

Critères d'évaluation des performances et de l'utilité d'un test diagnostique

Mancini J, Gaudart J, Giorgi R

Laboratoire d'Enseignement et de Recherche sur le Traitement de l'Information Médicale, EA 3283, Aix-Marseille Université, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille Cedex, France.

Med Trop 2009 ; **69** : 78-82

RÉSUMÉ • La valeur informationnelle d'un test diagnostique correspond à son aptitude à discriminer les sujets atteints ou indemnes d'une maladie donnée. Plusieurs indicateurs ont été proposés pour étudier la performance d'un test, dont la sensibilité, la spécificité, la courbe ROC, les valeurs prédictives et le diagnostic odds ratio. Les rapports de vraisemblance sont des indicateurs pertinents pour apprécier son utilité. Dans cet article, nous définissons et présentons les propriétés et les avantages de ces différents indicateurs permettant d'évaluer un test diagnostique dans le cadre des infections systémiques sévères chez des patients neutropéniques fébriles.

MOTS-CLÉS • Tests diagnostique. Performance. Utilité.

CRITERIA FOR EVALUATING THE PERFORMANCE AND UTILITY OF A DIAGNOSTIC TEST

ABSTRACT • The informative value of a diagnostic test determines its usefulness in discriminating between subjects with or without the suspected disease. Various indicators have been proposed for assessment of diagnostic test performance including sensitivity, specificity, ROC curve, predictive values, and diagnostics odds ratio. Likelihood ratios are useful indicators of diagnostic utility of a test. The purpose of this article is to describe and present the properties and advantages of the different indicators used to evaluate diagnostic tests for severe systemic infection in patients with febrile neutropenia.

KEY WORDS • Diagnostic tests. Performance. Utility.

Le nombre d'innovations diagnostiques, ou le nombre de nouvelles indications concernant un test diagnostique déjà utilisé, est en constante augmentation dans le domaine médical. Un nouveau test diagnostique sera accepté en pratique courante s'il apporte une réelle plus-value dans l'aide à la décision médicale en permettant d'établir un nouveau diagnostic, ou un diagnostic plus précis, plus rapide, ou encore en proposant une alternative diagnostique moins invasive, ou complémentaire...(1, 2). Cependant, toutes les innovations diagnostiques ne sont pas concluantes et certains tests diagnostiques prometteurs peuvent se révéler sans intérêt : incapables de séparer correctement les sujets atteints ou non atteints d'une maladie, ou encore inapplicables dans la pratique médicale quotidienne (3).

Pour déterminer si un test diagnostique est performant et utile il faut réaliser une évaluation comparative par rapport au test de référence et par rapport au(x) test(s) habituel(s). Massaro *et al.* ont évalué les performances de la procalcitonine (PCT) dans le cadre des infections systémiques sévères chez des patients neutropéniques fébriles (4). Pour cela, une étude de cohorte a été réalisée et a porté sur 52 adultes fébriles (température corporelle avec un pic > 38,3 °C, ou > 38,0 °C pendant plus d'une heure et n'étant pas la conséquence de l'administration d'un produit dérivé du sang) ayant une neutropénie (neutrophiles < 500/mm³ ou < 1 000/mm³ avec une baisse attendue à 500/mm³ (5)) et hospitalisés dans un hôpital Universitaire de São Paulo pendant la période d'août 2004 à septembre 2006. Sur chaque patient, des échantillons de sang pour l'analyse de la PCT ont été prélevés avant ou jusqu'à 12 heures après l'introduction d'une antibiothérapie. Les informations provenant de l'examen clinique, paraclinique (biologie, culture, radiologie...) et du suivi de chaque patient jusqu'à sa sortie de l'hôpital ou son décès ont permis de les classer, sans commettre d'erreurs, dans l'un des deux

groupes suivants : patients présentant une infection systémique sévère (n = 26) ou patients présentant une fièvre d'origine indéterminée (n = 26). Les auteurs ont conclu que la PCT semblait être un marqueur utile pour le diagnostic d'infection systémique sévère chez des patients neutropéniques fébriles.

L'objectif de cet article est de présenter, à l'aide de cet exemple introductif, les indicateurs permettant de juger de la performance et de l'utilité d'un test diagnostique.

Indicateurs de la performance et de l'utilité d'un test

De vraies et de fausses informations

Intéressons nous à l'évaluation de la performance de la PCT dans le contexte de cette étude. En considérant que l'on évoquera le diagnostic d'infection systémique sévère chez les patients ayant une valeur de PCT ≥ 0,490 ng/mL (valeur seuil) et que la certitude

Tableau 1. Performance de la procalcitonine (PCT) (valeur seuil 0,490 ng/mL) pour le diagnostic d'infection systémique sévère comparée au test de référence*.

	Test de référence*		Total
	Infection systémique sévère (Maladie présente)	Fièvre d'origine indéterminée (Maladie absente)	
Test-index	PCT ≥ 0,490 ng/mL (Test positif) (vrais positifs)	13 (faux positifs)	35
	PCT < 0,490 ng/mL (Test négatif) (faux négatifs)	13 (vrais négatifs)	17
	Total	26	52

* Référence établie à partir des informations provenant de l'examen clinique, paraclinique (biologie, culture, radiologie,...) et du suivi de chaque patients jusqu'à sa sortie de l'hôpital ou son décès.

• Correspondance : roch.giorgi@ap-hm.fr

• Article reçu le 4/07/2008, définitivement accepté le 24/12/08.

Tableau II. Définitions, formules de calcul avec intervalles de confiance à 95 %, des principaux indicateurs de la performance et de l'utilité d'un test diagnostique.

Indicateur	Définition	Formule	Intervalle de confiance à 95 %
Exactitude (E)	Proportion de sujets correctement identifiés par le test	$\frac{VP + VN}{VP + FP + FN + VN}$	$E \pm 1,96 \sqrt{\frac{E(1-E)}{VP + FP + FN + VN}}$
Sensibilité (Se)	Proportion de sujets ayant un test positif parmi les malades	$\frac{VP}{VP + FN}$	$Se \pm 1,96 \sqrt{\frac{Se(1-Se)}{VP + FN}}$
Spécificité (Sp)	Proportion de sujets ayant un test négatif parmi les non malades	$\frac{VN}{VN + FP}$	$Sp \pm 1,96 \sqrt{\frac{Sp(1-Sp)}{VN + FP}}$
Valeur prédictive positive (VPP)	Proportion de sujets malades parmi ceux ayant un test positif	$\frac{VP}{VP + FP}$	$VPP \pm 1,96 \sqrt{\frac{VPP(1-VPP)}{VP + FP}}$
Valeur prédictive négative (VPN)	Proportion de sujets non malades parmi ceux ayant un test négatif	$\frac{VN}{VN + FN}$	$VPN \pm 1,96 \sqrt{\frac{VPN(1-VPN)}{VN + FN}}$
Rapport de vraisemblance positif (RVP)	Rapport des tests positifs chez les malades sur les tests positifs chez les non-malades	$\frac{Se}{1-Sp}$	$\exp \left[\ln RVP \pm 1,96 \sqrt{\frac{1-Se}{Sp} + \frac{Sp}{FP}} \right]$
Rapport de vraisemblance négatif (RVN)	Rapport des tests négatifs chez les malades sur les tests négatifs chez les non-malades	$\frac{1-Se}{Sp}$	$\exp \left[\ln RVN \pm 1,96 \sqrt{\frac{Se}{FN} + \frac{1-Sp}{VP}} \right]$
Diagnostic odds ratio (DOR)	Rapport de la cote d'avoir un test positif chez les malades sur la cote d'avoir un test positif chez les non malades	$\frac{RVP}{RVN}$	$\exp \left[\ln DOR \pm 1,96 \sqrt{\frac{1}{VP + FP + FN + FP}} \right]$

Remarque : E, Se, Sp, VPP et VPN correspondent à des estimations de proportions obtenues sur un certain effectif (n), qui leur est propre. Les formules de leurs intervalles de confiance données dans ce tableau ne sont applicables que si pn et (1-p)n sont supérieurs ou égaux à 5, où p représente les bornes inférieure ou supérieure de l'intervalle de confiance de la proportion estimée.

diagnostique sera apportée par les informations médicales et le suivi, ce qui constitue ce que l'on appelle le test de référence, ou le gold-standard, nous pouvons construire le tableau I. Ainsi, avec une valeur seuil de 0,490 ng/mL, le dosage de la PCT a permis de diagnostiquer correctement 22 patients présentant une infection systémique sévère (vrais positifs, VP) et 13 patients ne présentant pas une infection systémique sévère (vrais négatifs, VN). En revanche, le dosage de la PCT n'a pas permis de diagnostiquer une infection systémique sévère chez 4 patients (faux négatifs, FN) et a diagnostiqué à tort une infection systémique sévère chez 13 patients (faux positifs, FP).

La valeur informationnelle unitaire de ces simples valeurs est importante. D'une manière générale, un test donnant trop de FP est mauvais car cela peut avoir des conséquences directes sur le patient : inquiétude à tort, instauration d'un traitement inutile, prescription d'examen complémentaires injustifiés,... De même, un test donnant trop de FN est mauvais puisqu'un traitement nécessaire ne sera pas instauré, les examens complémentaires justifiés ne seront pas prescrits... La probabilité de ne pas commettre une erreur, ou exactitude, s'estime comme étant le rapport du nombre de VP plus le nombre de VN par le nombre total de patients dans l'échantillon étudié (soit, avec cette valeur seuil, 0,67 dans cette étude).

Sensibilité, spécificité et courbe ROC

Ces quatre quantités (VP, VN, FP, FN) nous permettent de déterminer la sensibilité (Se) et la spécificité (Sp) de la PCT (cf. Tableau II pour une définition de ces indicateurs). La sensibilité mesure la capacité du test à détecter tous les sujets malades. Elle représente, parmi les malades, la proportion de sujets considérés comme étant malades au vu du résultat du test évalué. Elle se calcule en faisant le rapport du nombre de vrais positifs sur le nombre total de malades ($Se = VP/(VP + FN)$). La spécificité est au contraire la capacité du test à écarter les sujets non malades et elle représente, parmi les sujets non malades, la proportion de sujets qui est correctement considérée comme étant non malades au vu du résultat du test. Elle se calcule en faisant le rapport du nombre de vrais négatifs

sur le nombre total de sujets indemnes ($Sp = VN/(VN + FP)$). La Se et la Sp évaluent donc le caractère discriminant d'un test, c'est-à-dire sa capacité à distinguer les malades des non malades. Elles s'expriment généralement en pourcentage.

Ainsi, une PCT $\geq 0,490$ ng/mL a permis d'identifier 80,4 % des infections systémiques sévères ($Se = 0,804$) et une PCT $< 0,490$ ng/mL a permis d'éliminer le diagnostic d'infection systémique sévère chez 50 % des patients ayant une fièvre d'origine indéterminée ($Sp = 0,500$).

Il n'existe pas de valeurs seuils de Se ou de Sp pour décider si un test est intéressant ou non puisque cela va dépendre des conséquences des erreurs engendrées. Dans une approche diagnostique, on aura tendance à privilégier un test ayant une Sp élevée, le nombre de faux positifs étant alors minimisé ; la valeur informationnelle d'un test spécifique est élevée quand le résultat est positif. Dans une approche de dépistage, on aura tendance à privilégier un test ayant une Se élevée, le nombre de faux négatifs étant alors minimisé ; la valeur informationnelle d'un test sensible est élevée quand son résultat est négatif. On utilise souvent les dénominations « SnNout » (quand un test a une Se élevée, un résultat Négatif permet de rejeter le diagnostic, ou out en anglais) et « SpPin » (quand un test a une Sp élevée, un résultat Positif permet d'accepter le diagnostic, ou in en anglais). Ce raisonnement est à nuancer en fonction du contexte clinique. Considérons le dépistage de masse organisé du cancer du colon par le test Hemocult II®. Dans une population de plus de 50 ans, le taux de positivité moyen de ce test a été estimé à 2 %, avec une sensibilité entre 50 et 60 % et une spécificité proche de 99 % (6). Dans une population d'un million d'adultes de plus de 50 ans dans laquelle le taux de prévalence du cancer du colon serait de 5/100 000 cela reviendrait à pratiquer 20000 coloscopies. Une faible diminution de la spécificité augmenterait fortement le nombre de faux positifs et par conséquent le nombre de coloscopies inutiles, coûteuses, désagréables pour les individus et non dépourvues de risque iatrogène (tel que la perforation digestive). Pour une diminution de la spécificité de l'ordre de 4 %, il y aurait 30 000 coloscopies supplémentaires à pratiquer.

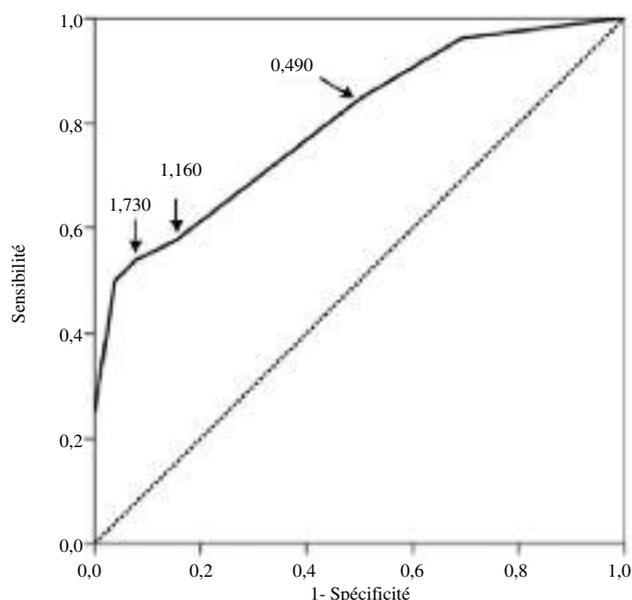


Figure 1. Courbe ROC de la procalcitonine (PCT) pour le diagnostic d'infection systémique sévère (trois valeurs seuils de PCT sont indiquées sur la courbe).

Dans notre exemple, le dosage de la PCT plasmatique fournit un résultat quantitatif, exprimé en ng/mL (il s'agit d'une réponse continue). Il est alors possible de considérer d'autres valeurs seuils que celle prise initialement (0,490 ng/mL) et de déterminer pour chacune d'elles la Se et la Sp correspondantes (il faut constituer autant de tableaux de structure identique à celle du Tableau I). Dans leur article, Massaro *et al.* ont ainsi pu considérer 7 valeurs seuils différentes (0,245 ; 0,275 ; 0,490 ; 1,160 ; 1,730 ; 2,145 ; 2,390) et calculer les valeurs de Se et Sp correspondantes. La représentation graphique des différents couples (Se ; 1 - Sp) constitue ce que l'on appelle une courbe ROC (Receiver-Operating Characteristic) qui permet d'identifier une éventuelle valeur seuil particulièrement pertinente en terme de Se et Sp. La figure 1 représente la courbe ROC de la PCT pour le diagnostic d'infection systémique sévère et montre que l'augmentation du seuil a pour conséquence de diminuer la Se et d'augmenter la Sp (le diagnostic est éliminé chez des patients ne présentant pas d'infection systémique sévère).

L'autre intérêt d'une courbe ROC est de fournir un indicateur global (pour toutes les valeurs seuils possibles) des capacités diagnostiques intrinsèques d'un test quantitatif grâce au calcul de son aire sous la courbe (Area Under the Curve, AUC). Cette aire fournit une estimation de la probabilité de bien classer 2 individus, un malade et un sain, et varie donc entre 0 et 1. Un test classant les individus au hasard donnerait une probabilité de 0,5 (un sujet sur deux va être bien classé du fait du hasard) qui correspond sur la courbe ROC, représentée dans un carré, à la diagonale première bissectrice (Se = 1 - Sp). Plus la courbe ROC tend vers le coin supérieur gauche (Se = Sp = 1) et s'éloigne de cette diagonale, plus l'aire sous la courbe se rapproche de 1 et la performance du test augmente. L'interprétation habituelle de l'AUC dans le domaine biomédical est la suivante : un test quantitatif a une performance modérée ou élevée si son AUC dépasse respectivement 0,7 ou 0,9 (7). L'aire sous la courbe calculée pour la PCT était de 0,80 (intervalle de confiance à 95 % : 0,69 - 0,92) et statistiquement différente de 0,5 ($p < 0,001$).

Valeurs prédictives

La Se et la Sp évaluent les capacités diagnostiques intrinsèques d'un test mais elles ne constituent pas une aide dans la

démarche décisionnelle faisant suite à la réalisation du test : quelle est la probabilité post-test (probabilité *a posteriori*) d'être atteint de la maladie étudiée. Sur l'échantillon de 52 adultes fébriles ayant une neutropénie, la prévalence des infections systémiques sévères était de 50 % (probabilité *a priori* de présenter une infection systémique sévère ; Tableau I). Parmi les 35 patients ayant une PCT $\geq 0,490$ ng/mL, 22 ont présenté une infection systémique sévère et sur les 17 patients ayant une PCT $< 0,490$ ng/mL, il y en a eu 13 qui n'ont pas présenté une infection systémique sévère (tableau I). Nous pouvons donc dire que, si un patient a une PCT $\geq 0,490$ ng/mL la probabilité qu'il présente une infection systémique sévère est de 0,628 (22/35). Par rapport à la probabilité *a priori* de présenter une infection systémique sévère une PCT $\geq 0,490$ ng/mL augmente la probabilité de 12,8 % (62,8 % - 50,0 %). Par ailleurs, si un patient a une PCT $< 0,490$ ng/mL la probabilité qu'il ne présente pas une infection systémique sévère est de 0,765 (13/17), soit une augmentation de 26,5 % (76,5 % - 50,0 %) par rapport à la probabilité *a priori* de ne pas présenter une infection systémique sévère.

On appelle valeur prédictive positive (VPP) d'un test la probabilité qu'un individu ayant un résultat positif au test soit effectivement malade et valeur prédictive négative (VPN) la probabilité d'être indemne si le résultat du test est négatif (Tableau II). La VPP se calcule en faisant le rapport du nombre de vrais positifs sur le nombre total de sujets ayant un résultat positif au test ($VPP = VP/(VP + FP)$) et la VPN se calcule en faisant le rapport du nombre de vrais négatifs sur le nombre total de sujets ayant un résultat négatif au test ($VPN = VN/(VN + FN)$). Elles s'expriment généralement en pourcentage. L'interprétation des valeurs prédictives doit être sujette à caution dans la mesure où elles dépendent de la prévalence de la maladie étudiée dans la population, et varient donc d'une population à l'autre. Une prévalence élevée augmente la VPP et diminue la VPN, et inversement. Il faut notamment faire attention au fait que dans certaines études le nombre de patients malades et le nombre de patients ne présentant pas la maladie étudiée est fixé *a priori* et que l'estimation des valeurs prédictives n'a alors pas de sens dans ce contexte. Les valeurs prédictives dépendent également de la Se et de la Sp du test.

Les rapports de vraisemblance

Un test diagnostique doit être considéré comme utile lorsqu'il modifie la probabilité du diagnostic après le résultat du test, permettant de passer d'un état décisionnel sous une certaine incertitude à un état décisionnel plus vraisemblable. Le résultat d'un test doit permettre de confirmer ou d'exclure le diagnostic avec la plus grande certitude possible.

Les rapports de vraisemblance (positifs : RVP et négatifs : RVN) intègrent dans un seul indicateur les qualités de Se et de Sp d'un test permettant de juger à la fois des capacités diagnostiques intrinsèques d'un test, d'apprécier l'intérêt clinique du test pour un patient donné tout en étant indépendants de la prévalence. Ils apportent ainsi des informations sur l'utilité du test. Le RVP ($Se/[1 - Sp]$) représente la crédibilité d'un résultat positif au test en mesurant, en cas de test positif, combien de fois il y a plus de malades que de non malades dans la population globale (on parle également de cote). Le RVN ($[1 - Se]/Sp$) représente la crédibilité d'un résultat négatif au test en mesurant combien de fois il y a moins de malades que de non malades en cas de test négatif que dans la population globale (Tableau II).

Le rapport de vraisemblance d'un test est compris entre 0 et l'infini. On considère généralement qu'un test est intéressant quand $RVP > 10$ ou quand $RVN < 0,1$ (8).

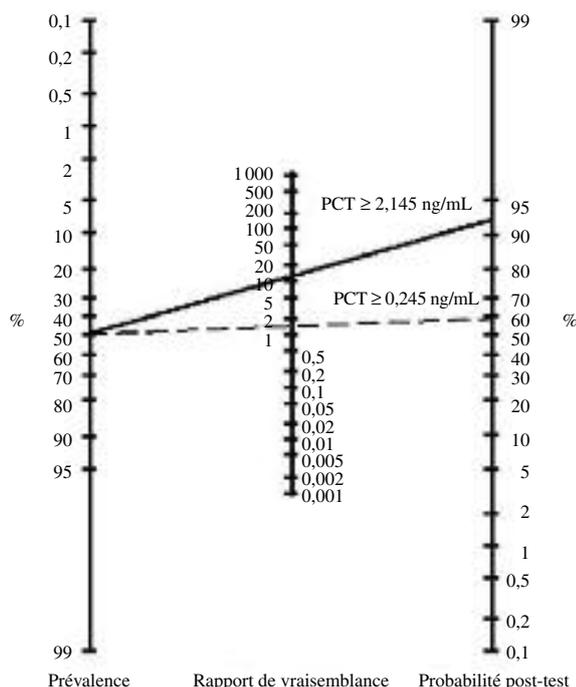


Figure 2. Nomogramme de Fagan permettant de déterminer la probabilité post-test à partir de la probabilité pré-test et du rapport de vraisemblance. Dans l'exemple, une procalcitonine (PCT) supérieure à 0,245 fait passer d'une probabilité de 50 % à 59 % (rapport de vraisemblance positif, RVP = 1,44) et une PCT supérieure à 2,145 fait passer d'une probabilité de 50 % à 93 % (RVP = 13,0).

Dans leur étude, Massaro *et al.* rapportent des valeurs de RVP allant de 1,44 (PCT ≥ 0,245 ng/mL) à 13,00 (PCT ≥ 2,145 ng/mL). Le RVP pour la valeur seuil de 0,490 ng/mL était de 1,69. Parmi les patients ayant une PCT ≥ 0,490 ng/mL, la présence d'une infection systémique sévère était 1,69 fois plus fréquente.

Plutôt que de raisonner en terme de côtes, et ayant connaissance de la probabilité pré-test d'une infection systémique sévère, il est possible d'utiliser le nomogramme de Fagan (9) qui est un outil graphique permettant de relier la probabilité post-test à la probabilité pré-test en passant par la valeur du RV. Ainsi, comme le montre la figure 2, une PCT ≥ 0,245 ng/mL fait passer d'une probabilité pré-test d'infection systémique sévère de 50 % à une probabilité post-test de 59 % (pour un RVP = 1,44) et une PCT ≥ 2,145 ng/mL fait passer d'une probabilité pré-test d'infection systémique sévère de 50 % à une probabilité post-test de 93 % (pour un RVP = 13,0).

Le diagnostic odds ratio

L'odds ratio (OR) est classiquement utilisé en épidémiologie pour mesurer la force de l'association entre un facteur de risque et une maladie (10). Par analogie, le diagnostic odds ratio (DOR) a été proposé pour mesurer la force de l'association entre le résultat d'un test diagnostique et une maladie (11). Contrairement aux indicateurs qui s'interprètent par paire (Se-Sp, RVP-RVN), le DOR offre l'avantage, à l'aide d'une valeur unique, de fournir une information sur la performance du test étudié et de ne pas dépendre de la prévalence. Il se calcule en faisant le rapport de la côte d'avoir un test positif chez les malades sur la côte d'avoir un test positif chez les non malades, ce qui correspond à $DOR = RVP/RVN = (Se/[1-Se])/([1-Sp]/Sp)$.

Le DOR varie de 0 à l'infini et s'interprète, comme l'OR, par rapport à la valeur 1 (test non discriminant) : un DOR supérieur à 1

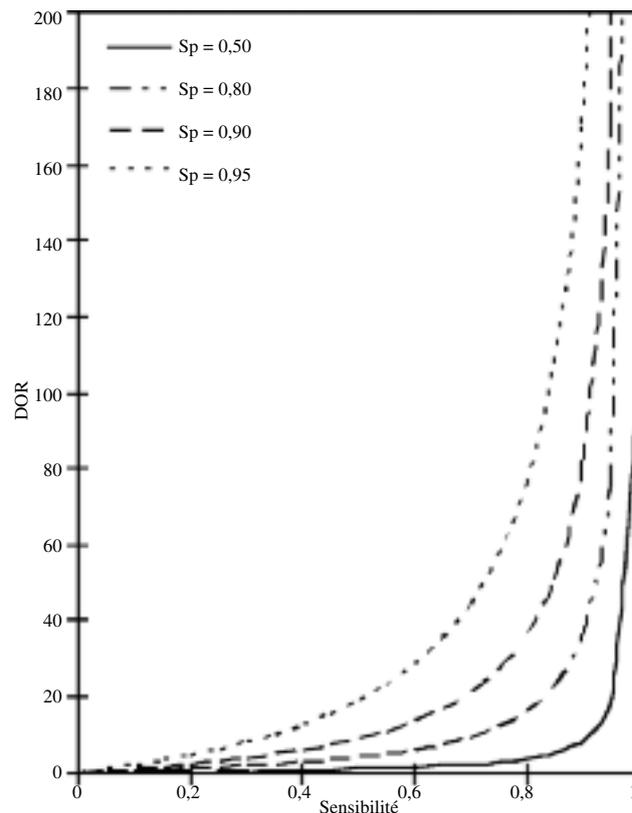


Figure 3. Evolution du diagnostic odds ratio (DOR) en fonction de la sensibilité et de certaines valeurs de spécificité (Sp).

signifie que le test est plus souvent positif chez les malades que chez les non malades (plus sa valeur augmente, plus le test est performant) ; un DOR inférieur à 1 signifie que le test est plus souvent positif chez les sujets non malades que chez les malades. Le DOR augmente fortement quand la Se ou la Sp se rapprochent de 1 (Fig. 3). Il faut cependant faire remarquer que l'avantage de résumer sensibilité et spécificité dans une même valeur perd tout son intérêt si une de ces deux caractéristiques du test doit être privilégiée.

Ainsi, avec les données de Massaro *et al.* on obtient un DOR de 4,10 (1,61/0,39) avec une valeur seuil de PCT à 0,490 ng/mL. Ce qui veut dire qu'il y avait dans cette étude 4,10 patients ayant une PCT ≥ 0,490 ng/mL et présentant une infection systémique sévère pour 1 patient ayant une PCT ≥ 0,490 ng/mL et ne présentant pas une infection systémique sévère.

Indicateur de précision

Comme la plupart des études de recherche clinique et épidémiologique, les études portant sur les tests diagnostiques sont réalisées sur un échantillon de patients issu de la population cible. On obtient alors des estimations (de la Se, Sp,...) soumises aux fluctuations d'échantillonnage et il est recommandé de calculer systématiquement leur intervalle de confiance (calculé le plus souvent avec un seuil de confiance de 5 %) afin d'avoir une idée de la précision de l'estimation. Des formules permettant le calcul de l'intervalle de confiance à 95 % des différents indicateurs traités dans cet article sont données dans le Tableau II. Ils peuvent être calculés sur les données de Massaro *et al.*

Pertinence clinique d'un test

Applicabilité

Pour pouvoir être utilisé en pratique, un test doit être accessible pour la société, en terme de disponibilité mais aussi de coût. Les arguments concernant l'applicabilité proviennent de faits scientifiques et il convient de distinguer les faits provenant de travaux de recherche très théoriques des faits provenant de travaux répondant à des questions plus pragmatiques. Des arguments en faveur de l'utilisation en pratique courante sont la validation du test dans des contextes différents (étude d'emblée multicentrique ou validation secondaire du test par une autre équipe clinique). Les études doivent aussi systématiquement s'intéresser à la faisabilité du test et décrire précisément les situations dans lequel le test n'a pas pu être réalisé. Pour pouvoir être utilisé en pratique, un test doit être d'utilisation simple et facilement acceptable par le patient.

Impact clinique

Il correspond à l'impact potentiel du test sur la prise en charge des patients dans le but d'améliorer leur état de santé. Une fois le résultat du test connu, la probabilité d'être atteint de la maladie obtenue influence-t-elle la prise en charge (mise en route d'un traitement, prescription d'un second examen plus invasif,...) ? L'impact clinique est généralement étudié comparativement aux autres tests existants afin de déterminer si le nouveau test apporte un bénéfice en terme de capacité diagnostique, de rapidité, de rapport coût/bénéfice et coût/utilité... Il existe différents tests statistiques permettant de comparer entre elles les performances de deux ou plusieurs tests évalués par rapport au même test de référence. De telles comparaisons peuvent être effectuées pour tous les indicateurs de performance présentés dans cet article (11-13). Par exemple, Massaro *et al.* ont comparé l'AUC de la PCT à celle de la protéine C réactive et ont montré sa probable supériorité.

Qualité des études de tests diagnostiques

Des critères ont été établis pour juger de la qualité des études portant sur les tests diagnostiques (14) : comparaison indépendante, en aveugle, avec un gold-standard ; spectre suffisamment large de patients pour lesquels ce test est utilisé en pratique clinique ; absence d'influence du résultat du test sur la décision de pratiquer le gold-standard ; description en détail de la méthode du test ; validation externe indépendante ; caractère discriminant du test ; applicabilité à nos patients. De plus, un certain nombre de revues

médicales scientifiques ont adopté la norme STARD (Standard for Reporting of Diagnostic Accuracy) pour la rédaction de telles études (15).

Conclusion

La sensibilité, la spécificité, les valeurs prédictives, ainsi que la courbe ROC et l'aire sous la courbe en cas de test à réponse continue, sont des indicateurs permettant de juger de la performance d'un test diagnostique. Bien que moins fréquemment utilisés à ce jour, le diagnostic odds ratio et les rapports de vraisemblance sont des indicateurs appréciant respectivement la performance et l'utilité d'un test. Ils offrent l'avantage d'intégrer les qualités de la sensibilité et de la spécificité.

Références

- Hamer DH, Ndhlovu M, Zurovac D, Fox M, Yeboah-Antwi K, Chanda P, *et al.* Improved diagnostic testing and malaria treatment practices in Zambia. *JAMA* 2007 ; 297 : 2227-31.
- Dosanjh DP, Hinks TS, Innes JA, Deeks JJ, Pasvol G, Hackforth S, *et al.* Improved diagnostic evaluation of suspected tuberculosis. *Ann Intern Med* 2008 ; 148 : 398-9.
- Vriesema JL, Atsma F, Kiemeny LA, Peelen WP, Witjes JA, Schalken JA. Diagnostic efficacy of the ImmunoCyt test to detect superficial bladder cancer recurrence. *Urology* 2001 ; 58 : 367-71.
- Massaro KS, Costa SF, Leone C, Chamone DA. Procalcitonin (PCT) and C-reactive protein (CRP) as severe systemic infection markers in febrile neutropenic adults. *BMC Infect Dis* 2007 ; 7 : 137.
- Hughes WT, Armstrong D, Bodey GP, Bow EJ, Brown AE, Calandra T, *et al.* 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer. *Clin Infect Dis* 2002 ; 34 : 730-51.
- Mitry E. Rationnel du dépistage. In : Le dépistage du cancer colorectal. État des lieux et perspectives. É. Dorval. Ed. Springer-Verlag ed, France, 2006.
- Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988 ; 240 : 1285-93.
- Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004 ; 329 : 168-9.
- Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975 ; 293 : 257.
- Ollivier L, Michel R, Spiegel A, Boutin JP. Les mesures d'association en épidémiologie. *Med Trop* 2003 ; 63 : 75-8.
- Glas AS, Lijmer JG, Prins MH, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003 ; 56 : 1129-35.
- Bennett BM. On comparisons of sensitivity, specificity and predictive value of a number of diagnostic procedures. *Biometrics* 1972 ; 28 : 793-800.
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983 ; 148 : 839-43.
- Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994 ; 271 : 703-7.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy : the STARD initiative. *BMJ* 2003 ; 326 : 41-4.