



Evaluation de la performance d'un test diagnostique: Reproductibilité, Validité interne et Validité externe

Evaluation of the performance of a diagnostic test: Reproducibility, Internal validity and External validity

تقييم أداء الاختبار التشخيصي: الاستنساخ والصلاحية الداخلية و الخارجية

Donia Ben Hassine¹, Dhekra Chebil², Sarra Melki³, Sarra Nouira⁴, Asma Ben Abdelaziz⁵, Kamel Ben Salem⁶, Ahmed Ben Abdelaziz⁷, Réseau Maghrébin PRP2S*

* Réseau- Maghrébin: Pédagogie- Recherche- Publications en Sciences de Santé (PRP2S)

1. *Direction des Systèmes d'Information du CHU Sahloul de Sousse. Laboratoire de Recherche LR19SP01. Université de Sousse (Tunisie)*
2. *Service de Médecine Préventive et Communautaire. Faculté de Médecine de Sousse (Tunisie)*
3. *CHU Sup de Sahloul de Sousse. Laboratoire de Recherche LR19SP01. Université de Sousse (Tunisie)*
4. *Laboratoire de Recherche LR19SP01. Faculté de Médecine de Monastir. Université de Monastir (Tunisie)*
5. *Service de Biochimie du CHU Sahloul de Sousse. Faculté de Pharmacie de Monastir. Université de Monastir (Tunisie)*
6. *Service de Médecine Préventive et Communautaire. Faculté de Médecine de Monastir. Université de Monastir (Tunisie)*
7. *Service de Médecine Préventive et Communautaire. Direction des Systèmes d'Information du CHU Sahloul. Laboratoire de Recherche LR19SP01. Université de Sousse*

CETTE SÉRIE...

Le Réseau Maghrébin PRP2S et la Rédaction de la revue «La Tunisie Médicale» ont l'honneur de continuer d'une manière régulière, à partir du numéro de janvier 2022, et pour la troisième année successive, la série des fiches techniques en épidémiologie, en bio statistique et en rédaction médicale scientifique.

Cette série a eu un grand succès au cours de sa première année d'édition en 2020, comme indique le nombre de téléchargements dépassant significativement celui des articles originaux et illustrant un besoin très manifeste des jeunes chercheurs, au renforcement de leurs capacités en méthodologie de recherche scientifique en sciences de santé, selon une pédagogie centrée sur l'acquisition des compétences pratiques de recherche biomédicale.

En effet, nos fiches méthodologiques décrivent, d'une manière standardisée, les modes d'usage des concepts, des outils et des méthodes, utilisés d'une part lors du continuum de la recherche biomédicale scientifique, dès la phase conceptuelle jusqu'à la phase rédactionnelle et d'autre part lors des différentes phases de la rédaction médicale scientifique, depuis l'étape de la recherche documentaire jusqu'à l'étape de la communication médicale scientifique.

Cette série est rédigée par les experts du Réseau Maghrébin PRP2S, en méthodologie de recherche, exerçant dans les universités du Grand Maghreb et les facultés sœurs au Nord de la Méditerranée. Chaque fiche répond à trois questions essentielles (Quoi ? Pourquoi ? Comment) du concept étudié, en se basant sur un article publié dans la revue «La Tunisie Médicale».

Le coordinateur de la série «Fiches Méthodologiques»

Professeur Ahmed Ben Abdelaziz (Président du Réseau Maghrébin PRP2S)

Email: ahmedbenabdelaziz.prp2s@gmail.com

Correspondance

Ahmed Ben Abdelaziz

Direction des Systèmes d'Information du CHU Sahloul de Sousse.

Email: ahmedbenabdelaziz.prp2s@gmail.com

**Série des Fiches Méthodologiques
Sommaire**

Année 2020

Fiche n°1 (janvier 2020):

Comment calculer la taille d'un échantillon pour une étude observationnelle. Serhier Z, *et al.* (Faculté de Médecine et de Pharmacie de Casablanca. Maroc)

Fiche n°2 (février 2020):

La recherche qualitative: méthodes, outils, analyse. Soulimane A. (Faculté de Médecine, Université Djillali Liabes, Sidi Bel Abbès, Algérie)

Fiche n°3 (mars 2020)

Et Allah ...créa la variabilité. Barhoumi T, *et al* (Réseau Maghrébin PRP2S)

Fiche n°4 (mai 2020)

Réussir votre recherche bibliographique sur *PubMed*. Ben Abdelaziz A, *et al* (Réseau Maghrébin PRP2S)

Fiche n°5 (juin 2020)

Réussir la rédaction de votre «Protocole de Recherche» en sciences de la santé. Ben Abdelaziz A, *et al* (Réseau Maghrébin PRP2S)

Fiche n°6 (juillet 2020)

Analyse multi variée par régression logistique. Ben Salem K, *et al* (Réseau Maghrébin PRP2S)

Fiche n°7 (aout/septembre 2020)

Tests non paramétriques pour comparer deux ou plusieurs moyennes sur des échantillons indépendants. Bezzaoucha A, *et al* (Réseau Maghrébin PRP2S)

Fiche n°8 (septembre 2020)

Comment évaluer la concordance entre deux mesures qualitatives par le test Kappa? Mellakh R, *et al* (Réseau Maghrébin PRP2S)

Fiche n°9 (octobre 2020)

Comment comparer plusieurs moyennes par le test d'Analyse de Variance (ANOVA) ? Khiari H, *et al* (Réseau Maghrébin PRP2S)

Fiche n°10 (novembre 2020)

Tests non paramétriques sur SPSS pour comparer deux ou plusieurs moyennes sur des échantillons appariés. (test de Wilcoxon et test de Friedman). Bezzaoucha A *et al* (Réseau Maghrébin PRP2S)

Année 2021

Fiche n°1 (mars 2021):

Comment calculer et interpréter la valeur de «p» dans une étude épidémiologique. Ladner J *et al.* (Faculté de Médecine et de Pharmacie de Rouen. France)

Fiche n°2 (avril 2021)

La Charge globale de morbidité (Global Burden of Diseases): c'est Quoi? Pourquoi? Comment? Hsairi M *et al* (Faculté de Médecine de Tunis, Tunisie)

Fiche n°3 (mai 2021)

Analyse en Composantes Principales (ACP). Ben Salem K *et al* (Faculté de Médecine de Monastir, Tunisie)

Fiche n°4 (juin 2021)

L'approche 5x5 en Rédaction biomédicale scientifique. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°5 (juillet 2021)

Elaboration du Résumé Structuré d'un article biomédical scientifique. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°6 (aout/septembre 2021)

Comment critiquer un Article Médical Scientifique ? Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°7 (octobre)

Réussir l'élaboration d'une liste d'auteurs d'un manuscrit scientifique. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°8 (novembre)

L'Evidence-based Medicine. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°9 : (décembre)

Les Statistiques descriptives. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Année 2022

Fiche n°1

Evaluation de la performance d'un test diagnostique. Ben Hassine D, *et al.*

ETUDE DE CAS

La revue «La Tunisie Médicale» a publié en 2013 un article original intitulé «Validation de la section G du Composite International Diagnostic Interview, relative à la schizophrénie, traduite au dialecte tunisien» (1). La schizophrénie est un problème de santé publique universel qui entraîne un retentissement lourd et grave sur la morbidité et la mortalité, pouvant justifier la mise en place des stratégies de dépistage. La prévalence mondiale de la schizophrénie varie entre 0,19% et 1%. L'objectif de ce travail était de traduire, au dialecte tunisien, la section G du Composite International Diagnostic Interview (CIDI), relative à la schizophrénie et de vérifier la validité et la fiabilité de cette section traduite. Afin d'atteindre ce but, deux étudiants en psychologie ont administré l'instrument aux patients et aux témoins qui ont été recrutés par deux psychiatres. Le texte intégral de cet article est actuellement en accès libre soit sur le site de la revue soit via PubMed. Ci-dessous un tableau (encadré 1) extrait de l'article, intitulé «Validité externe de la section G du CIDI: confrontation des résultats du Gold standard et du CIDI», résumant les résultats observés par le test (instrument CIDI) par rapport à ceux de la réalité (gold standard).

Encadré 1: Tableau de contingence des résultats du test CIDI et du Gold Standard

RESULTATS DU GOLD STANDARD				
		Patient	Témoin	Total
RESULTATS DU TEST CIDI	Schizophrénie (+)	55	8	63
	Schizophrénie (-)	59	103	162
	Total	114	111	225

QUIZZ

Répondre aux questions suivantes en choisissant la(es) réponse(s) exacte(s) :

1. La sensibilité est:

- la capacité d'un test à détecter les sujets malades.
- la probabilité qu'une personne ayant un test positif soit malade.
- la capacité d'un test à détecter les sujets non malades.
- égale à 48,2%.
- égale à 92,8%.

2. La spécificité est:

- la capacité d'un test à détecter les sujets malades.
- la probabilité qu'une personne ayant un test négatif ne soit pas malade.

c. la capacité d'un test à détecter les sujets non malades.

d. égale à 48,2%.

e. égale à 92,8%.

3. La Valeur Prédictive Positive est:

a. la probabilité qu'une personne ayant un test positif soit malade.

b. la probabilité qu'une personne ayant un test négatif ne soit pas malade.

c. la capacité d'un test à détecter les sujets malades.

d. égale à 87,3%.

e. égale à 63,6%.

4. La Valeur Prédictive Négative:

a. est la probabilité qu'une personne ayant un test positif soit malade.

b. est la probabilité qu'une personne ayant un test négatif ne soit pas malade.

c. dépend de la prévalence de la schizophrénie.

d. est égale à 87,3%.

e. est égale à 63,6%.

INTRODUCTION

Un test diagnostique est un examen médical, clinique ou para-clinique, permettant d'aider le clinicien dans sa démarche des soins. Il est important en raison de son implication pour dépister un facteur de risque (test de dépistage), diagnostiquer une maladie (test de diagnostic) ou estimer le pronostic d'un patient (test pronostique) (encadré n°2). La certitude diagnostique est rarement atteinte en pratique clinique et la notion de probabilité diagnostique et sa révision en raison de preuves diagnostiques supplémentaires sont au cœur du raisonnement diagnostique. Face au développement continu de nouvelles méthodes diagnostiques, les exigences de haute qualité et les pressions de maîtrise des coûts des tests cliniques, à savoir leurs indicateurs de performance, semblent nécessaires aussi bien pour les chercheurs que pour les patients et leurs familles. Il existe des indicateurs synthétiques fournissant une seule valeur de performance «globale» qui sont probablement plus connues par les chercheurs que par les cliniciens. Ainsi, au terme de cette fiche méthodologique, le professionnel de la santé doit être capable de poser un diagnostic à l'aide de ces tests avec plus au moins de certitude. Ce diagnostic oriente vers la stratégie thérapeutique rationnelle (abstention, traitement médical, traitement chirurgical).

Encadré n°2 : Les différences entre le test diagnostique et le test de dépistage

Test diagnostique	Test de dépistage
Doit réduire l'incertitude diagnostique (examens spécifiques) : ✓ Il est appliqué aux personnes présentant des troubles définis ✓ Il constitue une base du traitement ✓ Il est plus précis ✓ Il coûte éventuellement plus cher ✓ Essentiellement individuel	Appliqué aux personnes apparemment en bonne santé : ✓ Praticué sur des groupes d'individus ✓ Il ne constitue pas une base de traitement ✓ Moins précis ✓ Coûte moins cher

LA PERFORMANCE D'UN TEST DIAGNOSTIQUE: C'EST QUOI ?

La performance d'un test diagnostique est reflétée par sa capacité de discrimination entre plusieurs classes diagnostiques, plus particulièrement la présence ou l'absence du diagnostic dans le cas d'un test binaire. L'utilisation d'un test diagnostique pour l'activité de soins suppose une évaluation préalable de sa performance de la même façon qu'un médicament ou un acte thérapeutique. En effet, un test diagnostique peu performant est susceptible d'être associé à de mauvais rapports bénéfice/risque, pour le patient, et coût/efficacité, pour la société. L'évaluation de la performance d'un test diagnostique confronte en général les résultats de ce test à la classification des sujets réalisée par un «gold standard». On appelle «gold standard», le meilleur test diagnostique (examen de référence) déjà disponible pour la pathologie d'intérêt et utilisable en situation «raisonnable» en pratique clinique ou en recherche étant donné qu'il est douloureux, risqué, coûteux et peut causer un inconfort pour les patients. Ce terme a déjà été soumis à controverses dans la mesure où l'on pourrait croire qu'il est équivalent à «la réalité», alors qu'il ne s'agit que d'un point de référence correspondant aux connaissances dans leur état actuel. Malgré qu'il est réalisable en routine, moins douloureux et moins coûteux par rapport à l'examen de référence, le test diagnostique peut se tromper et sa réponse peut être soit sous forme binaire (malade ou non-malade), ordinale (probabilité du diagnostic forte/ intermédiaire/ faible) ou quantitative (valeur continue ou discrète).

LA PERFORMANCE D'UN TEST DIAGNOSTIQUE: POURQUOI?

Apprécier la valeur d'un signe pour une maladie, c'est juger la capacité du signe à reconnaître la maladie, ou encore à classer les malades des non-malades. L'utilisation d'un test diagnostique pour l'activité de soins suppose une évaluation préalable de sa performance. Le but d'un test diagnostique est de réduire l'incertitude clinique (la notion de «probabilité diagnostique»). Il permet d'affiner la probabilité que le patient ait ou non la maladie suspectée pour améliorer la prise en charge et l'évolution clinique. Le but est d'avoir le moins des sujets «mal-classés» possible: faux positifs (tests positifs chez les non malades) ou faux négatifs (tests négatifs chez les malades).

LA PERFORMANCE D'UN TEST DIAGNOSTIQUE: COMMENT?

La performance d'un test diagnostique permet de mesurer la reproductibilité, la validité interne et la validité externe.

1. La reproductibilité

La reproductibilité, appelée aussi fiabilité, est définie par la propriété d'un test dont les résultats peuvent être reproduits aussi longtemps que les conditions du test ne changent pas. C'est la mesure de la concordance entre deux mesures répétées: deux mesures de deux observateurs différents (reproductibilité inter-observateurs) ou deux mesures d'un même observateur (reproductibilité intra-observateur). Plus un examen est reproductible, plus il est fiable. La reproductibilité peut être mesurée par le coefficient de Kappa de Cohen (κ). Ce coefficient a été développé pour évaluer la concordance entre deux juges:

une concordance ajustée sur une concordance obtenue par hasard. La concordance attendue, «par hasard», aléatoire (CA) est déterminée à partir des valeurs attendues (A) pour les vrais positifs et les vrais négatifs. La concordance observée (CO) traduit la précision «brute», non ajustée. Elle correspond à la proportion de sujets bien classés, qu'ils soient positifs ou négatifs. On prend l'exemple de deux médecins qui ont lu 100 lames d'anatomopathologie. Le critère de jugement est binaire: présence de cancer Oui/ Non. Quatre combinaisons possibles sont résumées dans un tableau de contingence à quatre cases (tableau 1).

Tableau 1. Exemple d'un tableau de contingence de la concordance entre deux médecins dans la lecture des lames d'anatomopathologie.

		Médecin n°1		Total
		Cancer +	Cancer -	
Médecin n°2	Cancer +	18 (A)	2 (B)	20 (A+B)
	Cancer -	4 (C)	76 (D)	80 (C+D)
Total		22 (A+C)	78 (B+D)	100 (N=A+B+C+D)

La Concordance Observée (CO) = concordance réelle: $CO = (A+D)/N = 0,94$ (94%). La Concordance Aléatoire (CA) est calculée sous l'hypothèse d'indépendance des jugements des deux médecins. L'Effectif théorique = Total ligne * Total colonne / N : $A' = (A+B)*(A+C)/N$ et $D' = (C+D)*(B+D)/N$. $CA = A'+D'/N = [(22*20)/100 + (78*80)/100]/100 = 0,668$ (66,8%).

La formule suivante détaille le calcul du κ :

$$\kappa = \frac{CO - CA}{1 - CA}$$

L'étendue du coefficient kappa est entre -1 et 1. On parle d'un coefficient kappa excellent lorsque l'index kappa $\geq 0,81$ (voir tableau 2).

Tableau 2. Degré d'accord et valeur de Kappa proposé par Landiset Koch

Accord	Index de Kappa
Excellent	$\geq 0,81$
Bon	0,61 -0,80
Modéré	0,41 -0,60
Médiocre	0,21 -0,40
Mauvais	0,01 -0,20
Très mauvais	$< 0,0$

2. La validité interne

Nous étudierons le cas où le résultat du test diagnostique est binaire, ne pouvant prendre que deux valeurs exclusives, «positive» et «négative». Par exemple, ces couples de valeurs peuvent être «présence»/«absence» ou «évolution» / «stabilité» d'une pathologie. Les performances d'un test se mesurent par rapport à une référence «gold standard» (abusivement désigné par «réalité» pour la suite) qui peut être un examen de référence ou une expertise permettant de catégoriser les patients en «malade» ou «non malade». Les résultats des tests en fonction de la classification des patients, en malades et non malades, sont résumés dans un tableau de contingence (tableau 3). Les sujets bien classés par le test sont ceux qui sont réellement positifs et évalués positifs par le test, notés Vrais Positifs (VP), ou ceux qui sont réellement négatifs et évalués négatifs, notés Vrais Négatifs (VN). À l'opposé, les sujets mal classés par le test sont les Faux Positifs (FP), réellement négatifs mais évalués positifs, et les Faux Négatifs (FN), réellement positifs mais évalués négatifs. Les valeurs marginales représentent la division de la population (de taille N) en «Réellement Positifs» (RP = VP + FN) et «Réellement Négatifs» (RN = FP + VN) par le *gold standard*, et en «Evalués Positifs» (EP = VP + FP) et «Evalués Négatifs» (EN = FN + VN) par le test diagnostique. Le taux de prévalence de la pathologie, c'est-à-dire de la situation «positive», correspond donc à $P = RP/N$.

Tableau 3. Tableau de contingence entre les résultats observés d'un test par rapport à ceux de la réalité (Gold standard)

		Examen de référence /Gold standard « Réellement »		Examens inutiles : inquiétude
		Positif (RP)	Négatif (RN)	
Nouveau Test « évalué »	Positif (EP)	Vrai Positif (VP)	Faux Positif (FP)	
	Négatif (EN)	Faux Négatif (FN)	Vrai Négatif (VN)	

Les valeurs du tableau de contingence peuvent être combinées pour la construction de différents indicateurs visant à donner une évaluation partielle ou globale de la performance du test diagnostique. Les indicateurs partiels de performance d'un test diagnostique ne s'intéressent qu'à l'évaluation de la discrimination d'une des deux classes diagnostiques, positive ou négative.

Les deux principales qualités d'un test, qui définissent la validité interne de l'instrument de mesure, sont la sensibilité (Se) qui est la capacité d'identifier

correctement comme malades ceux qui ont la maladie et la spécificité (Sp) qui est la capacité d'identifier correctement les non malades parmi ceux qui n'ont pas la maladie. Ces indicateurs sont fixes, dits « intrinsèques » dans la mesure où ils ne dépendent pas du taux de prévalence de la pathologie dans la population sur laquelle est réalisé le test. Ces indicateurs sont calculés par les formules suivantes en lisant le tableau de contingence de façon verticale (Tableau 4).

Tableau 4. Tableau de contingence entre les résultats observés d'un test par rapport à ceux de la réalité (gold standard)

		Examen de référence /Gold standard « Réellement »	
		Positif (RP)	Négatif (RN)
Nouveau Test « évalué »	Positif (EP)	Vrai Positif (VP)	Faux Positif (FP)
	Négatif (EN)	Faux Négatif (FN)	Vrai Négatif (VN)

		Examen de référence /Gold standard « Réellement »	
		Positif (RP)	Négatif (RN)
Nouveau Test « évalué »	Positif (EP)	Vrai Positif (VP)	Faux Positif (FP)
	Négatif (EN)	Faux Négatif (FN)	Vrai Négatif (VN)

$$Se = \frac{VP}{VP+FN} = \frac{VP}{RP}$$

$$Sp = \frac{VN}{VN+FP} = \frac{VN}{RN}$$



Se



Sp

Un test très sensible (Se≈1) est utile pour exclure la maladie, lorsqu'il est négatif. De même, un test très spécifique (Sp≈1) est utile pour confirmer la maladie lorsqu'il est positif.

Deux indicateurs moins connus mais cliniquement pertinents sont les Rapports de Vraisemblance Positif (RVP) et Négatif (RVN), en anglais: *LR (Likelihood Ratio)*. Ces indicateurs s'appuient sur les deux caractéristiques d'un test qui sont sa sensibilité et sa spécificité. Ils ne dépendent pas du taux de prévalence et peuvent à ce titre être qualifiés d'intrinsèques également. Le RVP est défini par le rapport de la probabilité d'un test positif chez les malades (RP) sur celle chez les non malades (RN).

On privilégie les tests avec un rapport de vraisemblance positif le plus élevé possible. Le RVN est défini par le ratio de la probabilité d'un test négatif chez les malades sur celle chez les non malades. On privilégie les tests avec un rapport de vraisemblance négatif le plus faible possible (proche de 0). De manière générale, on utilise le RVP si le test est associé positivement à la maladie et le RVN si le test est associé négativement à la maladie. Ces rapports sont calculés par les formules suivantes:

$$RVP = \frac{Se}{1 - Sp} = \frac{VP/RP}{1 - VN/RN} = \frac{VP/(VP + FN)}{1 - VN/(VN + FP)}$$

$$1 \leq RVP \leq +\infty$$

$$RVN = \frac{1 - Se}{Sp} = \frac{1 - VP/RP}{VN/RN} = \frac{1 - VP/(VP + FN)}{VN/(VN + FP)}$$

$$0 \leq RVN \leq 1$$

Dans le cas où le résultat brut d'un test n'est pas qualitatif binaire mais quantitatif, la distribution des valeurs du test pour les malades et les non-malades est d'allure bimodale. Dans une situation idéale, les distributions du résultat du test sont bien distinctes, et cela permet ainsi de classer sans erreur les sujets en malades ou non malades. Mais dans la vraie vie, les valeurs de résultat du test s'échelonnent et se superposent entre les malades et les non-malades, il y aura ainsi des faux positifs et des faux négatifs. Il n'est plus possible donc de dresser un tableau de contingence. Le chevauchement des valeurs du test pour les malades et les non malades permet de déterminer un seuil (*cut-off point*). La meilleure valeur seuil est celle qui entraîne le minimum de mauvaises classifications. Il y a différentes façons de déterminer les seuils, la façon la plus courante est de déterminer le seuil qui maximise à la fois la spécificité et la sensibilité, mais on peut aussi choisir de maximiser la sensibilité ou la spécificité en fonction des situations. En fonction du seuil choisi, quand la sensibilité augmente, la spécificité diminue et inversement. En pratique, il y a toujours une part d'arbitraire dans la notion de seuil de décision. Cependant, le choix de la valeur seuil dépend du contexte et des caractéristiques de la maladie étudiée. Quand le résultat du test s'exprime par une variable continue et en faisant varier le seuil, on peut tracer l'ensemble des points à partir des couples (Se, Sp) qui représente la courbe ROC pour «Receiver Operating Curve». Cet indicateur synthétique permet de visualiser et d'évaluer la capacité diagnostique d'un test en déterminant la valeur seuil optimale de la Se et de la Sp du test. L'Aire Sous la

Courbe (*Area Under Curve*, AUC) correspond à la probabilité qu'un individu présentant le diagnostic, tiré au hasard, ait une valeur plus élevée qu'un individu ne présentant pas le diagnostic (en supposant que le diagnostic soit associé aux valeurs élevées). Plus le test est informatif, plus la courbe ROC (AUC) se situera dans le quadrant supérieur gauche du graphique et plus la ROC (AUC) est proche de 1. Une AUC > 0.7 est considérée comme bonne. Un test qui a une AUC de 0,5 (correspond à la diagonale) ne fait pas mieux que le hasard. Il s'agit d'un test non discriminant sans intérêt qui correspond à la situation où la Se et la Sp sont égales à 50%. Un AUC < 0,5 signifie que la probabilité d'un test positif est plus faible chez les malades. Les résultats du test sont inversés d'où son inutilité.

3. La validité externe

La sensibilité et la spécificité apportent peu d'information pour le clinicien en situation réelle de soins. En effet, ces indicateurs sont des probabilités conditionnelles, sachant que le cas est positif ou négatif. Or le clinicien ne sait pas si le patient présente ou non le diagnostic, mais connaît au contraire le résultat du test, qu'il cherche à interpréter. Cette information est fournie par une autre paire d'indicateurs, les Valeurs Prédictives Positives (VPP) et Négatives (VPN) correspondant respectivement à la probabilité d'être malade lorsque le résultat est positif et la probabilité d'être indemne de la maladie lorsque le résultat est négatif. Ils sont qualifiés d'«extrinsèques» puisque leur valeur dépend du taux de prévalence du diagnostic dans la population où est réalisé le test, ce qui limite leur généralisation en pratique car un même test est susceptible d'être utilisé pour des populations différentes ne correspondant pas forcément au contexte de l'étude. Les VPP et VPN sont calculées par les formules suivantes en lisant le tableau de contingence de façon horizontale (Tableau 5).

Tableau 5. Tableau de contingence entre les résultats observés d'un test par rapport à ceux de la réalité (gold standard)

		Examen de référence /Gold standard		
		« Réellement »		
		Positif (RP)	Négatif (RN)	
Nouveau Test « évalué »	Positif (EP)	Vrai Positif (VP)	Faux Positif (FP)	VPP
	Négatif (EN)	Faux Négatif (FN)	Vrai Négatif (VN)	VPN

$$VPP = \frac{VP}{VP+FP} = \frac{VP}{EP} = \frac{Se \times P}{(Se \times P) + (1-P)(1-Sp)} \quad 0 \leq VPP \leq 1$$

$$VPN = \frac{VN}{VN+FN} = \frac{VN}{EN} = \frac{Sp(1-P)}{Sp(1-P) + P(1-Se)} \quad 0 \leq VPN \leq 1$$

avec P : la prévalence de la maladie.

CONCLUSION

Ainsi, nous distinguons deux types des critères de performance des tests diagnostiques:

Les critères intrinsèques du test: sont d'une part la Sensibilité (Se) qui est la probabilité que le test soit positif si on est malade et d'autre part la Spécificité (Sp) qui est la probabilité que le test soit négatif si on est non malade. Ces indicateurs ne dépendent pas de la prévalence. Ils dépendent des conditions de réalisation de l'examen et des caractéristiques des patients.

Les critères extrinsèques du test: soit la Valeur Prédictive Positive (VPP) qui est la probabilité d'être malade si le test est positif et la Valeur Prédictive Négative (VPN) qui est la probabilité d'être non malade si le test est négatif. Ils dépendent de la prévalence de la maladie, de la Se et de la Sp.

RÉPONSES AUX QUIZZ

1. a-d
2. c-e
3. a-d
4. b-c-e

RÉFÉRENCES

1. Rafrafi R, Namouri N, Ghaouar M, Hsairi M, Melki W, El Hechmi Z. Validation de la section G du «Composite International Diagnostic Interview», relative à la schizophrénie, traduite au dialecte tunisien. Tunis Med. 2013; 91(11): 648-54.