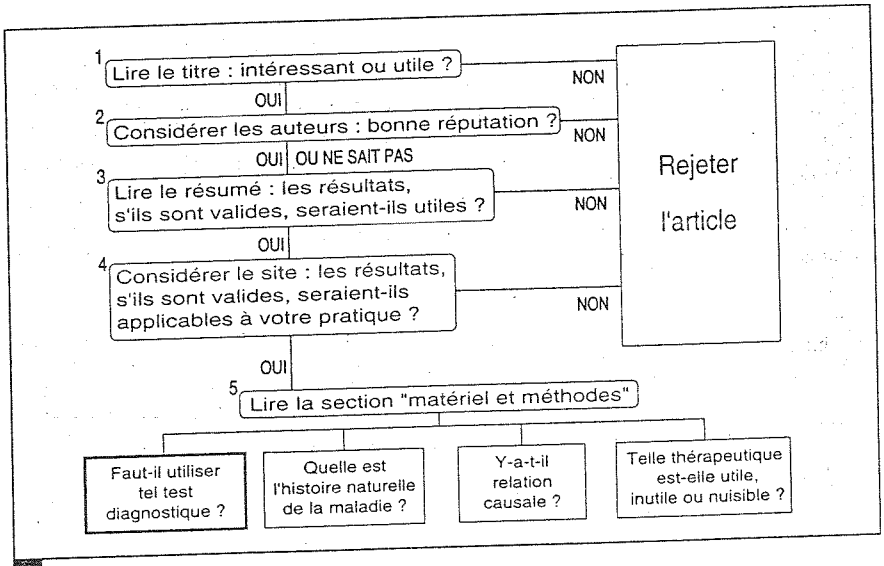
# LECTURE CRITIQUE DES ARTICLES MÉDICAUX

## II. Juger de l'intérêt d'un test diagnostique

L.R. SALMI [1]
J.P. COLLET [2]
ET LE DÉPARTEMENT D'ÉPIDÉMIOLOGIE ET DE BIOSTATISTIQUE DE L'UNIVERSITÉ McMASTER [3]

Le premier article de cette série (1) présentait 10 raisons pour lire les revues médicales et proposait une série d'étapes pour en faire une lecture critique et rapide (fig. 1) : prendre connaissance du titre, du nom des auteurs, du résumé et du site de l'étude. Lorsque vous rencontrez un article qui décrit un test diagnostique potentiellement utile (c'est-à-dire que le titre est intéressant, les auteurs ont bonne réputation, le résumé montre que le test serait très utile s'il marche comme cela est prévu, et le site est similaire au vôtre), que faut-il rechercher dans la section des méthodes de l'article pour juger rapidement de sa qualité ?




1 *Les premières étapes de la lecture critique d'un article médical.*

Les 10 éléments d'une évaluation critique d'un test diagnostique peuvent être regroupés en quatre sections (tableau I). Une première série de critères vise à juger de la qualité des conditions d'évaluation du test diagnostique ; elle consiste à vérifier que la présence ou l'absence de la maladie peut être confirmée par un test de référence, c'est-à-dire d'une autre manière que par le test évalué, et que le test diagnostique évalué et ce test de référence sont appliqués à un échantillon représentatif de patients. Par exemple, une évaluation de l'électrocardiogramme (ECG) d'effort comme test diagnostique de la sténose coronaire nécessite que la présence ou l'absence de sténose puisse être confirmée par le test de référence qu'est la coronarographie. La deuxième série de critères permet de mesurer la performance du test à prédire la présence ou l'absence de la maladie, telle que confirmée par le test de référence. La troi-

TABLEAU I

### Éléments de l'évaluation critique d'un article sur un test diagnostique

**A. L'évaluation du test a-t-elle été faite dans de bonnes conditions ?**
1. Y a-t-il eu une comparaison avec un test de référence ?
2. Le terme "normal" (non-malade) a- t-il été correctement défini ?
3. La comparaison avec le test de référence s'est-elle faite "en aveugle" ?
4. L'échantillon de patients représente-t-il un panorama complet des formes de la maladie ?
5. Le lieu de l'étude et la sélection des patients sont-ils bien décrits ?

**B. Quelle est la performance du test ?**

**C. Le test est-il applicable dans votre pratique ?**
1. La technique d'utilisation est-elle bien décrite ?
2. La reproductibilité de la technique est-elle acceptable ?

**D. Qu'apportera ce test à vos patients ?**
1. Qu'apporte-t-il de nouveau ?
2. Quelle est son utilité ?

*1. Département d'épidémiologie et de biostatistique, Pasteur-Mérieux sérums et vaccins, 15-41, avenue Marcel-Mérieux, 69280 Marcy-l'Étoile*
*2. Unité de pharmacologie clinique, université Claude-Bernard, 162, avenue Lacassagne, 69424 Lyon*
*3. Cette série d'articles est une traduction et une adaptation de la série «How to read a clinical journal» du Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre, Hamilton, Ontario, Canada, publiée originellement dans le Journal de l'Association Médicale Canadienne en 1981*

sième série de critères permet de juger de l'applicabilité du test à votre pratique. Finalement, un test bien évalué, performant et applicable ne vous intéressera que s'il apporte quelque chose de nouveau et s'il est utile.

## L'ÉVALUATION DU TEST A-T-ELLE ÉTÉ FAITE DANS DE BONNES CONDITIONS ?

### Y a-t-il eu une comparaison avec un test de référence ?

Deux groupes de sujets doivent subir le même test diagnostique :
- des sujets chez qui on a montré, par l'application d'un test de référence, la présence d'une maladie ;
- un second groupe de sujets chez qui on a montré, par l'application du même test de référence, l'absence de cette maladie.

Ultérieurement, les résultats du test diagnostique évalué sont comparés avec ceux du test de référence. Le test de référence doit être un test de diagnostic définitif tel que la biopsie, la chirurgie, l'autopsie, le suivi à long terme ou toute autre référence acceptée. Évidemment, le test de référence ne doit pas inclure le test diagnostique évalué comme un de ses composants ; le "biais d'incorporation" qui en résulterait rendrait alors la comparaison invalide.

Si vous ne pouvez pas accepter le test de référence (dans les limites du raisonnable, bien sûr), rejetez l'article.

### Le terme "normal" (non-malade) a-t-il été correctement défini ?

Si l'article utilise le terme "normal" pour décrire l'absence de la maladie étudiée, vous devez être certain que les auteurs ont défini ce qu'ils entendent par "normal", et que cette définition est utile au clinicien.

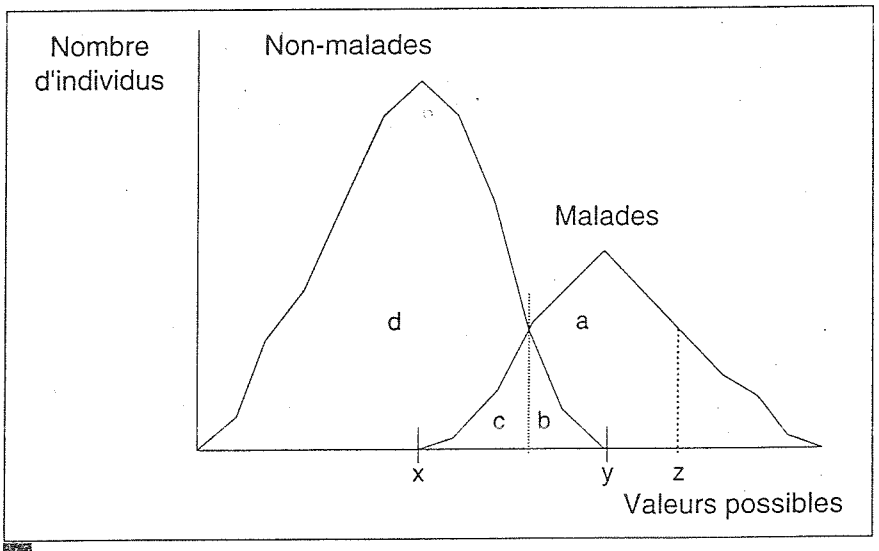
*La définition gaussienne* de "normal" suppose que les résultats du test diagnostique pour une population de référence suivent une distribution théorique connue comme la distribution normale ou gaussienne. Une des propriétés de la distribution gaussienne est que 95 % de la population étudiée a des valeurs situées dans l'intervallle défini par la moyenne plus ou moins 2 écarts-types. Cette définition n'est pas satisfaisante dans de nombreux cas pour deux raisons : premièrement, la distribution normale s'étend à l'infini dans les deux directions et suppose qu'occasionnellement des sujets aient des valeurs impossibles, par exemple une pression artérielle négative. Deuxièmement, si les 2,5 % les plus hauts et les plus bas des résultats d'un test diagnostique sont appelés "anormaux", toutes les maladies auraient la même fréquence, 5 %. Cette critique de la définition gaussienne du "normal" s'applique également au percentile : quand les résultats des tests diagnostiques ne suivent pas une distribution gaussienne, on peut ignorer la forme de la distribution et faire référence aux 95 % des valeurs les plus basses (ou les plus hautes) des résultats du test comme étant normales.

Une conséquence de cette fréquence du normal fixée à 95 % est que la probabilité d'être normal s'amenuise avec la multiplication des tests diagnostiques. La probabilité qu'un patient soit appelé "normal" est en fait égale à 0,95 élevé à une puissance égale au nombre de tests diagnostiques indépendants appliqués. Par exemple, un patient qui a subi 100 tests a seulement 6 chances sur 1 000 d'être appelé "normal" à la fin de son bilan.

*La définition* du "normal" *basée sur la notion de facteur de risque*, si elle évite les erreurs décrites plus haut, présente d'autres problèmes. Cette définition repose sur l'étude des facteurs de prédiction statistique de l'apparition ultérieure d'événements cliniques. Selon cette définition, les valeurs normales du cholestérol sérique ou de la pression artérielle seraient celles qui ne sont pas associées à un risque accru de mortalité ou de morbidité. Malheureusement, la plupart des facteurs sont associés au risque selon une fonction croissant tout au long de leurs valeurs.

Deux définitions sont plus utiles pour le clinicien parce qu'elles se concentrent sur les actes cliniques diagnostiques et thérapeutiques.

*La définition diagnostique* identifie un seuil au-delà duquel une maladie est présente avec une probabilité connue. Cette définition est illustrée par la figure 2, sur laquelle nous voyons le chevauchement habituel des valeurs du test entre patients malades ou non. La "probabilité connue" de présence ou d'absence de la maladie dépend du niveau choisi pour définir l'étendue des valeurs normales du test diagnostique. Si nous voulions obtenir un maximum de résultats corrects, nous mettrions la limite là où les courbes se croisent (ligne pointillée), mais cela ne serait pas forcément utile cliniquement. Si nous baissions cette limite au point x, la région c tendrait vers 0, et nous pourrions utiliser le test pour éliminer la présence de la maladie, car aucun des malades n'a une valeur sous x. À l'opposé, si nous montions la limite au point y, la région b tendrait vers 0, et nous pourrions utiliser le test pour affirmer la présence de la maladie car aucun des non-malades n'a une valeur au-dessus de y. Cette définition a donc une utilité clinique mais nécessite que vous connaissiez le niveau des points x et y de votre clientèle.




2 *Définitions diagnostique et thérapeutique du "normal".*

*La définition thérapeutique* du "normal" met comme limite le point, indiqué par la lettre z sur la figure 2, au-delà duquel on a démontré que les traitements spécifiques font plus de bien que de mal. La définition thérapeutique des valeurs normales de la pression artérielle, par exemple, évite de qualifier des sujets comme étant malades s'ils ne doivent pas être traités. L'utilisation de cette définition nécessite que les cliniciens restent à jour sur les progrès en thérapeutique, et qu'ils sachent différencier les bonnes des mauvaises thérapeutiques ; un prochain article discutera cet aspect de la lecture des articles médicaux.

### La comparaison avec le test de référence s'est-elle faite en aveugle ?

La comparaison avec le test de référence est aveugle si les personnes qui appliquent ou qui interprètent les résultats du test diagnostique évalué ne savent pas si le patient testé a ou n'a pas la maladie considérée.

De la même manière, les personnes qui appliquent le test de référence ne doivent pas connaître les résultats du test diagnostique évalué chez les patients. Ce n'est que lorsque le test diagnostique et le test de référence ont été appliqués d'une manière aveugle que nous avons la garantie qu'aucun biais conscient ou inconscient n'est apparu.

### L'échantillon de patients comporte-t-il des patients atteints de formes sévères ou non sévères, traitées ou non traitées, ainsi que des patients ayant des affections différentes de la maladie cible mais souvent confondues avec celle-ci ?

Une maladie ou syndrome clinique complet (telle une polyarthrite rhumatoïde évoluée) pose en général un problème diagnostique bien plus simple que la même maladie dans une forme précoce ou bénigne ; la vraie valeur clinique d'un nouveau test diagnostique est souvent liée à sa performance chez les cas équivoques. De même que l'on n'a pas besoin d'une analyse chromosomique pour différencier une souris d'un éléphant, la capacité d'un test diagnostique à différencier certaines maladies qui sont de toutes façons rarement confondues, est une piètre justification de son utilisation. La valeur principale d'un test diagnostique réside donc souvent dans sa capacité à distinguer des maladies qui sont couramment confondues, en particulier quand leurs pronostics ou leurs traitements diffèrent de manière importante. C'est ce caractère discriminant qui fait du dosage de la thyroxine un test si utile pour différencier parmi les patients tendus, anxieux, trémulants et transpirants, ceux qui ont une fonction thyroïdienne anormale et ceux qui ont d'autres affections.

Le choix des témoins est également important. Bien que les techniciens de laboratoire et les étudiants en médecine puissent être des témoins appropriés dans le développement initial d'un nouveau test diagnostique, la comparaison définitive avec un test de référence nécessite que les témoins représentent un panorama complet de patients sans la maladie.

### Le site de l'étude et les critères de sélection des patients ont-ils été décrits de manière adéquate ?

Dans l'article précédent (1), nous avons vu comment la proportion de malades hypertendus ayant une lésion curable chirurgicalement varie du simple au décuple selon que les mêmes tests diagnostiques sont appliqués dans une pratique générale ou dans un hôpital universitaire.

Du fait des variations de la performance du test avec la prévalence de la maladie, l'article doit décrire le lieu de l'étude et les critères de sélection des patients. Vous devez être assuré que les différences dans les résultats du test diagnostique sont réellement dues au mécanisme de la maladie et ne sont pas simplement liées à des différences de caractéristiques telles que l'âge, le sexe, l'alimentation et la mobilité des cas et des témoins sélectionnés. La méthode de sélection des sujets témoins (qui n'ont pas la maladie) devrait être aussi bien décrite que celle des sujets malades.

## QUELLE EST LA PERFORMANCE DU TEST ?

### Mesures de la performance d'un test diagnostique

Les mots clés de la mesure de la performance d'un test sont la sensibilité, la spécificité et les valeurs prédictives. Si les deux premiers mots ne sont pas dans le résumé, méfiez-vous. Si vous ne trouvez pas ou ne pouvez pas construire un tableau 2x2 (tableau II) à partir d'une lecture rapide de la section des résultats, il n'est probablement pas nécessaire de continuer votre lecture.

Si l'article survit à ce test rapide, une grande quantité d'informations utiles peut être dérivée de cette comparaison entre le test diagnostique évalué et le test de référence. Les cases *a*, *b*, *c*, et *d* du tableau II correspondent aux aires *a*, *b*, *c* et *d* de la figure 2. Le test de référence a identifié (*a* + *c*) patients comme

TABLEAU II

**Comparaison en aveugle d'un test diagnostique et d'un test de référence : analyse de la performance du test**

| Résultats du test diagnostique | Résultats du test de référence | | |
|---|---|---|---|
| | Maladie présente | Maladie absente | Total |
| positif | a | b | a + b |
| négatif | c | d | c + d |
| total | a + c | b + d | a + b + c + d |

Propriétés stables du test :
a / (a + c) = sensibilité
d / (b + d) = spécificité

Propriétés dépendant de la fréquence de la maladie :
a / (a + b) = valeur prédictive positive
d / (c + d) = valeur prédictive négative
(a + c) / (a + b + c + d) = prévalence

ayant la maladie d'intérêt ; les patients a présentent un test diagnostique positif. Un index de la capactié du test à détecter la maladie quand elle est présente est $a / (a + c)$, habituellement exprimé comme un pourcentage et appelé la sensibilité. De même, le test de référence a identifié $(b + d)$ individus comme n'ayant pas la maladie d'intérêt ; les patients d présentent un test diagnostique négatif. La capacité du test diagnostique à identifier correctement l'absence de la maladie est le rapport $d / (b + d)$ ; cet index est la spécificité. Sensibilité et spécificité sont les propriétés stables du test car elles ne changent pas quand des proportions différentes de malades et de non-malades sont testées.

Dans la situtation clinique usuelle, cependant, nous voulons savoir comment les résultats du test vont prédire la présence et l'absence de la maladie. Si cette prédiction est suffisamment bonne, nous allons ajouter ce test diagnostique à notre arsenal.

Parmi les $(a + b)$ patients qui présentent un test diagnostique positif, dans quelle proportion, $a / (a + b)$, avons-nous correctement prédit la présence de la maladie ? Cette proportion, souvent exprimée comme un pourcentage, s'appelle la valeur prédictive positive ; la valeur prédictive positive est la probabilité d'avoir la maladie quand le test est positif. De même, nous voulons savoir comment un résultat de test négatif prédit correctement l'absence de la maladie en question. La proportion correspondante est $d / (c + d)$ et est appelée la valeur prédictive négative.

## Rôle de la prévalence de la maladie

Si les valeurs prédictives d'un test diagnostique constituent l'objet principal de notre intérêt clinique, pourquoi perdre du temps à considérer la sensibilité et la spécificité ? De manière simplifiée, les valeurs prédictives positives et négatives d'un test diagnostique fluctuent largement en fonction de la proportion d'individus réellement malades parmi les patients à qui le test est appliqué. Cette proportion $(a + c) / (a + b + c + d)$ dans le tableau II, est la prévalence de la maladie.

Les valeurs prédictives peuvent changer de manière marquée (et de ce fait sont instables) quand la prévalence de la maladie change. En situation réelle, la prévalence d'une maladie donnée varie considérablement entre les patients testés en pratique générale et dans un hôpital universitaire, comme nous l'avons vu dans l'exemple de l'hypertension (1). De plus, pendant leur développement initial, la plupart des tests diagnostiques sont évalués sur des groupes de taille égale d'individus avec et sans maladie d'intérêt (c'est-à-dire, avec une prévalence artificielle de 50 %). Cela correspond pratiquement toujours à une prévalence plus haute que celle vue en pratique clinique, même dans les hôpitaux universitaires.

Il est fondamental de comprendre l'effet de la prévalence sur les propriétés des tests diagnostiques pour les utiliser de manière rationnelle ; nous vous suggérons donc d'évaluer vous-même les exemples suivants à titre d'exercice.

*Exemple 1 :* Des chercheurs ont étudié, en consultation spécialisée, un groupe d'hommes adressés pour douleurs thoraciques (2). L'application d'un électrocardiogramme (ECG) d'effort (le test diagnostique) et d'une artériographie coronarienne (le test de référence) a donné les résultats indiqués dans le tableau III.

La capacité de l'ECG d'effort à prédire la présence ou l'absence d'une sténose de 75 % ou plus est reflétée par la valeur prédictive positive de 89 % et par la valeur prédictive négative de 63 %. Les auteurs ont conclu : "Chez les hommes, un ECG d'effort positif est utile pour prédire la présence d'une maladie coronaire, bien qu'un test négatif ne permette pas d'éliminer la présence d'une maladie notable."

D'après les résultats de l'artériographie, 53 % des patients (104/195) avaient une sténose marquée de l'artère coronaire, ce qui représente un groupe très sélectionné de patients. Qu'arriverait-il si des médecins enthousiasmés par la lecture de cet article décidaient d'utiliser plus largement l'ECG d'effort afin de détecter une maladie coronaire en l'absence de douleurs thoraciques ? (Ce que d'ailleurs les auteurs de cet exemple ne recommandèrent pas.) Si la prévalence de sténose coronaire chez les hommes adressés, à partir d'une pratique générale, pour une coronarographie est seulement de 1 sur 6 (17 %), plutôt que les 53 % de l'étude (tableau IV), il y aurait 5 fois plus de patients n'ayant pas de maladie coronaire (5 x 104 = 520) que de patients ayant cette maladie. La sensibilité restant constante, 55 (53 %) des 104 malades auraient un ECG d'effort positif. De même, la spécificité restant de 92 %, 478 des 520 hommes non malades auraient un test négatif. Le reste du tableau peut être complété par simple addition ou soustraction, et les valeurs prédictives peuvent ainsi être calculées : la valeur prédictive positive chute de 89 % à 57 % et

TABLEAU III

**Valeurs prédictives de l'électrocardiogramme d'effort quand la moitié des sujets testés sont porteurs d'une sténose des artères coronaires (2)**

| Résultats de l'ECG d'effort | Résultats de l'artériographie | | |
|---|---|---|---|
| | Sténose présente | Sténose absente | Total |
| positif | 55 | 7 | 62 |
| négatif | 49 | 84 | 133 |
| total | 104 | 91 | 195 |

sensibilité = 55/104 = 53 %
spécificité = 84/91 = 92 %
valeur prédictive positive = 55/62 = 89 %
valeur prédictive négative = 84/133 = 63 %
prévalence = 104/195 = 53 %

TABLEAU IV

**Valeurs prédictives de l'électrocardiogramme d'effort quand un sixième des sujets testés est porteur d'une sténose des artères coronaires**

| Résultats de l'ECG d'effort | Résultats de l'artériographie | | |
|---|---|---|---|
| | Sténose présente | Sténose absente | Total |
| positif | 55 | 42 | 97 |
| négatif | 49 | 478 | 527 |
| total | 104 | 520 | 624 |

sensibilité = 55/104 = 53 %
spécificité = 478/520 = 92 %
valeur prédictive positive = 55/97 = 57 %
valeur prédictive négative = 478/527 = 91 %
prévalence = 104/624 = 17 %

la valeur prédictive négative augmente de 63 % à 91 %, une situation inverse de la situation originale.

Ces résultats sont complètement prévisibles ; c'est cette possibilité de prédiction qui permet au lecteur d'adapter les résultats de l'évaluation d'un test diagnostique au contexte de sa propre pratique. Les seuls outils nécessaires sont l'arithmétique et une estimation grossière de la prévalence de la maladie dans sa propre clinique, "à partir de son expérience personnelle" par exemple.

De manière générale, la valeur prédictive positive chute et la valeur prédictive négative augmente quand le test diagnostique, initialement développé dans un groupe de patients dans lequel la prévalence de la maladie cible est élevée, est appliqué par la suite à un groupe de patients dans lequel la prévalence est plus faible.

*Exemple 2 :* Les patients présentent des caractéristiques facilement discernables qui permettent une estimation assez précise du diagnostic avant même que les tests diagnostiques ne soient appliqués. Par exemple, la prévalence de la sténose coronaire variant avec l'âge, l'utilité de l'ECG d'effort dépendra beaucoup de l'âge de vos patients. Ainsi, un homme de 30 ans qui présente des douleurs thoraciques non angineuses a une probabilité faible de maladie coronaire ; la prévalence est d'environ 5 % chez les hommes de 30 ans. Au contraire, un homme de 62 ans présentant une douleur angineuse typique a une très haute probabilité de sténose coronaire ; la prévalence à cet âge est d'environ 94 % (3).

Pour l'homme jeune, on peut calculer que la valeur prédictive positive est de 26 % et la valeur prédictive négative de 97 %. L'ECG d'effort est peu utile dans ce cas : un test négatif nous donne une information que nous connaissons déjà (la maladie coronaire est peu probable) et un test positif ne justifie pas la prescription de tests complémentaires plus agressifs.

L'ECG d'effort est également peu utile chez l'homme de 62 ans. Dans le cas d'un test positif, la probabilité de la maladie augmente seulement de 95 % à 99 %. La probabilité d'être malade quand l'ECG est normal chute seulement à 89 %, ce qui ne nous rassure pas suffisamment pour éviter des tests complémentaires. Les indications de l'ECG d'effort (ou de tout autre test diagnostique) concernent donc les cas où l'incertitude est grande a priori. Prenons le cas d'un homme de 45 ans avec une douleur angineuse atypique, chez qui la probabilité de sténose coronaire est d'environ 46 % (2). La valeur prédictive positive est de 85 %, et l'on prescrira alors facilement une angiographie. Si l'ECG d'effort est négatif, la probabilité d'une sténose coronaire chute à 30 % et une exploration complémentaire n'est plus nécessaire immédiatement.

## LE TEST EST-IL APPLICABLE DANS VOTRE PRATIQUE ?

### ■ Les méthodes exactes d'application du test sont-elles décrites avec suffisamment de détails pour permettre sa reproduction ?

Si les auteurs ont conclu que vous deviez utiliser leur test diagnostique, ils doivent vous dire comment l'utiliser. Cette description doit couvrir les aspects de sélection et de préparation des patients, ainsi que la "mécanique" d'application du test et d'interprétation des résultats. Y a-t-il des exigences particulières de régime ou d'activité physique ? Quels médicaments doivent être utilisés ? La procédure est-elle douloureuse ? Que fait-on pour calmer cette douleur ? Quelles précautions doivent être prises pendant et après le test ? Comment les échantillons doivent-ils être transportés et conservés pour analyse ultérieure ? Tout cela doit être parfaitement décrit pour que vos patients tirent un bénéfice de l'utilisation du test.

### ■ La reproductibilité des résultats du test (précision) et leur interprétation (variation interobservateur) ont-elles été déterminées ?

La validité d'un test diagnostique dépend de l'absence de biais (déviation systématique des vraies valeurs) et de sa précision (le même test appliqué aux mêmes patients doit donner le même résultat).

La description d'un test diagnostique doit indiquer si les résultats sont reproductibles. Cela est particulièrement vrai quand une compétence particulière ou l'expérience sont nécessaires pour l'application du test (par exemple, l'interprétation des échographies présente encore d'énormes variations d'un opérateur à l'autre). La variation interobservateur est un problème majeur pour les tests impliquant la lecture de radiographies, d'ECG ou d'électroencéphalogrammes (4, 5).

## QU'APPORTERA CE TEST À VOS PATIENTS ?

### Le test diagnostique apporte-t-il quelque chose de nouveau ?

Si vous considérez que le test diagnostique a passé les étapes précédentes : a-t-il quelque chose à offrir que le test de référence n'offre pas ? Par exemple, est-il moins risqué, plus confortable ou moins embarrassant pour le patient, moins coûteux ou applicable à un stade plus précoce de la maladie ? Si le test diagnostique proposé n'offre aucun avantage théorique par rapport au test de référence, pourquoi continuer à lire l'article ?

### L'utilité du test est-elle déterminée ?

Le dernier critère d'évaluation d'un test diagnostique ou de toute autre manœuvre clinique concerne le bénéfice que le patient doit en tirer. Si vous êtes d'accord avec ce point, vous devez scruter l'article pour voir si les auteurs sont allés au-delà des aspects de précision et de validité pour explorer les conséquences à long terme de l'utilisation de leur test diagnostique.

En plus de décrire le devenir des patients qui ont été correctement classés par le test diagnostique, les auteurs doivent vous informer de ce que sont devenus les patients qui avaient un résultat faussement positif ou faussement négatif. Si l'application du test entraîne une attente avant la mise en route de la thérapeutique finale (par exemple, s'il y a une liste d'attente avant d'être examiné, ou si l'analyse des échantillons prend du temps), les conséquences de ce délai doivent être discutées.

Dans le cas de l'évaluation des méthodes non sanglantes du diagnostic de la thrombose veineuse profonde, par exemple, la scintigraphie au fibrinogène marqué met de plusieurs heures à plusieurs jours avant de se positiver, quand le thrombus est petit ou limité au mollet (6). Il est donc important de surveiller les patients et de signaler si certains font des accidents emboliques pendant cette période (heureusement, cela n'arrive pas). Il faut enfin décrire les conséquences du traitement des sujets aux résultats faussement positifs et celles de l'absence de traitement des sujets aux résultats faussement négatifs (6).

## APPLICATION DE CES CRITÈRES À LA LECTURE

L'application de ces critères devrait vous permettre de décider si un test diagnostique est utile dans le cadre de votre pratique clinique.

Le critère majeur pour l'évaluation de ces articles sera toujours la comparaison avec un test de référence. Il est également indispensable que l'article fournisse l'information nécessaire au calcul de la sensibilité et de la spécificité du test. En fonction du contexte dans lequel vous évaluerez l'article, l'un ou l'autre des autres critères sera le plus important, et vous pourrez aller directement à ce critère. Si l'article a rempli ce critère, vous pourrez vérifier les autres critères ; au contraire, si le critère qui vous paraît le plus important n'est pas rempli, vous pourrez rejeter l'article. Cette méthode permet ainsi d'améliorer le rendement du temps que vous consacrez à la lecture. De même, quand vous essayez de choisir le meilleur test pour vos patients, ces critères vous aideront à comparer les tests et à choisir celui qui est le mieux adapté à vos exigences cliniques.

L'article suivant de cette série concerne la lecture des articles qui décrivent l'histoire naturelle et le pronostic des maladies. ■

Remerciements : Les auteurs remercient H. Maisonneuve et C. Dupuy pour leurs critiques et commentaires, et M. Portier pour la dactylographie.

## RÉFÉRENCES

1. Salmi LR, Collet JP et le département d'épidémiologie et de biostatistique de l'université McMaster. Lecture des articles médicaux : I. Introduction. Rev Prat (Paris 1991 ; 41 : 25 : 2598-605.

2. Sketch MH, Mohiuddin SM, Lynch JD, Zencka AE, Runco V. Significant sex differences in the correlation of electrocardiographic exercise testing and coronary arteriograms. Am J Cardiol 1975 ; 36 : 169-73.

3. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med 1979 ; 300 : 1350-8.

4. Fermanian J. Mesure de l'accord entre deux juges. Cas qualitatif. Rev Epidémiol Santé Publique 1984 ; 32 : 140-7.

5. Fermanian J. Mesure de l'accord entre deux juges. Cas quantitatif. Rev Epidémiol Santé Publique 1984 ; 32 : 408-13.

6. Hull R, Hirsh J, Sackett DL, Powers P, Turpie AGG, Walker I. Combined use of leg scanning and impedance plethysmography in suspected venous thrombosis. An alternative to venography. N Engl J Med 1977 ; 296 : 1497-500.

## LECTURE CRITIQUE DES ARTICLES MÉDICAUX

### Articles déjà parus

Éditorial N° 25 du 1er/12/91
page 2597

I. Introduction N° 25 du 1er/12/91
page 2605