

---

Description d'un phénomène de santé

Evaluation d'une méthode diagnostique

---

1

1

---

## Plan du cours

- Domaines de recherche scientifique
    - Évaluation d'une méthode diagnostique
      - Phases
        - I: validité, fiabilité, K, ICC
        - II: intrinsèque: Se, Sp, LR+, LR-
          - Courbe ROC
        - III: extrinsèque: VPP, VPN
        - IV: cout – bénéfice
      - Bland Altman plot
      - Evaluation de la validité des tests diagnostiques
  - Exemples des études
  - Exemples des exercices
- 

2

2

## 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**
  - identifier la gravité/stade d'un maladie afin de **choisir le meilleur traitement**
- **Exemples** des méthodes diagnostiques: signes (fièvre), symptômes (douleur des dents), imagerie médicale (radiographie du crane, tomodensitométrie, radiographie dentaire), examen anatomopathologique, biochimie du sang/ urine... (glycémie, glycosurie, ...),

3

3

## 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

4

4

## Phase I:

Le **but** - de déterminer:

- **La Validité** de la méthode (qualité technique)
- **La Fiabilité** de la méthode (**ex.** la lecture d'une radiographie du crane):
  - **Reproductibilité** = le degré dont des sujets observées peuvent être distinguées entre eux malgré l'erreur de mesure
  - **Reproductibilité inter-observationnelle/inter-opérateur**
    - Définition: interprétation similaire par deux observateurs différents
  - **Reproductibilité intra-observationnelle/intra-opérateur**
    - Définition: interprétation similaire du même observateur, à différentes lectures
  - **On mesure la reproductibilité** en fonction des résultats possible des tests:
    - Si le test a deux catégories: positive/négative
      - => on le mesure avec le coefficient Kappa
    - Si le test offre un résultat quantitative
      - => on le mesure avec le coefficient de corrélation interclasse (ICC)

Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012.

5

5

## Phase I:

Le **coefficient/test Kappa de Cohen** de reproductibilité

- $>0,8$  – parfait reproductibilité
- $0,6 - 0,8$  - bonne
- $0,4 - 0,6$  - modérée
- $0,2 - 0,4$  – réduite
- $0 - 0,2$  – presque absente
- $< 0$  – mauvaise reproductibilité

Tab. II Concordance et reproductibilité de différentes méthodes utilisées pour le diagnostic de la carie des fissures, sur des dents saines et des dents cariées (Lussi 1993, Lussi et coll. 1995, Lussi 1996)

	Concordance entre praticiens		Reproductibilité	
	sans cavité	avec cavité	sans cavité	avec cavité
Examen visuel	18% <sup>1</sup>	61% <sup>2</sup>	49% <sup>3</sup>	51% <sup>4</sup>
Examen visuel plus sonde	24% <sup>1</sup>	45% <sup>2</sup>	–	–
Examen visuel avec loupe	18% <sup>1</sup>	39% <sup>2</sup>	–	–
Radiographies «bite-wing»	45% <sup>1</sup>	84% <sup>2</sup>	55% <sup>3</sup>	67% <sup>4</sup>
Examen visuel plus «bite-wing»	46% <sup>1</sup>	85% <sup>2</sup>	–	–
Résistance électrique	95% <sup>1</sup>	–	85% <sup>3</sup>	–

[http://www.swissdentaljournal.org/fileadmin/upload\\_sso/2\\_Zahnaerzte/2\\_SDJ/SMIZ\\_1998/SMIZ\\_04\\_1998/1-smfz-98-04-bildun](http://www.swissdentaljournal.org/fileadmin/upload_sso/2_Zahnaerzte/2_SDJ/SMIZ_1998/SMIZ_04_1998/1-smfz-98-04-bildun)

Le **coefficient de corrélation interclasse** (en anglais interclass correlation coefficient - **ICC**)

- **Reproductibilité inter-observationnelle/inter-opérateur**
- **Reproductibilité intra-observationnelle/intra-opérateur**
- La reproductibilité est réduite si  $<0,50$ ; modérée si  $[0,50 - 0,75]$ ; bonne si  $[0,75 - 0,90]$ ; excellente si  $>=0,90$

6

6

## Phase II: qualités intrinsèques du test

- Le test est appliqué chez des patients pour lesquels le **statut malade ou non-malade** est clairement précisé avec un test de référence = "Norme d'or" ou "Gold standard", ou « Standard d'or".

		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

Berlin-based Pharmact, rapid test - 20-minute immunoassay containing three SARS-CoV-2 antigens

The test scored highly in terms of specificity. "We had a true negative rate of 100% — zero false positives," says Gunther Burgard, medical director at Pharmact. Its sensitivity was lower, however, as the IgM response does not offer a strong initial signal. During the early stage of the infection (days 4–10), the IgM component of the test provides a sensitivity of just 70%. This rises rapidly to 92.3% between days 11 and 24, and the IgG component of the test offers a sensitivity of 98.6% during this phase of the infection

- Les résultats sont généralisables à la population seulement si il est représentatif à la population cible

Al T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Shi 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;295:200942. doi: 10.1148/radiol.2020200942

7

7

## Phase II: qualités intrinsèques du test

- Le test est appliqué chez des patients pour lesquels le **statut malade ou non-malade** est clairement précisé avec un test de référence = "Norme d'or" ou "Gold standard", ou « Standard d'or".

		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

Tab. 1 Spécificités et sensibilités respectives de différentes méthodes utilisées pour le diagnostic de la carie occlusale (Lussi 1993, Lussi et coll. 1995, Lussi 1996)

Dents exemptes de carie ou avec carie de l'émail		Spécificité (%)	
Examen visuel	(26 praticiens)	93	
Examen visuel plus sonde	(23 praticiens)	93	
Examen visuel avec loupe	(26 praticiens)	89	
Radiographies «bite-wing»	(24 praticiens)	83	
Examen visuel plus «bite-wing»	(10 praticiens)	87	
Résistance électrique		77	

Dents avec carie ayant atteint la dentine		Sensibilité (%)	
		sans cavité	avec cavité
Examen visuel	(26 praticiens)	12	62
Examen visuel plus sonde	(23 praticiens)	14	82
Examen visuel avec loupe	(26 praticiens)	20	75
Radiographies «bite-wing»	(24 praticiens)	45	79
Examen visuel plus «bite-wing»	(10 praticiens)	49	90
Résistance électrique		93	-

- Les résultats sont généralisables à la population seulement si il est représentatif à la population cible

[https://www.swissdentaljournal.org/fileadmin/upload\\_sso2\\_Zahnaerzte2\\_SDJ/SMZ\\_1998/SMZ\\_04\\_1998/f-smfz-98-04-bildung2.pdf](https://www.swissdentaljournal.org/fileadmin/upload_sso2_Zahnaerzte2_SDJ/SMZ_1998/SMZ_04_1998/f-smfz-98-04-bildung2.pdf)

8

8

## Phase II: qualités intrinsèques du test

- Le test est appliqué chez des patients pour lesquels le **statut malade ou non-malade** est clairement précisé avec un test de référence = "Norme d'or" ou "**Gold standard**", ou « **Standard d'or** ».

		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

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 [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
Overall									
580	105	308	21		97 (580/601) [95-98]	25 (105/413) [22-30]	65 (580/888) [62-68]	83 (105/126) [76-89]	68 (685/1014) [65-70]

- Les résultats sont généralisables à la population seulement si il est représentatif à la population cible

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;295:200942. doi: 10.1148/radiol.2020200942

9

9

## Indicateurs évaluable

		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

10

10

## Indicateurs évaluable

		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

11

11

## Phase II: Indicateurs

Definitions, formules et interpretations statistique de la Sensibilite, specificite:

- **Sensibilité (Se)** - anglais – **sensitivity**

- Probabilité d'un sujet malade d'avoir un test positif
- Probabilité d'un sujet d'avoir un test positif sachant qu'il est malade
  - Ex. La probabilité d'un sujet d'avoir un test de coloration avec toluidine positif sachant qu'il a un cancer de lèvre est de 78%
  - Ex. La probabilité d'un sujet d'avoir un test Covid 19 rapide positif sachant qu'il a réellement Covid 19 est de 70%
- Proportion de sujets qui ont été testés positifs parmi les malades

$$Se = VP / (VP + FN)$$

- **Spécificité (Sp)** - anglais – **specificity**

- Probabilité d'un sujet non-malade d'avoir un test négatif
- Probabilité d'un sujet d'avoir un test négatif sachant qu'il est non-malade
  - Ex. La probabilité d'un sujet d'avoir un test de coloration avec toluidine négatif sachant qu'il n'a pas un cancer de lèvre est de 89%
  - Ex. La probabilité d'un sujet d'avoir un test Covid 19 rapide négatif sachant qu'il n'a pas réellement Covid 19 est de 98%

- ~~Proportion de sujets ayant des résultats négatifs chez les non-malades~~

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

12

12

## Phase II:

- Haute sensibilité → élimine un diagnostic (SnOut = Sensibility rules out the disease – la sensibilité élimine la maladie)
  - Se = 100% + test négatif → certainement le patient n'a pas la maladie
- Haute spécificité → confirme un diagnostic (SpIn = Specificity rules in the disease – la spécificité confirme la maladie)
  - Sp = 100% + test positif → le patient a certainement la maladie

13

13

## Phase II:

Indicateurs globales de qualité diagnostique:

- **Valeur globale du test (anglais – accuracy) =  $(VP + VN) / (VP + FP + VN + FN)$** 
  - (proportion des sujets bien classées)
  - De 0% (absente) - 100% (très bien)
  - Le pourcentage des vrai réponses du test
- **Indice Youden =  $Se + Sp - 1$** 
  - De 0 (absente) – 1 (très bien)
- Si ils ne sont proche a 100% ou 1 on ne sait pas si la Se ou la Sp est réduite ou les deux
- On recommande de regarder Se et Sp plus que précision ou l' indice Youden

14

14

## Phase II:

La sensibilité et la spécificité **ne dépendent pas** de la **prévalence** de la maladie dans l'échantillon étudié. Par conséquent, Se et Sp auront les mêmes valeurs, n'importe quelle soit la population à laquelle le test est appliqué.

- Un **test idéal** a la **Se et Sp = 100%**, ce qui est rare dans la pratique:
  - Plus haute la Se, plus basse la Sp, et vice-versa.

15

15

## Phase II:

- **Interprétation clinique de la sensibilité, spécificité:**
  - Il n'y a pas une règle universelle. Ca dépend de la maladie à évaluer, de la gravité, des possibilités d'être traitée
  - Un indicateur de diagnostic pour Se, Sp, précision/Le pourcentage des biens classées (accuracy en anglais)
    - proche de 100% signifie un diagnostic de haute qualité
    - <80% - mauvaise qualité,
    - de >=80-90% - modéré
    - >= 90% bonne qualité - mais dans certains cas, il doit prendre des valeurs proches de 100% - soit testé pour le VIH
  - (p. ex Se = 95% signifie que le nouveau test de diagnostic a une sensibilité élevée).

16

16

## Phase II:

Définition, formules, interprétation statistique du likelihood ratio – rapport de vraisemblance - RVP, RVN +LR, -LR

- Rapport de vraisemblance positif (RVP) ou **(+LR)=Se/(1-Sp)**
  - - anglais – **+LR – positive likelihood ratio**
  - combien de fois un **résultat positif du test** est plus susceptible d'apparaître chez un sujet malade que chez un non-malade.
  - $+LR = \frac{a}{\frac{a+c}{b+d}} \sim$  il est similaire a le Risque Relatif sur les colonnes
- Rapport de vraisemblance négatif (RVN) ou **(-LR)=(1-Se)/Sp**
  - - anglais – **-LR – negative likelihood ratio**
  - combien de fois un **résultat négatif du test** est plus susceptible d'apparaître chez un sujet malade que chez un non-malade.
  - $-LR = \frac{c}{\frac{a+c}{b+d}} \sim$  il est similaire a le Risque Relatif sur les colonnes inversée

17

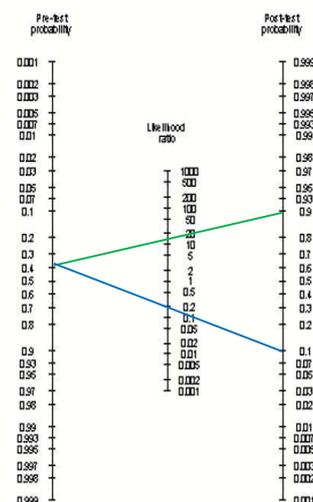
17

## Probabilité pré-test / post-test d'avoir la maladie

### Nomogramme Fagan:

#### Utilisation:

- On fixe sur l'axe pré-test (avant le test) la probabilité avant faire une test
- On fait le test, et si le résultat du test est positif on utilise LR +  
négative on utilise LR –
- On met un point avec la valeur du LR sur la deuxième axe (likelihood ratio)
- On dessine une ligne entre les deux points vers la troisième axe et on obtient la probabilité après le test



Ex. P avant=40%, test + => LR+=13 => P après = 90%

P avant=40%, test - => LR-=0.2 => P après = 10%

18

18

## Probabilité pré-test / post-test d'avoir la maladie

- **Interprétation clinique des rapports de vraisemblance:**
  - certaines valeurs du LR peuvent entraîner des changements plus ou moins importante de la probabilité d'avoir une maladie. Ainsi les valeurs LR dans les gammes:
    - • LR > 10 ou < 0,1 entraînent des changements
      - importants et souvent concluante de la probabilité d'un diagnostic
    - • LR entre 5 et 10 ou entre 0,1 (=1/10) et 0,2 (=1/0,5) entraînent des changements
      - modéré et utile de la probabilité d'un diagnostic
    - • LR entre 2 et 5 ou entre 0,5(=1/2) et 0,2(=1/5) entraînent des changements
      - réduits/faible mais parfois concluante/cliniquement significatif de la probabilité d'un diagnostic
    - • LR entre 1 et 2 ou entre 0,5 et 1 conduit à des modifications
      - de faible probabilité, et rarement cliniquement important.
  - LR = 1 – pas de changement
  - Donc plus le LR+ est grand plus il permet de confirmer la maladie
  - Plus le LR- est petit, plus il permet de exclure la maladie

19

19

## Pour des tests diagnostiques qui ont une résultat quantitative (non dichotomique)

- On utilisent les **courbes ROC (Receiver Operator Characteristic)**
  - Pour trouver la valeur seuil (cut-off – en anglais) 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 capacité diagnostiques statistiquement différentes ou pas)

**Définition de l'aire en dessous la courbe** (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 qui a comme résultat une valeur quantitative.

20

20

## Pour des tests diagnostiques qui ont une résultat quantitative

### ■ La méthode de construction de la courbe ROC

Chaque valeur numérique observe 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 **une tableau de contingence**, on **calcule Se, Sp**, et on **met un point** sur le graphique ROC.

On **répète l'opération** pour toutes les valeurs observe 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

Hirotsu Y, Maejima M, Shibusawa M, Nagakubo Y, Hosaka K, Amemiya K, et al. Comparison of automated SARS-CoV-2 antigen test for COVID-19 infection with quantitative RT-PCR using 313 nasopharyngeal swabs, including from seven serially followed patients. Int J Infect Dis. 2020 Oct;99:397-402.

21

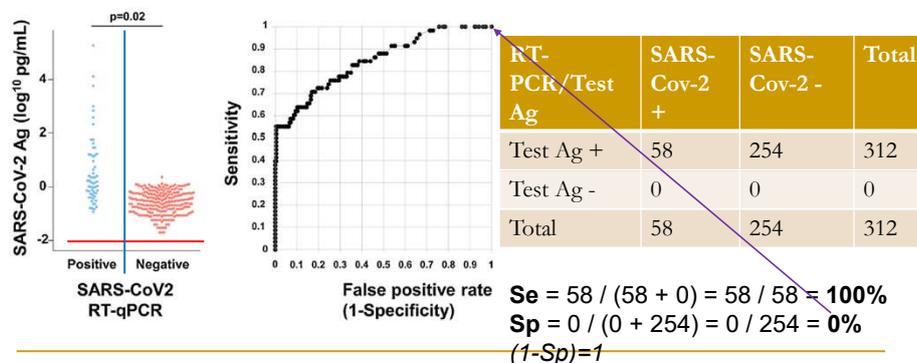
21

## Pour des tests diagnostiques qui ont une résultat quantitative

### ■ La méthode de construction de la courbe ROC

Valeur seuil du test antigène = -2:

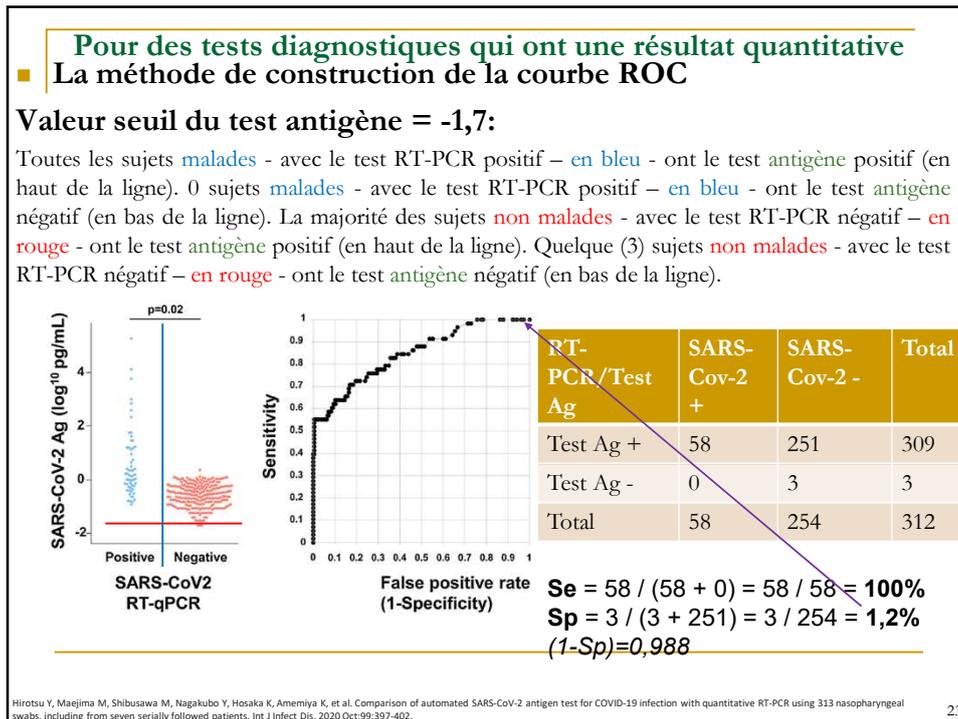
Toutes les sujets **malades** - avec le test RT-PCR positif – **en bleu** - ont le test **antigène** positif (en haut de la ligne). 0 sujets **malades** - avec le test RT-PCR positif – **en bleu** - ont le test **antigène** négatif (en bas de la ligne). Toutes les sujets **non malades** - avec le test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



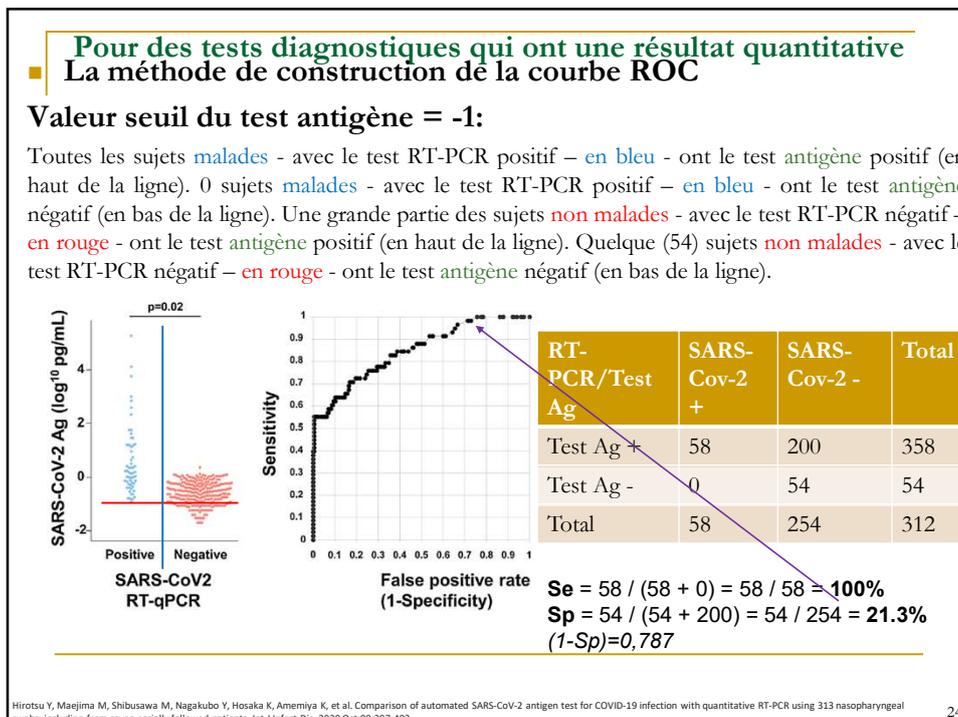
Hirotsu Y, Maejima M, Shibusawa M, Nagakubo Y, Hosaka K, Amemiya K, et al. Comparison of automated SARS-CoV-2 antigen test for COVID-19 infection with quantitative RT-PCR using 313 nasopharyngeal swabs, including from seven serially followed patients. Int J Infect Dis. 2020 Oct;99:397-402.

22

22



23



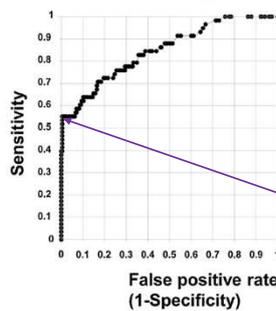
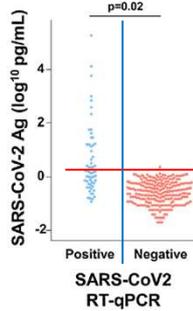
24

## Pour des tests diagnostiques qui ont une résultat quantitative

### ■ La méthode de construction de la courbe ROC

Valeur seuil du test antigène = 0,3:

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



RT-PCR/Test Ag	SARS-Cov-2 +	SARS-Cov-2 -	Total
Test Ag +	26	1	27
Test Ag -	32	253	285
Total	58	254	312

$$Se = 32 / (26 + 32) = 58 / 58 = 55,2\%$$

$$Sp = 54 / (54 + 200) = 54 / 254 = 99,8\%$$

$$(1-Sp)=0,012$$

Hirotsu Y, Maejima M, Shibusawa M, Nagakubo Y, Hosaka K, Amemiya K, et al. Comparison of automated SARS-CoV-2 antigen test for COVID-19 infection with quantitative RT-PCR using 313 nasopharyngeal swabs, including from seven serially followed patients. Int J Infect Dis. 2020 Oct;99:397-402.

25

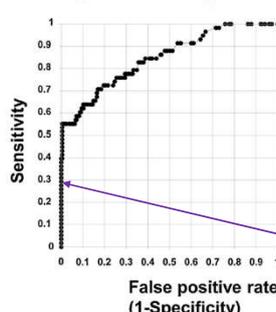
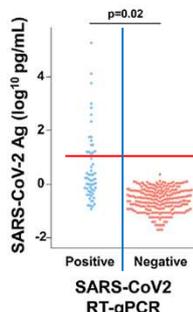
25

## Pour des tests diagnostiques qui ont une résultat quantitative

### ■ La méthode de construction de la courbe ROC

Valeur seuil du test antigène = 1:

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



RT-PCR/Test Ag	SARS-Cov-2 +	SARS-Cov-2 -	Total
Test Ag +	17	0	17
Test Ag -	41	254	295
Total	58	254	312

$$Se = 17 / (17 + 41) = 17 / 58 = 29,3\%$$

$$Sp = 54 / (54 + 200) = 254 / 254 = 100\%$$

$$(1-Sp)=0,0$$

Hirotsu Y, Maejima M, Shibusawa M, Nagakubo Y, Hosaka K, Amemiya K, et al. Comparison of automated SARS-CoV-2 antigen test for COVID-19 infection with quantitative RT-PCR using 313 nasopharyngeal swabs, including from seven serially followed patients. Int J Infect Dis. 2020 Oct;99:397-402.

26

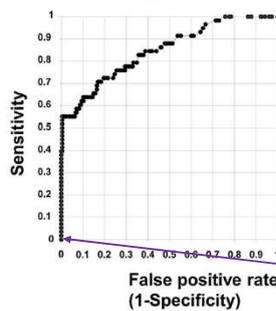
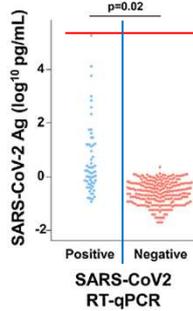
26

## Pour des tests diagnostiques qui ont une résultat quantitative

### ■ La méthode de construction de la courbe ROC

Valeur seuil du test antigène = 4,8:

Tous les sujets **malades** - avec le test RT-PCR positif – **en bleu** - ont le test antigène positif (en haut de la ligne). 0 sujets **malades** - avec le test RT-PCR positif – **en bleu** - ont le test antigène négatif (en bas de la ligne). 0 sujets **non malades** - avec le test RT-PCR négatif – **en rouge** - ont le test antigène positif (en haut de la ligne). Tous les sujets **non malades** - avec le test RT-PCR négatif – **en rouge** - ont le test antigène 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	254	312
Total	58	254	312

$$Se = 0 / (0 + 58) = 0 / 58 = 0\%$$

$$Sp = 254 / (0 + 254) = 254 / 254 = 100\%$$

$$(1-Sp)=0,0$$

Hirotsu Y, Maejima M, Shibusawa M, Nagakubo Y, Hosaka K, Amemiya K, et al. Comparison of automated SARS-CoV-2 antigen test for COVID-19 infection with quantitative RT-PCR using 313 nasopharyngeal swabs, including from seven serially followed patients. Int J Infect Dis. 2020 Oct;99:397-402.

27

27

## Pour des tests diagnostiques qui ont une résultat quantitative (non dichotomique)

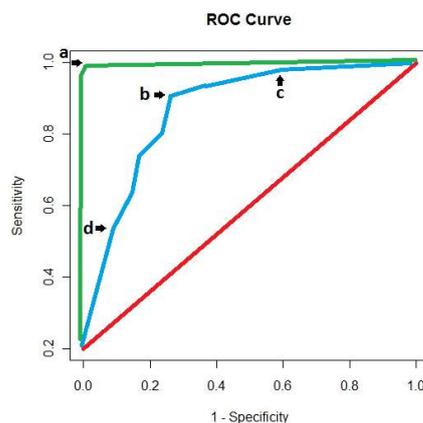
Le graphique ROC - Receiver Operator Characteristic

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

**Caractéristiques**

- axe horizontale = 1 – 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
  - **a le coin supérieur gauche** – valeur diagnostique élevée – test discriminant (voir la courbe verte) – AUROC/AUC proche à 1
  - **a 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



28

28

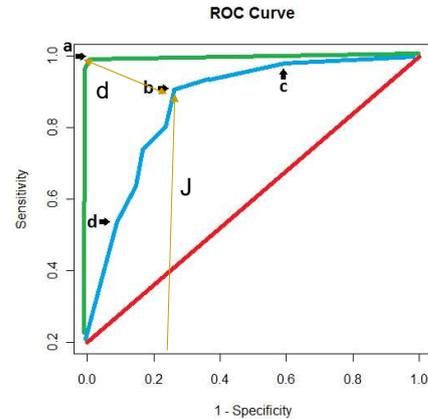
## Pour des tests diagnostiques qui ont une résultat quantitative (non dichotomique)

Le graphique ROC - Receiver Operator Characteristic

La chois de la valeur seuil (cut-off)

Le but du chois est pour avoir la meilleur combinaison de Se et Sp élevée

- En tenant compte du but de l'étude (voir l'exemple pour la courbe bleu)
  - **Diagnostic 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 a le coin supérieur gauche - 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  $j = \max(\text{indice Youden})$ .  
Le point trouvée, est le point avec la plus petite distance (d) vers le point (0,1) – supérieur gauche.



29

29

## Pour des tests qui ont une résultat quantitative (non dichotomique)

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** voir s'il est statistiquement significative.
  - L' hypothèse nulle du test statistique pour l'AUROC est que l'AUROC est égale a 0,5
  - L' hypothèse alternative du test statistique pour l'AUROC est que l'AUROC n'est pas égale a 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 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 **contienne 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 contienne 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**

30

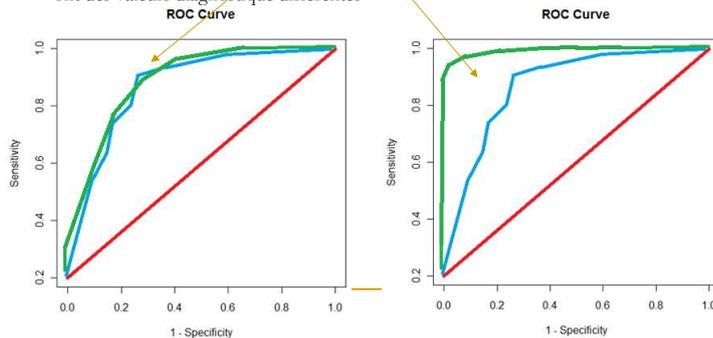
30

## Pour des tests diagnostiques qui ont une résultat quantitative

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

- La comparaison peut être faite initiale d'une manière graphique.
  - La courbe qui est **plus proche** a le 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



31

31

## Pour des tests diagnostiques qui ont une résultat quantitative

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

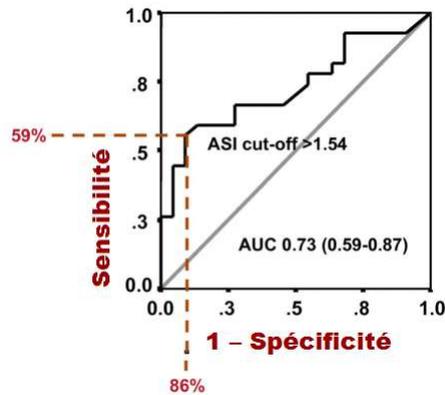
- Ensuite, certaines courbes peuvent avoir des valeur seuils, qui ont un certaine avantage par rapport a les autres
- Enfin, il y a des **tests statistiques** qui nous permettent de vérifier si la **différence est statistiquement significative**.
  - Les hypothèses statistiques du test:
    - L'hypothèse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

32

32

## Pour des tests qui ont un résultat quantitative (non dichotomique)

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



Lehrke S, Egenlauf B, Steen H, Lossnitzer D, Korosoglou G, Merten C, et al. Prediction of coronary artery disease by a systemic atherosclerosis score index derived from whole-body MR angiography. *Journal of Cardiovascular Magnetic Resonance* 2009; 11

33

33

## Pour des tests qui ont un résultat quantitative (non dichotomique)

- Interprétation clinique de AUC / AUROC, pour un graphique ROC - Receiver Operator Characteristic
  - Comme pour Se, Sp, précision (des suggestions approximatives d'interprétation clinique):
    - proche de 100% signifie un diagnostic de haute qualité – plus le test est discriminant
    - <0.8 - mauvaise qualité,
    - de  $\geq 0.8$  - 0.9 - modéré
    - $\geq 0.9$  - bonne qualité
    - 0.5 - sans aucune utilité diagnostique – le test n'est pas discriminant

34

34

## 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		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

35

35

## Phase III: Indicateurs

Definitions, formules 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 qui a été testé positif
  - Ex. la probabilité pour un sujet d'avoir cancer de langue sachant qui a été testé positif a le test de coloration avec toluidine est de 78%
  - Ex. la probabilité pour un sujet d'avoir Covid 19 sachant qui a été testé positif a le test de rapide pour Covid 19 est de 78%
- proportion des sujets réellement malades parmi les sujets qui ont été testés positifs

$$VPP = VP / (VP + FP)$$

Valeur prédictive négative (VPN)

- en anglais – positive predictive value - PPV
- 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 qui a été testé négatif
  - Ex. la probabilité pour un sujet de n'avoir pas cancer de langue sachant qui a été testé négatif a le test de coloration avec toluidine est de 89%
  - Ex. la probabilité pour un sujet de n'avoir pas Covid 19 sachant qui a été testé négatif a le test de rapide pour Covid 19 est de 93%
- proportion des sujets vraiment non-malades parmi les personnes testées négatives

$$VPN = VN / (VN + FN)$$

36

36

## 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.

37

37

## Phase III:

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

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

On ne doit pas apprendre les deux formules

38

38

## 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

39

39

## 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)

40

40

## 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 ou les patients ont été **recrute** - la location, le cadre, ou ont été réalise les deux tests diagnostiques – essentiel pour évaluer la generalisabilite des résultats

■ **Recrutement** (en fonction des symptômes, résultats d'autres tests ou si le sujet a reçu le nouveau test ou le test standard)

41

41

## Phase II et III

### ■ Méthodologie d'étude diagnostique (pour phase II et III)

#### □ Sujets

##### ■ Sélection des sujets:

□ **échantillon représentatif** – meilleure / difficile a faire

□ **série consécutif des patients** – l'un des meilleure si on ne peut pas faire représentatif

- Les deux permet que toutes les situations sont prise en compte (malades, non malades, différentes 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 appliqué les deux tests diagnostiques

■ idéale

□ Collection **retrospective** des données enregistrées dans les feuilles d'observation des hôpitaux

42

42

## 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 invasif (moins des risques)**, moins chère, plus simple, mieux acceptée par le patient - en principe
    - **Les tests doivent être reproductibles** (intra/ inter observationnel – on doit préciser le coefficient Kappa, ICC 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**

43

43

## 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 **évalués avec les deux tests** diagnostiques (si un **grand nombre** des sujets n'est pas évalués avec un test ça peut poser des problèmes. Si l'étude est prospective et la perte de vue est importante ça peut poser des problèmes - biais d'attrition)
  - Le **temps passé entre les deux tests** (doit être **précises** avec date du début 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

44

44

## 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
  - âge, 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

45

45

## 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, Zhai C, Chen C, Li W, Tan 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;295(2):200542. doi: 10.1148/radiol.2020200542

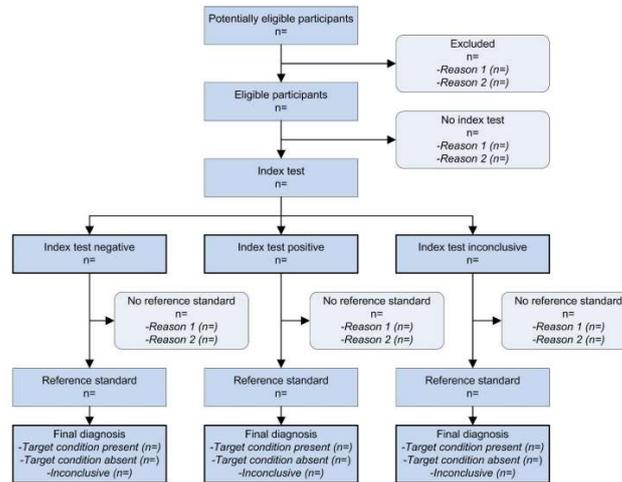
46

46

## Phase II et III

### ■ Méthodologie d'étude diagnostique (pour phase II et III)

- Les résultats doit préciser:
  - dans le texte ou diagramme du flux (en: flow-chart)



<http://www.equator-network.org/wp-content/uploads/2015/03/STARD-2015-flow-diagram.pdf>

47

47

## 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]	[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]	

Al T, Yang Z, Nour H, Zhai C, Chen C, Lu 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;20200442. doi: 10.1148/radiol.2020200442

48

48

## Phase IV:

- **Évaluation 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?

49

49

## 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 (fréquence des toutes les malades) de la maladie dans la population cible – pour échantillon représentatif/série consécutive des cas – dépends ...

50

50

## 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

51

51

## Évaluation d'une méthode diagnostique

### Collection des données - cas-témoins

(approche sémiologique)

- 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

52

52

## Le choix des tests en fonction de Se et Sp

- Un **test sensible** doit être préférée pour:
  - Pour des **maladies graves** mais **traitables**
  - Si le résultats sont faux positifs les répercussions psychologiques et financières sur les sujets ne sont pas très graves
  - Ex. maladies infectieuses, a transmission sexuelle.
- Un **test spécifique** doit être préférée pour:
  - Maladie grave **non traitable** ou **difficile a traiter**
  - Si on connait le fait qu'on n'a pas la maladie a une utilité psychologique et sanitaire
  - Si le résultats sont faux positifs les répercussions psychologiques et financières sur les sujets sont très importantes
  - Ex. des cancers

53

53

## Le choix des tests en fonction de VPP et VPN

- Un **test avec bon VPP** doit être préférée pour:
  - Si le résultats sont faux positifs les traitement peut avoir des conséquences graves
  - Ex. radiothérapie/ chimiothérapie pour une cancer mal diagnostique
- Un **test avec un bon VPN** doit être préférée pour:
  - Maladie **grave** mais bien **traitable**
  - Si on connait le fait qu'on n'a pas la maladie a une utilité psychologique et sanitaire
  - Si le résultats faux positifs et faux négatifs ont les répercussions psychologiques et financières sur les sujets sont très importantes
  - Ex. certaines formes des cancers bien traitables, maladies cardiovasculaires graves traitables.

54

54

## Evaluation de la concordance entre les mesures

### ■ L'évaluation de la concordance entre les mesures

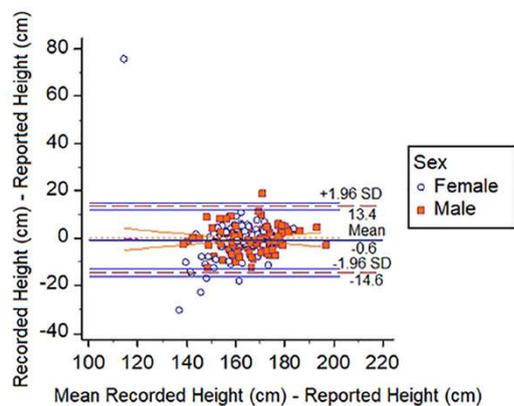
- Les indicateurs de concordance
  - représentent des mesures du l'instrument de mesure et ne dépend pas de la variabilité entre les sujets a mesurer.
- Utilité:
  - peuvent évaluer à quel point deux techniques de mesures différents sont rapprochées
  - peuvent évaluer à quel point les mesures répétées sont rapprochées
    - peuvent être utilisée pour évaluer par exemple si l'effet du traitement (le critère de jugement) a changée réellement ou non

55

55

## Evaluation de la concordance entre les mesures

### ■ La diagramme Bland Altman



Gaskin PS, Chami P, Nancoo T, Warner P, Barrett P, Mayers V. Electronic based reported anthropometry-A useful tool for interim monitoring of obesity prevalence in developing states. PLoS One. 2020 Dec 7;15(12):e0243202. doi: 10.1371/journal.pone.0243202.

Fig 1. Bland-Altman plot showing recorded and reported height vs. mean recorded and reported height. Ninety-nine percent (99%) of all height measures fell within the 95% CI.

56

56

## Evaluation de la précision/fiabilité des mesures faits dans les laboratoires

- Quant on est intéressée si la technique pour mesurer la concentration d'une substance dans le sang/sérum (e.g. la technique ELISA de mesure) a une bonne précision, on doit regarder le coefficient de variance (CV = déviation standard/moyenne \*100).
  - Les valeurs < 10% sont considérées comme bonne précision

57

57

## Evaluation de la validité des études diagnostiques

1. Le **spectre** des patients était-il **représentatif**? (échantillon consécutive ou représentative - choisi par hasard)
2. La **test de référence acceptable** (identifie bien la présence/absence de la maladie)
3. La période de **temps** entre le test de référence et le test d'index est-elle **suffisamment courte** pour être raisonnablement sûr que la condition cible n'a pas changé 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)

58

58

## Evaluation de la validité des études diagnostiques

QUADAS 2 – partie qui évalue le risque de biais (sans l' applicabilité) <https://www.bristol.ac.uk/media-library/sites/quadas/migrated/documents/quadas2.pdf> + <https://www.bristol.ac.uk/media-library/sites/quadas/migrated/documents/background-doc.pdf>

### Domain 1: Sélection des patients

- Un échantillon consécutif ou aléatoire de patients a-t-il été inscrit ? (Oui/Non/Pas clair)
- Une design cas-témoins a-t-elle été évitée ?
- L'étude a-t-elle évité les exclusions inappropriées ?

### Domain 2: Test diagnostique d' intérêt/nouveau test (index test – en anglais)

- Les résultats des tests index ont-ils été interprétés sans connaissance des résultats de test de référence ? (en aveugle par rapport a le test standard)
- Si un seuil a été utilisé, a-t-il été pré-spécifié ?

### Domain 3: Test diagnostique standard d'or / de référence

- Le test standard est-elle susceptible de classer correctement la maladie d' intérêt? (bon test standard?)(identifie bien la présence/absence de la maladie)
- Les résultats des tests standard ont-ils été interprétés sans connaissance des résultats de test de référence ? (en aveugle par rapport a le test d' intérêt?)

### Domain 4: Flux et temps

- Y avait-il un intervalle approprié (court) de temps entre le(s) test(s) d' intérêt et le test standard pour être raisonnablement sûr que la maladie n'a pas changé entre les deux tests?
- Tous les patients ont-ils reçu le test de référence ?
- Les patients ont-ils reçu la même test de référence ?
- Tous les patients ont-ils été inclus dans l'analyse ?
- Les résultats des tests non interprétables / intermédiaires ont-ils été rapportés?

59

59

## Exemple article - MD

Br Dent J. 2001 Apr 28;190(8):440-3. Available from: <http://www.nature.com/bdj/journal/v190/n8/full/4800998a.html>

**Occlusal caries detection in primary teeth: a comparison of DIAGNOdent with conventional methods.**

Attrill DC<sup>1</sup>, Ashley PF.

**OBJECTIVE:** To compare the accuracy and repeatability of three diagnostic systems (DIAGNOdent, visual and radiographic) for occlusal caries diagnosis in primary molars.

**DESIGN:** Two examiner, in vitro, blinded study. Histological gold standard.

**MATERIALS AND METHODS:** 58 occlusal surfaces of primary molars examined in turn by two examiners using each of three diagnostic systems (DIAGNOdent, visual and radiographic). These results were compared with a histological gold standard. Sensitivity and specificity were calculated for each diagnostic system at a range of thresholds. Inter- and intra- examiner repeatability were calculated for each diagnostic system using the kappa statistic.

**MAIN OUTCOME MEASURES:** Sensitivity, specificity, inter and intra diagnostic system.

**RESULTS:** The highest sensitivity values were provided by DIAGNOdent (0.77 and 0.80, examiners 1 and 2 respectively) however this was offset by a lower specificity (0.82 and 0.85) than all other systems with the exception of examiner 1 at V1 visual threshold. The DIAGNOdent gave the highest values of kappa for intra- and inter-examiner repeatability with the exception of intra-examiner repeatability for examiner 2 where visual diagnosis had the highest value of kappa.

**CONCLUSION:** The DIAGNOdent was the most accurate system tested for the detection of occlusal dentine caries in primary molars. The performance of the DIAGNOdent systems was not statistically significantly better than that achieved using visual examination for non-cavitated teeth (V1 and V2 threshold). DIAGNOdent may prove useful as a predictive clinical tool, however with appropriate training visual examination may offer similar results without the need for additional equipment.

60

60

### 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

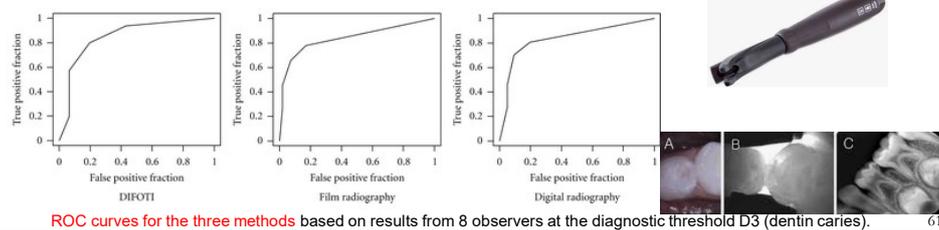
Á. Ástvaldsdóttir,<sup>1,2</sup> K. Áhlund,<sup>1</sup> W. P. Holbrook,<sup>2</sup> B. de Verdier,<sup>1</sup> and S. Tranæus<sup>1,3</sup>

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 is greater than that of film and digital radiography, while the potential for detecting lesions in dentin is similar for all three



61

### Exemple article - MG

BMC Res Notes. 2011 Mar 24;4:79. doi: 10.1186/1756-0500-4-79. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC21429231/>

#### MPT 64 Antigen detection for Rapid confirmation of M.tuberculosis isolates.

Kumar VG<sup>1</sup>, Urs TA, Ranganath RR.

##### BACKGROUND:

A new rapid Immunochromatographic test kit (SD MPT64TB Ag Kit) for detection of MPT 64 Antigen in M. tuberculosis isolates using mouse monoclonal MPT 64 Antibody developed by SD Bioline, South Korea was evaluated for rapid identification of M. tuberculosis isolates. We also assessed the sensitivity, specificity and predictive values of this kit. The test kit has an excellent sensitivity, specificity, negative predictive value & positive predictive value. This rapid method is found to be a reliable, rapid and cheaper method for confirming MTB culture isolates in resource poor laboratories.

##### MATERIAL/METHODS:

54 culture isolates of M. tuberculosis in broth & on LJ medium, 12 Non mycobacterial isolates, 10 Non tubercular (NTM) rapidly growing Mycobacteria isolated from pus & 5 smear positive sputum samples were tested for detection of MPT64 antigen using the SD Bioline immunochromatography (ICT) test kit. H37 RV strain was employed as the positive reference control.

##### FINDINGS:

H37 RV strain showed the presence of MPT64 antigen band. Similar band was formed in all the 54 MTB isolates tested **proving 100% sensitivity**. MPT64 band formation was not detected in any of the other test isolates which **proved the 100% specificity** of the test kit. **Both PPV & NPV were 100%**.

##### CONCLUSION:

Tuberculosis is a global pandemic. Rapid identification of MTB culture isolate is very important for drug susceptibility testing. MPT 64 TB Ag detection ICT kit is a rapid, reliable method; it can be a substitute for the molecular identification methods.

62

**Exemple article - MG**

Radiology, 200642 2020 Feb 26[Online ahead of print] DOI: [10.1148/radiol.2020200642](https://doi.org/10.1148/radiol.2020200642)

**Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases**

Tao Ai<sup>1</sup>, Zhenlu Yang<sup>1</sup>, Hongyan Hou<sup>1</sup>, Chenao Zhan<sup>1</sup>, Chong Chen<sup>1</sup>, Wenzhi Lv<sup>1</sup>, Qian Tao<sup>1</sup>, Ziyong Sun<sup>1</sup>, Liming Xia<sup>1</sup>

**Abstract**

Background Chest CT is used for diagnosis of 2019 novel coronavirus disease (COVID-19), as an important complement to the reverse-transcription polymerase chain reaction (RT-PCR) tests. Purpose To investigate the diagnostic value and consistency of chest CT as compared with comparison to RT-PCR assay in COVID-19. Methods From January 6 to February 6, 2020, 1014 patients in Wuhan, China who underwent both chest CT and RT-PCR tests were included. With RT-PCR as reference standard, the performance of chest CT in diagnosing COVID-19 was assessed. Besides, for patients with multiple RT-PCR assays, the dynamic conversion of RT-PCR results (negative to positive, positive to negative, respectively) was analyzed as compared with serial chest CT scans for those with time-interval of 4 days or more. Results Of 1014 patients, 59% (601/1014) had positive RT-PCR results, and 88% (888/1014) had positive chest CT scans. **The sensitivity of chest CT** in suggesting COVID-19 **was 97% (95%CI, 95-98%, 580/601 patients)** based on positive RT-PCR results. In patients with negative RT-PCR results, 75% (308/413) had positive chest CT findings; of 308, 48% were considered as highly likely cases, with 33% as probable cases. By analysis of serial RT-PCR assays and CT scans, the mean interval time between the initial negative to positive RT-PCR results was  $5.1 \pm 1.5$  days; the initial positive to subsequent negative RT-PCR result was  $6.9 \pm 2.3$  days). 60% to 93% of cases had initial positive CT consistent with COVID-19 prior (or parallel) to the initial positive RT-PCR results. 42% (24/57) cases showed improvement in follow-up chest CT scans before the RT-PCR results turning negative. Conclusion Chest CT has a high sensitivity for diagnosis of COVID-19. Chest CT may be considered as a primary tool for the current COVID-19 detection in epidemic areas.

**Keywords:** 2019-nCoV pneumonia; chest CT imaging; diagnostic value; positive rate; reverse transcription polymerase chain reaction.

63

63

**Exemple article - MG**

Radiology, 200642 2020 Feb 26[Online ahead of print] DOI: [10.1148/radiol.2020200642](https://doi.org/10.1148/radiol.2020200642)

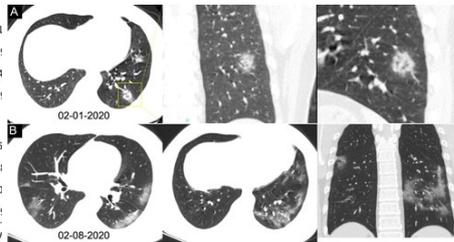
**Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases**

Tao Ai<sup>1</sup>, Zhenlu Yang<sup>1</sup>, Hongyan Hou<sup>1</sup>, Chenao Zhan<sup>1</sup>, Chong Chen<sup>1</sup>, Wenzhi Lv<sup>1</sup>, Qian Tao<sup>1</sup>, Ziyong Sun<sup>1</sup>, Liming Xia<sup>1</sup>

**Abstract**

**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 [95% CI]	Specificity [95% CI]	PPV [95% CI]	NPV [95% CI]	Accuracy [95% CI]
<b>Overall</b>	580	105	308	21	<b>97</b> (580/601) [95-98]	<b>25</b> (105/413) [22-30]	<b>65</b> (580/888) [62-68]	<b>83</b> (105/126) [76-89]	<b>68</b> (685/1014) [65-70]
<b>Age</b>									
< 60 years	362	81	225	15	96 (362/377) [94-98]	27 (81/306) [22-32]	62 (362/587) [58-66]	84 (81/97) [76-91]	68 (685/1014) [65-70]
≥ 60 years	218	24	83	6	97 (218/224) [94-99]	22 (24/107) [16-31]	72 (218/301) [67-77]	80 (24/30) [63-97]	
<b>Sex</b>									
Male	272	35	148	12	96 (272/284) [93-98]	19 (35/183) [14-25]	65 (272/420) [60-69]	75 (35/46) [61-89]	
Female	308	70	160	9	97 (308/317) [95-99]	30 (70/230) [25-37]	66 (308/468) [61-70]	89 (70/79) [80-97]	



TP= true positive, TN=true negative, FP=false positive, FN=false negative predictive value, NPV=negative predictive value, RT-PCR= reverse transcription polymerase chain reaction.

Data in parentheses are numbers of patients used to calculate percentages. Data in brackets are 95% confidence intervals.

Ai 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.2020200642

64

64

## 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.

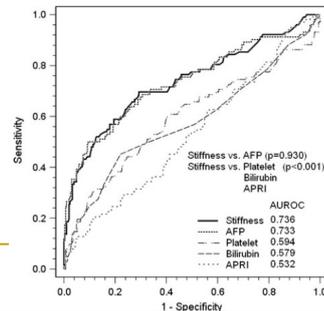
**BACKGROUND/AIMS:** [corrected] The risk of hepatocellular carcinoma (HCC) increased with progression of hepatic fibrosis as assessed by liver stiffness measurement (LSM). This study used LSM to assess the risk of HCC presence in patients with chronic hepatitis.

**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.



65

## Exemples des questions pour l'examen

1) Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a la sensibilité de 0,99, et la spécificité de 0,91. Lesquelles des suivantes affirmations sont correctes ?

- le test est meilleur comme test de screening
- le test est meilleur comme test de précision
- la sensibilité est très bonne
- 9% des sujets non malades ont le test positif
- 1% des malades ont le test négatif

Réponse: a, c, d, e

### Explications

Les tests de screening sont des tests avec une grande sensibilité. Les test de diagnostique de précision sont avec une grande spécificité. Ici la sensibilité est plus grande que la spécificité, donc le réponse a) est correcte

Pour la Se, Sp, les valeurs >=90%, sont cliniquement bonnes, les valeurs entre >=80% et <90% sont modérées, les valeurs inférieure a <80% sont cliniquement avec une qualité réduite. Ici Se=99%, donc le réponse c) est correcte.

Dans le tableau de contingence sur la colonne des non malades on peut calculer la spécificité ( $d/(b+d)$ ) – qui représente la probabilité d'avoir une test négative sachant que le sujet est non malade. Dans la même colonne on peut la probabilité d'avoir une test positif sachant que le sujet est non malade ( $b/(b+d)$ ). La somme de ces deux probabilités est 1. Ici la  $Sp=0,91$ , donc,  $1-0,91=0,09$ , qui est 9%. Le réponse d) est correcte.

Dans le tableau de contingence sur la colonne des malades on peut calculer la sensibilité ( $a/(a+c)$ ) – qui représente la probabilité d'avoir une test positif sachant que le sujet est malade. Dans la même colonne on peut la probabilité d'avoir une test négatif sachant que le sujet est malade ( $c/(a+c)$ ). La somme de ces deux probabilités est 1. Ici la  $Se=0,99$ , donc,  $1-0,99=0,01$ , qui est 1%. Le réponse e) est correcte.

66

## Exemples des questions pour l'examen

2) Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique. Le test a la sensibilité de 0,91, et la spécificité de 0,99. Lesquelles des suivantes affirmations sont correctes ?

- a) le test est meilleur comme test de screening
- b) le test est meilleur comme test de précision
- c) la spécificité est très bonne
- d) 99% des sujets non malades ont le test négatif
- e) 99% des sujets non malades ont le test positif

Réponse: b, c, d

2) Dans une collection des données exposée-non exposée pour une test diagnostique on peut calculer les suivantes indicateurs médicales

- a) Se
- b) Sp
- c) VPP
- d) VPN
- e) le pourcentage des bien classes

Réponse: c, d

67

## Exemples des questions pour l'examen

4) Dans une collection des données cas témoin pour une test diagnostique on peut calculer les suivantes indicateurs médicales

- a) Se
- b) Sp
- c) +LR
- d) -LR
- e) le pourcentage des bien classes

Réponse: a, b, c, d

5) Dans une collection des données échantillon représentatif pour une test diagnostique on peut calculer les suivantes indicateurs médicales

- a) Se
- b) Sp
- c) la prevalence de la maladie
- d) VPN
- e) le pourcentage des bien classes

Réponse: a, b, c, d, e

6) Dans une collection des données cas témoin pour une test diagnostique on peut calculer les suivantes indicateurs médicales

- a) Indice Youden
- b) VPP
- c) VNP
- d) -LR
- e) le pourcentage des bien classes

Réponse: a, d

68

## Exemples des questions pour l'examen

7) Lesquelles des suivantes affirmations sont correctes ?

- a) Une bonne sensibilité permet de limiter le nombre de faux positifs
- b) Une bonne spécificité permet de limiter le nombre de faux positifs
- c) Une bonne valeur prédictive positive permet de limiter le nombre de faux positifs
- d) Une bonne valeur prédictive négative permet de limiter le nombre de faux positifs
- e) Une bonne sensibilité permet de limiter le nombre de faux négatifs

Réponse: b, c, e

8) 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

9) 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 significatif
- 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% contient la valeur 0,5
- e) L'hypothèse nulle du test statistique pour l'AUROC est que l'AUROC est égale a 0,5

Réponse: a, b, c, d, e

69

## Exemples des questions pour l'examen

11) 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

12) 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

13) 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

70

## Exemples des questions pour l'examen

14) \* Dans une étude d'évaluation de la radiographie comme test de diagnostic pour la détection de la tuberculose, réalisé avec un collection cas témoin , parmi les 100 malades 90 ont été détectés comme positifs par le nouveau test et parmi les 100 patients en bonne santé 85 ont été qualifiées comme pas affectée à juste titre par le nouveau test. La VPP et la VPN sont:

- a) 0,857 et 0,895
- b) 0,90 et 0,895
- c) 0,85 et 0,857
- d) 0,857 et 0,857
- e) Ne peuvent pas être calculée

Réponse: e

15) Vous travaillez dans un hôpital, et vous avez a votre disposition trois test diagnostique. Le test A a la sensibilité de 0,55, et la spécificité de 0,85; le test B a la sensibilité de 0,82, et la spécificité de 0,97; le test C a la sensibilité de 0,96, et la spécificité de 0,83. Lesquelles des suivantes affirmations sont correctes ?

- a) le test C est meilleur comme test de screening
- b) le test B est meilleur comme test de précision
- c) la sensibilité du test B est cliniquement modérée
- d) la sensibilité du test C est cliniquement bonne
- e) la sensibilité du test A est cliniquement modérée

Réponse: a, b, c, d

71

## Exemples des questions pour l'examen

16) Vous suspectez un de vos patients d'avoir la maladie testée par le nouveau test diagnostique. De l'anamnèse vous assumez une probabilité de 55% pour votre patient d'avoir la maladie. LR+ = 3,5, LR- = 0,49. Vous appliquez le nouveau test et le patient a un résultat positif. Utilisez le nomogramme de Fagan pour trouver la probabilité post-teste que votre patient soit réellement malade. Lesquelles des suivantes affirmations sont correctes ?

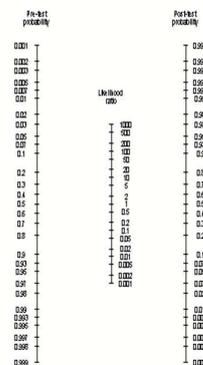
- a) la probabilité d'avoir la maladie post test est ~80%
- b) la probabilité d'avoir la maladie post test est ~40%
- c) le test diagnostique positif entraînent des changements réduits/faible mais parfois concluante/cliniquement significatif de la probabilité d'un diagnostic
- d) le test diagnostique négatif conduit à des modifications de faible probabilité, et rarement cliniquement important
- e) on doit utiliser LR+ sur la deuxième axe pour trouver la probabilité post test

Réponse: a, c, d, e

17) Vous suspectez un de vos patients d'avoir la maladie testée par le nouveau test diagnostique. De l'anamnèse vous assumez une probabilité de 0,7 pour votre patient d'avoir la maladie. LR+ = 10,1, LR- = 0,09. Vous appliquez le nouveau test et le patient a un résultat négatif. Utilisez le nomogramme de Fagan pour trouver la probabilité post-teste que votre patient soit réellement malade. Lesquelles des suivantes affirmations sont correctes ?

- a) la probabilité d'avoir la maladie post test est ~ 97%
- b) la probabilité d'avoir la maladie post test est ~ 0,2
- c) le test diagnostique positif entraînent des changements importants et souvent concluante de la probabilité d'un diagnostic
- d) le test diagnostique négatif entraînent des changements réduits/faible mais parfois concluante/cliniquement significatif de la probabilité d'un diagnostic
- e) on doit utiliser LR+ sur la deuxième axe pour trouver la probabilité post test

Réponse: b, c



DL1

72

DL1 user, 4/11/2017

## Exemples des questions pour l'examen

### 17) Explications

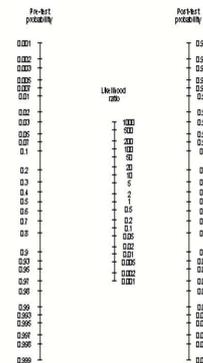
Pour trouver la probabilité post test, sur la première axe on va identifier une point avec la probabilité donnée dans la question:0,7, puis sur la deuxième axe (Likelihood ratio), si le résultat du test est positif on va identifier un point avec la valeur du +LR (positive likelihood ratio/ rapport de vraisemblance positif), et si le résultat du test est négatif on va identifier un point avec la valeur du -LR (negative likelihood ratio/ rapport de vraisemblance négatif) – ici, dans le scenario il est indiquée que le résultat a été négatif, donc on va utiliser  $-LR=0,09$ , et on va tirer une ligne entre ces deux points (pretest et likelihood ratio), qui on va le prolonger jusqu'à-a la troisième axe (posttest probability), ou on va trouve la probabilité post test. Donc le réponse b) est correcte, et les réponses a), et e) sont incorrectes.

Concernant le niveau des changements entrainées par des tests positifs et négatifs, on doit regarder les valeurs du LR.

- • LR > 10 ou < 0,1 entraînent des changements importants et souvent concluante de la probabilité d'un diagnostic
- • LR entre 5 et 10 ou entre 0,1 (=1/10) et 0,2 (=1/0,5) entraînent des changements importants et souvent concluante de la probabilité d'un diagnostic
- • LR entre 2 et 5 ou entre 0,5(=1/2) et 0,2(=1/5) entraînent des changements réduits/fiables mais parfois concluante/cliniquement significatif de la probabilité d'un diagnostic
- • LR entre 1 et 2 ou entre 0,5 et 1 conduit à des modifications de faible probabilité, et rarement cliniquement important
- • LR = 1 – pas de changement

Le LR+ = 10, 1 > 10 => changements importants et souvent concluante => réponse c est correcte

Le LR- = 0,09 < 0,1 => changements importants et souvent concluante => réponse e est incorrecte



73

## Exemples des questions pour l'examen

18) \* Dans une étude d'évaluation d'une technique radiographique comme test de diagnostic pour la détection de la sarcoïdose thoracique sur un échantillon représentatif, parmi les 200 malades 110 ont été détectés comme positifs par le nouveau test et parmi les 200 patients en bonne santé 105 ont été qualifiés comme pas affectée à juste titre par le nouveau test. La Sensibilité et la VPP sont:

- a) ne peuvent pas être calculées
- b) 0,55 et 0,537
- c) 0,525 et 0,537
- d) 0,55 et 0,538
- e) 0,538 et 0,525

Réponse: b

19) \* Dans une étude d'évaluation d'une technique radiographique comme test de diagnostic pour la détection de la sarcoïdose thoracique sur un échantillon représentatif, parmi les 200 malades 110 ont été détectés comme positifs par le nouveau test et parmi les 200 patients en bonne santé 105 ont été qualifiés comme pas affectée à juste titre par le nouveau test. Les VP, VN, FP, FN sont:

- a) 110, 105, 95, 90
- b) 95, 90, 110, 105
- c) 110, 95, 105, 90
- d) 105, 110, 105, 95
- e) 110, 90, 95, 105

Réponse: a

74

## Exemples des questions pour l'examen

20) \* Dans une étude d'évaluation d'une technique radiographique comme test de diagnostic pour la détection de la sarcoïdose thoracique sur un échantillon représentatif, parmi les 200 malades 110 ont été détectés comme positifs par le nouveau test et parmi les 200 patients en bonne santé 105 ont été qualifiés comme pas affectée à juste titre par le nouveau test. La Sensibilité et la VPP sont:

- a) la probabilité d'un patient présentant un nouveau test positif d'avoir la maladie testée représente le VPP
- b) la probabilité d'un patient non-malade à tester négatif en utilisant le nouveau test représente la spécificité
- c) la probabilité d'un patient malade à tester positif en utilisant le nouveau test représente la sensibilité
- d) la probabilité d'un patient présentant un nouveau test négatif d'avoir la maladie testée représente la spécificité
- e) la probabilité d'un patient malade à tester négatif en utilisant le nouveau test représente la sensibilité

Réponse: a, b, c

75

## Exemples des questions pour l'examen

21) **Explications (continuation)**

Si on regarde sur l'axe verticale depuis haut en bas, la première valeur est celui qui correspond à 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 à 1 (en haut), et les plus petites valeurs sont proche à 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 à 1 indique une qualité très bonne. Donc les réponses f), g), h) sont correctes;

22) Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique pour la maladie M. En cherchant sur Pubmed vous avez trouvée deux test diagnostiques. Le test diagnostique A a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,66 (95% CI 0,46 – 0,86), et le test diagnostique B, a l'AUC égale a 0.89 (95% CI 0,79 – 0,99). Dans l'article il y a une comparaison statistique entre le test diagnostique A et B, qui a offert la valeur  $p=0,03$ . Lesquelles des suivantes affirmations sont correctes ?

- a) le test statistique pour l'AUC du test diagnostique A est statistiquement signifiant
- b) le test statistique pour l'AUC du test diagnostique A n'est pas statistiquement signifiant
- c) Il y a une différence statistiquement significative entre le test diagnostique A et le test diagnostique B
- d) le test statistique pour l'AUC du test diagnostique B est statistiquement signifiant
- e) le test statistique pour l'AUC du test diagnostique B n'est pas statistiquement signifiant

Réponse: b, c, d

76

## Exemples des questions pour l'examen

24) Vous travaillez dans un hôpital, et vous voulez utiliser un nouveau test diagnostique pour la maladie M. En cherchant sur Pubmed vous avez trouvée deux test diagnostiques. Le test diagnostique A a l'aire en dessous la courbe receiver operator characteristic (AUROC ou AUC) égale à 0,72 (95% CI 0,47 – 0,89), et le test diagnostique B, a l'AUC égale a 0.58 (95% CI 0,39 – 0,76). Dans l'article il y a une comparaison statistique entre le test diagnostique A et B, qui a offert la valeur  $p=0,36$ . Lesquelles des suivantes affirmations sont correctes ?

- a) le test statistique pour l'AUC du test diagnostique A est statistiquement significatif
- b) le test statistique pour l'AUC du test diagnostique A n'est pas statistiquement significatif
- c) il n'y a pas une différence statistiquement significative entre le test diagnostique A et le test diagnostique B
- d) le test statistique pour l'AUC du test diagnostique B est statistiquement significatif
- e) le test statistique pour l'AUC du test diagnostique B n'est pas statistiquement significatif

Réponse: b, c, e

25) Vous travaillez dans un hôpital, et vous voulez utiliser deux tests diagnostique pour la maladie M. En cherchant sur Pubmed vous avez trouvée un article qui présente la reproductibilité de ce test. Le test A est de type qualitatif (a comme des résultats: positif / négatif), le test B est de type quantitatif (a comme résultat un nombre). . Lesquelles des suivantes affirmations sont correctes ?

- a) 50 patients ont été évalués par le même médecin dans deux moments différents dans le temps en utilisant le test A. Pour évaluer la reproductibilité intra observateur on doit utiliser le coefficient kappa.
- b) 40 patients ont été évalués par deux médecins dans en utilisant le test A. Pour évaluer la reproductibilité inter observateur on doit utiliser le coefficient kappa.
- c) 50 patients ont été évalués par le même médecin dans deux moments différents dans le temps en utilisant le test B. Pour évaluer la reproductibilité intra observateur on doit utiliser le coefficient interclass correlation coefficient.
- d) 40 patients ont été évalués par deux médecins dans en utilisant le test B. Pour évaluer la reproductibilité inter observateur on doit utiliser le coefficient ICC.
- e) Un ICC de 0.6 indique une reproductibilité excellente.

Réponse: a,b,c,d

77

## Notions étudiées dans le cours

- Description d'un phénomène de santé
  - Temps, lieu, personne
  - Indicateurs
- Évaluation d'une méthode diagnostique
  - Phases
    - I: validité, fiabilité, le test kappa de Cohen
    - II: intrinsèque: Se, Sp, LR+, LR-
      - Courbe ROC
    - III: extrinsèque: VPP, VPN
    - IV: cout – bénéfice
- Exemples des études
- Exemples des exercices

78

78

Merci beaucoup!

79

79

## De modificat pe viitor

- Exercitii si explicatii pentru a alege intre 2 teste dg diferite in fct se, sp, de gandit cati FP, FN apar, snout spin
- Bland Altman, ICC, kappa
- Coefficient de variance pour laboratory assays!!!

80

80