

\* Médecine générale  
5060 Sambreville

# Petit lexique pour lecture critique...

par le Dr Elide Montesi\*

**L**ire un article de façon critique, interpréter les résultats d'une étude publiée dans la littérature ou présentée par un délégué pharmaceutique, comprendre l'utilité d'un test diagnostique nécessitent d'en connaître le jargon statistique et épidémiologique. Sans vouloir faire injure à tous les confrères qui jonglent déjà avec ce vocabulaire, nous trouvons utile de publier un petit lexique sans prétention rassemblant les définitions de la plupart des termes utilisés.

## PRÉTEST

- |   | Vrai                     | Faux                     |
|---|--------------------------|--------------------------|
| 1. On parle de faux positif lorsque la maladie est absente et le test est positif.  | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. La spécificité d'un test est le taux de tests positifs dans une population présentant effectivement la maladie testée. | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. Les méta analyses servent à comparer des études de puissance insuffisantes ou aux résultats discordants.               | <input type="checkbox"/> | <input type="checkbox"/> |

Réponses au prochain numéro.

## ÉVALUATION DE LA VALEUR DIAGNOSTIQUE D'UN TEST

Le mot test dans cette rubrique peut signifier autant un symptôme ou un examen clinique qu'un examen complémentaire.

### REPRODUCTIBILITÉ D'UN TEST

Permet d'évaluer les résultats donnés par un test. On parle de **reproductibilité intra-observateur** lorsque le test est répété dans le temps chez les mêmes sujets par le même observateur et de **reproductibilité inter-observateur** lorsque plusieurs observateurs utilisent le test chez les mêmes sujets.

Résultats test	Maladie présente	Maladie absente	Total
> 0	a	b	a + b
< 0	c	d	c + d
Total	a + c	b + d	a + b + c + d

### VRAI POSITIF

Le test est positif et la maladie est présente (a)

### VRAI NÉGATIF

Le test est négatif et la maladie est absente (d)

### FAUX POSITIF

Le test est positif et la maladie est absente (b)

### FAUX NÉGATIF

Le test est négatif et la maladie est présente (c)

### SENSIBILITÉ D'UN TEST

Probabilité d'avoir un test positif quand on est malade: % de tests positifs dans une population malade.  $a/a + c$

### SPÉCIFICITÉ D'UN TEST

Probabilité d'avoir un test négatif quand on n'est pas malade: % de tests négatifs dans une population non malade.  $d/b + d$

La sensibilité et la spécificité dépendent uniquement des qualités du test et de l'opérateur.

Un test avec un taux élevé de faux positifs est très sensible mais peut être peu spécifique.

Un test avec un taux élevé de faux négatifs est très spécifique mais peut être peu sensible.

Pour les tests diagnostiques, on privilégiera la spécificité pour limiter les faux positifs.

Pour un dépistage, on veut limiter les faux négatifs pour détecter un plus grand nombre de malades et on privilégiera la sensibilité du test. (Dans le cas du dépistage du SIDA où il est important de ne rater aucun diagnostic on privilégie la sensibilité dans le premier test de dépistage).

### VALEUR PRÉDICTIVE POSITIVE (VPP)

Probabilité d'avoir la maladie quand le test est positif:  $a/a + b$

### VALEUR PRÉDICTIVE NÉGATIVE (VPN)

Probabilité de ne pas avoir la maladie quand le test est négatif:  $d/b + d$

Les valeurs prédictives dépendent du contexte clinique (probabilité prétest) et des caractéristiques (sensibilité et spécificité) du test.

## ABSTRACT



The article provides the definition of the principal statistical terms used in the epidemiological and therapeutic studies.

**Keywords:**  
critical lecture, EBM,  
general medicine.

## RÉSUMÉ

L'article fournit la définition des principaux termes statistiques utilisés dans les études épidémiologiques et thérapeutiques.

**Mots clefs:**  
lecture critique, EBM,  
médecine générale.

Plus le test est sensible, plus la valeur prédictive négative est meilleure et plus augmente l'assurance du médecin que le patient avec un test normal n'a pas la maladie.

Plus le test est spécifique, plus la valeur prédictive positive est meilleure et plus augmente la certitude que le patient avec un test anormal est malade.

### PROBABILITÉ PRÉTEST

Probabilité d'existence de la maladie d'après le contexte clinique avant la réalisation du test. Un test négatif diminue la probabilité, un test positif l'augmente.

Un résultat de valeur prédictive dans une population donnée n'est pas applicable à une autre population où la probabilité pré-test serait différente. Un exemple est celui du diagnostic d'entorses de la cheville dont la probabilité pré-test est de 1 à 4% en médecine générale et 13% dans les services d'urgence. Une probabilité pré-test faible donne une valeur prédictive positive faible et une valeur prédictive négative augmentée.

### PROBABILITÉ POSTTEST

Probabilité d'existence de la maladie après le test. Si un test n'a aucun faux positif, la probabilité post test est de 100%

Le test idéal est celui qui est à la fois très sensible et très spécifique. Mais l'idéal étant rarement voire jamais atteint, on cherche des compromis et on utilise alors les éléments qui suivent.

### SEUIL DE POSITIVITÉ

Le meilleur compromis entre spécificité et sensibilité d'un test c'est de définir la valeur optimale du seuil de positivité d'un test, c'est-à-dire le choix du seuil déterminant le normal et le pathologique ou l'état de malade ou non malade.

Le choix de la valeur seuil aura une influence sur le taux de faux positifs et faux négatifs. Selon le seuil choisi, il y aura plus ou moins de valeurs normales chez les patients malades ou de valeurs pathologiques chez les patients sains (cf. tableau 1). Le seuil idéal est donc celui qui permet de séparer totalement tous les positifs des négatifs sans faux positifs ni faux négatifs.

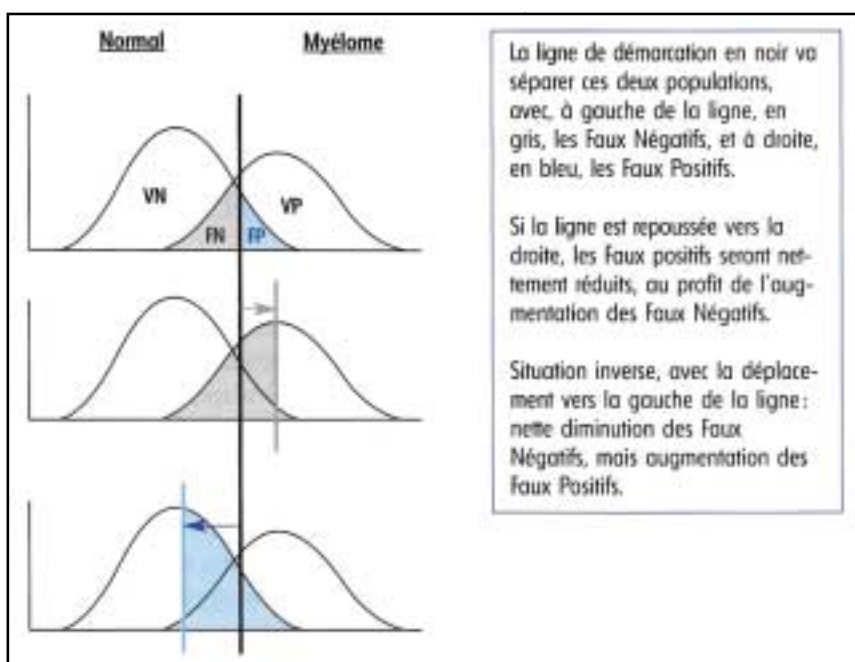


Tableau 1 : représentation des 2 populations de valeurs de protéines plasmatiques chez des patients normaux et atteints de myélome multiple. Effets du changement de la valeur seuil définissant les valeurs pathologiques sur les faux positifs et les faux négatifs.

Un test positif aura pour conséquence un examen complémentaire ou une attitude thérapeutique.

Ici intervient le **rapport coût-bénéfice**. Le **coût** est le fait de traiter à tort un individu non malade, le **bénéfice** est le fait de traiter à bon escient un individu malade.

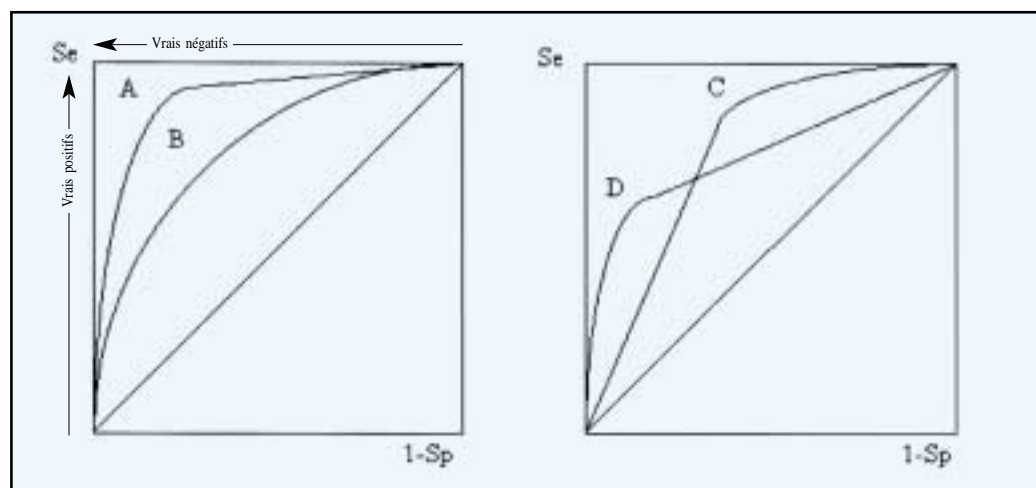
Si les faux positifs et faux négatifs ont le même coût, on choisit comme seuil celui qui donne globalement le plus de diagnostic exact: vrais positifs + vrais négatifs/total patients.

Si les faux positifs ont un coût supérieur, on choisira un test de sensibilité plus faible même au risque d'un plus grand nombre de faux négatifs.

### COURBE ROC

(Receiver Operating Characteristics) est une courbe qui exprime la relation entre la sensibilité et la spécificité d'un test diagnostique. Cette courbe montre la relation entre le % de faux positifs et de vrais positifs suivant les diffé-

Tableau 2 : courbe ROC.



Les tableaux 1 et 2 sont extraits de Junod A. F.: Décision médicale ou la quête de l'explicite Ed Med et Hyg Paris 2003

rentes valeurs de seuils de positivités possibles. Cette courbe se construit de façon empirique en calculant la sensibilité puis la spécificité d'un test pour différents niveaux de seuils de sensibilité. Sur l'axe des abscisses, la variable 1-spécificité est égale au nombre de faux positifs parmi les non-malades.

Lorsque la courbe couvre la diagonale, le test est non informatif, le seuil ne permet pas de discriminer les positifs des négatifs, la proportion de tests positifs est la même chez les malades et les non-malades.

Plus la surface est grande, plus le test est puissant (cf. courbe ROC). Pour le dire autrement : plus la courbe ROC s'éloigne de la diagonale pour rejoindre l'angle supérieur gauche, plus le test est globalement puissant. Mais un test ne doit pas nécessairement être le meilleur toutes catégories confondues (fig. b: A et B sont les courbes ROC schématiques de deux examens entre lesquels on souhaite choisir. A a une meilleure spécificité et sensibilité que B. A sera préféré à B s'il n'a pas d'effets secondaires. C et D sont les courbes ROC schématiques de deux autres examens paracliniques. C est préféré si on veut privilégier la sensibilité, D est choisi si c'est la spécificité qui prime même au prix d'un plus grand nombre de négatifs...).

### PRÉCISION DU TEST

Ensemble des vrais positifs et vrais négatifs par rapport au total des tests effectués ( $a + d/a + b + c + d$ )

Cette valeur est particulièrement utile pour comparer différents tests pour une même maladie.

### RAPPORT DE VRAISEMBLANCE (RV)

Ce terme ne s'adresse qu'aux individus qui ont la maladie.

Le **RV pour un test positif** correspond au rapport du % des vrais positifs sur celui des faux positifs ou rapport sensibilité/faux positifs ou sensibilité/(1-spécificité).

$RV+ = 1$  le test est inutile

$RV+ > 2$  ou  $3$  : test utile pour confirmer le diagnostic

Le **RV pour un test négatif** est le rapport du % de faux négatifs sur celui des vrais négatifs ou % faux négatifs sur spécificité (1-sensibilité/spécificité).

$RV- < 0.5$  test utile pour exclure le diagnostic.

## ÉVALUATION D'UNE ÉTUDE

### 1) DIFFÉRENTS TYPES D'ÉTUDES

#### Essai clinique

Il s'agit d'une étude d'intervention où l'on compare un groupe de sujets exposés à un facteur ou traités par la molécule à évaluer par rapport à un groupe contrôle.

#### Étude prospective

Essai clinique réalisé dans le temps après réalisation d'un protocole d'étude.

#### Étude rétrospective

Analyses de résultats basée sur les dossiers antérieurs de patients.

#### Case report

Présentation de cas cliniques de patients traités ou présentant un problème déterminé sans comparaison avec un groupe témoins. Leur utilité est surtout de poser les questions permettant de créer des protocoles d'étude.

#### Étude contrôlée

Le groupe traité par le médicament est comparé à un groupe témoin traité par placebo ou par molécule de référence. Ceci permet de tenir compte de l'évolution naturelle de la maladie et de l'effet placebo en cas de comparaison à un placebo.

#### Étude comparative

La comparaison doit se faire versus placebo et accessoirement versus une molécule de référence. Des raisons éthiques empêchent parfois la comparaison au placebo et on doit se contenter alors de l'étude comparative avec une molécule de référence. Mais les deux traitements pourraient dans ce cas se révéler également inactifs.

#### Étude randomisée

Les groupes comparés doivent être comparables et pour cela les patients doivent être pris par tirage au sort pour **éviter les biais de sélection**.

#### Étude en aveugle

Les études menées en double-aveugle permettent d'**éviter le biais d'observation** favorable au nouveau médicament. Ni l'investigateur ni le patient ne savent si le produit administré est le médicament testé ou le témoin (placebo ou médicament de référence).

Dans l'étude en simple aveugle, seul le patient ignore ce qu'on lui administre.

On fait aussi des études en « ouvert ». Pour un même médicament, les résultats des open label studies sont en général meilleurs que ceux des études en double-blind.

#### Étude transversale

On étudie un échantillon de patient sur une courte période pour un problème de santé déterminé. La fréquence de ce problème est **calculé par la prévalence** (cf. ultra).

#### Étude longitudinale

L'échantillon de patient est étudié sur une longue période et les résultats en début et en fin d'étude sont comparés entre eux

#### Étude de cohorte

Elle peut être prospective ou rétrospective et observe un groupe d'individus exposés à un ou des facteurs de risque pendant une période donnée pour voir si le groupe exposé court plus de risque que le groupe témoin. On **calcule le risque par l'incidence et le risque relatif** (cf. ultra).

#### Étude cas-témoin

Une étude cas témoin sélectionne deux groupes de sujets malades et non malades puis étudie

leur exposition à un facteur de risque pour déterminer son rôle dans la survenue de la maladie.

## 2) BIAIS ET ERREURS

La validité d'une étude est fonction des auteurs à éviter les biais ou erreurs.

Le **biais de sélection** (ou de diversité) : application de l'étude à des patients répartis de façon non strictement aléatoire dans les différents bras de l'étude.

Le **biais de vérification** : le test de référence n'est pas réalisé lorsque le test est négatif.

Les **fluctuations d'échantillonnage** : la répétition de la même étude sur un même échantillon et dans les mêmes conditions ne donnera pas nécessairement le même résultat du fait du hasard. On peut le réduire en augmentant la taille de l'échantillon et en utilisant des tests statistiques appropriés.

Les **biais de classement** sont liés à des erreurs d'observation.

## 3) ANALYSE STATISTIQUE

Elle est indispensable pour évaluer la pertinence de l'étude en validant les résultats.

### Valeur p

Cette valeur exprime la probabilité que la différence observée entre le groupe contrôle et le groupe traité soit due au hasard. Elle est comprise entre 0 et 1. Plus elle se rapproche de 0 et plus elle est significative car le risque d'erreur est plus faible. Un  $p < 0.05$  est généralement exigé (moins de 5% de risque d'erreur).

### Intervalle de confiance (IC)

Il s'agit d'une fourchette de valeurs contenant la valeur réelle du paramètre considéré avec une probabilité donnée.

On l'exprime en % ou en déviations standards. L'intervalle [a, b] est centré sur la valeur numérique estimée du paramètre inconnu et contient la valeur vraie avec une probabilité  $\alpha$  fixée *a priori*. Cette probabilité permet de s'adapter aux exigences de l'application.

	Malades Cas dans les enquêtes cas témoins	Non malades Témoins dans les enquêtes cas témoins
Exposés	a	b
Non exposés	c	d

$\alpha$  est le **coefficient de confiance**. Une estimation par intervalle de confiance sera d'autant meilleure que l'intervalle sera petit pour un coefficient de confiance grand.

Cet intervalle de confiance peut être calculé pour toutes les valeurs citées.

Une valeur de 95% est souvent choisie pour l'IC, 95% de chances que la valeur réelle soit bien dans la fourchette des valeurs données.

### Prévalence d'une maladie (P)

C'est la proportion des individus atteints de la maladie étudiée par rapport au nombre total d'individus dans la population étudiée.

$$P = \frac{a}{a + b + c + d}$$

### Incidence d'une maladie (I)

C'est la proportion d'individus atteints de la maladie par rapport au nombre total d'individus présents dans la population étudiée et non malades au début de l'intervalle de temps donné pendant une période déterminée.

$$I = \frac{c}{a + b + c + d} \text{ par période définie (an, mois, décennie)}$$

### Risque attribuable

C'est la différence entre les taux de maladie chez les individus exposés et non exposés.

### Risque absolu

Il s'agit d'une **proportion**, un rapport entre le nombre d'événements observés et le nombre de patients, de sujets testés. Cette proportion est le plus souvent exprimée en pour-cent (%).

### Risque relatif

Cette valeur correspond au rapport des taux de maladie chez les individus exposés et non exposés.

Si  $RR = 1$ , cela signifie qu'il n'y a pas d'association entre l'exposition et la maladie. S'il s'agit de l'exposition à un traitement,  $RR = 1$  signifie que le traitement est inefficace.

$RR < 1$ , le traitement sera jugé efficace

$RR > 1$ , le traitement est nocif.

### Odd ratio

Cette valeur permet d'estimer le risque relatif lorsqu'il est impossible de mesurer les risques de la maladie chez les sujets exposés et non exposés.

L'odd ratio (en français rapport de cote) est toujours supérieur au risque relatif sauf si la maladie étudiée est rare et le risque relatif faible. Dans ce cas les deux valeurs sont proches.

$$OR = \frac{AD}{BC}$$

Il s'agit d'une valeur utilisée surtout dans les études épidémiologiques rétrospectives, dans les enquêtes de cohorte ou dans les enquêtes de type cas/témoins.

### Réduction du risque relatif

C'est l'expression en miroir du risque relatif. Elle ne tient pas compte de la fréquence réelle du risque dans le groupe témoin et ne permet donc pas de mesurer le bénéfice absolu d'une intervention. (Cette valeur est celle qui est le plus souvent présentée par les laboratoires pour présenter les résultats de leurs études...)

$$RRR = 1 - RR$$

### Réduction du risque absolu

Elle exprime la différence entre le risque absolu dans le groupe témoin et celui du groupe traité. On l'exprime en % ou en différence d'incidence par 1000 années-patients. Cette valeur permet de relativiser la réduction du risque relatif. **Cette notion est capitale**. C'est elle que nous devons traquer à la lecture d'une étude.

### Nombre de sujets à traiter

Number needed to treat: NST

Il s'agit du nombre de patients à traiter pour éviter **un seul événement** pathologique en traitant pendant **le même temps** qu'aura duré l'étude.

Elle est fonction inverse de la réduction du risque absolu.

**NST = 1/ARR**

Cette notion est particulièrement utile pour connaître l'ampleur de l'utilité d'un traitement. Toutes ces valeurs RR, OR, RRR et RAR doivent en principe être présentées munies de leur intervalle de confiance.

## 4) MÉTA-ANALYSE

Il s'agit d'une méthode quantitative pour combiner les résultats d'études indépendantes (habituellement tirées de la littérature) et synthétiser les sommaires et les conclusions qui peuvent être employées pour évaluer l'efficacité thérapeutique, planifier de nouvelles études, etc.

On procède à l'analyse statistiques de données provenant de plusieurs analyses.

Les méta analyses sont surtout utiles pour comparer des études aux résultats discordants ou de puissance insuffisante.

La méta-analyse est elle-même exposée à différents types de biais :

- Le **biais de publication** : toutes les études sur le sujet n'ont pas été publiées
- Le **biais de sélection**
- Le **biais d'extraction**

Les résultats sont présentés comme dans le tableau 3.

La longueur des rectangles est proportionnelle à l'intervalle de confiance à 95 %.

Le trait vertical dans chaque rectangle a pour abscisse la valeur du RR ou de l'OR : si le trait est proche de la verticale d'abscisse 1,  $RR < 1$ . Si le rectangle coupe la verticale :  $RR = 1$  est contenu dans l'intervalle de confiance et donc la différence n'est pas significative.

Le rectangle le plus bas est calculé sur l'ensemble des essais c'est-à-dire sur le total des patients. À gauche de la verticale d'abscisse 1, les études sont positives, à droites elles sont négatives.

(Dans l'exemple défini par le tableau 3 le résultat global de la méta-analyse est positif).

## CONCLUSIONS

Cette rubrique avait donc pour seul objectif de fournir une définition non exhaustive de termes dont chacun mériterait un article en lui-même. ■

## POUR EN SAVOIR PLUS

1. Pestiaux D, Bouilliez DJ, Bouniton Marc, Denis B et al: L'information médicale, une jungle à déchiffrer. Ed Quorum Ottignies 1997
2. Junod A. F.: Décision médicale ou la quête de l'explicite. Ed. Med & Hyg Paris 2003
3. Van Driel M, Chevalier P. Minerva Glossaire des termes utilisés en Evidence Based Medicine 2004
4. Devresse P-Y: Le goût du risque *RMG* 2004; (27) 217: 455-7

*Nous remercions le Dr Pierre Chevalier pour la relecture et les corrections de cet article.*

Tableau 3: présentation résultats méta-analyse.

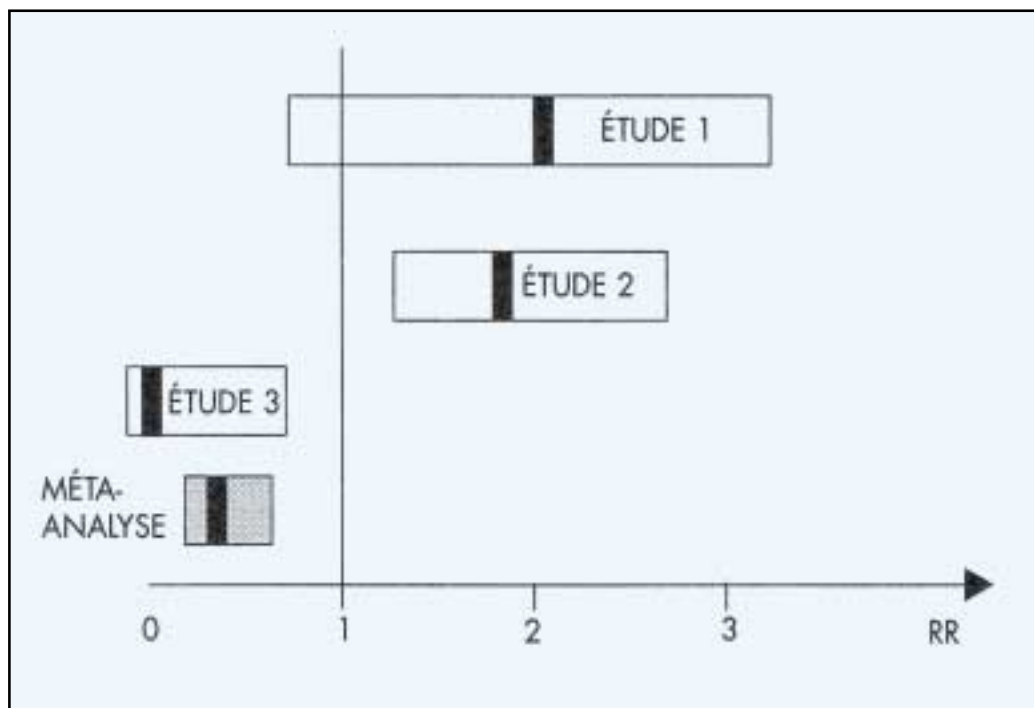


Tableau 3 tiré de Pestiaux D. L'information médicale, une jungle à déchiffrer, 1997.