujets avec le test RT-PCR positif – en bleu – ont le test Ag négatif (en bas de la ligne). Une grande partie des sujets avec le test RT-PCR négatif – en rouge – ont le test Ag positif (en haut de la ligne). Quelque (55) sujets avec le test RT-PCR négatif – en rouge – ont le test Ag négatif (en bas de la ligne).



RT-PCR vs.Test Ag	SARS- Cov-2 +	SARS- Cov-2–	Total
Test Ag +	58	200	258
Test Ag –	0	55	55
Total	58	255	313

$$Se = 58 / (58 + 0) = 58 / 58 = 1 \text{ (100\%)}$$

$$Sp = 55 / (55 + 200) = 55 / 255 = 0,216 \text{ (21,6\%)}$$

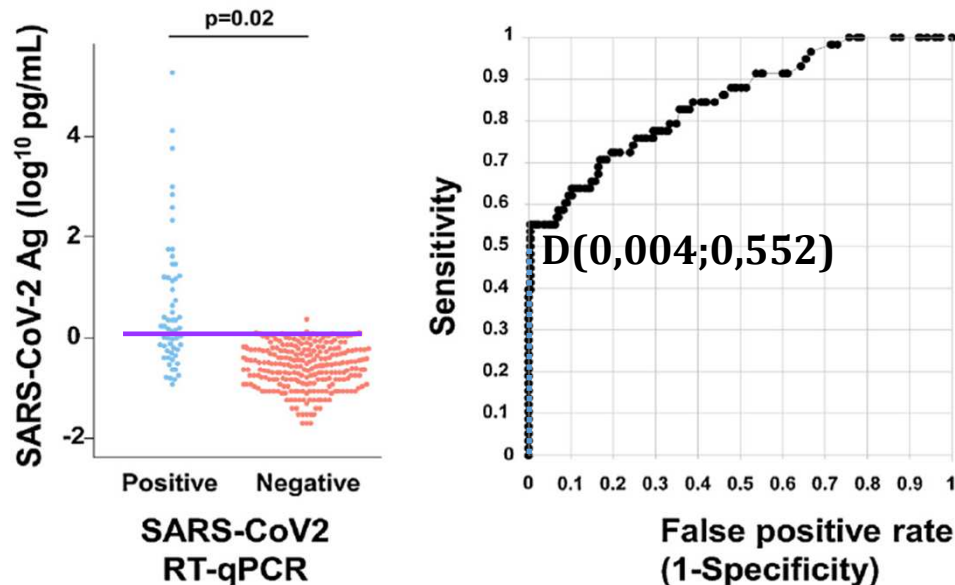
$$1-Sp=0,784 \text{ (78,4\%)}$$

# Pour des tests diagnostiques quantitatifs

## • La méthode de construction de la courbe ROC

**Valeur seuil du test antigène = 0,12 (1,318 pg/mL):**

Approximative une demie des sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag positif (en haut de la ligne).  
Approximative une (26) demie des sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag négatif (en bas de la ligne).  
Une seul sujet avec le test RT-PCR négatif – **en rouge** – a eu le test Ag positif (en haut de la ligne). Presque toutes les sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag négatif (en bas de la ligne).



RT-PCR vs. Test Ag	SARS- Cov-2 +	SARS- Cov-2–	Total
Test Ag +	32	1	33
Test Ag–	26	254	280
Total	58	255	313

$$Se = 32 / (26 + 32) = 32 / 58 = 0,552 \text{ (55,2\%)}$$

$$Sp = 254 / (1 + 254) = 254 / 255 = 0,996 \text{ (99,6\%)}$$

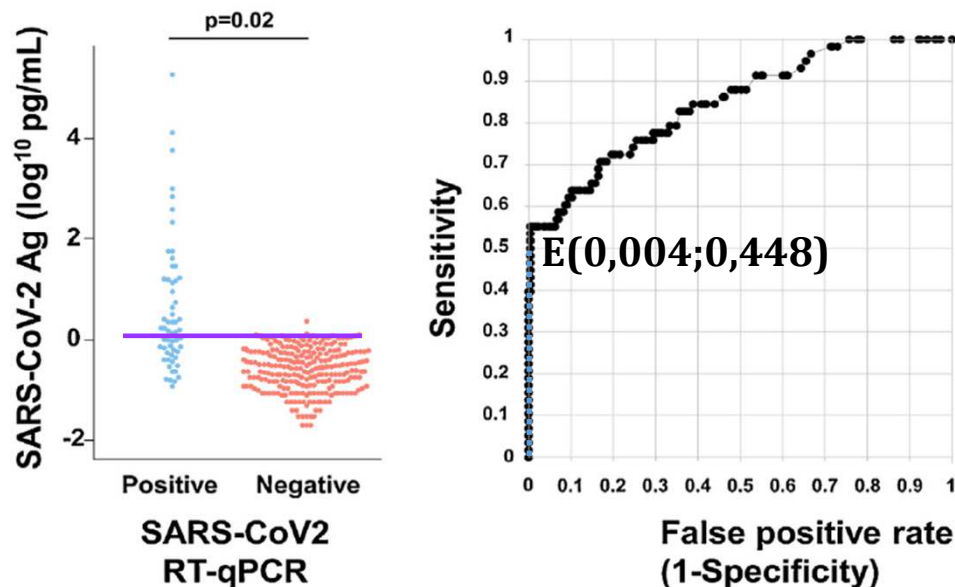
$$1-Sp=0,004 \text{ (0,4\%)}$$

# Pour des tests diagnostiques quantitatifs

## • La méthode de construction de la courbe ROC

**Valeur seuil du test antigène = 0,3 (1,995 pg/mL):**

Approximative une demie des sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag positif (en haut de la ligne). Approximative une (32) demie des sujets avec le test RT-PCR positif – **en bleu** - ont le test négatif (en bas de la ligne). Une seul sujet avec le test RT-PCR négatif – **en rouge** – a eu le test Ag positif (en haut de la ligne). Presque toutes les sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag négatif (en bas de la ligne).



RT-PCR vs. Test Ag	SARS- Cov-2 +	SARS- Cov-2–	Total
Test Ag +	26	1	27
Test Ag–	32	254	286
Total	58	255	313

$$Se = 26 / (26 + 32) = 26 / 58 = \mathbf{0,448} \text{ (44,8\%)}$$

$$Sp = 54 / (54 + 200) = 254 / 255 = \mathbf{0,996} \text{ (99,6\%)}$$

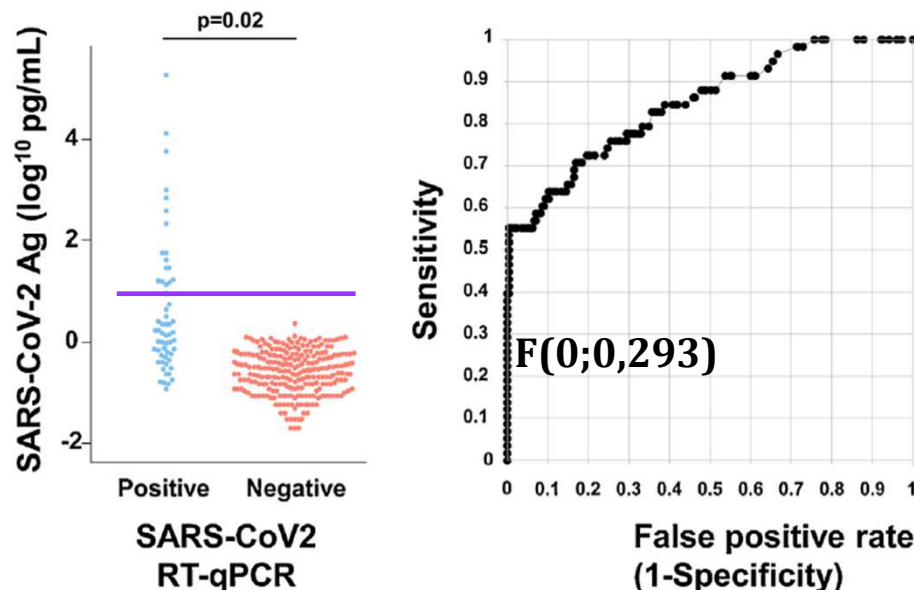
$$1-Sp = \mathbf{0,004} \text{ (0,4\%)}$$

# Pour des tests diagnostiques quantitatifs

- La méthode de construction de la courbe ROC

Valeur seuil du test antigène = 1 (10 pg/mL):

Approximative une tier (17) des sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag positif (en haut de la ligne). Approximative deux tiers des sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag négatif (en bas de la ligne). 0 sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag positif (en haut de la ligne). Tous les sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag négatif (en bas de la ligne).



RT-PCR vs. Test Ag	SARS- Cov-2 +	SARS- Cov-2 -	Total
Test Ag +	17	0	17
Test Ag -	41	255	296
Total	58	255	313

$$Se = 17 / (17 + 41) = 17 / 58 = 0,293 \text{ (29,3\%)}$$

$$Sp = 255 / (0 + 255) = 255 / 255 = 1 \text{ (100\%)}$$

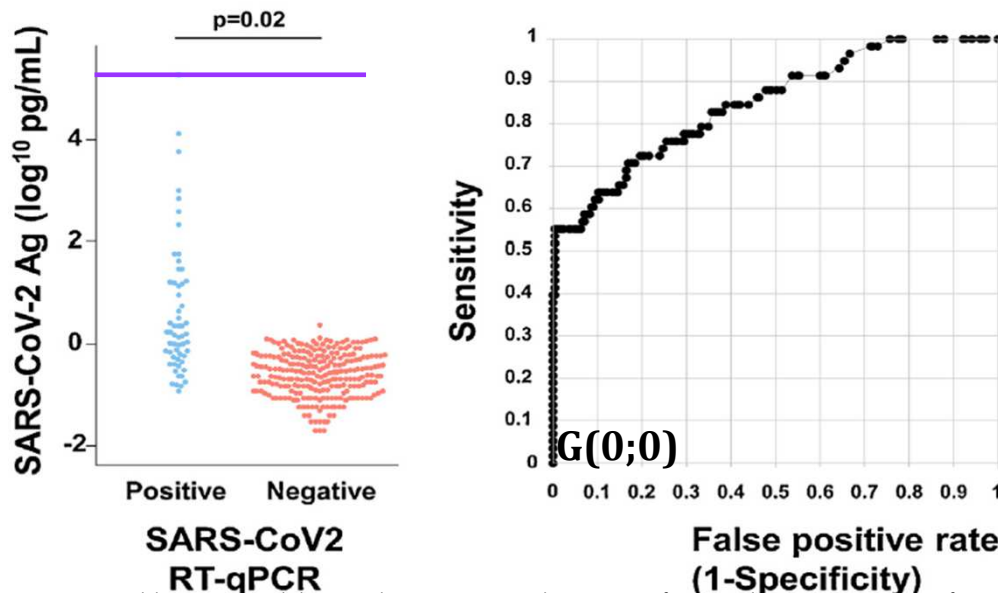
$$1-Sp=0 \text{ (0\%)}$$

# Pour des tests diagnostiques qui ont un résultat quantitatif

- La méthode de construction de la courbe ROC

Valeur seuil du test antigène = 4,8 (63095,7 pg/mL):

Tous les sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag positif (en haut de la ligne). 0 sujets avec le test RT-PCR positif – **en bleu** - ont le test Ag négatif (en bas de la ligne). 0 sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag positif (en haut de la ligne). Tous les sujets avec le test RT-PCR négatif – **en rouge** - ont le test Ag négatif (en bas de la ligne).



RT-PCR/Test Ag	SARS-Cov-2 +	SARS-Cov-2 -	Total
Test Ag +	0	0	0
Test Ag -	58	255	312
Total	58	255	312

$$Se = 0 / (0 + 58) = 0 / 58 = 0 \text{ (0\%)}$$

$$Sp = 255 / (0 + 255) = 255 / 255 = 1 \text{ (100\%)}$$

$$1-Sp=0 \text{ (0\%)}$$

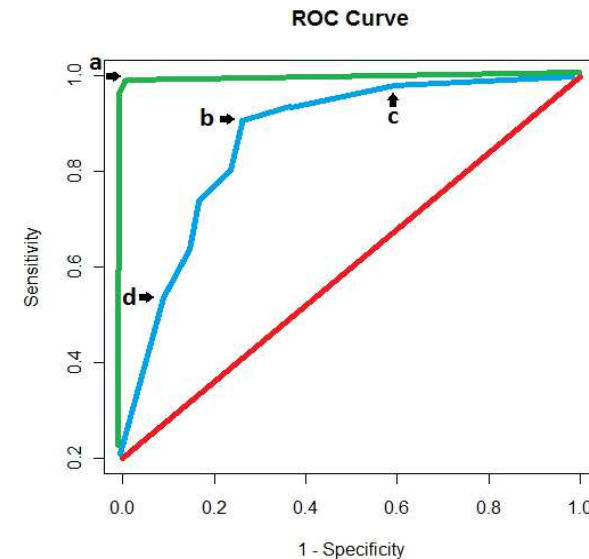
# Pour des tests diagnostiques qui ont un résultat quantitatif

## Le graphique ROC - Receiver Operator Characteristic

**Définition:** graphique qui aide à évaluer la qualité diagnostique d'un test diagnostique qui a comme résultat une variable quantitative (au lieu d'un résultat dichotomique: positif/ négatif)

### Caractéristiques

- axe horizontale =  $1 - \text{Spécificité}$
- axe verticale = Sensibilité
- la courbe représente **toutes les valeurs de Se, 1-Sp pour toutes les valeurs possibles du cut-off** pour le résultat quantitative du test
- plus la courbe est proche
  - **du coin supérieur gauche** – valeur diagnostique élevée – test discriminant (voir la courbe verte) – AUROC/AUC proche à 1
  - **de la diagonale** – valeur diagnostique absente – test non discriminant (voir la courbe rouge) – AUROC/AUC proche à 0,5
- **Exemple:** évaluer la présence d'un cancer du poumon en fonction du diamètre d'un nodule pulmonaire (le nouveau test), le test standard étant l'examen d'histopathologie





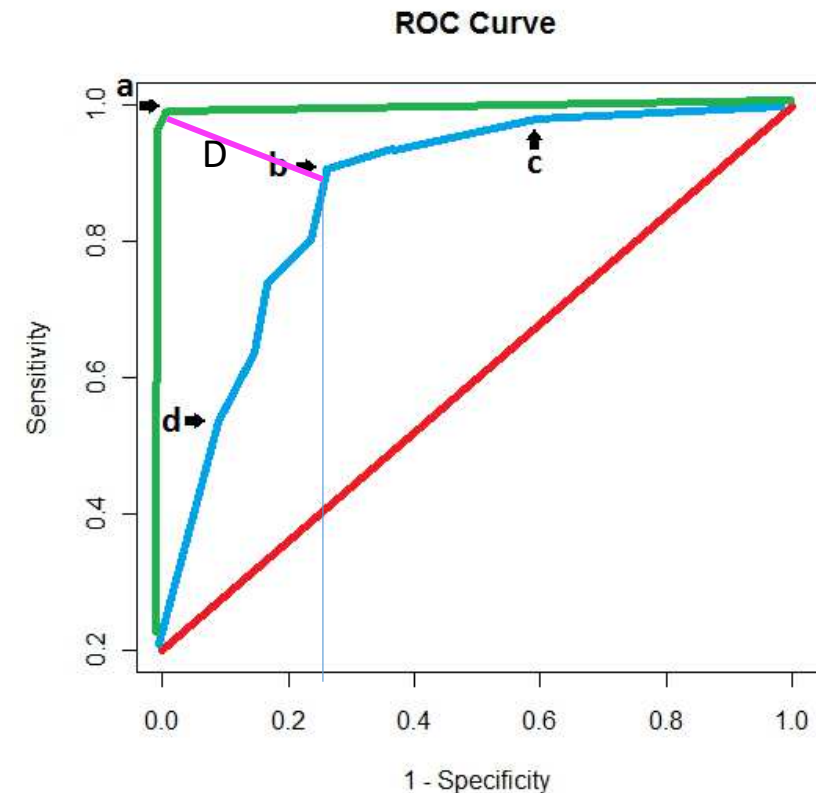
# Pour des tests diagnostiques quantitatifs

Le graphique ROC - Receiver Operator Characteristic

Le choix de la valeur seuil (cut-off)

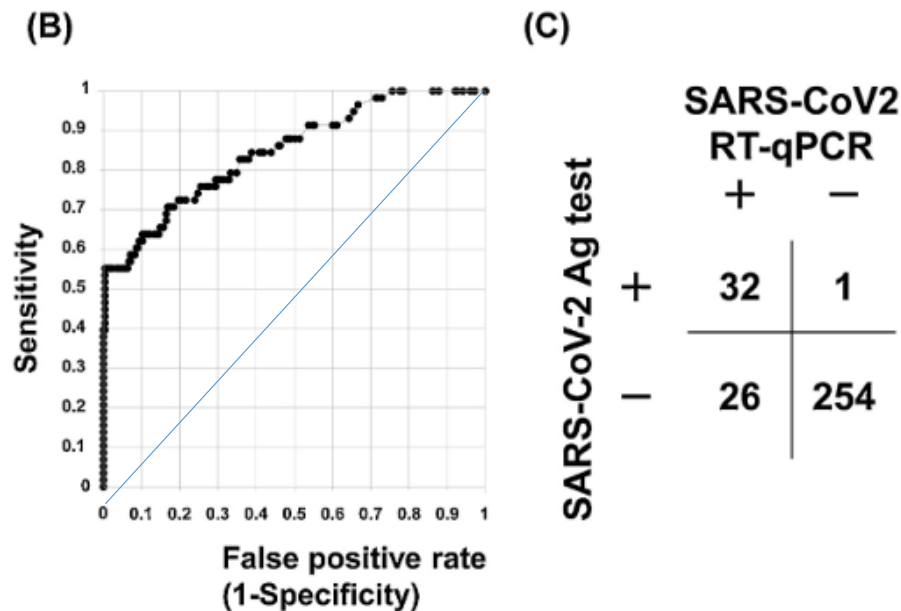
Le but du choix est pour avoir la meilleure combinaison de Se et Sp élevée

- En tenant compte du but de l'étude (voir l'exemple pour la courbe bleu)
  - **Diagnostic de screening** - Se élevée - **voir (c)** -  $Se = 0,95 = 95\%$ ,  $Sp = 1 - 0,6 = 0,4 = 40\%$
  - **Diagnostic de précision** - Sp élevée - **voir (d)** -  $Se = 0,52 = 52\%$ ,  $Sp = 1 - 0,1 = 0,9 = 90\%$
  - **Meilleur rapport Se/Sp** - le point le plus proche au coin supérieur gauche de la courbe - **voir (b)** -  $Se = 0,9 = 90\%$ ,  $Sp = 1 - 0,25 = 0,75 = 75\%$
- Il y a plusieurs méthodes pour identifier cette valeur seuil optimale (la meilleure combinaison des plus grandes valeurs de Se et Sp):
- Une technique est de trouver la valeur du indice Youden maximale (indice Youden =  $Se + Sp - 1$ ):
  - On cherche la valeur  $j = \max (Se+Sp-1)$ .
- Le point le plus proche du coin supérieur gauche de la courbe ROC:
  - le point avec la plus petite distance (D) vers le point (0,1) - supérieur gauche
    - On cherche la valeur  $j = \min [ (1-Se)^2 + (1-Sp)^2 ]$





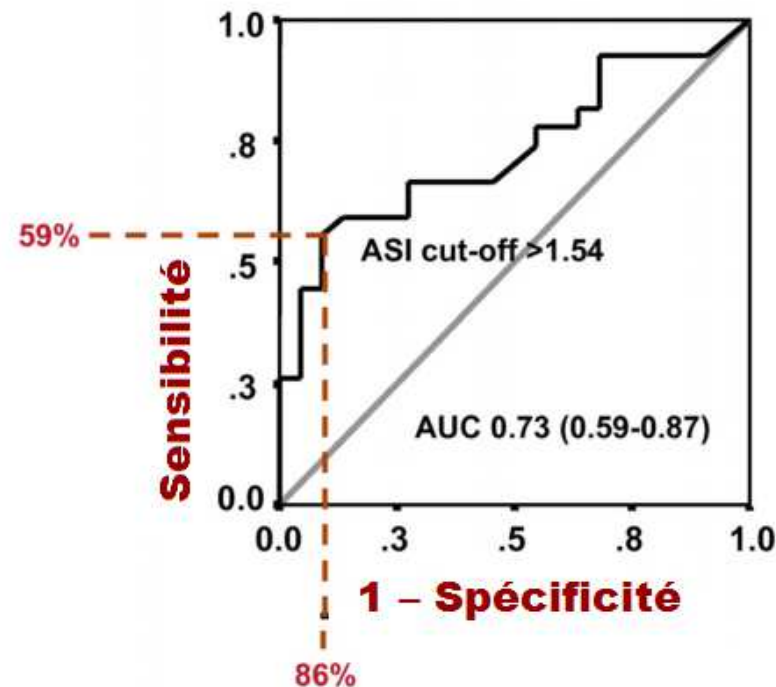
# La courbe ROC et AUC



- Valeur seuil: 0,12 (pour log Sars – Cov-2 Ag)= 1,31 pg/ml
- AUC=0,848 (dans 84,8 % des cas, un individu tiré au sort de la population des individus Sars-cov-2 positifs va avoir une valeur du test Ag supérieure (plus élevée) à celle d'un individu tiré au sort de la population des individus Sars-cov-2 négatifs )

## Exemple: Pour des tests diagnostiques quantitatifs

- Exemple: ASI - Atherosclerosis Score Index (score d'athérosclérose) pour le diagnostic de la présence de la maladie artérielle coronaire



## Pour des tests diagnostiques quantitatifs

- **Interprétation clinique de AUC / AUROC**, pour un graphique ROC - Receiver Operator Characteristic:
  - Comme pour les Se, Sp, précision (des suggestions approximatives d'interprétation clinique):
    - 0,5 - sans aucune utilité diagnostique – le test n'est pas discriminant
    - 0,5- 0,8 - mauvaise qualité
    - 0,8-0,9 - modéré
    - 0,9-1- bonne qualité
  - proche de 1 (100%) signifie un diagnostic de haute qualité – plus le test est discriminant

# Pour des tests diagnostiques quantitatifs

## Le graphique ROC - Receiver Operator Characteristic

### La relation entre l'intervalle de confiance de AUROC (AUC) et la valeur du p

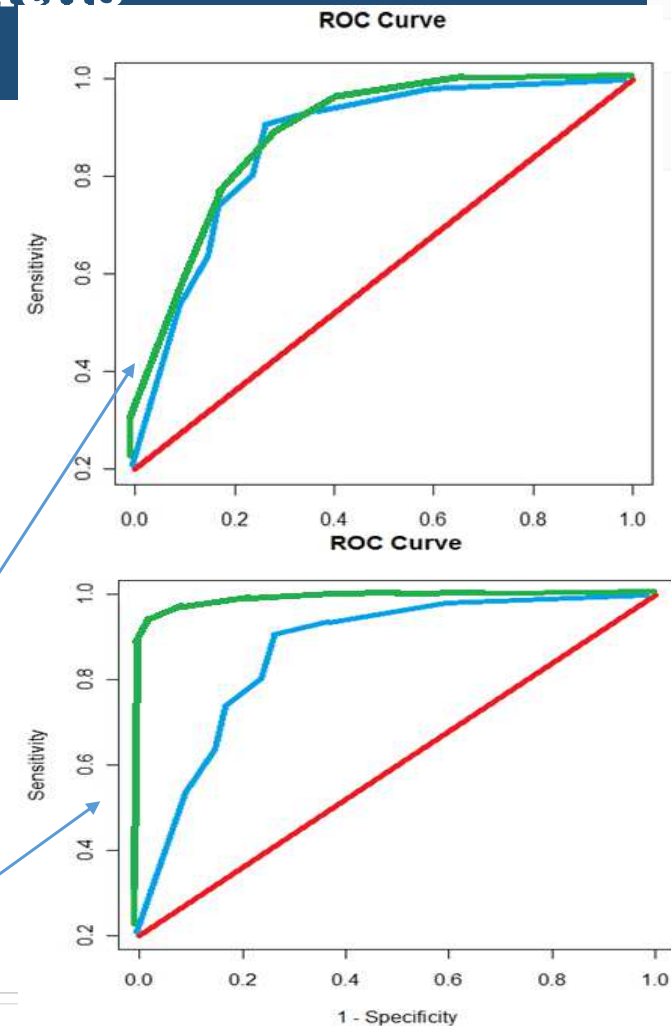
- Pour l'AUROC on peut faire un **test statistique pour AUROC** pour voir s'il est statistiquement significative.
  - L' hypothèse nulle du test statistique pour l'AUROC est que l'AUROC est égale à 0,5
  - L' hypothèse alternative du test statistique pour l'AUROC est que l'AUROC n'est pas égale à 0,5
  - Si la valeur p du test est  $< 0,05$  → le résultat est statistiquement significative
  - Si la valeur p du test est  $> 0,05$  → le résultat n'est pas statistiquement significative
- Il y a une relation entre l'intervalle de confiance 95% de l'AUROC et la valeur p du test pour l'AUROC
  - si **l'intervalle de confiance 95% de l'AUROC contient la valeur 0,5**, - on n'a pas pu rejeter l' hypothèse nulle, la **valeur p** du test est  $> 0,05$ , le résultat du test statistique pour l'AUROC **n'est pas statistiquement significative**
  - si l'intervalle de confiance 95% de l'AUROC **ne contient pas la valeur 0,5**, - on peut rejeter l' hypothèse nulle, accepter l' hypothèse alternative, la **valeur p** du test est  $< 0,05$ , le résultat du test statistique pour l'AUROC est **statistiquement significative**

# Pour des tests diagnostiques quantitatifs

## Le graphique ROC - Receiver Operator Characteristic

### La comparaison entre deux courbes ROC des deux tests diagnostiques quantitatifs par rapport a un test standard de reference

- La comparaison peut être faite initiale d'une manière graphique.
  - La courbe qui est **plus proche** du point **supérieur gauche** – va avoir la plus grande valeur diagnostique globale.
  - La courbe qui est plus proche a la **première diagonale**, va avoir la pire valeur diagnostique globale
  - Si les deux courbes sont **proche une a l'autre** (on s'intersectant ou non), ca est une suggestion que les deux tests diagnostiques ont une valeur diagnostique similaire
  - Si les deux courbes **ne sont pas proche une a l'autre**, ca est une suggestion que les deux tests diagnostiques ont des valeurs diagnostique différentes



# Comparaison des plusieurs tests diagnostiques quantitatifs

## Le graphique ROC - Receiver Operator Characteristic

### La comparaison entre deux courbes ROC des deux tests diagnostiques quantitatives par rapport a un test standard d'ore

- Ensuite, certaines courbes peuvent avoir des valeur seuils, qui ont un certaine avantage par rapport a les autres
- Enfin, il y a des **tests statistiques** qui nous permettent de vérifier si la **différence est statistiquement significative**.
  - Les hypothèses statistiques du test:
    - L'hypothèse nulle: il n'y a pas de différence statistiquement significative entre le test diagnostique A et le test diagnostique B (par rapport au test standard C), en ce qui concerne la performance diagnostique globale
    - L'hypothèse alternative: il y a une différence statistiquement significative entre le test diagnostique A et le test diagnostique B (par rapport au test standard C), en ce qui concerne la performance diagnostique globale
  - La décision:
    - Si  $p < 0,05 \Rightarrow$  on rejette  $H_0$ , on accepte  $H_1$ , il y a une différence statistiquement significative ...
    - Si  $p > 0,05 \Rightarrow$  on ne peut pas rejeter  $H_0$ , on ne peut pas dire qu'il y a une différence ...

## Phase III: les qualités extrinsèques du test

- Dans la pratique, le statut malade ou non-malade des sujets soumis au test n'est pas connue.

		Test standard		
		Maladie présente	Maladie absente	Total
Nouveau test	Test positif	VP a Vrai Positif	FP b Faux Positif	a+b
	Test négatif	FN c Faux Négatif	VN d Vrai Négatif	c+d
Total		a+c	b+d	a+b+c+d

# Phase III: Indicateurs

## Definitions et interpretations statistique de la VPP, VPN

### Valeur prédictive positive (VPP)

- en anglais – positive predictive value - PPV
- la probabilité pour un sujet qui a été testé positif de vraiment avoir la maladie
- la probabilité pour un sujet d'avoir la maladie sachant qu'il a été testé positif
  - Ex. la probabilité pour un sujet d'avoir cancer de langue sachant qu'il a été testé positif à le test de coloration avec toluidine est de 78%
  - Ex. la probabilité pour un sujet d'avoir Covid 19 sachant qu'il a été testé positif à le test de rapide pour Covid 19 est de 78%
- proportion des sujets réellement malades parmi les sujets qui ont été testés positifs

$$\text{VPP} = \text{VP} / (\text{VP} + \text{FP})$$

### Valeur prédictive négative (VPN)

- en anglais – negative predictive value - NPV
- la probabilité pour un sujet ayant un résultat négatif de vraiment n'avoir pas la maladie
- la probabilité pour un sujet n'avoir pas la maladie sachant qu'il a été testé négatif
  - Ex. la probabilité pour un sujet de n'avoir pas cancer de langue sachant qu'il a été testé négatif à le test de coloration avec toluidine est de 89%
  - Ex. la probabilité pour un sujet de n'avoir pas Covid 19 sachant qu'il a été testé négatif à le test de rapide pour Covid 19 est de 93%
- proportion des sujets vraiment non-malades parmi les personnes testées négatives:  $\text{VPN} = \text{VN} / (\text{VN} + \text{FN})$

		Test standard		Total
		Maladie présente	Maladie absente	
Nouveau test	Test positif	VP a Vrai Positif	FP b Faux Positif	a+b
	Test négatif	FN c Faux Négatif	VN d Vrai Négatif	c+d
	Total	a+c	b+d	a+b+c+d



## Phase III:

### Relations VPP et VPN avec Se, Sp et la prévalence:

- dépendent de la Se, Sp et de la prévalence (la fréquence de la maladie)
- étant fonction de la prévalence, les valeurs prédictives **ne peuvent être appliquée que dans une population ayant une prévalence identique** à celle dans la population étudiée lors du calcul de ces indicateurs.

## Phase III:

- VPP et VPN peuvent être calculé à partir de: Se, Sp et p (prévalence):

$$VPP = \frac{Se}{Se + (1 - Sp) \times \frac{1 - p}{p}} \quad \quad VPN = \frac{Sp}{Sp + (1 - Se) \times \frac{p}{1 - p}}$$

On doit pas apprendre les deux formules

## Phase III:

### Relations VPP et VPN avec Se et Sp et la prévalence

- $VPP = Se * \frac{\text{prévalence de la maladie}}{\text{prévalence du test positif}}$
- $VPN = Sp * \frac{1 - \text{prévalence de la maladie}}{\text{prévalence du test négatif}}$
- Plus la Sp est proche de 1 → le VPP est proche du 1
- Plus la Se est proche de 1 → le VPN est proche du 1

# Phase III:

## Relations VPP et VPN avec Prévalence

- $VPP = Se * \frac{\text{prévalence de la maladie}}{\text{prévalence du test positif}}$
- $VPN = Sp * \frac{1 - \text{prévalence de la maladie}}{\text{prévalence du test négatif}}$
- En tenant la:
  - Sensibilité égale –
    - VPP augmente avec une prévalence de la maladie plus grande
    - VPP diminue avec une prévalence du test positif petite
  - Spécificité égale –
    - VPN augmente avec une prévalence de la maladie plus petite
    - VPN diminue avec une prévalence du test positif plus grande (la prévalence du test négatif est petite)

## Phase IV:

- **Evaluation du test en termes de:**
  - Coût - Bénéfice
  - Coût - Efficacité
- **Réponses à des questions telles que:**
  - Est-ce que l'application du test va modifier le lieu de la pathologie dans la collectivité?
  - Est-ce que l'application du test va avoir un impact
    - sur le clinicien,
    - sur la morbidité et la mortalité de la population?
  - Quel est l'impact économique du test?

# Phase II et III

- **Méthodologie d'étude diagnostique (pour phase II et III)**
  - **Sujets**
    - **Population étudiée**
      - **Critères d'inclusion/exclusion** – essentiel pour limiter les biais de sélection (évaluation de la validité de l'étude)
      - Depuis où les patients ont été **recrute** - la location, le cadre, ou ont été réalisés les deux tests diagnostiques – essentiel pour évaluer la généralisabilité des résultats
    - **Recrutement** (en fonction des symptômes, résultats d'autres tests ou si le sujet a reçu le nouveau test ou le test standard)

## Phase II et III

- **Méthodologie d'étude diagnostique (pour phase II et III)**
  - **Sujets**
    - Sélection des sujets:
      - **échantillon représentatif** – meilleure / difficile à faire
      - **série consécutive des patients** – l'un des meilleurs si on ne peut pas faire représentatif
        - Les deux permettent que toutes les situations soient prises en compte (malades, non malades, différents stades de la maladie, avec/sans traitement, maladies similaires)
      - **ou série des patients non consécutifs** – on doit préciser comment – possibles biais de sélection
      - **nombre** des sujets suffisamment grand
      - on doit préciser la **période du temps du recrutement**
  - **Variantes des études:**
    - Collection **prospective** des patients sur qui on applique les deux tests diagnostiques
      - idéale
    - Collection **rétrospective** des données enregistrées dans les feuilles d'observation des hôpitaux

# Phase II et III

- Méthodologie d'étude diagnostique (pour phase II et III)
  - Les tests diagnostiques – méthodologie de travail
    - **Le test standard** doit être **acceptable** (le **meilleur** ou proche: un examen histologique, ou plusieurs tests, on doit argumenter qu'il est valide, préciser qu'il représente la référence), doit être bien décrit – critères pour les malades/ non malades.
    - **Le « nouveau » test** est parfois, **moins invasive (moins des risques)**, moins chère, plus simple, mieux acceptée par le patient - en principe
    - **Les tests doivent être reproductible** (intra/ inter observationnel – on doit préciser le coefficient Kappa ou autres mesures de reproductibilité)
    - **Evaluation** des toutes les patients
      - **Indépendante** (deux évaluateurs différents des deux tests)
      - En **aveugle** (les résultats ne sont pas connus pour l'autre évaluateur)
      - Répartition des tests diagnostiques – **randomisée**



## Phase II et III

- **Méthodologie d'étude diagnostique** (pour phase II et III)
  - Pour certaines tests – imagerie, anatomopathologique – le nombre et l' **expérience, formation** des médecins doit être précisée pour les deux tests diagnostiques (années d' expérience, spécialistes/primaire/ résident/ étudiant)
  - Tous les sujets doit être **évaluées avec les deux tests** diagnostiques (si un **grand nombre** des sujets n'est pas évaluées avec un test ca peut poser des problèmes. Si l' étude est prospective et la perte de vue est importante ca peut poser des problèmes - biais d'attrition)
  - Le **temps passée entre les deux tests** (doit être **précises** avec date du debout et du fin – il doit être **court**, sinon il y a un possible biais – modification de l' état du patient)
  - **Préciser** si les patients ont pris des **médicaments** pendant la période de l' étude – qui peuvent modifier les résultats des tests - biais

# Phase II et III

- Méthodologie d'étude diagnostique (pour phase II et III)
  - Les **résultats** doit préciser:
    - Description des **caractéristiques démographiques et cliniques** de l'échantillon
      - age, sexe, ...
      - pour évaluer la generalisabilite des résultats
    - Préciser
      - le **nombre de sujets qui n'ont pas eu le test** de référence, et le nouveau test
        - dans le texte ou diagramme du flux (en: flow-chart)
        - pour évaluer des biais
      - préciser des **données manquantes**
      - préciser si il y a des **situation pas claires** (pour le test standard et nouveau test), comme catégorie séparé, et de ne pas les exclure (biais de pureté diagnostique). Préciser diverses méthodes de tenir compte de ces situations.

# Phase II et III

- Méthodologie d'étude diagnostique (pour phase II et III)
  - Les **résultats** doit préciser:
    - Description des **caractéristiques démographiques et cliniques** de l'échantillon

**Table 1. Characteristics of the included 1014 patients.**

Characteristic	Results
<b>Age (years)</b>	
Mean age	51 ± 15, Range of 2 to 95
< 20	7 (1)
20-39	267 (26)
40-59	409 (40)
≥ 60	331 (33)
<b>Male</b>	467 (46)
<b>Median time-interval between chest CT scan and RT-PCR assay (days)</b>	1, Range of 0 to 7
<b>Results of RT-PCR assay</b>	
Positive	601 (59)
Negative	413 (41)
<b>Findings and manifestations of chest CT</b>	
Consistent with viral pneumonia (positive)	888 (88)
Ground-glass opacity	409/888 (46)
Consolidation	447/888 (50)
Reticulation/thickened interlobular septa	8/888 (1)
Nodular lesions	24/888 (3)
No CT findings of viral pneumonia	126 (12)

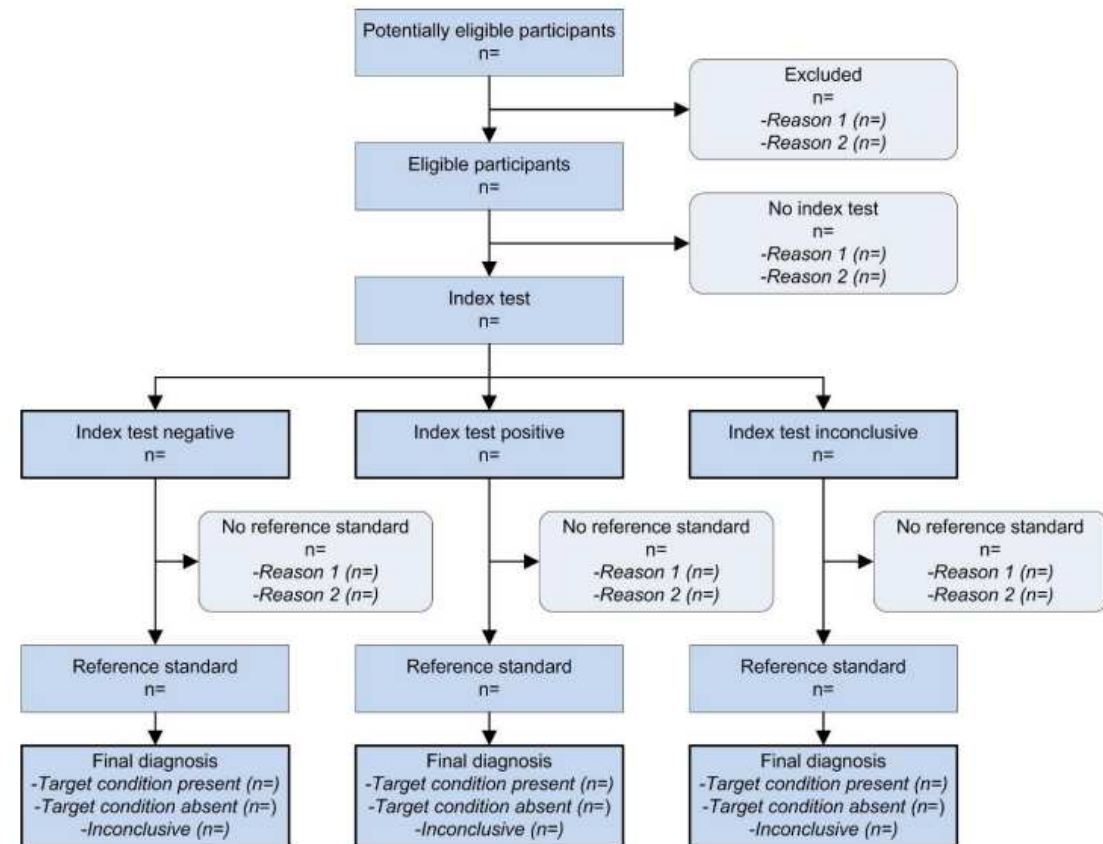
Note.-Data are patients (n =7) with percentages in parentheses. Age is mean ± standard deviation. Time-interval is shown as median.

RT-PCR= reverse transcription polymerase chain reaction.

Al T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology. 2020 Feb 26;200642. doi: 10.1148/radiol.20200642

# Phase II et III

- Méthodologie d'étude diagnostique (phase II et III)
  - Les **résultats** doit préciser:
    - dans le texte ou diagramme du flux
    - (en: flow-chart)



# Phase II et III

- Méthodologie d'étude diagnostique (pour phase II et III)
  - Les **résultats** doit préciser:
    - **tableau de contingence**
    - présentation des **Se, Sp, +LR, -LR, VPP, VPN, prévalence, précision**, avec des intervalles de confiance
    - présentation de **L'AUROC**, avec intervalle de confiance, et le **test statistique** pour la valeur diagnostique, ou tests statistiques pour comparer différentes méthodes diagnostiques. Présentation de la courbe ROC
    - **effets indésirables** - description

**Table 2: The performance of chest CT for COVID-19 infection with RT-PCR result as reference.**

Results (n)				Test performance (%)					
TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	Accuracy	[95% CI]
				[95% CI]	[95% CI]	[95% CI]	[95% CI]		
<b>Overall</b>									
580	105	308	21	97 (580/601)	25 (105/413)	65 (580/888)	83 (105/126)	68 (685/1014)	
				[95-98]	[22-30]	[62-68]	[76-89]	[65-70]	

## Phase IV:

- **Evaluation du test en termes de:**
  - Coût - Bénéfice
  - Coût - Efficacité
- **Réponses à des questions telles que:**
  - Est-ce que l'application du test va modifier le lieu de la pathologie dans la collectivité?
  - Est-ce que l'application du test va avoir un impact
    - sur le clinicien,
    - sur la morbidité et la mortalité de la population?
  - Quel est l'impact économique du test?

## Évaluation d'une méthode diagnostique Collection des données – un seul group

(Échantillon représentatif, cohorte, série consécutive des cas)

- la seule chose fixée par le chercheur est la **taille totale** de l'échantillon
- le résultat du test et le statut de la maladie résultent en manière aléatoire
- en utilisant la collection des données – avec un seul group de sujets (Échantillon représentatif, série consécutive des cas, série des cas, ...), on peut estimer/calculer:
  - Se et Sp
  - +LR (RVP), -LR (RVN)
  - Le pourcentage des biens classées (accuracy)
  - L' indice de Youden
  - VPP et VPN
  - Prévalence (frequence des toutes les malades) de la maladie dans la population cible – pour échantillon représentatif/série consécutive des cas – dépends ...

# Évaluation d'une méthode diagnostique

Collection des données - exposés-non exposés

i.e. résultat positif vs. négatif)

- approche prédictive
- **deux groupes de taille prédéterminée**
  - pas nécessairement égaux
  - sont constitués selon les résultats du test : (+) et (-)
- **l'état réel de la maladie** est le facteur aléatoire
- En utilisant la collection des données - exposés-non exposés, on peut estimer/calculer:
  - VPP et VPN



# Évaluation d'une méthode diagnostique

## Collection des données - cas-témoins

- **deux groupes de taille prédéterminée**
  - (pas nécessairement égaux → ils ne reflètent pas la prévalence de la maladie)
  - sont constitués, basée sur la présence ou l'absence de la maladie étudiée.
- **le résultat du test** est le facteur aléatoire
- En utilisant **la collection des données - cas-témoins on peut estimer/calculer:**
  - Se et Sp
  - +LR (RVP), -LR (RVN)
  - L'indice de Youden

# Evaluation de la validité des études diagnostiques

1. Le **spectre** des patients était-il **représentatif**? (échantillon consecutive ou representative - choisi par hasard)
2. La **test de référence acceptable** (identifie bien la présence/absence de la maladie)
3. pour être raisonnablement sûr que la condition cible n'a pas La période de **temps** entre le test de référence et le test d'index est-elle **suffisamment courte** hangé entre les deux tests?
4. **L'ensemble de l'échantillon** ou une sélection aléatoire de l'échantillon a-t-il reçu une **vérification à l'aide du test de référence** prévue? (vérification partielle évitée)
5. Les patients ont-ils reçu **le même test de référence** quel que soit le résultat du test d'index?
6. **Le test de référence était-il indépendant du test d'indice** (c'est-à-dire que le test d'index n'a pas partie de la norme de référence)? (incorporation évitée)
7. Les résultats du test de référence ont-ils été interprétés sans connaissance des résultats de la test d'index? (**résultats du test d'index en aveugle**)
8. Les résultats du test d'index ont-ils été interprétés sans connaissance des résultats du test de référence? (**résultats standard de référence aveuglés**)
9. **Les mêmes données cliniques étaient-elles disponibles** lorsque les résultats des tests ont été interprétés comme le seraient disponible lorsque le test est utilisé dans la pratique?
10. Les **résultats des tests non interprétables / intermédiaires ont-ils été rapportés**?
11. Les **retraits** de l'étude ont-ils été **expliqués**? (retraits expliqués)

[https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/ch09\\_Oct09.pdf](https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/ch09_Oct09.pdf)

## Validation of a new diagnostic method for quantification of sleep bruxism activity

Michelle Alicia Ommerborn<sup>1</sup>, Nicole Walentek<sup>2,3</sup>, Nora Bergmann<sup>2</sup>, Michael Franken<sup>2</sup>,  
Andreas Gotter<sup>4</sup>, Ralf Schäfer<sup>3</sup>

Affiliations + expand

PMID: 35195761 DOI: 10.1007/s00784-022-04398-w

FULL TEXT LINKS



ACTIONS

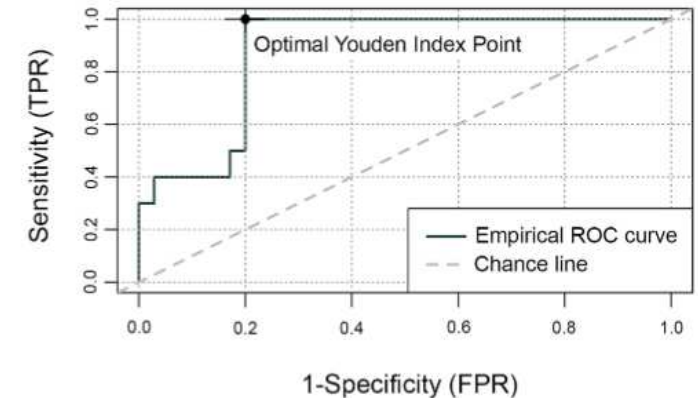
“ Cite

☆ Favorites

SHARE

# Example article MD

- Lien vers l'article: <https://pubmed.ncbi.nlm.nih.gov/35195761/>
- **Abstract**
- **Objectives:** To validate a new diagnostic method (DIABRUX) for quantifying sleep bruxism (SB) activity using the current gold standard, polysomnography (PSG), as a criterion in an adequate sample size investigation.
- **Materials and methods:** For SB diagnosis, each participant received a two-night ambulatory PSG including audio-video recordings. The 0.5-mm-thick sheet is produced in a thermoforming process. After diagnosis via PSG, each subject wore the diagnostic sheet for five consecutive nights. The resulting total abrasion on the surface was automatically quantified in pixels by a software specially designed for this purpose.
- **Results:** Forty-five participants (10 SB and 35 non-SB subjects) were included. The difference of the mean pixel score between the SB (M = 1,306, SD = 913) and the non-SB group (M = 381, SD = 483; 3.4 times higher for SB) was statistically significant ( $p < 0.001$ ). **The receiver operator characteristic (ROC) analysis revealed a value of 507 pixels as the most appropriate cut-off criterion with a sensitivity of 1.0, a specificity to 0.8, and an area under the curve (AUC) of 0.88. The positive and negative predictive value accounted for 0.59 and 1.0.**
- **Conclusions:** The present data confirm that the new diagnostic method is valid and user-friendly that may be used for therapeutic evaluation, and for the acquisition of larger sample sizes within sophisticated study designs.



**Fig. 3** Receiver operator characteristic curve. PSG classification (SB subjects, non-SB subjects) was used as binary outcome. Pixel score of the diagnostic sheet was used as the predictor. Area under the curve (AUC) fits to 0.88. Number of positives=10, number of negatives=35. The optimal ratio between sensitivity and specificity (Youden index) is marked by a dot and corresponds to a cut-off pixel score of 507. TPR=true positive rate or sensitivity, FPR=false positive rate or 1 – specificity

## Exemple article - MD

International Journal of Dentistry. Volume 2012 (2012), Article ID 326401, 8 pages. <http://dx.doi.org/10.1155/2012/326401>

### Approximal Caries Detection by DIFOTI: In Vitro Comparison of Diagnostic Accuracy/Efficacy with Film and Digital Radiography

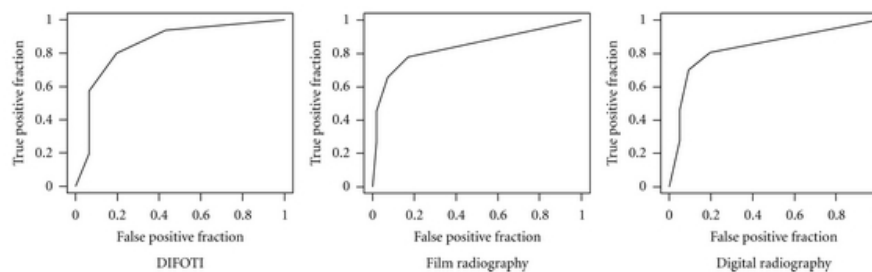
Á. Ástvaldsdóttir,<sup>1,2</sup> K. Åhlund,<sup>1</sup> W. P. Holbrook,<sup>2</sup> B. de Verdier,<sup>1</sup> and S. Tranæus<sup>1,3</sup>

The aim of the present study was to compare the diagnostic accuracy/efficacy of digital imaging fiber-optic transillumination (DIFOTI) with film and digital radiography, in detection of approximal caries lesions.

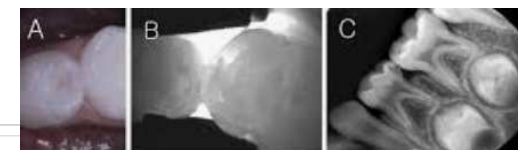
One hundred and twelve approximal surfaces were scored for caries, using DIFOTI images film and digital radiographs. All three sets of images were examined twice by 8 observers, with a minimal interval of one week between examinations. Validation of histological sections served as a reference standard. Reproducibility, based on intra- and interobserver agreement, was similar for all three methods.

At diagnostic threshold D1 (enamel and dentin caries), DIFOTI showed significantly higher sensitivity, but differences in specificity between methods were nonsignificant. Diagnostic accuracy in the form of **area under the receiver operating characteristic curve (AUC) was significantly higher for DIFOTI**. At diagnostic threshold D3 (dentin caries), the differences in sensitivity and AUC among methods were nonsignificant, but DIFOTI showed significantly lower specificity. Compared with the radiographs, DIFOTI showed closer agreement, expressed as weighted kappa values, with the reference standard.

The results show that under in vitro conditions, the diagnostic accuracy of DIFOTI in detecting early approximal enamel lesions was similar to that of film and digital radiography, while the potential for detecting lesions in dentin is similar for all three methods.



ROC curves for the three methods based on results from 8 observers at the diagnostic threshold D3 (dentin caries).



## Exemple article - MG

[Hepatology Int.](#) 2010 Oct 8;4(4):700-6. doi: 10.1007/s12072-010-9223-1. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC21286340/>

### Liver stiffness measurement in the risk assessment of hepatocellular carcinoma for patients with chronic hepatitis.

[Kuo YH<sup>1</sup>](#), [Lu SN](#), [Hung CH](#), [Kee KM](#), [Chen CH](#), [Hu TH](#), [Lee CM](#), [Changchien CS](#), [Wang JH](#).

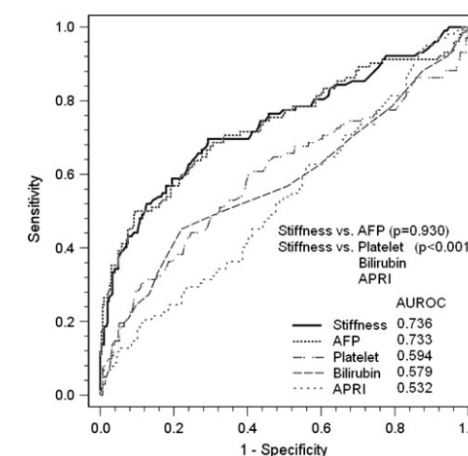
**BACKGROUND/AIMS:** [corrected] The risk of hepatocellular carcinoma (HCC) increased with progression of hepatic fibrosis as assessed by liver stiffness measurement (LSM). This study used LSM to assess the risk of HCC presence in patients with chronic hepatitis.

**METHODS:** The patients with liver tumor or chronic hepatitis indicated for biopsy were prospectively enrolled. LSM was performed on the same day as biopsy. The diagnostic performances of clinical parameters and LSM in predicting HCC presence were compared with the areas under receiver operating characteristics curves (AUROC). The risk of HCC presence was assessed with stratum-specific likelihood ratios (SSLR). The cut-off values and its diagnostic validity were calculated for LSM.

**RESULTS:** A total of 435 patients, including 106 HCC and 329 chronic hepatitis, were enrolled. **The AUROC in predicting HCC presence was 0.736, 0.733, 0.594, 0.579 and 0.532 for LSM, alpha-fetoprotein, platelet count, total bilirubin, and aspartate aminotransferase-platelet ratio index, respectively.** Multivariate analysis showed liver

stiffness was an independent factor for HCC presence (odds ratio 1.07, 95% confidence interval (CI) 1.05-1.09). SSLR for HCC presence by liver stiffness was 0.43 (95% CI 0.32-0.57) in <12 kPa, 1.28 (0.89-1.84) in 12-24 kPa, and 5.94 (3.77-9.35) in >24 kPa. With 12 and 24 kPa as the cut-offs in predicting HCC presence, the sensitivity was 69.8 and 41.5%, respectively. The specificity was 69.6 and 92.7%, respectively.

**CONCLUSIONS:** LSM identified the risk group for HCC presence in chronic hepatitis patients and had high specificity in the prediction of HCC with the cut-off of 24 kPa.



# Exemples des questions pour l'examen

**Q1.** Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,55. Lesquelles des suivantes affirmations sont correctes ?

- A. le test est de mauvaise qualité
- B. le test est de bonne qualité
- C. le test est de qualité modérée
- D. le test n'est pas statistiquement significatif parce que  $AUROC > 0,05$
- E. le courbe est très proche a la diagonale du carrée du graphique ROC

**Réponse: A, E**

**Q2.** Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,55 (95% CI 0,45 – 0,62). Lesquelles des suivantes affirmations sont correctes ?

- A. l'estimateur ponctuel du AUROC est 0,55
- B. le test statistique pour l'AUROC n'est pas statistiquement signifiant
- C. le test statistique pour l'AUROC a la valeur  $p > 0,05$
- D. le test de l'AUROC n'est pas statistiquement significatif parce que l'intervalle de confiance 95% contienne la valeur 0,5
- E. L'hypothese nule du test statistique pour l'AUROC est que l'AUROC est égale a 0,5

**Réponse: A, B, C, D, E**

# Exemples des questions pour l'examen

**Q3.** Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,95 (95% CI 0,92 – 0,99). Lesquelles des suivantes affirmations sont correctes ?

- A. la courbe est très proche du coin supérieur gauche de la courbe ROC
- B. le test statistique pour l'AUROC est statistiquement significatif
- C. le test statistique pour l'AUROC a la valeur  $p < 0,05$
- D. l'estimateur ponctuel de l'AUROC est cliniquement de qualité diagnostique modérée
- E. l'intervalle de confiance 95% de l'AUROC est cliniquement de bonne qualité diagnostique

**Réponse: A, B, C, E**

**Q4.** Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,85 (95% CI 0,82 – 0,89). Lesquelles des suivantes affirmations sont correctes ?

- a) le test statistique pour l'AUROC est statistiquement significatif
- b) le test statistique pour l'AUROC a la valeur  $p > 0,05$
- c) l'estimateur ponctuel de l'AUROC est cliniquement de qualité diagnostique modérée
- d) l'intervalle de confiance 95% de l'AUROC est cliniquement de bonne qualité diagnostique
- e) l'intervalle de confiance 95% de l'AUROC est cliniquement de qualité diagnostique modérée

**Réponse: A, C, E**

# Exemples des questions pour l'examen

**Q5.** Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,76 (95% CI 0,48 – 0,99). Lesquelles des suivantes affirmations sont correctes ?

- A. le test statistique pour l'AUROC est statistiquement significatif
- B. le test statistique pour l'AUROC a la valeur  $p > 0,05$
- C. l'estimateur ponctuel de l'AUROC est cliniquement de qualité diagnostique mauvaise
- D. l'intervalle de confiance 95% de l'AUROC cliniquement n'a pas une qualité diagnostique claire
- E. l'intervalle de confiance 95% de l'AUROC est cliniquement de qualité diagnostique modérée

**Réponse:** b, c, d



# Exemples des questions pour l'examen

**Q6.** Un étude diagnostique a montre le suivant graphique ROC avec

Trois tests diagnostiques évaluées par rapport a un test standard.

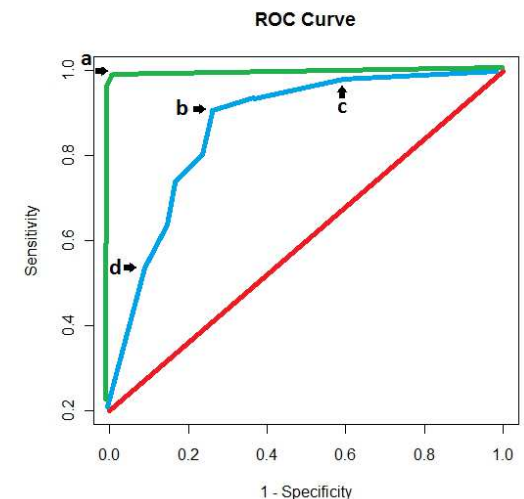
Les valeurs des tests de l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) sont: rouge=0,5, bleue=0,81, vert=0,99

Lesquelles des suivantes affirmations sont correctes ?

- A. La courbe rouge a une meilleure qualité diagnostique que la courbe bleue
- B. La courbe verte a une meilleure qualité diagnostique que la courbe bleue
- C. Le meilleur cut-off point pour la courbe bleue est le point b, entre b, c, d sur la courbe bleue
- D. Le cut-off d a une spécificité plus grand, entre b, c, d sur la courbe bleue
- E. Le cut-off c a une sensibilité plus grand, entre b, c, d sur la courbe bleue
- F. Le AUROC vert cliniquement indique une valeur diagnostique très bonne
- G. Le AUROC rouge cliniquement indique une valeur diagnostique absente
- H. Le AUROC bleue cliniquement indique une valeur diagnostique modérée

**Réponse: B, C, D, E, F, G, H**

**Explications: voir la diapo suivante**



# Exemples des questions pour l'examen

## Explications:

Plus les courbes ROC sont proches a la point supérieur gauche, plus leur qualité diagnostique est importante. Aussi, plus la AUC (AUROC), est proche a 1, plus la qualité diagnostique est importante. Donc la courbe vert a une valeur plus bonne que la courbe bleue, qui a une valeur plus bonne que la courbe rouge. Donc réponse a est incorrecte, réponse b) est correcte

La meilleur valeur seuil (cut-off point) est celui qui est plus proche au coin supérieur gauche du carrée, parce que dans cette position on a le maximum de la somme entre sensibilité et spécificité. Le point b est dans cette position. Donc le réponse c) est correcte

Si on regarde sur l'axe horizontale de gauche a droite, la première valeur est celui qui correspond a le point a, après d, après b, après c. L'axe horizontale représente 1-spécificite. Donc sur l'axe horizontale, les plus grands valeurs de la spécificité sont proche a 0 (pour une spécificité de  $1 - 100\%$  parce que  $1-1=0$ ), et les plus petites valeurs sont proche a 1 ( $1-0=1$ ). Ca signifie que le point d a une meilleur spécificité que b, et c => réponse d) est correcte.

Si on regarde sur l'axe verticale depuis haut en bas, la première valeur est celui qui correspond a le point a, après c, après b, après d. L'axe verticale représente la sensibilité. Donc sur l'axe verticale, les plus grands valeurs de la sensibilité sont proche a 1 (en haut), et les plus petites valeurs sont proche a 0 (en bas). Ca signifie que le point c a une meilleur spécificité que b, et d => réponse e) est correcte.

La valeur clinique d'un test diagnostique en fonction de la AUC est la suivante  $\geq 0,9$  – bonne qualité,  $\geq 0,8$  et  $< 0,9$  – qualité modérée;  $< 0,8$  – qualité réduite. Une AUC de 0 indique l'absence de la valeur clinique. Les valeurs très proches a 1 indique une qualité très bonne. Donc les réponses f), g), h) sont correctes;

Fin

