

# Sur la voie de l'intuition? Théorème de Bayes et diagnostic en médecine générale

## Partie I

D. Pewsner<sup>a</sup>, J. P. Bleuer<sup>b</sup>, H. C. Bucher<sup>c</sup>, M. Battaglia<sup>d</sup>, P. Jüni<sup>d,e</sup>, M. Egger<sup>d,f</sup>

### Introduction

L'incertitude est un compagnon permanent de l'activité médicale: dans le quotidien du médecin, nombreuses sont les décisions qui doivent être prises malgré l'insuffisance des informations disponibles [1]. Les médecins expérimentés se distinguent par leur capacité à évaluer correctement, souvent de manière tout à fait intuitive, la probabilité de certaines maladies à l'aide de l'anamnèse, des symptômes et des signes cliniques. Les inconvénients de la méthode intuitive sont une reproductibilité insuffisante, la difficulté de la transmission didactique et surtout la fiabilité incertaine de notre mémoire [1]. Nous mémorisons en effet mieux les événements récemment vécus que d'autres plus anciens. Les émotions fortes constituent d'autre part un liant très résistant de nos souvenirs. Ainsi, ne pas avoir détecté une tuberculose pulmonaire oligosymptomatique très contagieuse peut nous entraîner à surestimer pendant des années la prévalence de cette maladie, d'où une fréquence disproportionnée de la prescription de radiographies du thorax.

Nous nous efforçons de montrer dans les paragraphes suivants comment l'application conséquente d'un concept mathématique simple permet un diagnostic plus rationnel. Deux conditions préalables jouent un rôle prépondérant en matière de précision d'un diagnostic: la capacité d'évaluer la probabilité de la maladie présumée (prévalence ou probabilité avant test) et la connaissance du potentiel de performance (sensibilité et spécificité) du test utilisé. Le théorème de Bayes exprime ces relations de manière reproductible sous la forme d'un principe mathématique: la probabilité du diagnostic peut être calculée, si la probabilité avant test du diagnostic incertain ainsi que les caractéristiques du test sont connues. De même que les banques de données Cochrane constituent des bases essentielles dans l'objectif d'une thérapie rationnelle, de même devrions-nous pouvoir

nous référer en matière de diagnostic également à une banque de données réunissant avec clarté les propriétés des tests diagnostics et les probabilités avant test significatives.

Nécessité d'autant plus impérieuse que, dans le cadre du cabinet médical, le diagnostic revêt souvent une importance plus grande que la thérapie. Dans le domaine de la thérapie, nous sommes actuellement à même de recourir à des milliers de vues d'ensemble, de méta-analyses et d'études cliniques, alors que de tels efforts en sont encore au stade des balbutiements dans le domaine du diagnostic. Le potentiel d'une banque de données diagnostique est illustré ci-dessous à l'exemple d'un cas de présomption d'appendicite.

### Un agriculteur souffre de douleurs abdominales

Un agriculteur âgé de 25 ans consulte pour des douleurs abdominales relativement intenses, diffuses et jusque-là inconnues. Les douleurs sont apparues pour la première fois 48 heures plus tôt, et se sont progressivement déplacées pour se situer maintenant dans la région hypogastrique droite. Le patient se plaint de légères nausées, les dernières selles, quatre heures auparavant, étaient normales. L'appendice est encore présent. Vous mesurez une température rectale de 38,2°C et constatez un état général légèrement affaibli. Lors de l'examen, vous constatez un abdomen souple et l'absence de douleur au relâchement de la pression abdominale. L'examen rectal ne révèle rien de particulier, les bruits abdominaux sont normaux. L'examen en laboratoire révèle une protéine C-réactive de 25 mg/l, les leucocytes s'élèvent à  $13 \times 10^9/l$ . Le patient, qui vit seul avec sa femme dans une exploitation agricole, ne souhaite en aucun cas être hospitalisé.

Quelle est la probabilité d'une appendicite? Peut-on échapper à une hospitalisation?

<sup>a</sup> Praxis für Innere Medizin FMH, Bern

<sup>b</sup> Mediscope AG PF, 3000 Bern 23

<sup>c</sup> Medizinische Universitäts-Poliklinik, Kantonsspital, Basel

<sup>d</sup> Institut für Sozial- und Präventivmedizin, Universität Bern

<sup>e</sup> Rheumatologische Universitätsklinik, Inselspital, Bern

<sup>f</sup> Division of Health Services Research, Department of Social Medicine, University of Bristol

Correspondance:  
Dr Daniel Pewsner  
Case postale  
CH-3000 Berne 26

[daniel.pewsner@bluewin.ch](mailto:daniel.pewsner@bluewin.ch)

Tableau 1.

**Règles de l'acquisition d'information lors de l'interview anamnestique.**

L'examen de diagnostic différentiel est largement guidé par l'émission d'hypothèses [1, 6]. Ainsi, les données ne sont pas collectées systématiquement dès le début de l'interview. Des hypothèses sont sans cesse émises à partir de l'entretien, et ensuite infirmées ou confirmées dans le cadre d'un processus cyclique en trois étapes:

**Processus cyclique du diagnostic différentiel**

I Ecouter → émettre des hypothèses	
II Contrôler les hypothèses	
III Evaluation – Décision	
IV Thérapie	

Lors de l'émission d'hypothèses, conformément au théorème de Bayes, la fréquence (prévalence) d'une maladie possible joue un rôle prépondérant. Les hypothèses concernant des maladies plus rares, dont le diagnostic est impératif parce qu'elles nécessitent un traitement immédiat, doivent également être contrôlées. A cela s'ajoute le fait que chez un patient jusque-là en bonne santé, l'explication d'une symptomatologie donnée par un seul diagnostic est plus vraisemblable que la combinaison de deux ou de plusieurs diagnostics. Dans le cas d'une symptomatologie respectivement de résultats non-spécifiques, qui rendent plus difficile l'émission d'hypothèses, on utilise les questions de l'anamnèse systémique, qui se distinguent par une sensibilité élevée pour une spécificité la plupart du temps médiocre («Snout-Tests», tab. 2). Si ces questions ne permettent pas d'abandonner une hypothèse, suivent alors des questions de spécificité élevée, permettant l'inclusion de l'hypothèse («SpPin», tab. 2) en cas de résultat positif. Ces principes de l'émission et du contrôle des hypothèses sont valables également pour les examens cliniques et paracliniques.

**Tests diagnostiques**

Le diagnostic différentiel, qui permet de structurer hiérarchiquement et par ordre de probabilité décroissante les diagnostics entrant en ligne de compte, fait partie de tout diagnostic (tab. 1). Il est rarement possible d'évaluer l'état réel d'un patient avec une certitude absolue, dans la mesure où pour la plupart, les tests diagnostiques sont imparfaits. Pour des raisons pratiques, le diagnostic doit cependant être suffisamment sûr, autrement dit se distinguer suffisamment, dans sa probabilité, des autres diagnostics différentiels. Par tests diagnostics, nous n'entendons pas seulement les examens paracliniques réalisés à l'aide d'appareils: les indications issues de l'anamnèse et les résultats des examens cliniques ont eux aussi valeur de tests et permettent des conclusions importantes. Ainsi, la déclaration faite par le patient selon laquelle il ressent une douleur abdominale *d'un type nouveau, d'une qualité jamais encore éprouvée jusque-là*, augmente la probabilité d'une appendicite. La question ciblée de la qualité de la douleur peut donc tout à fait être considérée comme un test.

**Paramètres des performances des tests: sensibilité, spécificité**

Le potentiel de performance d'un test peut être mesuré de différentes manières. Les grandeurs de mesure que sont la *sensibilité* et la *spécifi-*

*cité* sont largement connues [1, 5]. Si nous supposons qu'un résultat de test positif signifie «malade», les principes suivants sont valables:

**Sensibilité:**

Probabilité qu'une personne malade soumise au test présente un résultat de test positif. Désigné comme «taux vrai positif» (True Positive Rate; TPR).

**Spécificité:**

Probabilité qu'une personne en bonne santé soumise au test présente un résultat de test négatif.

Désigné comme «taux vrai négatif» (True Negative Rate; TNR).

La sensibilité et la spécificité sont des paramètres qui décrivent objectivement le potentiel de performance d'un test de diagnostic (tab. 2). Ces paramètres sont déterminés de manière empirique par comparaison avec un test de référence que l'on peut appeler «test standard de référence» («golden standard»). Le test de référence est l'examen qui saisit l'état du patient de la manière la plus fidèle possible à la réalité. La sensibilité et la spécificité ne sont cependant que d'un secours limité en ce qui concerne la pratique: bien plus que la probabilité d'un résultat positif en cas de maladie, c'est la probabilité de la présence de la maladie en cas de résultat positif respectivement négatif qui est intéressante. Ces résultats qui renseignent sur ce que l'on appelle la «probabilité post-test», d'une importance bien supérieure dans la pra-

tique, sont appelés également «valeur prédictive positive» et «valeur prédictive négative» (en anglais «positive» et «negative predictive value»; PPV et NPV).

### Probabilité d'un diagnostic et prévalence de la maladie: le théorème de Bayes

Un bon diagnostic repose sur le principe que ce qui est rare arrive rarement, et ce qui est fré-

Figure 1.

Thomas Bayes (1702–1761) était pasteur de l'église presbytérienne de Tunbridge Wells, à 50 km au sud de Londres. Malgré une certaine notoriété acquise de son vivant comme mathématicien, son principal ouvrage «*Essay towards Solving a Problem in the Doctrine of Changes*» n'a paru qu'en 1764 à titre posthume dans les «*Philosophical Transactions of the Royal Society of London*».



quent arrive souvent: «If you hear hoofbeats, think of horses, not zebras» [1]. Ainsi, en matière de diagnostic différentiel, les priorités de la grippe et de la malaria ne seront pas les mêmes chez un patient vivant à Zurich et chez une personne rentrant d'un séjour à Abidjan. (tab. 1). La probabilité de la présence d'une certaine maladie peut ainsi être déduite aussi de la fréquence (prévalence) à laquelle cette maladie intervient dans un groupe d'individus comparables: les indications sur la prévalence, qui correspond à la probabilité avant test des maladies, se trouvent dans la littérature. Mais les articles significatifs ne sont pas toujours faciles à trouver. L'indexation de ces articles dans les banques de données bibliographiques (dans MEDLINE par exemple) n'est pas unitaire. Les probabilités avant test sont dépendantes du collectif exploré, et vous ne trouvez souvent qu'un petit nombre de travaux relatifs à un collectif de patients représentatif pour les médecins généralistes, même à l'aide de stratégies de recherche étendues.

Le théorème développé au 18<sup>e</sup> siècle par le prêtre anglais Thomas Bayes (fig. 1) énonce que pour des qualités de test donnée (sensibilité et spécificité), le potentiel informatif de ce test dépend de la probabilité du diagnostic recherché avant le test. Le principe est plus facile à comprendre si nous prenons conscience qu'un certain résultat de test ne confirme ou n'infirme pas un diagnostic avec certitude, mais seulement avec une certaine probabilité. La proba-

Tableau 2. «SpPins» et «SnNouts».

On peut déduire de la définition de la sensibilité la règle approximative selon laquelle un test négatif de sensibilité élevée (faible taux faux négatif) permet l'exclusion d'une maladie recherchée (par ex. D-Dimères [ELISA] lors de la recherche d'une embolie pulmonaire), mais n'autorise par contre pas l'inclusion en présence d'un résultat positif. Pour être plus explicite, imaginons un pêcheur qui utilise un filet aux mailles très serrées (= sensibilité élevée), afin de ne pas laisser échapper les poissons de petite taille. Il doit par contre accepter le risque de récolter une grande quantité de déchets dans son filet (un filet lourd ne doit pas induire la conclusion qu'un poisson a été pris). Cette règle est souvent appelée «SnNout» dans la littérature anglo-saxonne [2, 15].

«SnNout»: «*Sensitivity Negative rule the target disorder out*»: un résultat de test négatif d'un test de sensibilité élevée permet l'exclusion d'une maladie.

Inversement, la règle approximative selon laquelle un test positif de spécificité élevée (faible taux faux positif) permet l'inclusion d'une maladie recherchée (par ex. mise en culture de salive Tbc positive), mais n'autorise par contre pas son exclusion en présence d'un résultat négatif. De tels résultats sont également appelés «pathognomoniques». Un filet aux mailles larges ou mieux une nasse permet certes à de nombreux poissons de s'échapper, mais la probabilité est cependant élevée que lors d'une prise rare, un poisson recherché soit réellement capturé. On parle alors de «SpPin».

«SpPin»: «*Specificity Positive rule the target disorder in*»: un résultat de test positif d'un test de spécificité élevée permet l'inclusion d'une maladie.

Ajoutons à titre limitatif que le potentiel informatif d'un test – qu'il s'agisse d'exclusion ou d'inclusion – dépend à la fois de la sensibilité et de la spécificité. C'est pourquoi les règles de «SnNout» et de «SpPin» ne sont valables que si la spécificité respectivement la sensibilité correspondante ne se situent pas au-dessous de 40 à 50%; une condition que remplissent d'ailleurs la plupart des tests usuels.

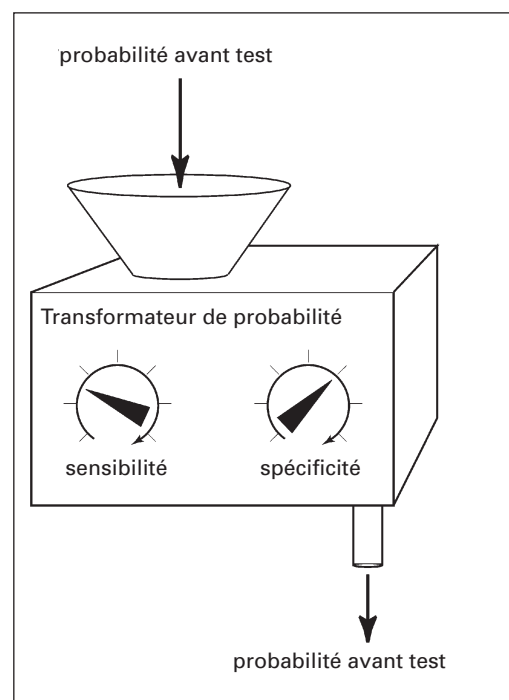


Figure 2.

En fonction de son potentiel de performance (sensibilité resp. spécificité), un test transforme une probabilité avant test donnée en probabilité post-test.

bilité après le test (probabilité post-test) se trouve être naturellement d'autant plus élevée qu'elle était plus grande avant le test. Autrement dit, un test qui transforme la probabilité avant le test en une probabilité après le test peut être qualifié de «*transformateur de probabilité*» (fig 2). Cette probabilité post-test correspond aux valeurs prédictives positive et négative (PPV resp. NPV). Dans le cas d'une probabilité avant test faible, un test doit présenter des qualités de test exceptionnelles pour générer une probabilité post-test élevée. En d'autres termes, peu nombreux sont les tests susceptibles de «changer l'eau en vin».

### Le coefficient de probabilité (Likelihood-Ratio)

Le coefficient de probabilité [1, 2, 5] décrit la performance des tests en fonction de la probabilité avant test. Il réunit les qualités de sensibilité et de spécificité du test en une seule valeur et représente ainsi un paramètre objectif de la performance du test. Celle-ci est définie comme le rapport des probabilités d'un résultat de test déterminé chez les malades et les personnes en bonne santé. *Le coefficient de probabilité pour un test positif* (CP+) indique combien de fois il est plus probable qu'un résultat de test déterminé (une indication d'anamnèse, un symptôme, un résultat de laboratoire ou tout autre test) soit *obtenu* chez un malade que chez une personne en bonne santé, elle constitue une valeur de mesure de la «puissance d'inclusion». *Le coefficient de probabilité pour un test négatif* (CP-) indique combien de fois il est plus probable qu'un résultat de test déterminé (une indication d'anamnèse, un symptôme, un résultat de laboratoire ou tout autre test) soit *absent* chez un malade que chez une personne en bonne santé, elle est une valeur de mesure de la «puissance d'exclusion».

LR+:	$\frac{\text{Taux des vrais positifs au test}}{\text{Taux des faux positifs au test}}$	=	$\frac{\text{TPR}}{\text{FPR}}$	=	$\frac{\text{Sensibilité}}{1-\text{Spécificité}}$
LR-:	$\frac{\text{Taux des faux négatifs au test}}{\text{Taux des vrais positifs au test}}$	=	$\frac{\text{FNR}}{\text{TNR}}$	=	$\frac{1-\text{Sensibilité}}{\text{Spécificité}}$

Dans le cas d'un résultat de test positif, le complément d'information est d'autant plus élevé que le CP+ du test est important. Inversement, un résultat négatif est d'autant plus expressif qu'il se rapproche plus du CP zéro. Un CP+ respectivement un CP- de 1 n'apporte aucune information. Un CP+ de plus de 10 permet de confirmer avec une probabilité élevée une maladie recherchée, sauf dans le cas d'une probabilité avant test très faible. Inversement, un CP- de moins de 0,1 permet d'exclure une maladie

recherchée avec une probabilité élevée, à condition que la probabilité avant test ne soit pas particulièrement élevée.

L'un des grands avantages du coefficient de probabilité (Likelihood Ratio) est que sa multiplication avec la probabilité avant test permet le calcul direct de la probabilité post-test (probabilité avant test  $\times$  coefficient de probabilité = probabilité post-test). Ceci présuppose cependant une autre «unité» de mesure de la probabilité: nous avons l'habitude de définir les probabilités en pourcentages. Ceux-ci expriment le rapport d'événements positifs à l'ensemble des événements (positifs et négatifs). En ce qui concerne la probabilité avant test, elle correspond à la relation entre les *malades* et l'ensemble des personnes examinées (*malades/personnes en bonne santé et malades*). Les statisticiens donnent la préférence à l'«unité» appelée Odds pour désigner la probabilité, dans la mesure où celle-ci est utilisable mathématiquement. Les Odds correspondent au rapport entre les événements positifs et négatifs, c'est-à-dire entre les *malades* et les *personnes en bonne santé* (*malades/personnes en bonne santé*). Nous connaissons cette manière d'exprimer la probabilité dans le jargon des courses de chevaux, où les chances de victoire d'un cheval déterminé ne sont pas indiquées par 20% mais par 4 contre 1 (correspondant à 20%/ [100%-20%] = 20%/80%). Cette «unité» n'est pas très courante chez nous, alors qu'elle est par exemple familière même aux jeunes enfants en Angleterre, où les paris sur les chevaux sont très populaires.

La relation suivante est donc valable: Odds avant test  $\times$  coefficient de probabilité = Odds de post-test.

Un autre avantage du coefficient de probabilité (Likelihood Ratio) réside dans la possibilité de multiplier entre eux les coefficients correspondants lors de la réalisation de plusieurs tests. On obtient ainsi le coefficient de probabilité (CP) total de la combinaison de tests. Si celui-ci est multiplié par l'Odds de pré-test, on obtient l'Odds de post-test pour l'ensemble des tests réalisés, l'ordre chronologique des tests ne jouant en l'occurrence aucun rôle.

Odds avant test  $\times$  CP<sub>Test1</sub>  $\times$  CP<sub>Test2</sub>  $\times$  CP<sub>Test3</sub> = Odds de post-test.

### Un agriculteur souffre de douleurs abdominales: quelle est la probabilité d'une appendicite?

Dans la deuxième partie de cet article, à paraître dans le prochain numéro de «Swiss Medical Forum», nous montrerons comment il est possible de calculer, à l'aide du théorème de Bayes, la probabilité post-test d'une appendi-

cite dans le cas décrit ci-dessus, en se basant sur la prévalence et les caractéristiques de test. Nous y expliquerons également le concept des seuils de décision, qui, dans le cas d'une maladie déterminée, permettent d'agir de manière appropriée en fonction de la probabilité post-test établie. Nous nous pencherons d'autre part sur le potentiel d'une banque de données diagnostiques, destinée à permettre aux médecins généralistes d'appliquer les principes du théorème de Bayes dans la pratique quotidienne du cabinet médical.

## Remerciements

Nous tenons à remercier les docteurs A. Jaggi et Ch. Junker, Berne, de la lecture critique de ce manuscrit.

## Références

- 1 Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision-Making. Boston: Butterworth-Heinemann; 1987.
- 2 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine. New York: Churchill Livingstone; 1997.
- 3 Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. JAMA 1995;274:1935-8.
- 4 Antes G, Oxman AD. The Cochrane Collaboration. In: Egger M, Smith GD, Altman DG, eds. Systematic Reviews in Health Care: Meta-Analysis in Context. London: Br Med J Books; 2000.
- 5 Bucher HC, Schmidt JG, Steurer J. Kritische Beurteilung einer Arbeit zu einem diagnostischen Test. Schweiz Rundsch Med Prax 1998; 87:1096-102.
- 6 Kassirer JP, Gorry GA. Clinical problem solving: a behavioral analysis. Ann Intern Med 1978;89:245-55.
- 7 Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13:313-21.