

Indices de performance diagnostique*

Isabelle Colombet¹, Emmanuel Touzé²

¹ Hôpital Cochin, Université Paris-Descartes, SPIM, 15, rue de l'école de médecine, 75006 France, Paris
<isabelle.colombet@parisdescartes.fr>

² Hôpital Sainte-Anne, université Paris-Descartes, service de neurologie, Inserm UMR S894, Paris, France

La série d'articles coordonnée par Emmanuel Touzé, publiée en 2011, avait soulevé l'enthousiasme de nos lecteurs. Il nous a donc paru intéressant de proposer à notre lectorat une actualisation de ces articles, auxquels s'ajouteront de nouveaux articles. Emmanuel Touzé s'est impliqué à nouveau dans cette deuxième édition, qui, nous l'espérons, continuera à passionner notre lectorat.

Résumé. Les propriétés permettant de caractériser un test diagnostique sont de trois ordres : qualités métrologiques, propriétés informationnelles ou discriminantes et impact sur la prise en charge ou sur l'issue clinique. Cet article se concentre sur la définition, l'évaluation et l'utilisation des indices informationnels d'un test diagnostique, en s'appuyant prioritairement sur le cas d'un test donnant un résultat qualitatif binaire. Selon le contexte de raisonnement, recherche en évaluation diagnostique ou aide à la décision clinique, on exprimera ces indices sous forme de fréquences observées ou de probabilités. Les « qualités informationnelles intrinsèques » du test (sensibilité et spécificité, dont les valeurs se conjuguent dans les rapports de vraisemblance) renseignent sur la capacité informative du test, quel que soit le contexte dans lequel il est utilisé. Les « qualités diagnostiques extrinsèques » (valeurs prédictives) reflètent la probabilité diagnostique à laquelle le résultat du test permet d'aboutir, probabilité qui dépend aussi forcément de la probabilité pré-test, c'est-à-dire de la situation clinique dans laquelle est utilisé le test. Les rapports de vraisemblance sont particulièrement intéressants car ils expriment le gain intrinsèquement apporté par un résultat du test ; ils permettent une représentation graphique de la probabilité post-test en fonction de la probabilité pré-test.

Mots clés: diagnostic, sensibilité, spécificité, rapports de vraisemblance, processus décisionnel

Abstract

Indices of diagnostic performance

Properties to characterize a diagnostic test are threefold: metrological, informational or diagnostic accuracy and impact on care or clinical outcome. This article focuses on the assessment and use of diagnostic accuracy, based primarily on the case of a test giving a qualitative result. Depending on the context of reasoning, research or clinical decision support, these indices will be expressed either observed frequencies or probabilities. Intrinsic diagnostic properties (sensitivity and specificity, which values are combined in the likelihood ratios) indicate the gain of information provided by the test, whatever the context in which it is used. Extrinsic diagnostic properties (predictive values) reflect the probability of the disease once the test result is known, which also depends on the pretest probability, e.g. the clinical situation in which it is used. Likelihood ratios are particularly interesting as they allow graphical representation of the post-test probability based on the pre-test probability.

Key words: diagnosis, sensitivity, specificity, likelihood ratios, medical decision-making

Tirés à part :

I. Colombet

* Actualisation de l'article paru dans *STV* en 2011 : Colombet I, Touzé E. Indices de performance diagnostique. *Sang Thrombose Vaisseaux* 2011 ; 23 (6) : 307-16. doi:10.1684/stv.2011.0619.

Les tests diagnostiques sont utilisés à tous les stades de l'investigation et du raisonnement médical : depuis l'examen clinique du patient, au moment de la mise en œuvre d'examen complémentaires (examens de laboratoire ou d'imagerie) pour le diagnostic initial, positif ou différentiel, ou au cours du suivi de l'évolution de la maladie. Les propriétés permettant de caractériser un test diagnostique sont de trois ordres :

– les qualités métrologiques du test correspondent aux qualités de la mesure sur laquelle repose le résultat du test. Il s'agit d'abord de sa fiabilité, qui résulte de l'exactitude et de la précision de la mesure. Elle est évaluée à partir d'un échantillon de mesures en répondant aux questions suivantes :

- existe-t-il une erreur systématique dans l'estimation de la mesure (la valeur moyenne de l'échantillon est-elle proche de la valeur réelle ? C'est l'exactitude) ;
- existe-t-il une erreur aléatoire dans l'estimation de la mesure (la dispersion des mesures autour de la moyenne est-elle importante ou large ? C'est la précision) ?

Il s'agit ensuite de la variabilité de la mesure, évaluée par sa reproductibilité [1] ;

– les qualités informationnelles du test représentent sa capacité à discriminer les patients atteints de la maladie des patients qui n'en sont pas atteints, ou, exprimé autrement, le gain d'information apporté par le résultat du test dans un processus d'élaboration du diagnostic. Ce gain fait progresser le raisonnement médical mais n'apporte pas toujours une certitude diagnostique. Ces notions de raisonnement incertain et de gain d'information conduisent, assez naturellement, à s'appuyer sur les probabilités pour définir et surtout exploiter les qualités d'un test afin de prendre des décisions en pratique clinique.

L'utilisation des propriétés diagnostiques d'un test diagnostique pour la décision médicale suppose évidemment de les avoir évaluées en amont, dans le cadre d'une démarche de recherche clinique. Cette démarche s'appuie sur l'observation d'un échantillon ; elle ne manipule pas des probabilités mais des fréquences d'observation. En tête de son chapitre sur les qualités diagnostiques d'un test, Grenier souligne la confusion fréquente des deux termes : une fréquence est une donnée d'observation alors qu'une probabilité est l'extrapolation de cette donnée projetée pour le temps présent (dans le cadre d'une démarche diagnostique) ou pour le futur (dans le cadre d'une démarche de prédiction) [2].

Il est en fait utile, pour appréhender les indices de performance diagnostiques d'un test, de « jongler » entre les deux registres de vocabulaire, c'est-à-dire entre les deux démarches (de recherche, en population, et de pratique médicale, en présence d'un individu) [3, 4] ;

– l'impact du test pour le patient est l'utilité du test pour améliorer l'état de santé du patient. L'évaluation de cet impact doit aussi prendre en compte l'innocuité du test pour le patient. Elle nécessite une étude expérimentale incluant des patients ayant une suspicion clinique de la maladie et alloués aléatoirement soit à une stratégie diagnostique incluant le nouveau test, soit à une stratégie standard ne l'incluant pas.

Nous nous concentrerons dans cet article sur la définition, l'évaluation et l'utilisation des qualités informationnelles d'un test diagnostique, en s'appuyant prioritairement sur le cas d'un test donnant un résultat qualitatif binaire (résultat positif ou négatif).

Définitions des indices

Exemple tiré de Oudkerk *et al.* Lancet 2002 ; 359 : 1643-47

L'objectif de l'étude est d'évaluer la performance diagnostique de l'angiographie par résonance magnétique avec produit de contraste (angio-IRM) par référence à l'angiographie pulmonaire conventionnelle pour le diagnostic d'embolie pulmonaire. Des résultats d'angio-IRM et d'angiographie conventionnelle sont disponibles pour 118 patients inclus dans l'étude sur une suspicion d'embolie pulmonaire (tableau 1).

La proportion de patients atteints de la maladie définit la prévalence de la maladie observée dans la population d'étude : elle est de $35/118 = 30\%$.

Si l'on considère l'angiographie conventionnelle comme méthode de référence pour affirmer le diagnostic d'embolie pulmonaire :

- les 27 patients positifs pour l'angio-IRM et malades (confirmés atteints d'embolie pulmonaire) sont les vrais positifs (VP) ;
- les 81 patients négatifs pour l'angio-IRM et non malades sont les vrais négatifs (VN) ;
- les deux patients positifs pour l'angio-IRM et néanmoins non malades sont les faux positifs (FP) ;

Tableau 1.

Angiographie conventionnelle			
Angio-IRM	EP présente	EP absente	Total
Positive	27	2	29
Négative	8	81	89
Total	35	83	118

EP : embolie pulmonaire.

– les huit patients négatifs pour l’angio-IRM et néanmoins malades sont les faux négatifs (FN).

Sensibilité et spécificité

La proportion de patients atteints d’embolie pulmonaire (selon l’angiographie conventionnelle), correctement identifiés par l’angio-IRM, rend compte de la sensibilité du test : $27/35 = 77\%$ (IC 95 % : 63-91 %). La proportion de patients non atteints (selon l’angiographie conventionnelle), correctement éliminés par l’angio-IRM, rend compte de la spécificité du test : $81/83 = 98\%$ (IC 95 % : 95-101 %). Comme toute proportion estimée à partir d’un échantillon, les estimations de sensibilité et spécificité doivent être présentées assorties de leurs intervalles de confiance (cf. méthode de calcul en *annexe 1*).

On définit ainsi :

- la sensibilité, comme la proportion de VP chez les malades : $VP/(VP + FN)$;
- la spécificité, comme la proportion de VN chez les non-malades : $VN/(FP + VN)$.

Un test ayant une sensibilité élevée sera performant pour éliminer le diagnostic lorsqu’il est négatif. Inversement, un test ayant une spécificité élevée sera performant pour affirmer le diagnostic lorsqu’il est positif.

Rapports de vraisemblance

Le rapport de vraisemblance positif (RV+) représente la vraisemblance d’obtenir un test positif chez des patients atteints de la maladie, rapporté à la vraisemblance d’obtenir un test positif chez des patients non atteints. C’est le rapport du taux de VP (chez les malades) sur le taux de FP (chez les non-malades), soit :

$$RV+ = \frac{VP/(VP + FN)}{FP/(FP + VN)} \quad \text{soit} \\ RV+ = \frac{\text{sensibilité}}{1 - \text{spécificité}}$$

Il prend donc ses valeurs dans l’intervalle $[1 ; +\infty]$, témoignant d’un test d’autant plus performant pour discriminer les malades des non-malades qu’il s’éloigne de 1. Le sujet a ainsi RV+ fois plus de risque d’avoir un test positif s’il est atteint de la maladie étudiée que s’il n’est pas atteint de la maladie.

De même, le rapport de vraisemblance négatif (RV-) représente la vraisemblance d’obtenir un test négatif chez des patients atteints de la maladie, rapporté à la vraisemblance d’obtenir un test négatif chez des patients non atteints. C’est le rapport du taux de FN (chez les malades) sur le taux de VN (chez les non-malades), soit :

$$RV- = \frac{FN/(VP + FN)}{VN/(FP + VN)} \quad \text{soit} \quad RV- = 1 - \frac{\text{sensibilité}}{\text{spécificité}}$$

Tableau 2. Ordre de grandeurs pour l’interprétation des rapports de vraisemblance (RV).

RV+	RV-	Propriétés discriminantes
> 10	< 0,1	Importante
5-10	0,1-0,2	Modérée
2-5	0,5-0,2	Faible
1-2	0,5-1	Très faible

Il prend donc ses valeurs dans l’intervalle $[1 ; 0]$, témoignant d’un test d’autant plus performant pour discriminer les non-malades des malades qu’il se rapproche de 0. Le sujet a ainsi RV fois plus de risque d’avoir un test négatif s’il est atteint de la maladie étudiée que s’il n’est pas atteint de la maladie. À titre indicatif, l’échelle suivante permet d’interpréter le niveau de propriété discriminante du test, à partir de la valeur de ses rapports de vraisemblance (*tableau 2*).

Exemple : calculés à partir du *tableau 1*, les rapports de vraisemblance positifs et négatifs de l’angio-IRM pour le diagnostic d’embolie pulmonaire sont respectivement (le calcul des intervalles de confiance est détaillé en *annexe 1*) :

$RV+ = 32$ [IC95 % : 8-127] et $RV- = 0,23$ [IC95 % : 0,130,43]
Une angio-IRM positive a donc une très forte valeur discriminante pour affirmer le diagnostic d’embolie pulmonaire. En revanche, une angio-IRM négative n’a qu’une faible valeur pour éliminer le diagnostic.

Valeurs prédictives

Toujours d’après les données du *tableau 1* :

- la proportion de patients atteints d’embolie pulmonaire parmi les patients positifs pour l’angio-IRM est $27/29 = 93\%$. Cela définit la valeur prédictive positive (VPP) = $VP/(VP + FP)$;
- la proportion de patients indemnes d’embolie pulmonaire parmi les patients négatifs pour l’angio-IRM est de $81/89 = 91\%$. Cela définit la valeur prédictive négative (VPN) = $VN/(VN + FN)$.

Utilisation des propriétés diagnostiques pour la décision médicale

En pratique, les médecins sont assez familiers des indices de sensibilité, spécificité, et valeurs prédictives mais beaucoup moins des rapports de vraisemblance. Pourtant, les rapports de vraisemblance présentent de nombreux avantages théoriques [5]. Comme nous allons le montrer, les valeurs

prédictives présente l'inconvénient majeur de dépendre à la fois de la sensibilité et de la spécificité du test, mais aussi de la prévalence de la maladie.

Des fréquences aux probabilités

En situation clinique, avant de pratiquer une angio-IRM chez un patient donné, le médecin formule plus ou moins explicitement, à partir des informations immédiatement disponibles (douleur thoracique, dyspnée, antécédent récent d'une chirurgie du genou, D-dimères positifs, etc.), une probabilité que le patient ait une embolie pulmonaire. Cette probabilité représente la probabilité pré-test ou *a priori*. Le passage à un raisonnement probabiliste va permettre d'exprimer concrètement l'apport de l'angio-IRM pour poser le diagnostic d'embolie pulmonaire.

Admettons que la probabilité pré-test que le patient ait une embolie pulmonaire soit estimée par le médecin à 60 %. Cela signifie que sur 1 000 patients présentant le même tableau clinique, 600 auraient une embolie pulmonaire. La connaissance des valeurs de sensibilité et spécificité, propriétés intrinsèques de l'angio-IRM (respectivement 77 % et 98 %), permet d'extrapoler, dans cet échantillon fictif, les résultats présentés dans le *tableau 3*.

La VPP, calculée à partir du *tableau 3*, est de 98 %, ce qui signifie qu'un patient suspect d'embolie pulmonaire avec une probabilité pré-test de 60 %, voit sa probabilité monter à 98 % si l'angio-IRM est positive. La valeur prédictive équivaut à la probabilité post-test. La VPN calculée à partir du *tableau 3* est de 74 %, ce qui signifie également qu'un patient suspect d'embolie pulmonaire avec une probabilité pré-test de 60 %, voit sa probabilité descendre à 26 % (100-74 %) si l'angio-IRM est négative. La valeur prédictive négative équivaut à (1-probabilité post-test).

Outre le fait de passer des fréquences observées dans un échantillon d'étude aux probabilités utilisables pour la décision, l'exemple ci-dessus illustre qu'un même test, dont la sensibilité et la spécificité sont intrinsèquement fixées, utilisé dans des conditions de prévalence différentes, donnera des valeurs prédictives différentes. Plus la prévalence est élevée, plus les probabilités post-test sont élevées.

Tableau 3.

Angiographie conventionnelle			
Angio-IRM	EP présente	EP absente	Total
Positive	462 (77 %)	8	470
Négative	138	392 (98 %)	530
Total	600	400	1 000

EP : embolie pulmonaire.

Tests diagnostiques et raisonnement bayésien

Rappel du théorème de Bayes

En théorie des probabilités, le théorème de Bayes énonce des probabilités conditionnelles. Soit A et B deux événements, p(A) et p(B) leurs probabilités de survenue et p(A ∩ B) la probabilité que A et B surviennent en même temps. On définit les probabilités conditionnelles de survenue de A sachant B et de B sachant A comme :

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

Le théorème de Bayes s'appuie sur cette définition pour déterminer la probabilité de A sachant B, si l'on connaît les probabilités de A, de B et de B sachant A.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Si l'on considère chaque case du tableau de contingence comme l'intersection de deux ensembles (par exemple : les VP représentent l'intersection de l'ensemble des patients malades avec l'ensemble des patients ayant un test positif), tous les indices informationnels peuvent ainsi s'exprimer sous la forme de probabilités conditionnelles (*tableau 4*).

Il découle de ce raisonnement deux propriétés importantes, l'une concernant les valeurs prédictives, l'autre impliquant les rapports de vraisemblance.

Expression des valeurs prédictives (ou probabilités post-test) en fonction de la prévalence (ou probabilité pré-test)

La première est l'expression des valeurs prédictives en fonction de la prévalence, de la sensibilité et de la spécificité :

$$\text{partant de : } p(M+/T+) = \frac{p(T+/M+) \times [p(M+)]}{p(T+)}$$

$$\text{on déduit : } VPP = \frac{Se \times p}{Se \times p + (1 - Sp) \times (1 - p)}$$

de même, partant de :

$$p(M-/T-) = \frac{p(T-/M-) \times [p(M-)]}{p(T-)}$$

$$\text{on déduit : } VPN = \frac{Sp \times (1 - p)}{Sp \times (1 - p) + (1 - Se) \times p}$$

Ces deux expressions des valeurs prédictives permettent de construire la courbe présentée dans la *figure 1*. Cette courbe représente, en fonction de la probabilité pré-test (équivalent à la prévalence), la valeur de probabilité post-test (équivalent à une VPP, si le résultat du test est positif, ou à une VPN si le résultat du test est négatif). Cette représentation permet donc de visualiser, pour toute valeur de prévalence estimée dans un contexte clinique particulier,

Tableau 4. Tableau récapitulatifs des indices de performances diagnostiques, exprimés en termes de fréquences ou de probabilités.

Prévalence de la maladie (p)	(VP + FN)/(VP + FN + VN + FP)	p(M +)
Sensibilité (Se)	VP/(VP + FN)	p (T + / M +)
Spécificité (Sp)	(VN)/(VN + FP)	p (T - / M -)
Valeur prédictive positive (VPP, probabilité post-test +)	(VP)/(VP + FP)	p(M + / T +)
Valeur prédictive négative (VPN)	(VN)/(VN + FN)	p(M - / T -)
1-VPN (probabilité post-test -)	(FN)/(VN + FN)	p(M + / T -)
Rapport de vraisemblance positif (RV +)	[VP/(VP + FN)]/[FP/(VN + FP)]	p (T + / M +)/p (T + / M -)
Rapport de vraisemblance négatif (RV-)	[FP/(VP + FN)]/[VN/(VN + FP)]	p (T - / M +)/p (T - / M -)

le gain d'information intrinsèquement apporté par le résultat du test. Elle constitue une aide à la décision adaptée en situation clinique.

Expression des probabilités post-test, connaissant la probabilité pré-test et les rapports de vraisemblance

Du même théorème de Bayes, et en passant par la transformation des probabilités en cotes (une cote représente le rapport d'une probabilité sur sa complémentaire), on montre que :

$$Cote_{post-test} = RV \times Cote_{pré-test}$$

Démonstration :

Rappel : une cote est égale au rapport d'une probabilité sur sa complémentaire (p/(1-p))

On a donc : $Cote_{pré-test} = p(M+)/p(M-)$

On sait, par le théorème de Bayes que :

$$p(M+/T+) = \frac{p(T+/M+) \times P(M+)}{p(T+)} \quad \text{et}$$

$$p(M-/T+) = \frac{p(T+/M-) \times P(M-)}{p(T+)}$$

On en déduit :

$$\frac{P(M+/T+)}{P(M-/T+)} = \frac{p(T+/M+)}{p(T+/M-)} \times \frac{P(M+)}{P(M-)}$$

$$Cote_{post-test} = RV \times Cote_{pré-test}$$

C'est cette relation qui est utilisée pour construire le nomogramme de Fagan (figure 2). Ce nomogramme permet de visualiser la VPP (probabilité post-test), en fonction de la prévalence (probabilité pré-test) et connaissant les rapports de vraisemblance, qualités intrinsèques du test, qui ne dépendent pas de la prévalence de la maladie au sein de l'échantillon qui a servi pour les estimer. Il représente

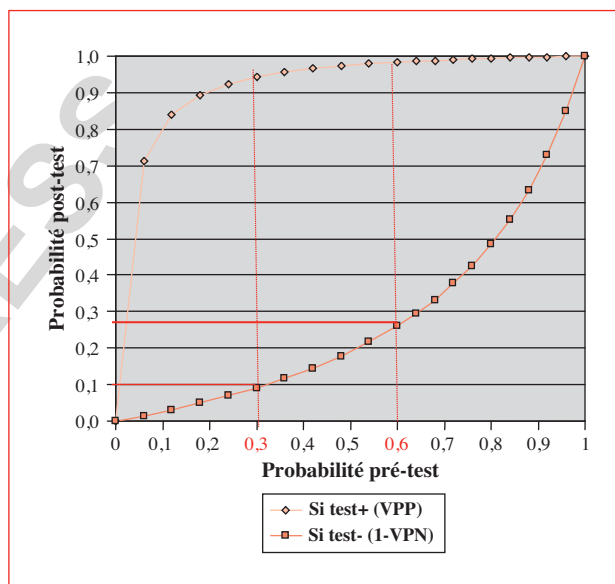


Figure 1. Probabilité d'embolie pulmonaire post-angio-IRM (résultat positif ou négatif) pour différentes valeurs de prévalence (probabilité pré-test d'embolie pulmonaire).

respectivement, sur trois axes verticaux, avec une échelle logarithmique : la probabilité pré-test, le rapport de vraisemblance positif (ou négatif) et la probabilité post-test positif (ou négatif).

Dans la figure 1 comme dans la figure 2, l'utilisation des qualités informationnelles de l'angio-IRM (respectivement les sensibilité, spécificité ou les rapports de vraisemblance) permet d'illustrer qu'une probabilité pré-test de 60 % peut être élevée à 98 % en cas de test positif, ou peut être abaissée à 26 % en cas de test négatif, alors qu'une probabilité pré-test de 30 % sera élevée à 93 % en cas de test positif et abaissée à 9 % en cas de test négatif.

L'angio-IRM est donc d'autant plus intéressante pour confirmer le diagnostic d'embolie pulmonaire que la présomption clinique de la maladie est déjà assez élevée, alors

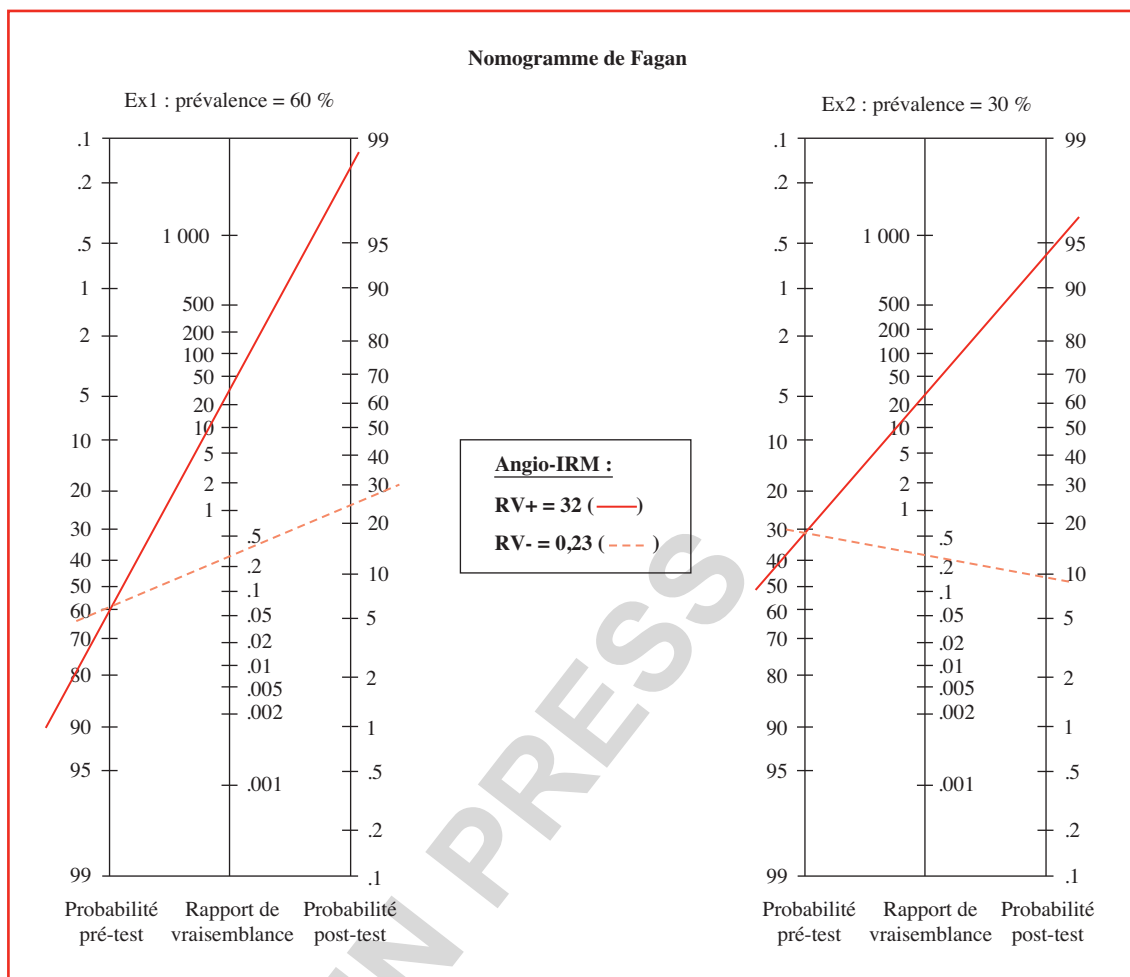


Figure 2. Application du nomogramme de Fagan.

qu'un résultat négatif de cet examen ne suffira pas à éliminer formellement le diagnostic, même si la présomption clinique de la maladie est déjà assez faible. Ainsi, le rapport de vraisemblance (*likelihood ratio*) est un indicateur plus utile pour la décision que la sensibilité et la spécificité, car il donne une mesure indirecte renforçant la probabilité de la maladie lorsqu'il est supérieur à 1 ou, inversement, la diminuant lorsqu'il est inférieur à 1.

Cas d'un test à réponse quantitative (ou variable continue)

Dans le cas d'un test diagnostique biologique, tel que les tests biochimiques ou hématologiques, la distribution des résultats du test dans un échantillon de sujets malades et dans un échantillon de sujets non malades peut être représentée comme sur la *figure 3*. La fixation d'un seuil de

normalité S (ou ligne de partage) définit la répartition des VP, VN, FP et FN.

Ainsi, le déplacement de la ligne de partage du test modifie les proportions de VP, VN, FP, FN. La représentation graphique des différents points (taux de VP, taux de FP) définis par tous les seuils possibles du test, permet de construire une courbe ROC (pour *receiving operating curve* [figure 3]) [6]. Cette courbe présente l'avantage de rendre compte des propriétés diagnostiques intrinsèques (sensibilité et spécificité) d'un test à réponse quantitative continue, quel que soit le seuil choisi pour définir une réponse normale (négative) ou anormale (positive).

Plus la courbe s'étire vers le haut à gauche, plus elle a tendance à passer par un point d'abscisse 0 et d'ordonnée 1, ce qui signifie qu'il existe un seuil du test pour lequel la sensibilité et la spécificité se rapprochent de 100 % (figure 4). Inversement, si la courbe se rapproche de la diagonale, jusqu'à s'y confondre, cela signifie que pour le plus

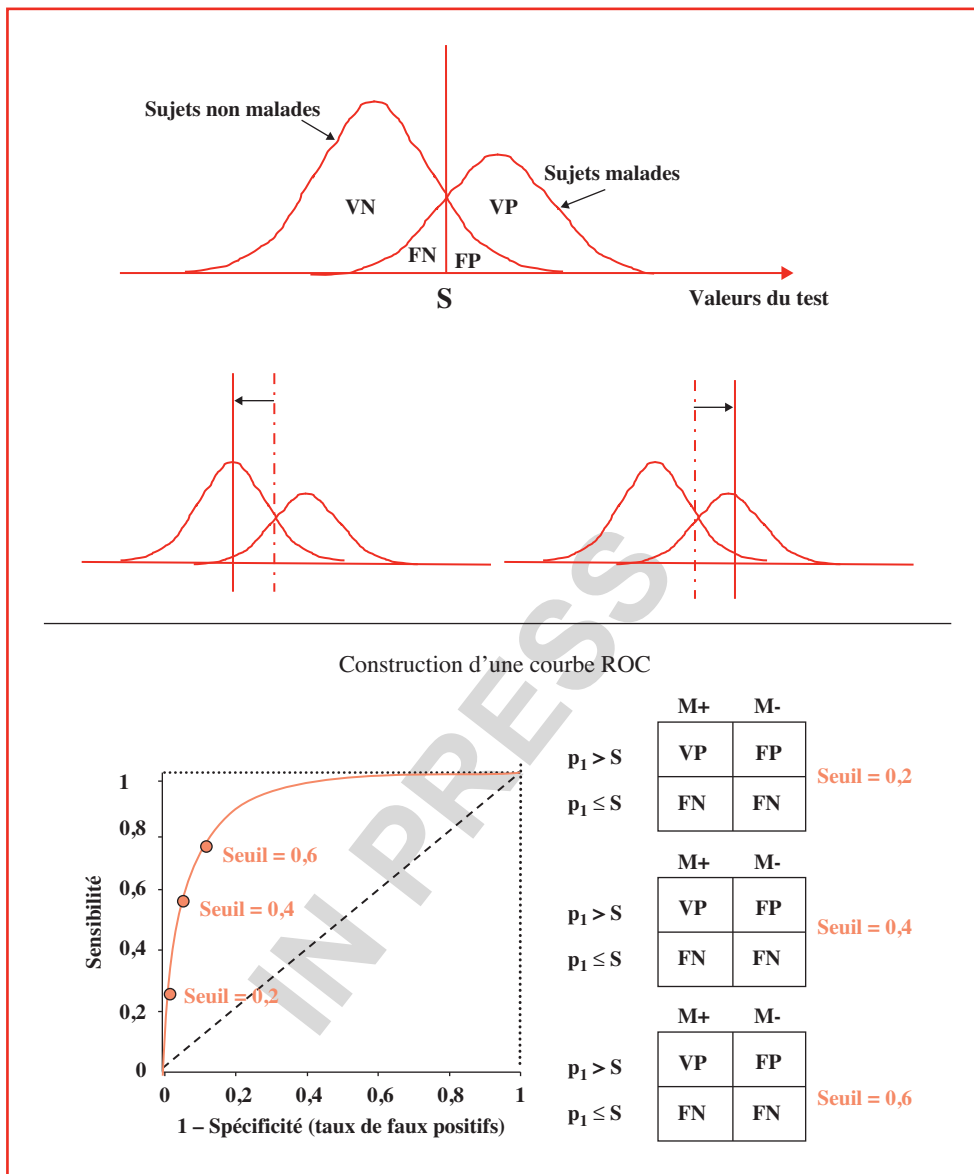


Figure 3. Représentation des propriétés discriminantes d'un test donnant une réponse quantitative (courbe ROC).

grand nombre de seuils possibles, la sensibilité est égale à (1-spécificité), ce qui équivaut à un rapport de vraisemblance positif = 1, donc un test non discriminant, dont le gain informationnel par rapport à la probabilité pré-test est nul.

L'aire sous la courbe ROC (ou *area under the curve* [AUC]), représente un indice global de la performance discriminante du test, quel que soit le seuil considéré.

En dessous de la diagonale, la courbe ROC serait celle d'un test dont la négativité (et non plus la positivité) serait liée à la maladie étudiée.

Synthèse

Pour la pratique : utilisation des propriétés diagnostiques

Pour utiliser les différents indices qui ont été définis, il est intéressant de revenir sur leurs caractères intrinsèques ou extrinsèques.

Les qualités informationnelles intrinsèques du test sont les sensibilité et spécificité. Leurs valeurs se conjuguent dans les rapports de vraisemblance. Elles renseignent sur la

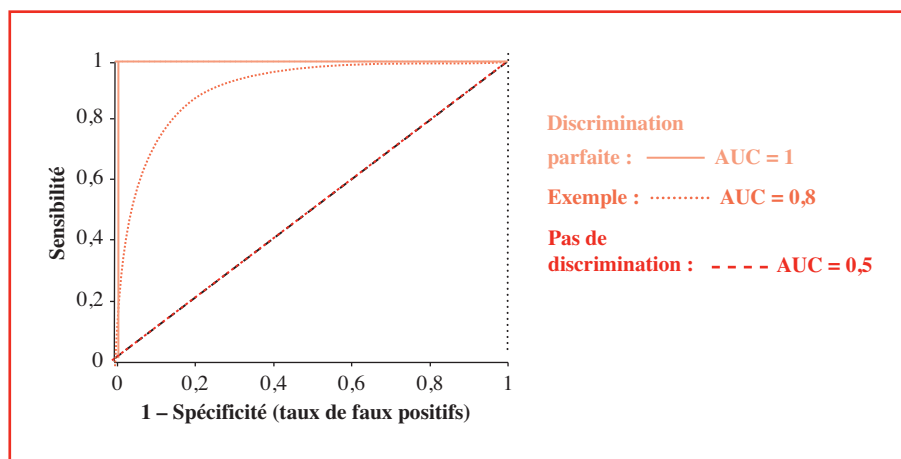


Figure 4. Signification de l'aire sous la courbe (*receiving operating curve* [ROC]) (AUC).

capacité informative du test, c'est-à-dire le degré auquel le test va modifier l'évaluation du diagnostic :

- un test dont le rapport de vraisemblance négatif est très bas (ce qui s'associe à une sensibilité élevée) est performant pour éliminer un diagnostic : son résultat, s'il est négatif, permettra de réduire drastiquement la probabilité de la maladie ;
- un test dont le rapport de vraisemblance positif est élevé (ce qui s'associe à une spécificité élevée) est performant pour confirmer un diagnostic : son résultat positif permettra d'approcher à 1 la probabilité de la maladie.

Les qualités diagnostiques extrinsèques sont les valeurs prédictives. Elles reflètent la probabilité diagnostique à

laquelle le résultat du test permet d'aboutir ; elles dépendent donc forcément de la probabilité pré-test, c'est-à-dire de la situation clinique dans laquelle est utilisé le test. À ce titre, les représentations graphiques des probabilités post-test en fonction de la probabilité pré-test sont sans doute les plus « parlantes » pour le médecin clinicien.

Ainsi, si dans le cadre d'une démarche de recherche, il est important de se concentrer sur l'estimation des qualités intrinsèques du test, qui seules seront généralisables (indépendantes de la prévalence), l'utilisation et l'interprétation du test en situation clinique va elle s'appuyer sur l'estimation pour un individu d'une probabilité pré-test et une probabilité pré-test. C'est pourquoi, les

Tableau 5. Différents types d'études d'évaluation diagnostique (exemple d'évaluation de la performance du *brain natriuretic peptide* (BNP) pour le diagnostic d'insuffisance cardiaque) et leur niveau d'interprétation (d'après [8]).

Types d'étude et de recrutement d'un échantillon de patients (exemple)	3 niveaux d'interprétation
Comparaison de 2 échantillons de cas et de témoins	Comparaison des distributions des résultats du test
Comparaison de cas et de témoins Mesure du BNP chez des contrôles normaux et chez des patients présentant une dysfonction VG à différents degrés	Les patients ayant un résultat élevé du test sont plus vraisemblablement atteints de la maladie. Définition de la valeur seuil qui sépare le mieux les malades des contrôles
Étude transversale (recrutement prospectif et consécutif de patients adressés à l'hôpital par leur médecin généraliste pour suspicion de dysfonction du ventricule gauche, test index et test de référence pratiqués indépendamment)	Les résultats du test permettent de distinguer les patients avec et sans la maladie, parmi ceux chez qui il est cliniquement plausible de la suspecter. Le mode de recrutement de la population de l'étude permet une estimation de la prévalence de la maladie
Essai contrôlé randomisé Suivi comparatif d'un groupe de patients ayant eu le test et d'un groupe de patients ne l'ayant pas eu, avec allocation aléatoire à l'une des 2 stratégies	Les patients testés évoluent plus favorablement que des patients similaires non testés ?

abaques, permettant de représenter ces probabilités l'une en fonction de l'autre, connaissant les sensibilité, spécificité ou rapports de vraisemblance du test, sont particulièrement intéressantes pour l'usage clinique.

Pour la recherche: estimation des propriétés diagnostiques

Dans le cadre d'une démarche de recherche pour l'estimation des propriétés diagnostiques intrinsèques d'un test, plusieurs types d'études sont appropriés, selon une gradation d'exigence méthodologique, qui a été décrite par Sackett et Haynes (*tableau 5*) [7]. Les niveaux de Sackett peuvent être en quelque sorte compris par analogie aux quatre phases de développement et d'évaluation d'un médicament.

Le troisième type d'évaluation est celui qui permet d'obtenir une estimation des propriétés diagnostiques dans les conditions les plus proches de la pratique réelle [8].

Les résultats de ce type d'étude doivent s'exprimer le plus logiquement en recourant aux propriétés intrinsèques du test, qui seules sont « transportables » à tout contexte d'utilisation. Un certain nombre de biais doivent être évités, notamment concernant les modalités de recrutement de patients pour l'étude (biais de sélection) et l'indépendance entre l'interprétation du test évalué et du test de référence. Enfin, comme toute estimation à partir d'un échantillon ou d'une population, les indices informationnels doivent être présentés assortis d'un intervalle de confiance (*cf.* calcul en *annexe 1*). ■

Liens d'intérêts : les auteurs déclarent ne pas avoir de lien d'intérêt.

Références

1. Elie C, Colombet I. Méthodes d'estimation de la reproductibilité. *Sang Thrombose Vaisseaux* 2011 ; 23 : 138-45.
2. Grenier B. *Évaluation de la décision médicale*. Paris : Éditions Masson, 1996.

3. Paolaggi JB, Coste J. *Le raisonnement médical*. Paris : Éditions Estem, 2001.
4. Junod AF. *Décision médicale ou la quête de l'explicite*. Paris : Éditions Médecine et Hygiène, 2003.
5. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005 ; 365 : 1500-5.
6. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007 ; 115 : 654-7.
7. Sackett DL, Haynes RB. The architecture of diagnostic research. *Br Med J* 2002 ; 324 : 539-41.
8. Knottnerus JA, Weel C, Muris JWM. Evaluation of diagnostic procedures. *Br Med J* 2002 ; 324 : 477-80.

Annexe 1. Intervalles de confiance des indices informationnels

1) Sensibilité et spécificité sont des pourcentages (ou valeurs comprises entre 0 et 1). Les bornes de leur intervalle de confiance à 95 % peuvent donc s'estimer, à partir de leur erreur standard (*standard error*) comme suit :

Sensibilité : $Se \pm 1,96 \times \text{racine carrée } [Se(1 - Se)/n_1]$

Spécificité : $Sp \pm 1,96 \times \text{racine carrée } [Sp(1 - Sp)/n_2]$

où : n_1 représente le nombre de malades (a + c)

n_2 représente le nombre de non-malades (b + d)

2) Pour le rapport de vraisemblance, l'erreur standard s'exprime pour le logarithme népérien du RV :

$RV^+ : SE_{\ln RV^+} = \text{racine carrée } \{[(1 - Se)/(n_1 \times Se)] + [Sp/(n_2 \times (1 - Sp))]\}$

$RV^- : SE_{\ln RV^-} = \text{racine carrée } \{[Se/(n_1 \times (1 - Se))] + [(1 - Sp)/(n_2 \times Sp)]\}$

Les bornes de l'intervalle de confiance des rapports de vraisemblance s'estiment donc par exponentiation :

$RV^+ : \text{borne inférieure} = e^{(1,96 \times SE_{\ln RV^+})} \times RV^+$ borne supérieure : $e^{(+1,96 \times SE_{\ln RV^+})} \times RV^+$

$RV^- : \text{borne inférieure} = e^{(-1,96 \times SE_{\ln RV^-})} \times RV^-$ borne supérieure : $e^{(+1,96 \times SE_{\ln RV^-})} \times RV^-$