

La rigueur dans l'élaboration d'un diagnostic

Rigour in Medical Diagnostic Accuracy

Michel Huguier

Mots clés

- ◆ Diagnostic
- ◆ Examens complémentaires
- ◆ Sensibilité
- ◆ Spécificité
- ◆ Courbes ROC
- ◆ Valeurs prédictives
- ◆ Probabilités bayésiennes
- ◆ Risque de première espèce
- ◆ Utilité décisionnelle
- ◆ Coût-utilité

Résumé

Il est nécessaire d'apprécier objectivement la valeur des examens complémentaires utiles à un diagnostic. Trois définitions sont fondamentales :

1. la population sur laquelle l'examen complémentaire a été évalué ;
2. les critères sur lesquels on a jugé que l'examen était anormal ou normal ;
3. le standard de référence externe.

Les outils mesurant la valeur de l'examen sont la sensibilité, la spécificité, des indices globaux et les courbes ROC. Mais ces outils ne répondent pas à l'interrogation du clinicien : si l'examen est anormal quelle est la probabilité que le sujet ait la maladie ? Inversement, si l'examen est normal quelle est la probabilité que le sujet n'ait pas la maladie ? La réponse à ces questions est donnée par les valeurs prédictives qui dérivent des probabilités bayésiennes.

Trois remarques complémentaires :

1. il y a un risque statistique d'erreur d'interprétation qui augmente avec le nombre d'examens que l'on demande (risque de première espèce) ;
2. au-delà de la valeur informationnelle d'un examen, il faut estimer son utilité décisionnelle ;
3. il faut tenir compte de la notion de coût-utilité.

Keywords

- ◆ Diagnosis
- ◆ Radiological investigations
- ◆ Laboratory investigations
- ◆ Sensitivity
- ◆ Specificity
- ◆ Receiving operative characteristics (ROC) curves
- ◆ Predictive values
- ◆ Type 1 of error
- ◆ Decisional-making value
- ◆ Cost-utility

Abstract

We need how to exactly evaluate the value of radiological and laboratory exams. Three basic definitions are:

1. Population on which the exam has been evaluated;
2. Criteria of normal or abnormal results;
3. The frame of reference.

Main tools are sensitivity, specificity, likelihood ratio and receiving operative characteristics (ROC) curves. However, these tools did not answer to the clinician question: if the exam is abnormal what is the probability that the patient has the disease, and if the exam is normal what is the probability that the patient has not the disease. Answers to these questions are supplied by bayesian predictive values.

It is pointed out that:

1. Increasing the number of exams increases the statistical risk of type 1 of error;
2. It is necessary to evaluate the decisional-making value of exams;
3. Physicians must take in account cost-utility.

Les malades nous sont, en général, adressés avec un diagnostic. Néanmoins, nous sommes parfois amenés à le vérifier et plus souvent encore à faire des investigations complémentaires, par exemple pour préciser l'extension locorégionale d'un cancer ou rechercher des métastases. Pour se faire, en plus de l'examen clinique, des examens complémentaires sont habituellement nécessaires. Leur prescription doit résulter d'une réflexion qui intègre plusieurs composantes que cette communication se propose d'illustrer par quelques exemples concrets.

La valeur informationnelle des examens

Rappelons quelques notions :

Les trois définitions fondamentales

Lorsque l'on cherche à apprécier la valeur d'un symptôme, d'un signe physique, d'un examen biologique, radiologique, isotopique, etc. (que nous avons appelé « signe » dans le ta-

Correspondance :

Professeur Michel Huguier, 10, passage Gambetta - 75020 Paris

Professeur honoraire de chirurgie digestive - Membre de l'Académie nationale de Médecine - Paris.

E-mail : michel.huguier@tnn.aphp.fr

	M +	M -	Total
S +	a	b	a + b
S -	c	d	c + d
Total	a + c	b + d	N

- N représente la population sur laquelle l'examen a été évalué.
- M+ représente, au sein de la population étudiée, les malades.
- M - les personnes qui n'ont pas la maladie.
- S + représente la présence du signe (ou le résultat anormal de l'examen).
- S - l'absence du signe (ou le résultat normal de l'examen).

Les quatre sous-groupes du champ du tableau correspondent ainsi :

- a. à l'effectif des signes présents chez les malades ;
- b. à l'effectif des signes présents chez des sujets qui n'ont pas la maladie (appelés « faux positifs ») ;
- c. à l'effectif des signes absents chez les malades (appelés « faux négatifs ») ;
- d. à celui des signes absents chez les sujets qui n'ont pas la maladie.

Tableau I. Les données

bleau) trois définitions sont fondamentales dans l'interprétation des résultats (tableau I).

La population (N) sur laquelle l'examen a été évalué

Prenons l'exemple de l'évaluation de l'échographie dans le diagnostic des métastases hépatiques. Les résultats ne sont pas les mêmes selon que la population étudiée a été :

- ou bien un ensemble de malades qui avaient un cancer primitif connu (1) ;
- ou bien des malades qui avaient un cancer primitif connu mais sans métastases cliniquement palpables (2).

Dans le premier cas, la population étudiée incluait des malades qui avaient des métastases volumineuses et palpables. Dans le second cas, par définition, cela n'avait pas été le cas. De ce fait, la première étude sur le tout venant des malades avec un cancer primitif a montré des performances de l'échographie supérieures à celles observées dans l'autre étude (pas de métastases cliniquement palpables). Cet exemple montre que si le lecteur ne sait pas très bien sur quelle population les auteurs ont cherché à apprécier la valeur d'un examen, il ne peut pas faire d'extrapolation sur la valeur de cet examen. On appelle ce phénomène le biais de spectre (*spectrum bias* en anglais).

Les critères sur lesquels on a jugé que l'examen était anormal (S+) ou normal (S-)

Par exemple, dans la valeur des lactico-deshydrogénases dans le dépistage des métastases hépatiques, une étude a pris comme limite supérieure de la normale celle indiquée par le laboratoire (3). Une autre étude a pris la valeur observée, chez les malades inclus dans l'étude et qui n'avaient pas de métastases plus deux écart-types (4). Les résultats, de ce fait, ont été différents entre les deux études.

Le critère de jugement de l'existence (M+) ou non (M-) de la maladie : le standard de référence externe

Ce qui vient d'être dit à propos de l'interprétation de l'examen s'applique aussi aux critères de jugement : présence ou absence de la maladie. C'est ce que l'on appelle le standard de référence externe ou le référentiel. Dans notre premier exemple, il s'agit de savoir sur quels critères (autres que le résultat de l'échographie qui peut y avoir des résultats faussement positifs et faussement négatifs) les auteurs ont affirmé qu'il y avait ou qu'il n'y avait pas de métastases hépatiques. En effet, même à l'intervention chirurgicale, une tumeur bénigne ou un kyste biliaire peuvent être pris pour une métastase (5). Le standard de référence externe qui apporte le meilleur niveau de preuve est, dans cet exemple, l'examen anatomo-pathologique de la lésion hépatique. Inversement, la constatation lors d'une laparotomie, d'un foie apparemment

normal ne suffit pas à affirmer l'absence de métastases : il peut exister une petite métastase en plein centre du lobe droit du foie qui n'est pas visible ni palpable. Il convient donc d'avoir d'autres preuves : par exemple, échographie peropératoire et recul d'observation pour s'assurer que les malades ne développent pas à brève échéance une métastase. Si ces trois définitions ne sont pas précisées dans un article, il est possible d'en interrompre la lecture sans beaucoup de risque (6)

Outils de mesure : sensibilité et spécificité

Il est alors possible d'apprécier la valeur informationnelle de l'examen de façon aussi précise que l'on mesure avec un centimètre.

Les deux principaux outils de mesure sont la sensibilité et la spécificité (tableau II).

- La sensibilité d'un signe est le pourcentage de cas où il est présent chez les malades, ce que l'on peut encore exprimer en disant que c'est la probabilité du signe s'il y a la maladie. Elle est estimée par le rapport du nombre de malades chez lesquels le signe est présent (a) au nombre de malades (a + c).

Un signe est d'autant plus sensible qu'il est souvent présent dans la maladie. Si c = 0, c'est-à-dire pas de « faux négatifs », la sensibilité du signe est égale à 1 (ou 100 %), ce qui signifie que le signe est toujours présent chez les malades. C'est le cas de la fièvre dans la typhoïde, ou de l'élévation des transaminases dans les hépatites. Beaucoup d'études qui concernent l'évaluation d'un moyen diagnostique sont rétrospectives et ne portent que sur des patients atteints d'une maladie, chez lesquels un examen complémentaire a été étudié. De ce fait, elles ne peuvent apprécier que la sensibilité de cet examen, ce qui en limite beaucoup l'intérêt.

- La spécificité d'un signe dans une population mesure le pourcentage de sujets chez lesquels il est absent parmi ceux qui n'ont pas la maladie. En termes de probabilité, la spécificité estime la probabilité de l'absence du signe en l'absence de maladie. La spécificité est, en effet, estimée par le rapport du nombre de sujets qui n'ont pas la maladie ni le signe (d), à l'ensemble des sujets qui n'ont pas la maladie dans la population étudiée (b + d).

Un signe est d'autant plus spécifique qu'il est rarement présent chez les personnes qui n'ont pas la maladie. S'il n'y a pas de « faux positifs » (b = 0), la spécificité du signe est égale à 1 (ou 100 %). C'est alors un signe pathognomonique de la maladie comme le signe de Koplik dans la rougeole. Ce type de signe, on le sait, est très rare.

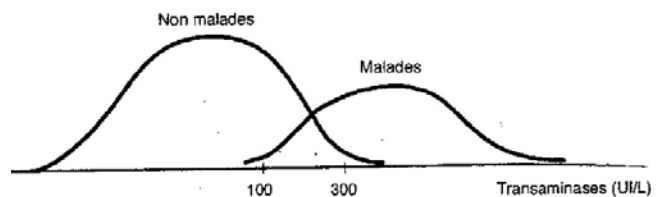


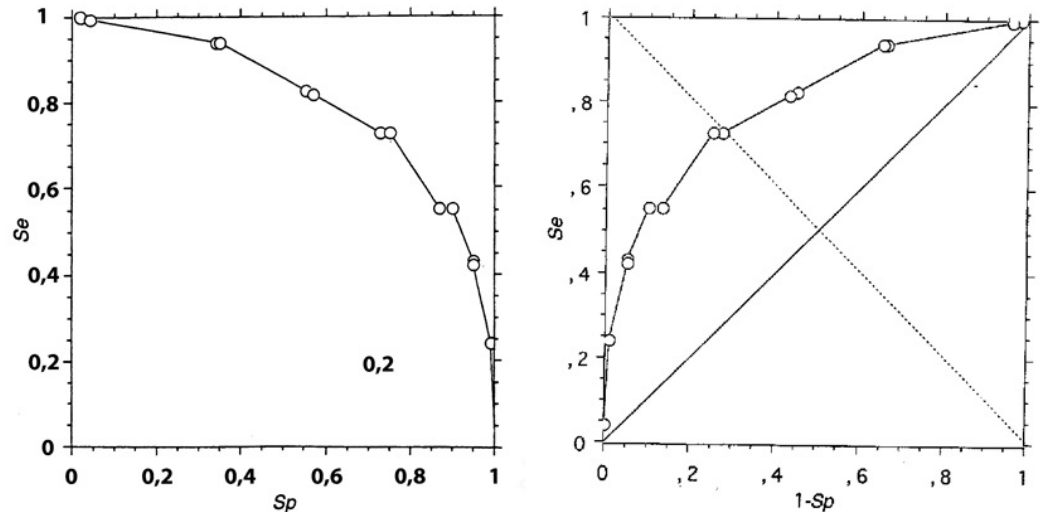
Figure 1. Résultats des mesures d'une variable quantitative dans une population comportant des malades et des non-malades. La sensibilité et la spécificité dépendent de la valeur limite de la normale qui a été choisie.

Tableau II. Sensibilité et spécificité d'un signe

	M +	M -	Total
S +	a	b	a + b
S -	c	d	c + d
Total	a + c	b + d	N

La sensibilité (Se) est égale à a / a + c.
La spécificité (Sp) est égale à d / b + d.

Figure 2. 2a. La figure de gauche montre les sensibilités (Se) et les spécificités (Sp) lorsque l'on fait varier la valeur seuil. 2b. En fait, la figure de droite montre une courbe Receiver Operating Characteristics (ROC). Par rapport à la courbe précédente, l'axe des abscisses est le complément de la spécificité (1 - Sp).



Problèmes posés par la sensibilité et la spécificité. Les courbes ROC

Le signe ou l'examen idéal serait celui qui aurait une sensibilité de 100 % et une spécificité de 100 %. En fait un examen très spécifique est souvent peu sensible et réciproquement : le *melaena* est très spécifique d'une hémorragie digestive, mais il est peu sensible : des hémorragies digestives peu abondantes n'entraînent pas de *melaena*. Inversement, l'hémocult® est plus sensible, pour dépister du sang dans les selles, mais il est peu spécifique d'hémorragie digestive : un petit saignement d'origine gingivale ou des facteurs alimentaires, par exemple, peuvent rendre un hémocult® positif.

Indices globaux

Pour essayer de concilier les termes de cette alternative, examen sensible mais peu spécifique ou spécifique mais peu sensible, des moyens prenant en compte à la fois la sensibilité et la spécificité ont été proposés.

Ainsi, la « fiabilité » d'un examen estime la somme des « vrais positifs » et des « vrais négatifs » sur l'ensemble des cas étudiés $(a + d) / N$.

Un autre outil est le rapport de vraisemblance de l'examen. C'est le rapport du pourcentage des « vrais positifs » chez les malades $(a / a + c)$ sur les « faux positifs » chez les sujets qui n'ont pas la maladie $(b / b + d)$. Par exemple, un rapport de vraisemblance égal à quatre signifie que l'examen est quatre fois plus souvent positif chez les malades que chez ceux qui n'ont pas la maladie. Par analogie, il est encore possible de mesurer le rapport de vraisemblance négatif que est le rapport du pourcentage des « faux négatifs » chez les malades $(c / a + c)$ sur les « vrais négatifs » chez les sujets qui n'ont pas la maladie $(d / b + d)$. Un rapport de 0,5 signifie que l'absence du signe est deux fois moins souvent observée chez les malades que chez les sujets qui n'ont pas la maladie.

Les Receiver Operating Characteristics curves ou courbes ROC

En fait, lorsque l'examen complémentaire s'exprime par un résultat quantitatif, comme c'est le cas de la plupart des examens biologiques, il est possible d'estimer sa sensibilité et sa spécificité pour des valeurs seuils variées. Prenons l'exemple du dosage des transaminases chez des personnes atteintes d'une hépatite et chez des sujets sains. La distribution du dosage dans l'un et dans l'autre groupe, en supposant leur distribution normale, peut être décrite par des courbes de Laplace-Gauss (fig. 1). Il est habituel que ces deux courbes se superposent partiellement avec un chevauchement entre les valeurs les plus élevées chez les non malades et les valeurs les

plus basses chez les malades. Si l'on prend une valeur seuil basse, 100 par exemple, il y aura des faux positifs (les sujets sans hépatite qui ont un taux de transaminases supérieur à 100) et très peu de faux négatifs : l'examen sera peu spécifique, mais très sensible. Si l'on prend, au contraire une valeur plus élevée, 300, il y aura très peu de faux positifs, mais des faux négatifs représentés par les malades atteints d'hépatite qui ont un taux de transaminases inférieur à 300 : l'examen sera considéré comme très spécifique, mais peu sensible. On comprend bien que lorsque que l'on fait varier le curseur de la valeur seuil en deçà de laquelle on considère que le résultat est normal et au-delà de laquelle on le considère comme étant anormal ou pathologique, on fera varier à chaque fois les valeurs estimées de la sensibilité et de la spécificité de l'examen. C'est ce que permettent de quantifier les courbes ROC.

Ces courbes ROC consistent à porter sur un graphique, pour différentes valeurs données de seuil entre le normal et l'anormal d'un examen, sa sensibilité (Se) en ordonnée et la spécificité correspondante (Sp) en abscisse. En répétant ces mesures pour de nombreuses valeurs seuil, on peut faire des courbes : c'est le principe des courbes ROC. Elles montrent qu'habituellement, une valeur seuil qui correspondrait à une sensibilité proche de 100 % aurait une spécificité proche de 0 %, alors qu'une valeur seuil proche de 100 % de spécificité aurait une très mauvaise sensibilité (fig. 2a). En fait, dans les courbes ROC, élaborées à l'aide de logiciels, l'axe des abscisses correspond, non pas à la spécificité, mais à son complément, $1 - Sp$ (fig. 2b).

Sous les conditions rappelées ci-dessous, la valeur seuil optimale est celle qui correspond sur la courbe à son point d'inflexion, soit encore au point d'intersection de la courbe ROC et de la deuxième bissectrice de l'axe des abscisses et des ordonnées. Ce choix s'applique dans l'hypothèse d'un coût affecté aux erreurs de diagnostic équivalent ou à peu près équivalent pour les « faux positifs » et les « faux négatifs ».

Un deuxième avantage des courbes ROC est de pouvoir comparer la valeur de deux examens entre eux en mesurant leurs aires sous la courbe (fig. 3). Ces comparaisons ont, par exemple, montré que pour évaluer la gravité des pancréatites aiguës, le score APACHE II était plus performant que le score de Glasgow (7).

Les valeurs prédictives

Pour intéressantes que soient les estimations de la sensibilité et de la spécificité (que l'on doit assortir de leur intervalle de confiance), ces moyens de mesure ne permettent pas de répondre aux deux principales questions qui intéressent le clinicien devant le résultat d'un examen complémentaire : Si

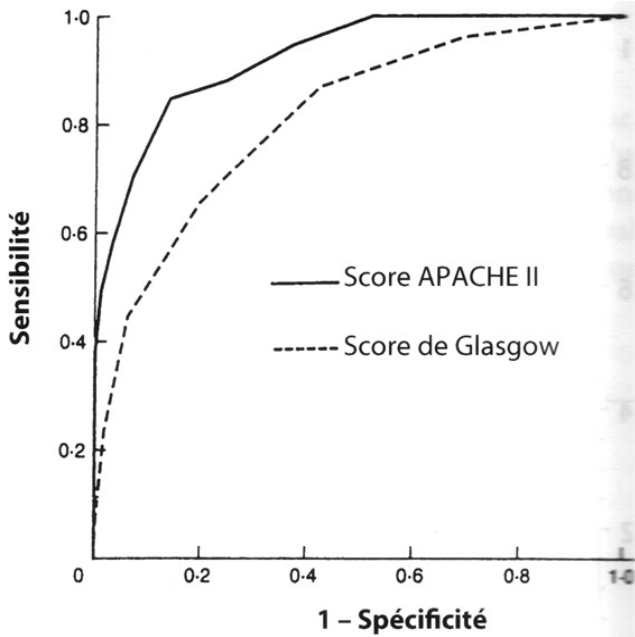


Figure 3. Les aires sous les courbes ROC permettent de comparer deux ou plusieurs examens entre eux. Dans cet exemple concernant l'évaluation de la gravité des pancréatites aiguës, le score APACHE (aire de 0,78) et le score de Glasgow (aire de 0,65). Un test (ici de McNemar) de $p = 0,005$ permet de conclure que la différence est statistiquement significative.

l'examen est anormal, quelle est la probabilité que le sujet ait la maladie que l'on cherche à diagnostiquer ? Et si l'examen est normal, quelle est la probabilité que le sujet n'ait pas la maladie que l'on a pu évoquer ?

Les valeurs prédictives répondent à ces interrogations. Ces valeurs, encore dénommées probabilités a posteriori, sont, comme la sensibilité et la spécificité, des probabilités conditionnelles dont les développements mathématiques ont été formulés par Thomas Bayes et rapportés par Price en 1763 (8).

Les valeurs prédictives (tableau III)

La valeur prédictive positive (VPP) d'un signe estime la probabilité de la maladie chez les personnes qui ont ce signe. Cette estimation correspond au rapport du nombre de sujets qui ont la maladie et chez lesquels le signe est présent (a) sur le nombre de sujets chez lesquels le signe est présent (a + b). La VPP d'un signe est d'autant plus grande que le signe est rarement présent chez les personnes qui n'ont pas la maladie, autrement dit que le nombre de « faux positifs » est faible.

La valeur prédictive négative (VPN) d'un signe estime la probabilité d'absence de la maladie chez les personnes qui n'ont pas ce signe. Elle est définie par le rapport du nombre de sujets qui n'ont pas la maladie et chez lesquels le signe est absent (d) sur le nombre de sujets chez lesquels le signe est absent (c + d). La VPN d'un signe est d'autant plus grande que l'absence du signe est rare chez les personnes qui ont la maladie, autrement dit, que le nombre de « faux négatifs » est faible.

Le théorème de Bayes

En fait, les valeurs prédictives sont des applications du théorème de Bayes qui permet, de façon générale, d'estimer la probabilité de survenue d'un événement, sachant qu'un autre événement est connu (tableau IV). Par exemple, il estime la probabilité d'une maladie lorsqu'un signe pathologique est présent. Ce théorème montre surtout que les valeurs prédictives, contrairement à la sensibilité et la spécificité, dépendent de la prévalence de la maladie dans la population étudiée.

	M +	M -	Total	
S +	a	b	a + b	La valeur prédictive positive (VPP) est égale à $a / a + b$.
S -	c	d	c + d	La valeur prédictive négative (VPN) est égale à $d / c + d$.
Total	a + c	b + d	N	

Tableau III. Les valeurs prédictives

Les autres données qui doivent intervenir dans la décision

Le risque de multiplier le nombre d'examen biologiques

La plupart des résultats des dosages biologiques suivent une loi normale (loi de Laplace-Gauss). Les valeurs normales indiquées par le laboratoire incluent la valeur moyenne plus ou moins deux écart-types (en fait 1,96 écarts-types), c'est-à-dire 95 % des sujets sains. Ce qui veut dire que 5 % de ces sujets sains ont des valeurs qui sont au-dessus ou au-dessous du seuil de « normalité ». Croire que ces sujets sont anormaux est le risque statistique appelé de première espèce. Plus on fait d'examen, plus on augmente ce risque. Si, par exemple, on prescrit dix examens biologiques dont la distribution est normale et qui sont indépendants entre eux, il y a près d'une probabilité sur cinq que l'un d'entre eux sorte des limites de la « normale » (tableau V).

Le théorème de Bayes permet d'écrire :

$$p(M+ / S+) = \frac{p(M+) \times p(S+ | M+)}{p(M+) \times p(S+ | M+) + p(M-) \times p(S+ | M-)}$$

dans lequel :

- p (M+) est la prévalence P de la maladie dans la population étudiée.
- p (M-) est le complément de la prévalence 1 - p (M+) ou 1 - P. (M+).
- Cette barre verticale exprime une probabilité conditionnelle :
- p (S+ | M+) est la probabilité du signe si la maladie est présente, c'est-à-dire la sensibilité (Se) du signe.
- p (S+ | M-) est la probabilité du signe s'il n'y a pas la maladie ; c'est le complément de la spécificité, c'est-à-dire 1 - Sp.

Appliqué aux valeurs prédictives positives (VPP) et négatives (VPN), le théorème de Bayes peut donc s'écrire :

$$VPP = p(M+ | S+) = \frac{P \times Se}{P \times Se + (1 - P) \times (1 - Sp)}$$

De façon analogue, la valeur prédictive négative, c'est-à-dire la probabilité d'absence de la maladie si le signe est absent, s'écrit :

$$VPN = p(M- | S-) = \frac{(1 - P) \times Sp}{(1 - P) \times Sp + (1 - P) \times (1 - Se)}$$

Il est possible, par analogie, de calculer de la même façon la probabilité de maladie si le signe est absent ou la probabilité d'absence de maladie si le signe est présent, mais ces probabilités sont, en général, moins utiles en pratique médicale.

Tableau IV. Valeurs prédictives et caractéristiques d'un test

(au seuil de 5%)		* Examens indépendants entre eux
Nombre d'examen réalisés*	Risque global d'erreur de 1ère espèce (%)	
1	5	
2	8	
5	14	
10	19	

Tableau V. Risque global de première espèce en fonction du nombre d'examen réalisés

Les contreparties pour le malade

Les progrès technologiques des examens, notamment morphologiques, expliquent qu'ils comportent de moins en moins de risques, voire de désagréments pour les patients. En contrepartie, cela contribue à diminuer la réflexion qui précède leurs prescriptions, aboutissant à les demander avec une insoutenable légèreté.

L'utilité décisionnelle

Dans la décision de prescrire un examen complémentaire son utilité décisionnelle doit être prise en compte. En voici un exemple. Une étude prospective a été faite chez 24 malades qui avaient un cancer de la tête du pancréas afin d'estimer la valeur du Pet-scan dans le diagnostic de métastases. Les résultats ont montré que le Pet-scan avait une sensibilité de 70 % et une spécificité de 83 % (9). Il était donc considéré par les biophysiciens qui avaient réalisé les examens comme un « très bon examen ». En fait, par rapport aux données de l'échographie, le Pet-scan n'avait modifié la décision que chez un seul malade. En admettant que le coût unitaire d'un Pet-scan qui, à l'époque, était d'environ 1 500 euros, le coût d'un examen utile sur le plan décisionnel pouvait être estimé à 40 000 euros, ce qui a paru exorbitant.

La notion de coût-utilité

L'exemple précédent introduit la notion de coût-utilité. Un autre exemple est celui de la radiographie pulmonaire avant une intervention chirurgicale. Il est caricatural. Une radiographie thoracique coûtait environ 20 euros en 2010. Mais, chez un malade qui n'a pas de symptomatologie clinique cardiopulmonaire, des études ont montré que la probabilité de dépistage d'une anomalie par la radiographie thoracique était nulle chez les patients de moins de 20 ans et de trois sur cent après cet âge (10). Le coût du dépistage chez l'adulte se monte ainsi à environ 665 euros. Ces mêmes études ont encore montré qu'une anomalie dépistée n'avait d'utilité décisionnelle que dans 0,2 % des cas. Le dépistage utile par une radiographie thoracique préopératoire revient ainsi à 333 000 euros. Le médecin est bien entendu, libre de sa décision. Mais il doit intégrer ce raisonnement dans la prise de celle-ci. En ce qui concerne la radiographie thoracique avant une intervention, les travaux cités ont fait abandonner sa prescription systématique au profit d'un interrogatoire et d'un examen clinique qui ont été ainsi revalorisés.

Conclusions

Il y a une tendance à multiplier les examens complémentaires chez les malades, en espérant diminuer le risque d'erreurs de diagnostic. Nous avons vu que pour les examens biologiques, cette attitude augmentait le risque statistique de première espèce. Dans tous les cas, ces multiplications ont des coûts qu'il convient toujours d'avoir à l'esprit.

Comme nous avons dit dans l'introduction, un raisonnement scientifique doit se substituer à des prescriptions insuffisamment réfléchies. Ce raisonnement doit prendre en compte la valeur informationnelle de chaque examen exprimé en termes de sensibilité, de spécificité, et surtout de valeurs prédictives, mais aussi l'utilité décisionnelle et le rapport coût-utilité. Cela implique un esprit critique scientifique et une bonne connaissance des études qui ont été réalisées avec rigueur. Cette démarche demande, au début, du travail et beaucoup de réflexion. Ensuite, avec l'entraînement, elle devient de plus en plus facile et fait de l'élaboration d'un diagnostic une démarche raisonnée et scientifique qui, se substituant à des réflexes pavloviens, fait un des intérêts de l'exercice médical (11).

Références

1. Alderson P, Adam DF, McNeil BJ et al. Computed tomography, ultrasound, and scintigraphy of the liver in patients with colon of breast carcinoma : a prospective comparison. *Radiology*. 1983;149:224-30.
2. Molkhou JM, Lacaine F, Houry S, Huguier M. Dépistage des métastases hépatiques des cancers digestifs. Place des dosages enzymatiques et de l'échographie. *Presse Med*. 1989;18:1370-4.
3. Adloff M, Arnaud JP. Etude prospective critique des différentes méthodes de détection des métastases hépatiques. *Ann Gastroenterol Hepatol*. 1985;21:31-4.
4. Smith TJ, Kemeny MM, Sugarbaker. A prospective study of hepatic imaging in the detection of metastatic disease. *Ann Surg*. 1982;33:519-23.
5. Raffaelsen SR, Kronborg O, Larsen C, Fenger C. Intraoperative ultrasonography in detection of hepatic metastases for colorectal cancer. *Dis Colon Rectum*. 1995;38:355-60.
6. Huguier M. Apprendre la lecture critique d'un article médical. Paris:Elsevier. 2003:83 pages.
7. Mofidi R, Madhavan KK, Garden OJ, Parks RW. An audit of the management of patients with acute pancreatitis against national standards practice. *Br J Surg*. 2007;94:844-8.
8. Price. An essay towards solving a problem in the doctrine of chances by the late Rev. Mr Bayes, F.R.S. *Philosophical transactions*. 1763:370-418.
9. Huguier M, Barrier A, Zacharias T, Valinas R. Résultats de la tomographie par émission de positons dans les cancers de l'appareil digestif. *Bull Acad Natl Med*. 2006;190:75-87.
10. Blery C. Examens paracliniques préopératoires. *Le Concours médical*. 1980;102:5607-10.
11. Huguier M, Boelle PY. *Biostatistiques pour le clinicien*. Paris:Springer. 2013:300 pages.