

## SEANCE DE L'A.E.S.A.

# L'évaluation des méthodes diagnostiques

E. THIRY, P.-P. PASTORET

Chaire de Virologie-Immunologie,  
Faculté de Médecine Vétérinaire,  
Université de Liège,  
Bât. B6 Sart Tilman  
B- 4000 LIEGE

Manuscrit déposé le 00/00/1992

### RESUME

Les méthodes qui permettent la validation des tests de diagnostic sont revues. Les valeurs intrinsèques des tests de diagnostic, la sensibilité et la spécificité, sont décrites et les conditions de leur variation sont présentées. Les valeurs prédictives du résultat positif et du résultat négatif sont définies et leur liaison avec la prévalence de la maladie dans la population est décrite. Le choix de la valeur seuil de la méthode de diagnostic est également discuté.

### 1. INTRODUCTION

Les tests de diagnostic revêtent une importance de plus en plus considérable en médecine vétérinaire. Les campagnes de contrôle et d'éradication de maladies contagieuses sont certainement concernées au premier chef par ces tests. A tous les niveaux cependant, et même si l'intérêt sanitaire ou économique ne prime pas, la nécessité d'un diagnostic de qualité et présentant une fiabilité suffisante se fait de plus en plus pressante.

Il existe des critères objectifs qui permettent d'évaluer les performances d'un test de diagnostic. La validation d'un test est une étape indispensable dans son développement. Néanmoins, cette étape est encore trop souvent négligée, principalement pour les tests qui ne sont pas utilisés dans le cadre d'une police sanitaire.

Lorsque les méthodes de validation sont appliquées, elles apportent une aide à la décision dans le choix du test de diagnostic le plus adéquat. Ce choix sera fonction de plusieurs éléments. La méthode utilisée dans le test de diagnostic sera d'abord examinée : s'agit-il d'une méthode

quantitative (titrage d'anticorps, par exemple) ou qualitative (absence ou présence de symptômes ou lésions) ? Ensuite, le cadre de l'application du test sera envisagé : diagnostic de routine ou programme d'éradication d'une maladie ? Enfin, le choix doit prendre en considération la population dans laquelle le test de diagnostic sera employé. En effet, la prévalence de la maladie, c'est-à-dire la proportion d'animaux malades dans la population à un moment donné, est un élément qui intervient dans les valeurs prédictives du test étudié.

Le test sérologique est un exemple particulièrement courant de test de diagnostic. Il est principalement réalisé par la méthode ELISA (Enzyme Linked ImmunoSorbent Assay) qui assigne une densité optique mesurée par un spectrophotomètre à une concentration d'anticorps spécifiques dans le sérum d'un animal donné. Les valeurs obtenues par un tel test ELISA sont continues. Pour attribuer à un animal un diagnostic positif ou négatif, c'est-à-dire un statut séropositif ou séronégatif, il faut définir une valeur seuil, appelée également cut-off, en dessous de laquelle le résultat est négatif et au-

dessus de laquelle le résultat est positif. La possibilité de conserver des résultats douteux ne sera pas abordée dans le cadre de cet article. La valeur seuil n'est pas définie de manière définitive. Elle peut en effet évoluer selon les performances que l'on attend du test.

Ces performances attendues, qui sont également les critères d'évaluation des tests de diagnostic, sont la sensibilité et la spécificité ainsi que les valeurs prédictives d'un résultat positif et d'un résultat négatif. Les méthodes de validation des tests utilisent ces valeurs en les exprimant ou les présentant de manière à faciliter la prise de décision. Par exemple, le coût (non seulement financier, mais aussi médical ou sanitaire) du faux positif ou du faux négatif peut y être introduit.

### 2. LES VALEURS INTRINSEQUES DU TEST

Quelles sont les capacités du test à diagnostiquer les vrais positifs et les vrais négatifs ? La sensibilité et la spécificité permettent de mesurer ces capacités. En effet, un test de diagnostic doit permettre de distin-

guer les animaux malades des non malades. Le test peut présenter des imperfections, et c'est généralement le cas. Certains animaux malades ne sont pas dépistés par le test et, inversément, des animaux sains sont reconnus à tort positifs par le test.

Ainsi, les résultats d'un test, confrontés au statut réel des animaux à l'étude, vont répartir les animaux dans quatre groupes : les vrais positifs et les vrais négatifs pour lesquels il y a accord entre le résultat du test et leur état réel ; les faux négatifs et les faux positifs, qui révèlent une discordance entre le résultat du test et le statut réel des animaux testés. La sensibilité est la probabilité conditionnelle de diagnostiquer les vrais positifs parmi les animaux malades. La spécificité est la probabilité conditionnelle de diagnostiquer les vrais négatifs parmi ces animaux non-malades (Bernard et Lapointe, 1987). Ce sont des valeurs intrinsèques du test parce qu'elles ne sont pas affectées par la prévalence de la maladie.

Pour étudier de manière valable ces deux valeurs, il faut posséder une méthode de référence pour le diagnostic de la maladie. Cette méthode permet de classer les animaux en groupe malade et groupe non-malade. Ces animaux sont alors soumis au test de diagnostic, ce qui permet d'obtenir les groupes positifs et négatifs selon le test. La prévalence dans l'échantillon d'animaux utilisé pour mesurer sensibilité et spécificité ne doit pas être représentative de la prévalence dans la population puisque ces deux valeurs sont indépendantes de la prévalence. La table de contingence qui décrit les différentes catégories qui viennent d'être définies peut être produite (Table 1).

La sensibilité (se) peut être exprimée par le rapport :

$$se = \frac{VP}{VP + FN} = \frac{a}{a + c}$$

La spécificité (sp) peut être exprimée par le rapport :

$$sp = \frac{VN}{VN + FP} = \frac{d}{b + d}$$

**Table 1**  
Table de contingence décrivant les différents effectifs d'un échantillon de population selon les critères de présence ou d'absence de maladie et de réaction positive ou négative à un test de diagnostic

		MALADIE		N = a + b + c + d
		OUI	NON	
TEST	+	a VP	b FP	a + b
	-	c FN	d VN	c + d
		a + c	b + d	

VP : vrais positifs ; FP : faux positifs ; FN : faux négatifs ; VN : vrais négatifs ; N : nombre total d'animaux.

Il faut insister sur le fait que, dans cette population, la maladie est diagnostiquée de manière non équivoque. Les résultats du test sont confrontés à la réalité : présence ou absence de maladie. Ils ne sont pas comparés aux résultats obtenus par un autre test de diagnostic.

La table de contingence décrit le cas d'un résultat de diagnostic dichotomique : positif ou négatif.

Quand les valeurs obtenues par un test sont continues, par exemple dans un test ELISA, la sensibilité et la spécificité sont fonction du seuil de positivité. Un seuil de positivité élevé augmente la spécificité du test alors qu'un seuil de positivité de valeur faible augmente la sensibilité du test. Cet exemple démontre clairement que la sensibilité et la spécificité varient en sens inverse.

Il faut considérer que la sensibilité et la spécificité doivent toutes deux atteindre la valeur 0,8 pour que le test soit valable (Jenicek et Cleroux, 1983). Néanmoins, d'autres facteurs peuvent influencer la décision d'utiliser un test qui présente des valeurs intrinsèques différentes.

Bien que l'on attende d'un test de diagnostic qu'il présente à la fois une bonne sensibilité et une bonne spécificité, la stratégie d'utilisation du test décidera de la valeur qu'il convient d'améliorer. En général, le dépistage initial d'une maladie demande une grande sensibilité (c'est-à-dire la détection du maximum de vrais positifs) aux dépens d'une moins bonne spécificité. On accepte une certaine proportion de faux positifs qui seront ensuite éliminés par

un diagnostic plus précis (Thrusfield, 1986 ; Jenicek et Cleroux, 1983). Le critère de spécificité devient prépondérant lorsque l'éradication d'une maladie arrive à son point ultime. Ces considérations montrent que le choix du test ou du seuil de détection du test est étroitement dépendant des caractéristiques de la population. Ce lien est objectivé par les valeurs prédictives.

### 3. LES VALEURS PREDICTIVES DU TEST

Quelles sont les capacités du test à dépister les vrais positifs parmi l'ensemble des résultats positifs et les vrais négatifs parmi l'ensemble des résultats négatifs ? Il y a en effet des faux positifs et des faux négatifs.

Les valeurs prédictives sont définies par les probabilités conditionnelles qu'un animal positif pour le test soit un vrai positif, donc un vrai malade (valeur prédictive du résultat positif, VPP) et qu'un animal négatif pour le test soit un vrai négatif, donc un animal indemne (valeur prédictive du résultat négatif, VPN).

Les valeurs prédictives dépendent des valeurs intrinsèques du test, sensibilité et spécificité, et de la prévalence de la maladie. Celle-ci affecte en effet la proportion d'animaux positifs qui sont réellement malades dans la population qui va être testée.

Les valeurs prédictives sont exprimées par les rapports suivants (qui font référence à la table de contingence, table 1) :

$$VPP = \frac{VP}{VP + FP} = \frac{a}{a + b}$$

$$VPN = \frac{VN}{VN + FN} = \frac{d}{c + d}$$

Le calcul des valeurs prédictives ne peut se réaliser que sur un échantillon représentatif de la population cible qui en possède donc la prévalence. En effet, les valeurs prédictives sont affectées par la prévalence. Le théorème de Bayes démontre ce lien avec la prévalence et les équations des valeurs prédictives peuvent être réécrites :

$$VPP = \frac{P \times se}{P \times se + (1-P) \times (1-sp)}$$

$$VPN = \frac{(1-P)sp}{(1-P)sp + P(1-se)}$$

où P est la prévalence de la maladie dans la population étudiée. (Bernard et Lapointe, 1987 ; Rumeau-Rouquette et al., 1990).

Une prévalence élevée augmente la VPP ; une faible prévalence diminue la VPN. Ainsi le même test de diagnostic présentera des valeurs prédictives différentes selon la prévalence de la maladie (André-Fontaine, 1990) (Figure 1). Cette notion a d'importantes implications dans les programmes de contrôle et d'éradication de maladie où la prévalence de la maladie combattue di-

minue au cours du programme de lutte.

La sensibilité et la spécificité font également varier les valeurs prédictives. Si sensibilité et spécificité sont supérieures à 0,5, la VPP augmente lorsque la spécificité augmente (Figure 2) ; de même, la VPN augmente, mais dans une moindre mesure, lorsque la sensibilité augmente (Bernard et Lapointe, 1987 ; André-Fontaine, 1990).

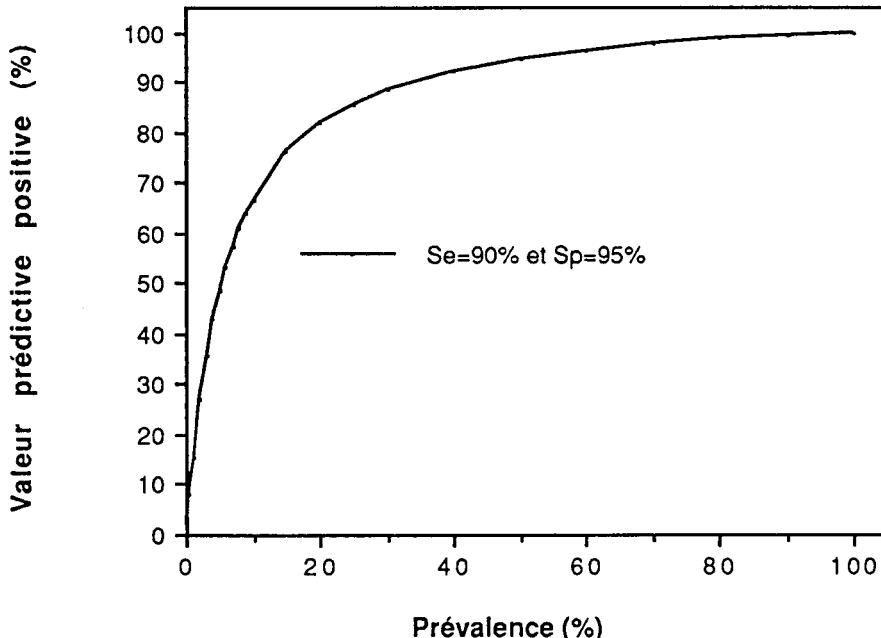


Figure 1

Variation de la valeur prédictive du résultat positif en fonction de la prévalence de la maladie.  
Cette simulation est donnée pour une sensibilité de 90 % et une spécificité de 95 %.

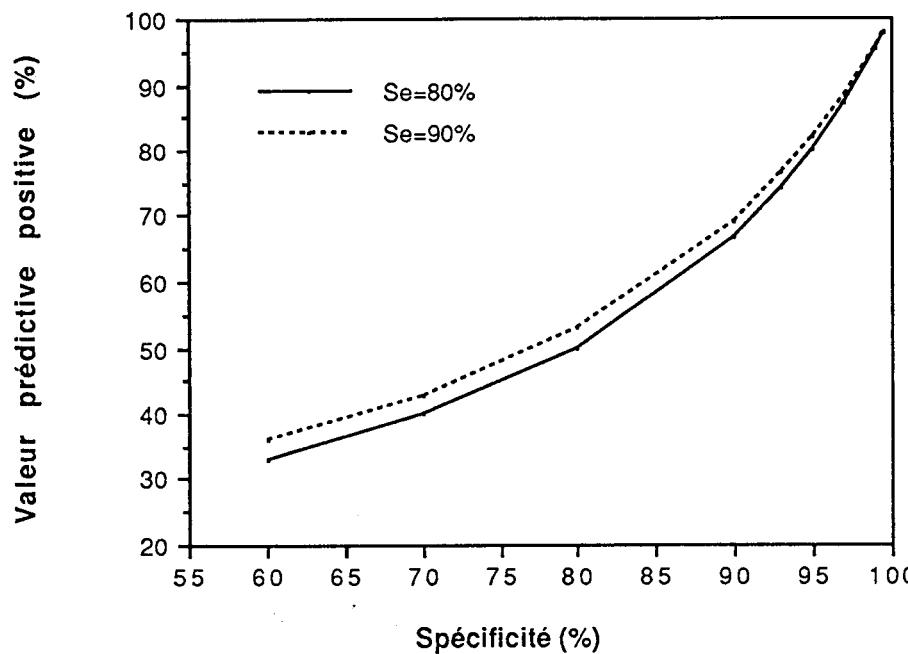


Figure 2

Variation de la valeur prédictive du résultat positif en fonction de la spécificité.  
Deux simulations sont proposées pour les valeurs de sensibilité de 80 % et de 90 %.

Une difficulté rencontrée dans la mesure des valeurs prédictives est la valeur de la prévalence. Celle-ci est souvent obtenue par le même test qui est étudié. Cette prévalence apparente est la somme de deux composantes : les vrais positifs et les faux positifs. Elle est donc surestimée par rapport à la vraie prévalence, et cela d'autant plus que la prévalence est faible. Une correction peut être apportée à la prévalence apparente lorsque sensibilité et spécificité sont connues :

$$P_a = (P \times se) + (1-P) \times (1-sp)$$

avec  $P_a$  qui est la prévalence apparente où le premier terme représente les vrais positifs et le second terme, les faux positifs.

Cette équation permet d'obtenir la prévalence corrigée :

$$P = \frac{P_a + sp - 1}{se + sp - 1}$$

qui est une valeur particulièrement utile à connaître lorsque la prévalence apparente est faible (Thrusfield, 1986).

#### 4. LE CHOIX DU SEUIL DE DETECTION

La mise au point d'un test sérologique, par exemple un test ELISA, implique le choix du seuil. Le calcul des valeurs intrinsèques du test sera appliqué pour ce choix. La première étape consiste donc à calculer sensibilité et spécificité pour différentes valeurs de seuil.

Une courbe R.O.C. (Receiver Operating Curve) peut être déduite, qui représente chaque seuil par un point ayant  $(1-sp)$  comme abscisse et  $se$  comme ordonnée. Le choix du seuil se fait sur la courbe en fonction

de la meilleure combinaison de sensibilité et de spécificité (Laplanche et al., 1987).

L'indice de Youden,  $1 - (se + sp)$ , combine sensibilité et spécificité. Le seuil choisi est celui qui minimise l'indice de Youden. Il faut néanmoins remarquer que cet indice ne tient pas compte de la prévalence (Pumeau-Rouquette et al., 1990).

Ces méthodes de décision du seuil ne prennent pas en considération le poids que l'on veut donner aux faux positifs et aux faux négatifs. Or le coût des erreurs résultant du test est un élément important dans la stratégie du choix du seuil. Cette notion de coût déborde l'aspect financier. Elle rend aussi compte de l'importance qu'on donne soit au faux positif, soit au faux négatif. Il est d'ailleurs plus aisés d'estimer un coût relatif. Par exemple, le faux négatif peut représenter un coût double du faux positif dans la stratégie suivie.

Il est très difficile d'avoir une bonne estimation du coût total des erreurs. Par contre, le coût relatif est plus aisés à obtenir. Il sera exprimé sous la forme d'un rapport entre le coût des faux négatifs et le coût des faux positifs. Sa valeur sera fonction de l'importance relative des deux types d'erreur.

$$\text{coût total des faux positifs} = (1 - P)(1 - sp) \times n \times C_1$$

$$\text{coût total des faux négatifs} = P(1 - se) \times n \times C_2$$

où  $n$  est le nombre d'animaux,  $C_1$  le coût attribué aux faux positifs,  $C_2$  le coût attribué aux faux négatifs.

Les coûts peuvent être exprimés sous la forme d'un rapport :  $C_1/C_2$ . Si  $C_1$  a une importance relative deux fois supérieure à  $C_2$ , le rapport vaut 2 et  $C_1 = 2C_2$  (Rumeau-Rouquette et al., 1990).

## 5. CONCLUSIONS

Les qualités d'un test de diagnostic dépassent de loin ses performances biologiques ou biochimiques. Ces dernières sont évidemment un prérequis et l'utilisation de produits issus des biotechnologies tels que les anticorps monoclonaux, les sondes nucléiques et l'amplification sélective du DNA améliore grandement la qualité du test. Néanmoins, le test doit être validé par la mesure de paramètres tels que sensibilité, spécificité et valeurs prédictives qui étudient son comportement vis-à-vis de la population cible du diagnostic. A cet égard, la variation des valeurs prédictives en fonction de la prévalence est particulièrement démonstrative.

Un test de diagnostic n'est donc pas un outil désincarné : les performances qu'il manifeste au cours de

sa mise au point au laboratoire doivent être réévaluées lors de son développement, quand il est confronté à la population dans laquelle il doit être utilisé.

## Remerciements

Tous nos remerciements vont à Mylène Lemaire, Bernard Mignon et Bernard Limbourg pour les stimulantes discussions. Ce travail est subsidié par le Ministère de l'Agriculture de la Région Wallonne, dans le cadre d'une mission sur le contrôle de la rhinotrachéite infectieuse bovine.

## SUMMARY

### Evaluation of diagnostic methods

Methods allowing the validation of diagnostic tests are reviewed. Intrinsic values of diagnostic tests, sensitivity and specificity, are described and factors affecting their variation are presented. Predictive values of positive and negative results are defined and their relationship with disease prevalence in the population is described. The choice of the cut-off value of the diagnostic method is also discussed.

## REFERENCES

- André-Fontaine G. Sensibilité, spécificité, valeurs prédictives : impact de la qualité des analyses en épidémiologie, in : Épidémiologie et Santé Animale, Bulletin de l'Association pour l'Etude de l'Epidémiologie des Maladies Animales, n° 17, 1990, p. 29-40.
- Bernard P.-M., Lapointe C. Mesure de validité des tests diagnostiques (ou de dépistage), in : Mesures statistiques en épidémiologie, Presses de l'Université du Québec, 1987, p. 176-190.
- Jenicek M., Cleroux R. Epidémiologie. Principes. Techniques. Applications. Edisem Inc., Ste Hyacinthe, Québec, 1983.
- Laplanche A., Com-Noughé C., Flamant R. Evaluation des méthodes diagnostiques, in : Méthodes statistiques appliquées à la recherche clinique, Médecine-Sciences Flammarion, Paris, 1987, p. 145-155.
- Rumeau-Rouquette C., Breart G., Padieu R. Analyse des études d'évaluation, in : Méthodes en Épidémiologie, Médecine-Sciences Flammarion, Paris, 1990, p. 376-388.
- Thrusfield M. Serological epidemiology, in : Veterinary epidemiology, Butterworths, London, 1986, p. 175-186.