

MÉTHODES D'ÉVALUATION D'UN TEST DE DÉPISTAGE EN L'ABSENCE DE RÉFÉRENCE *

Anne Praud¹, Barbara Dufour¹, Laurence Meyer² et Bruno Garin-Bastuji³

RÉSUMÉ

L'étude et le choix des stratégies décisionnelles en santé animale nécessitent de disposer d'informations sur l'efficacité des outils disponibles, qu'il s'agisse d'un test unique, d'une association de plusieurs tests ou d'un ensemble de critères cliniques et épidémiologiques. Lorsqu'il existe une référence permettant de connaître le statut infectieux des individus étudiés, l'estimation des caractéristiques des tests peut être effectuée de manière directe. En santé animale, il est toutefois fréquent de ne pas disposer de cette information, lorsqu'aucun test de référence n'est disponible ou lorsqu'il n'a pas pu être mis en œuvre, pour des raisons pratiques, économiques ou éthiques. Dans ce cas, l'étude des tests requiert l'utilisation de méthodes statistiques adaptées, telles que l'utilisation de modèles à classe latente implémentés par une approche bayésienne.

Mots-clés : test de dépistage, approche bayésienne, référence, sensibilité, spécificité.

SUMMARY

Assessing the accuracy of available diagnostic tools, whether a single test, a combination of tests or a group of clinical and epidemiological criteria, is essential to elaborate and eventually make decision strategies. Whenever there exists a reference providing information on the infection status of individual animals, the evaluation of the tests can be made directly. However, in many instances, the true infection status of individual animals is unknown, particularly in the absence of a gold standard or when the gold standard cannot be applied because of financial, practical or ethical constraints.

Keywords: Screening test, Bayesian approach, Gold standard, Sensitivity, Specificity.



* Texte de la communication orale présentée au cours des Journées scientifiques AEEMA, 31 mai 2013

¹ Université Paris-Est, Ecole nationale vétérinaire d'Alfort, Unité épidémiologie des maladies animales Infectieuses, ENVA USC ANSES, Maisons-Alfort, F-94704, France

² INSERM, Centre de recherche en épidémiologie et santé des populations (CESP), U1018, Faculté de médecine Paris-Sud, Le Kremlin-Bicêtre ; AP-HP, Hôpital Bicêtre, Service d'épidémiologie et de santé publique, Université Paris-Sud, Le Kremlin-Bicêtre, 94276, France

³ Anses, Laboratoire de santé animale, Unité zoonoses bactériennes, Maisons-Alfort, F-94706, France

I - INTRODUCTION

Les plans de lutte collective contre les maladies infectieuses animales reposent en partie sur l'utilisation de tests fournissant des informations sur le statut des animaux vis-à-vis de la maladie. Les tests de dépistage et/ou de diagnostic peuvent reposer sur la détection de marqueurs de

l'infection (présence d'immunoglobulines sériques, réaction allergique) ou sur la recherche de l'agent pathogène incriminé. La connaissance de leurs limites est indispensable à la rationalisation des décisions sanitaires prises sur la base des résultats qu'ils fournissent.

II - CARACTÉRISTIQUES INTRINSÈQUES D'UN TEST

L'interprétation des résultats fournis par un test de dépistage nécessite de connaître à la fois les caractéristiques du test utilisé, mais également le contexte épidémiologique dans lequel le dépistage est réalisé. La qualité d'un test de dépistage dépend de deux caractéristiques :

- la sensibilité (Se) est l'aptitude du test à donner un résultat positif chez un individu malade ;

- la spécificité (Sp) est l'aptitude du test à donner un résultat négatif chez un individu sain.

Le tableau 1 indique les différents cas de figure possibles, en fonction du résultat du test (positif : T+ ou négatif : T-) et du statut de l'animal vis-à-vis de la maladie (animal malade : M+ ou indemne : M-).

Tableau 1

Combinaisons possibles à l'issue d'un test de dépistage,
selon le résultat du test et le statut du sujet vis-à-vis de la maladie

	Malade (M+)	Sain (M-)	Total
Résultat positif (T+)	Vrai positif (VP)	Faux positif (FP)	VP + FP
Résultat négatif (T-)	Faux négatif (FN)	Vrai négatif (VN)	VN + FN
Total	VP + FN	VN + FP	VP + FP + VN + FN

La sensibilité et la spécificité sont donc définies comme suit :

$$Se = P(T+ | M+) = VP / (VP + FN)$$

$$Sp = P(T- | M-) = VN / (VN + FP)$$

La sensibilité et la spécificité d'un test sont des valeurs dites « intrinsèques » : elles sont censées demeurer constantes quelle que soit la prévalence de la maladie dans la population d'individus soumis au test [Toma *et al.*, 2010].

III - MÉTHODES D'ESTIMATION DES CARACTÉRISTIQUES D'UN TEST

Lorsqu'un test de référence (ou « gold standard ») est disponible pour dépister une maladie, la sensibilité et la spécificité du test en cours de validation peuvent être déterminées de manière directe. Dans les conditions idéales, le protocole doit alors respecter les conditions suivantes : le

test de référence ne doit conduire à aucune erreur de classement (faux positifs ou faux négatifs), les résultats du test étudié doivent tous être comparés au même test de référence et les deux tests doivent être réalisés au même moment. Ces conditions sont toutefois rarement réunies : en

l'absence d'information complète et fiable sur le statut réel des individus vis-à-vis de l'infection, il est nécessaire de recourir à d'autres méthodes d'estimation des caractéristiques intrinsèques du test étudié [Rutjes *et al.*, 2007].

1. ESTIMATION DIRECTE DES CARACTÉRISTIQUES DES TESTS : CHOISIR UNE RÉFÉRENCE

L'estimation directe nécessite de choisir une référence positive pour le calcul de la sensibilité et une référence négative pour le calcul de la spécificité.

Ce choix peut être fait dans un contexte expérimental : certaines études de validation des tests font appel à des individus élevés dans des conditions garantissant l'absence d'agents pathogènes dans leur environnement (animaux axéniques ou « germ free »), ou au contraire infectés expérimentalement. Il est évident que ces animaux ne sont en aucun cas représentatifs de la population réelle dans laquelle le test a pour vocation d'être appliqué. En effet, sur le terrain la population d'animaux est soumise à des variations environnementales (agents pathogènes environnementaux, climat, autres animaux), individuelles (diversité génétique), immunitaires (maladies intercurrentes, stress, gestation, traitements) qui n'existent pas dans des conditions expérimentales. De plus, ces études sont lourdes en termes de coût, de risques pour les personnels exposés si la maladie étudiée est une zoonose et peuvent poser un problème éthique. Les résultats obtenus constituent les caractéristiques intrinsèques du test au sens propre du terme mais sont totalement déconnectés de la réalité du terrain.

Il est également possible de choisir une référence dans des conditions de terrain. La population de référence positive est souvent constituée par des animaux sur lesquels on a isolé l'agent pathogène. Ces individus ne sont pas toujours représentatifs de la population d'animaux infectés (notamment si la recherche de l'agent présente elle-même un défaut de sensibilité) et en raison de l'existence de différents stades d'infection, de différentes localisations possibles (plus ou moins accessibles) de l'agent pathogène recherché et d'un biais de sélection des animaux sur lesquels la recherche est effectuée (dans le cas d'une découverte fortuite d'abattoir : animaux abattus plus ou moins tardivement selon le type de production ; dans le cas d'un protocole d'assainissement par abattage partiel : les premiers animaux abattus sont les

meilleurs répondants aux tests). L'échantillon constituant la référence négative provient en règle générale d'une population très particulière d'élevages officiellement indemnes de la maladie étudiée, ou situés dans une zone indemne, n'ayant pas été liés épidémiologiquement à un foyer. La validité de cette référence est bien sûr conditionnée par une bonne connaissance de la situation épidémiologique permettant de garantir que la population est réellement indemne.

L'étude directe des caractéristiques d'un test, sur le terrain comme au laboratoire, est donc parfois réalisée à partir de populations particulières, dans un contexte qui peut être différent de celui dans lequel le test sera utilisé après sa validation. Pour les services vétérinaires départementaux, ou pour les cliniciens, se pose alors le problème d'extrapoler les résultats obtenus dans de tels échantillons à une population réelle. Afin de faciliter ce type de démarche, l'utilisation de méthodes de construction d'une référence permet de conduire des études d'estimation des caractéristiques de tests dans un échantillon représentatif de la population dans lequel le test sera utilisé.

2. ESTIMATION DES CARACTÉRISTIQUES D'UN TEST EN L'ABSENCE DE RÉFÉRENCE

Plusieurs groupes de méthodes permettent l'estimation des caractéristiques d'un test en l'absence d'information sur le statut infectieux individuel des animaux [Rutjes *et al.*, 2007] : l'ajout (ou l'ajustement) de données manquantes, la correction d'une référence imparfaite et la construction d'une référence (*cf. infra*).

2.1. AJOUT OU AJUSTEMENT DE DONNÉES MANQUANTES

Ce groupe de méthodes est applicable si l'on dispose d'un test de référence acceptable, mais dont le résultat est indisponible pour plusieurs individus. L'information manquante peut être ajoutée à la base de données, à condition de prendre en compte la proportion et le profil de ces individus.

2.2. CORRECTION D'UNE RÉFÉRENCE IMPARFAITE

Ces méthodes sont utilisables s'il existe un test de référence dont on sait qu'il est imparfait et si l'on dispose de connaissances fiables sur son degré d'imperfection et sur le degré de corrélation entre

les « erreurs » commises par chacun des deux tests étudiés.

2.3. CONSTRUCTION D'UNE RÉFÉRENCE STANDARD

Cette approche combine les résultats d'individus soumis parallèlement à plusieurs tests. Les individus peuvent alors être classés comme « infectés » ou « indemnes » de trois manières différentes :

- en utilisant une règle prédéfinie selon les profils de résultats aux tests : référence dite « composite ». La conception d'une échelle composite nécessite de corrélérer les résultats du test étudié à d'autres indicateurs, notamment des critères cliniques ou la survenue d'événements cliniques ultérieurs. Cette démarche conduit à « abandonner » l'évaluation des caractéristiques intrinsèques du test seul au profit d'un ensemble de critères ;
- à l'aide de l'avis d'experts ;
- à l'aide de modèles à classe latente.

Les deux premières approches présentent l'inconvénient de comporter un biais dû aux erreurs de classification. La troisième pose le problème de l'absence de définition du « cas », *i.e.* de l'individu infecté.

3. ÉTUDE D'UNE MÉTHODE D'ESTIMATION DES CARACTÉRISTIQUES D'UN TEST EN L'ABSENCE DE GOLD STANDARD : LES MODÈLES À CLASSE LATENTE

Parmi les trois groupes de méthodes présentées précédemment, les approches par construction d'une référence sont particulièrement intéressantes car elles sont le seul moyen d'estimer les caractéristiques des tests en l'absence de toute information sur le statut individuel des sujets vis-à-vis de la maladie étudiée. Comme évoqué précédemment, les modèles à classe latente présentent l'avantage de ne pas engendrer de biais dû aux erreurs de classification mais leur utilisation peut s'avérer complexe.

3.1. PRINCIPE DES MODÈLES À CLASSE LATENTE

Les modèles à classe latente permettent de combiner différentes sources d'informations intégrant les résultats aux tests d'un grand nombre

d'individus testés afin de « construire » une référence et d'estimer la sensibilité et la spécificité des tests évalués, ainsi que le taux de prévalence de la maladie dans l'échantillon étudié [Rutjes *et al.*, 2007]. Ils peuvent être utilisés pour évaluer les caractéristiques d'un ou de plusieurs nouveaux tests, comparées à celles d'un test déjà connu mais imparfait, ou pour estimer les caractéristiques de plusieurs tests mal connus. Pour ce faire, il est nécessaire de disposer des résultats croisés de plusieurs individus à au moins deux tests.

Prenons l'exemple de deux tests dont les résultats sont binaires (négatif ou positif) appliqués à un groupe d'individus dans lequel la prévalence de la maladie est inconnue. Aucun de ces deux tests n'est parfait.

Pour le premier test (T_1), de sensibilité Se_1 et de spécificité Sp_1 , la probabilité d'obtenir un résultat positif chez un individu est égale à la somme de la probabilité d'obtenir un résultat positif chez un individu atteint ou « vrai positif » ($T_{1+} | M+$) et de la probabilité d'obtenir un résultat positif chez un individu indemne ou « faux positif » $P(T_{1+} | M-)$.

À l'aide de la formule de Bayes, les probabilités de survenue d'un résultat positif ou d'un résultat négatif peuvent être écrites comme des fonctions des caractéristiques intrinsèques du test (Se_1 et Sp_1) et de la prévalence de la maladie au sein du groupe étudié ($prev$) :

$$P(T_{1+}) = P(M+) * P(T_{1+} | M+) + P(M-) * P(T_{1+} | M-)$$

$$P(T_{1+}) = prev * Se_1 + (1 - prev) * (1 - Sp_1) \quad [1]$$

De la même manière, la probabilité d'obtenir un résultat négatif chez un individu est égale à la somme de la probabilité d'obtenir un résultat négatif chez un individu indemne ou « vrai négatif » ($T_{1-} | M-$) et de la probabilité d'obtenir un résultat négatif chez un individu infecté ou « faux négatif » $P(T_{1-} | M+)$:

$$P(T_{1-}) = P(M+) * P(T_{1-} | M+) + P(M-) * P(T_{1-} | M-)$$

$$P(T_{1-}) = prev * (1 - Se_1) + (1 - prev) * Sp_1 \quad [2]$$

Les probabilités d'obtenir un résultat positif ou négatif au second test (T_2) sont définies de même :

$$P(T_{2+}) = prev * Se_2 + (1 - prev) * (1 - Sp_2) \quad [3]$$

$$P(T_{2-}) = prev * (1 - Se_2) + (1 - prev) * Sp_2 \quad [4]$$

Dans le système d'équations [1, 2, 3, 4], cinq paramètres sont donc inconnus ($prev$, Se_1 , Sp_1 , Se_2 et Sp_2). Pour un individu, il existe 4 possibilités de combinaison des résultats aux tests : [$T_1 +$; $T_2 +$], [$T_1 +$; $T_2 -$], [$T_1 -$; $T_2 +$] et [$T_1 -$; $T_2 -$]. On peut donc écrire la vraisemblance de la survenue de chacune

des combinaisons de résultats aux tests à l'aide des équations [1], [2], [3] et [4].

Si les deux tests sont considérés comme statistiquement indépendants (*i.e.* si le résultat obtenu au premier test conditionnellement au statut de l'animal vis-à-vis de la maladie ne fournit pas d'information sur la probabilité d'obtention d'un résultat donné au second test [Enøe *et al.*, 2000]), les probabilités d'observer chacune des combinaisons de résultats aux tests sont les suivantes [Vose, 2008]:

$$P(T_{1-}; T_{2-}) = \text{prev} (1-Se_1) (1-Se_2) + (1-\text{prev}) Sp_1 Sp_2$$

$$P(T_{1-}; T_{2+}) = \text{prev} (1-Se_1) Se_2 + (1-\text{prev}) Sp_1 (1-Sp_2)$$

$$P(T_{1+}; T_{2-}) = \text{prev} Se_1 (1-Se_2) + (1-\text{prev}) (1-Sp_1) Sp_2$$

$$P(T_{1+}; T_{2+}) = \text{prev} Se_1 Se_2 + (1-\text{prev}) (1-Sp_1) (1-Sp_2)$$

Des variations peuvent être apportées au modèle présenté ci-dessus, afin de tenir compte du contexte de l'étude.

Les modèles à classe latente peuvent être utilisés pour l'évaluation simultanée de deux tests ou plus. L'inclusion de tests supplémentaires est intéressante car elle augmente le nombre de degrés de liberté (d.d.l.): un nombre plus élevé de paramètres peuvent être estimés, et éventuellement utilisés pour de plus amples vérifications concernant l'adaptation du modèle [Rutjes *et al.*, 2007].

Le modèle de base présenté ci-dessus suppose que les résultats des tests évalués soient indépendants conditionnellement à l'état sanitaire de l'individu concerné (par exemple : si un individu infecté est considéré à tort comme indemne par le test 1, cela ne modifie pas la probabilité pour ce même individu d'être considéré à tort comme indemne par le test 2) [Enøe *et al.*, 2000]. Cette hypothèse est valable lorsque les tests évalués reposent sur des processus biologiques différents (tels que la mise en évidence d'une infection bactérienne par sérologie d'une part et par culture ou amplification génique d'autre part). Dans de multiples situations cependant, l'hypothèse d'indépendance des tests n'est pas vérifiée (par exemple : lors de l'étude de plusieurs tests sérologiques reposant sur la détection des anticorps dirigés contre le lipopolysaccharide (LPS) de surface des *Brucella*). Pour résoudre ce problème, il est possible d'estimer la corrélation entre les différents tests et d'en tenir compte pour le calcul des caractéristiques des deux tests. Le nombre de paramètres du modèle est alors augmenté et leur estimation nécessite l'adjonction d'informations supplémentaires pour augmenter le nombre de

d.d.l., au moyen de l'inclusion d'autres tests, d'une autre population de prévalence différente ou en incorporant au modèle des informations *a priori* à l'aide d'une approche bayésienne.

3.2. ESTIMATION DES PARAMÈTRES DU MODÈLE : EXEMPLE DE L'APPROCHE BAYÉSIENNE

Lorsque suffisamment d'informations sont disponibles pour que le nombre de degrés de liberté autorisé par les données soit supérieur au nombre de paramètres à estimer, le modèle est dit « identifiable » et ses paramètres peuvent être estimés par une approche fréquentiste. Dans le cas contraire, il est nécessaire de recourir à des méthodes d'estimation statistiques telles qu'une approche bayésienne, afin d'obtenir un système de solutions unique.

L'approche bayésienne est couramment utilisée pour la détermination des caractéristiques de tests diagnostiques en l'absence de gold standard en médecine humaine, et de plus en plus fréquemment en médecine vétérinaire [Spangler *et al.*, 1992 ; Joseph *et al.*, 1995 ; Grenier, 1999 ; Enøe *et al.*, 2000 ; Branscum *et al.*, 2005 ; Rutjes *et al.*, 2007 ; Meyer *et al.*, 2009]. Elle peut également être utilisée pour la modélisation de processus biologiques et environnementaux [Bernier *et al.*, 2000 ; Clark, 2007]. Son intérêt principal réside dans la possibilité de synthétiser les informations déjà disponibles *a priori* sur les paramètres à estimer (ou « priors ») et les données collectées sur le terrain, pour aboutir à l'estimation *a posteriori* des paramètres d'intérêt (« posteriors ») : Posterior = Prior * Vraisemblance.

L'estimation des posteriors est obtenue mathématiquement à l'aide du théorème de Bayes : étant donnés deux événements A et B, dont les probabilités de survenue respectives sont P(A) et P(B), la probabilité de survenue de l'évènement B sachant que A est survenu est définie comme suit :

$$P(B|A) = P(A|B) * P(B) / P(A)$$

La démarche bayésienne consiste à probabiliser le paramètre inconnu en lui associant une loi de probabilité dite loi *a priori*. Cette loi représente l'ensemble des informations *a priori* disponibles sur le paramètre θ ainsi que les imprécisions s'y rattachant.

Soit θ le paramètre dont la loi *a priori* est notée π :

$$[\theta] \sim \pi(\theta)$$

Soient X les données dont la loi conditionnelle est notée : $[X|\theta] = V$ (vraisemblance)

La loi jointe de (θ, X) est : $[\theta, X] = [\theta] [X|\theta] = \pi(\theta) V$.

La loi marginale de X est :

$$m(x) = \int [\theta, X] d\theta = \int [\theta] [X|\theta] d\theta$$

La loi *a posteriori* de θ est donc : $[\theta|X] =$

$$[\theta] [X|\theta] / m(x)$$

$$\propto [\theta] [X|\theta]$$

Les connaissances sur le paramètre θ sont actualisées au travers du théorème de Bayes, c'est-à-dire via la loi *a posteriori* $[\theta|X]$. À l'accumulation d'observations X_1, \dots, X_n correspond donc une suite d'actualisations $[\theta|X_1], \dots, [\theta|X_1, \dots, X_n]$, qui sont mutuellement compatibles et dont l'ordre d'apparition n'a pas d'importance.

Lorsque cette approche est appliquée à l'étude des caractéristiques de tests, les paramètres « d'intérêt » (*i.e.* sensibilités et spécificités des tests et prévalences dans les populations étudiées) sont donc considérés comme des variables qui suivent une distribution donnée. En règle générale, les sensibilités, spécificités et prévalences sont modélisées sous la forme de lois beta de paramètres a et b , dans l'intervalle $[0,1]$. Le choix de la loi *a priori* est crucial pour la suite de l'analyse, même si son impact est moindre confronté à un très grand nombre de données collectées sur le terrain. Les informations *a priori* incluses dans le modèle peuvent être non informatives : dans ce cas les paramètres sont modélisés suivant des lois beta $(1,1)$, c'est-à-dire que toutes les valeurs prises par le paramètre sont équiprobables dans l'intervalle $[0,1]$. Il est également possible d'utiliser des priors dits « informatifs ». Ils sont modélisés à l'aide de distributions beta (a,b) et sont déterminés à l'aide d'informations bibliographiques, d'avis d'experts ou de données provenant d'une étude antérieure effectuée dans des conditions similaires [Enøe *et al.*, 2000]. Pour un paramètre d'intérêt donné, les paramètres a et b de la loi beta (a,b) sont alors déterminés à l'aide de la valeur la plus probable du paramètre d'intérêt (θ_0) et de sa limite inférieure à

95 % (θ_L : la valeur du paramètre d'intérêt est supérieure à ce seuil avec un degré de certitude de 95 %).

Les données sont ensuite simulées par un processus itératif (algorithme Monte Carlo par chaîne de Markov ; MCMC), en utilisant l'échantillonneur de Gibbs [Gelman *et al.*, 1995 ; Andersen, 1997]. L'échantillon simulé est utilisé pour estimer la distribution des paramètres d'intérêt *a posteriori*, leurs fonctions de densité de probabilité, leurs valeurs moyennes et un intervalle de probabilité à 95 % (ou intervalle de crédibilité ; ICr 95 %) pour chacune d'entre elles [Enøe *et al.*, 2000]. Les valeurs d'initiation des chaînes peuvent être déterminées au départ (en prenant, par exemple, la valeur la plus probable pour chaque paramètre) ou choisies au hasard dans les distributions *a priori*.

Le compromis entre adaptation du modèle aux données et complexité du modèle peut être estimé à l'aide du Deviance Information Criterion [DIC ; Spiegelhalter, 2002] :

$$DIC(M) = D(\bar{\theta}) + 2c$$

Où c mesure la complexité du modèle M et $D(\bar{\theta})$ est l'éloignement du modèle aux données ou déviance en la moyenne *a posteriori* du paramètre. En règle générale, plus le modèle est rendu complexe par l'addition de paramètres et meilleure est son adaptation aux données. Plus la valeur du DIC est basse, meilleur est le compromis adaptation/complexité. L'un des avantages du DIC est qu'il est calculé directement par le logiciel WinBUGS [Lunn *et al.*, 2000].

Les estimations *a posteriori* peuvent varier selon le type d'informations *a priori* introduites : on parle de sensibilité du modèle, qui doit être évaluée de manière systématique. D'autres critères peuvent être utilisés pour vérifier l'adaptation du modèle aux données et la fiabilité du résultat : stabilité des indicateurs (moyenne, variance, quantiles) par rapport aux itérations, MC error suffisamment basse, diagnostic de convergence de Brooks ou Gelman et Rubin...

III - QUAND LA RÉFÉRENCE MANQUE :

AVANTAGES ET LIMITES DE L'ESTIMATION DES CARACTÉRISTIQUES DES TESTS PAR UN MODÈLE À CLASSE LATENTE IMPLÉMENTÉ PAR UNE APPROCHE BAYÉSIENNE

Comme nous l'avons vu précédemment, la particularité des modèles à classe latente est de permettre la construction d'une référence lorsqu'aucune information n'est disponible concernant le statut infectieux individuel des animaux de l'échantillon. Ce type de situation est fréquemment rencontré en médecine vétérinaire, en raison de l'approche collective de la lutte contre les maladies du bétail. Dans la majorité des cas, les décisions sanitaires concernant les animaux de rente sont prises à l'échelle du troupeau et le statut sanitaire individuel des animaux constituant le troupeau est inconnu. Par exemple, lorsque la présence d'une maladie réglementée est confirmée dans un cheptel, il est fréquent d'abattre tous les animaux de ce troupeau mais la mise en évidence de l'agent pathogène sur chaque individu n'est pas toujours possible. Dans ce cas, aucune référence positive n'est disponible. Pour d'autres maladies, le contexte épidémiologique général est connu mais le statut individuel des animaux ne fait pas l'objet d'une recherche systématique : aucun statut « officiellement indemne » n'existe et aucune référence négative n'est disponible.

L'estimation des paramètres d'un modèle à classe latente peut être effectuée de manière fréquentiste, par la méthode du maximum de vraisemblance. Contrairement à l'approche bayésienne, elle fournit une estimation ponctuelle des paramètres d'intérêt [Hui et Walter, 1980] et nécessite de disposer d'un grand nombre de données afin que le degré de liberté permis par les données soit supérieur au nombre de paramètres à estimer. Ce point est souvent problématique. Il peut notamment être résolu en utilisant plusieurs populations dans lesquelles la prévalence de la maladie est différente, sans toutefois tendre vers 0 ou 1 [Nielsen et Toft, 2002]. Lorsqu'on ne dispose pas de telles populations, ou lorsque les tests sont liés par une relation de dépendance conditionnelle, les modèles sont souvent non identifiables et leur résolution nécessite d'utiliser une approche différente.

L'approche bayésienne pour l'estimation des caractéristiques de tests de dépistage en l'absence de test de référence est de plus en plus fréquemment utilisée en médecine humaine [Meyer *et al.*, 2009] comme en médecine vétérinaire. Son principal intérêt réside dans la

possibilité d'intégrer dans l'estimation les résultats d'enquêtes préliminaires, les données bibliographiques disponibles sur le sujet et les avis d'experts afin de rendre identifiable un modèle qui ne l'était pas. Ces informations *a priori* doivent être générées indépendamment de l'étude en cours [Enøe *et al.*, 2000]. Lorsque le nombre d'unités de l'échantillon étudié est faible, le choix de la loi *a priori* est crucial pour la suite de l'analyse : deux lois *a priori* différentes conduisent logiquement à deux systèmes d'inférence différents, mais l'accumulation d'un grand nombre de données finit par gommer cette différence. Le paradoxe de l'approche bayésienne est qu'il n'existe pratiquement jamais une unique loi *a priori*, mais plutôt un ensemble de lois compatibles avec les informations disponibles et les opinions des experts. Cette imprécision des informations cliniques et épidémiologiques disponibles est souvent assimilée à la subjectivité de l'analyse bayésienne et décrite comme une limite par certains auteurs [Berwick *et al.*, 1981]. On peut au contraire considérer que cette imprécision reflète la réalité des choses : lorsque les informations disponibles sur un test sont imparfaites et incomplètes, cette « imperfection » est retrouvée dans l'inférence résultante après intégration des résultats observés.

Par ailleurs, les études préliminaires nécessaires au choix des hypothèses *a priori* présentent l'avantage de rendre explicites les informations manquantes, incertaines ou approximatives et sont autant d'incitations à la mise en œuvre de nouvelles recherches cliniques et/ou épidémiologiques. Il conviendra dans tous les cas d'estimer l'impact de ces « priors » dans l'estimation des paramètres *a posteriori* par une étude de sensibilité.

Cette approche présente toutefois l'inconvénient inhérent à toute modélisation : le compromis entre la simplicité du modèle et son aptitude à refléter des situations parfois complexes est difficile à trouver. Un réalisme poussé implique d'utiliser des modèles compliqués : le nombre de paramètres à renseigner devient vite important et accroît les risques d'erreurs dans l'écriture du modèle. Au contraire, un modèle simple ne doit pas être simpliste. Le risque majeur serait alors le manque de finesse pour l'évaluation des scénarios [Seegers *et al.*, 2011].

IV - CONCLUSION

Les tests de dépistage (ou de diagnostic) constituent des outils indispensables dans la lutte contre les maladies animales réglementées. L'utilisation des tests comme outils d'aide à la prise de décision sanitaire nécessite donc l'étude de leurs caractéristiques intrinsèques et la connaissance du contexte épidémiologique dans lequel ils seront utilisés. Lorsqu'aucun test de référence n'est disponible, des méthodes statistiques adaptées doivent être employées pour évaluer la qualité des tests.

L'estimation directe présente l'intérêt d'être relativement simple à appliquer, à condition de disposer d'une référence valable concernant le statut des individus vis-à-vis de la maladie. Son inconvénient majeur réside dans le fait d'être fondée sur des populations très particulières, souvent peu représentatives de la population générale. Ce type d'approche convient bien à la réalisation d'études préliminaires en laboratoire ou sur le terrain.

L'estimation des caractéristiques des tests par une modélisation avec construction d'une référence permet de s'affranchir de la connaissance du statut individuel des animaux et de conduire l'étude dans une population d'animaux tout venant, tirés au

sort et donc représentatifs de la population d'origine mais est plus complexe à mettre en œuvre.

En dehors de ces points techniques, l'évolution de la maladie dans la population étudiée peut influencer sur l'approche choisie. Dans le cas d'une maladie d'évolution épizootique dans une population jusque-là indemne, que les pouvoirs publics souhaitent maîtriser le plus rapidement possible, l'estimation directe des caractéristiques présente l'avantage d'être rapide et de ne pas nécessiter d'avoir de recul par rapport à l'utilisation de ce test. Au contraire, lorsqu'une maladie évolue depuis plusieurs années avant le début de la lutte, il semble acceptable de prendre le temps d'utiliser des méthodes plus élaborées reposant sur les données de terrain et sur les informations *a priori* disponibles. Dans ce second cas, il est également fréquent que la circulation de la maladie à bas bruit complique sa détection (contrairement à la survenue de foyers dans une population jusque-là indemne, souvent accompagnée d'une expression clinique facilement repérable), ce qui plaide en faveur de l'utilisation de méthodes indirectes en raison de l'absence de référence facilement utilisable.

BIBLIOGRAPHIE

Andersen S. - Re: Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of gold standard. *American Journal of Epidemiology*, 1997, **145**, 290-291.

Bernier J., Parent E., Boreux J.J. - *Statistique pour l'environnement. Traitement bayésien de l'incertitude*. Paris : Tech & Doc, 2000.

Berwick D.M., Fineberg H.V., Weinstein M.C. - When doctors meet numbers. *American Journal of Medicine*, 1981, **71**, 991-998.

Branscum A.J., Gardner I.A., Johnson W.O. - Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. *Preventive Veterinary Medicine*, 2005, **68**, 145-163.

Clark J.S. - *Statistical computation for environmental sciences in R, lab manual for ecological data*. Princeton: Princeton University Press, 2007.

Enøe C., Georgiadis M.P., Johnson W.O. - Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease is unknown. *Preventive Veterinary Medicine*, 2000, **45**, 61-81.

Gelman A., Carlin J.B., Sterns H.S., Rubin D.B. - *Bayesian data analysis*. Londres: Chapman & Hall, 1995.

Grenier B. - *Evaluation de la décision médicale, introduction à l'analyse médico-économique*, 3ème éd., Paris : Masson, 1999.

- Hui S.L., Walter S.D. - Estimating the error rates of diagnostic tests. *Biometrics*, 1980, **36**, 167-171.
- Johnson W.O., Gastwirth J.L. - Bayesian inference for medical screening tests: approximations useful for the analysis of acquired immune deficiency syndrome. *Journal of Royal Statistic Society B.*, 1991, **53**, 427-439.
- Joseph L., Gyorkos T.W., Coupal L. - Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, 1995, **141**, 263-272.
- Lunn D.J., Thomas A., Best N., Spiegelhalter D. WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 2000, **10**(4), 325-337.
- Meyer N., Vinzio S., Goichot B. - La statistique bayésienne: une approche des statistiques adaptée la clinique. *La Revue de médecine interne*, 2009, **30**(3), 242-249.
- Nielsen S.S. and Toft N. - Optimisation of the validity of ELISA and faecal culture tests for paratuberculosis: Selection of population or correction by population characteristics? Proc. 7th International Colloquium on Paratuberculosis, Bilbao, Spain, June 11-14, 2002, 400-405.
- Rutjes A.W.S., Reitsma J.B., Coomarasamy A., Khan K.S., Bossuyt P.M.M. - Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment*, 2007, **11** (50), 72 p.
- Seegers H., Ezanno P., Krebs S., Rat-Aspert O., Viet A.F., Belloc C., Charron M., Malher X., Fourichon C. - Modélisation et aide aux décisions en gestion de la santé animale. *Rencontres Recherche Ruminants*, 2011, **18**, 4-10.
- Spangler E., Bech-Nielsen S., Heider L.E. - Diagnostic performance of two tests and faecal culture for subclinical paratuberculosis and associations with production. *Preventive Veterinary Medicine*, 1992, **13**, 185-195.
- Spiegelhalter D.J., Best N.G., Carlin B.P., Van der Linde A. - Bayesian measures of model complexity and fit. *Journal of Royal Statistic Society B.*, 2002, **64**, 583-639.
- Toma B., Dufour B., Bénet J.J., Sanaa M., Shaw A., Moutou F. - *Épidémiologie appliquée à la lutte contre les maladies animales transmissibles majeures* 3^{ème} éd., Maisons-Alfort : AEEMA, 2010.
- Vose D. - *Risk analysis : a quantitative guide*, 3rd ed. Chichester : John Wiley & sons, 2008.

