

UTILISATION DES TECHNIQUES BAYESIENNES POUR ESTIMER LES CARACTERISTIQUES DE DEUX TESTS DE DIAGNOSTIC DE LA PARATUBERCULOSE CAPRINE *

**Pascale Mercier¹, Christian Baudry¹, Jennifer Martin¹,
Claire Bertin², Karine Laroucau², François Beaudeau³,
Henri Seegers³ et Xavier Malher³**

RESUME : En l'absence de test parfait, la sensibilité (Se) et la spécificité (Sp) d'un test de diagnostic peuvent être estimées par des modèles probabilistes. Afin d'estimer les caractéristiques (Se et Sp) d'un test sérologique (ELISA) et de la culture fécale (FC) pour détecter l'infection par *Mycobacterium avium paratuberculosis* (Map), deux méthodes ont été mises en œuvre : la méthode du maximum de vraisemblance et l'inférence bayésienne.

Des échantillons de sang et de fèces ont été prélevés sur 532 chèvres appartenant à 15 troupeaux. Les résultats des deux modèles sont comparés et discutés. Le modèle bayésien avec « priors » informatifs permet d'obtenir les meilleures estimations. La Se et la Sp de l'ELISA ont été estimées respectivement à 34% et 99%, tandis que les valeurs pour la culture fécale ont été estimées à 53% et 100%.

Mots-clés : Paratuberculose, caprin, sensibilité, spécificité, méthodes bayésiennes.

SUMMARY: Two latent class models [maximum likelihood (ML) methods and Bayesian inference (BI)] were used to estimate the sensitivity (Se) and the specificity (Sp) of a serum ELISA test and of faecal culture in the detection of infection with Map in French dairy goats. Samples of blood and faeces were collected from 532 goats in 15 herds.

Estimates according to ML and BI methods were compared. The more accurate of the 2 tests was found to be the BI (WinBUGS) model with informative priors. The Se and Sp of the ELISA were estimated to be 34% and 99% respectively. For faecal culture, the Se was 53% and the Sp was 100%.

Keywords : Paratuberculosis, goat, sensitivity, specificity, bayesian estimations.



* Texte de la communication orale présentée lors de la Journée AEEMA, 1er juin 2007

¹ Afssa Niort – Laboratoire d'études et de recherches caprines, France

² Afssa Laboratoire d'études et de recherches en pathologie animale et zoonoses, Maisons-Alfort, France

³ UMR708 Gestion de la santé animale INRA ENVN, Ecole nationale vétérinaire, Nantes, France

I - INTRODUCTION

Pour conduire des études épidémiologiques ou évaluer des plans de contrôle, il est important de disposer de tests de diagnostic de caractéristiques (sensibilité et spécificité) connues. La sensibilité et la spécificité d'un test sont déterminées par comparaison avec un test de référence (ou « gold standard ») qui détermine sans ambiguïté le statut sanitaire des animaux. Mais, pour de nombreuses maladies, et en particulier la paratuberculose, il n'existe pas de test de référence parfait. Dès lors, des approches alternatives doivent être utilisées pour estimer les performances du test en l'absence de gold standard. L'objectif de ce travail est de décrire deux de ces approches pour définir les caractéristiques (sensibilité et spécificité) de deux tests de diagnostic de

l'infection à *Mycobacterium avium paratuberculosis* chez la chèvre.

Mycobacterium avium subsp *paratuberculosis* (Map) est l'agent de la paratuberculose qui se traduit cliniquement chez les caprins par un amaigrissement progressif des animaux de plus de 2 ans. On considère que la principale modalité de contamination est l'ingestion de bactéries pendant les premières semaines de vie. Il existe donc une période assez longue pendant laquelle l'infection est asymptomatique. Cependant, il est important d'un point de vue épidémiologique, de pouvoir diagnostiquer précocement les animaux infectés par Map. Il s'agit donc d'évaluer des tests permettant de réaliser le diagnostic de l'infection par Map.

II - MATERIEL ET METHODES

1. ANIMAUX ET PRELEVEMENTS

Des prélèvements de sang et de fèces ont été effectués sur 600 chèvres issues de 15 troupeaux de la Région Poitou-Charentes. Deux populations ont été identifiées : une population de troupeaux caprins avec des signes cliniques évocateurs de la paratuberculose et une population de troupeaux caprins sans signes cliniques. Dans chaque population, un échantillonnage a été réalisé dans respectivement neuf et six troupeaux. Dans chaque troupeau, des prélèvements ont été effectués sur 40 animaux sans signes cliniques et appartenant à différentes classes d'âge.

2. TESTS DE DIAGNOSTIC

Deux tests de diagnostic ont été mis en oeuvre : un test direct, reposant sur la mise en évidence de la bactérie (culture fécale) et un test indirect, reposant sur la mise en évidence des anticorps avec un test ELISA commercial (Institut Pourquier, Montpellier).

La culture fécale a été réalisée selon la norme en vigueur (NF U47-103). Les analyses sérologiques ont été effectuées conformément aux indications du producteur (test bicupule) : le seuil de positivité a été fixé à 60%.

3. L'APPROCHE BAYESIENNE

Il s'agit d'une approche probabiliste. Le statut des animaux vis-à-vis de l'infection paratuberculeuse est considéré comme « latent » c'est-à-dire qu'il existe (infecté ou non-infecté) mais qu'il ne peut pas être déterminé [Enøe *et al.*, 2000].

L'utilisation de cette approche repose sur le paradigme de Hui et Walter. Trois conditions doivent être vérifiées :

- les deux populations ont des prévalences différentes,
- les deux tests sont indépendants,
- les caractéristiques des tests (sensibilité et spécificité) sont identiques dans les deux populations.

Le calcul des différentes probabilités est lourd et le développement des méthodes bayésiennes a été rendu possible par la mise au point de logiciels de calcul en libre accès.

4. LOGICIELS

Les modèles « latent class » peuvent être séparés en deux approches :

- une approche fréquentielle : estimation du maximum de vraisemblance (Maximum Likelihood ou ML),
- une approche bayésienne qui permet d'estimer le paramètre recherché à partir des données récoltées sur le terrain et des informations d'experts [Enøe *et al.*, 2000].

Deux logiciels correspondant à chacune des deux approches ont été utilisés.

4.1. LE LOGICIEL TAGS (TESTS IN ABSENCE OF A GOLD STANDARD)

Il s'agit d'un logiciel développé par Pouillot *et al.* sous R (www.afssa.fr/interne/tags.htm) et qui permet d'estimer le maximum de vraisemblance. Ce modèle peut produire les valeurs les plus probables des paramètres qui auraient mené aux données observées. Le paramètre estimé est obtenu en maximisant la vraisemblance par des algorithmes. Deux algorithmes sont utilisés successivement : l'algorithme de Newton-Raphson (NR) et l'algorithme EM (Expectation-Maximisation).

4.2. LE LOGICIEL WINBUGS

Il s'agit d'un logiciel développé par Spiegelhalter *et al.*, pour faciliter l'utilisation de l'approche bayésienne (<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>). BUGS correspond à « Bayesian inference Using Gibbs Sampling ». Ce logiciel utilise la méthode de l'échantillonneur de Gibbs (Gibbs sampling) avec des algorithmes Monte Carlo par Chaîne de Markov (MCMC). Cet algorithme (MCMC) est un algorithme d'échantillonnage à partir d'une distribution de probabilités afin de construire une ou plusieurs chaînes d'itérations, celles-ci étant le résultat de chaque étape de la simulation. WinBUGS amorce une valeur arbitraire pour chaque chaîne et simule les différents paramètres à partir de cette valeur. Ceux-ci sont alors utilisés dans l'itération suivante comme valeur de départ [Praet *et al.*, 2006]. Si on réalise un grand nombre d'itérations, alors on pourra avoir une bonne estimation de la loi *a posteriori* [Enøe *et al.*, 2000].

Dans cette approche, les paramètres à évaluer sont des variables aléatoires dont il faut préciser la distribution. L'estimation de la valeur est la probabilité *a posteriori* (ou « Posterior »), qui est déterminée à partir d'informations fournies par des experts (ou « Priors ») et les données récoltées sur le terrain (vraisemblance) [Branscum *et al.*, 2005].

Les informations *a priori* sont considérées comme des variables binomiales dont la distribution suit une loi béta de paramètres a et b , dans l'intervalle $[0,1]$ et notée $Be(a,b)$ [Johnson *et al.*, 2001]. Pour construire les lois béta (logiciel *betabuster* : <http://www.epi.ucdavis.edu/diagnostictests/betabuster.html>), les experts doivent fournir pour chaque paramètre la valeur la plus probable (ou la meilleure estimation) et une valeur limite [Enøe *et al.*, 2000].

4.3. LES DIFFERENTS « PRIORS »

Les « priors » ont été définis à partir des données bibliographiques pour la sensibilité et la spécificité des deux tests et à partir de l'expérience des auteurs pour les prévalences.

Quatre modèles correspondant à différents *a priori* ont été testés (tableau 1). Le modèle II est une adaptation du modèle I dans laquelle la sensibilité du test ELISA a été modifiée pour tenir compte d'une étude réalisée sur des caprins et utilisant le même test ELISA [Gumber *et al.*, 2006]. Le modèle III reprend les « priors » utilisés par Kostoulas *et al.*. Ils ont considéré que les spécificités des deux tests étaient proches de 100% mais qu'ils ne disposaient pas d'informations concernant les sensibilités. Les lois $Be(1,1)$ correspondant à une distribution uniforme du « prior » ont donc été utilisées. Pour les prévalences, les lois $Be(0,1/0,9)$ et $(0,9/0,1)$ sont utilisées [Enøe *et al.*, 2000]. On parle alors de modèle non informatif (NI).

Le logiciel WinBUGS permet de calculer, pour chaque modèle, la valeur médiane de la distribution et de déterminer un intervalle de probabilité à 95% (2,5 et 97,5 percentiles) [Branscum *et al.*, 2005].

Tableau 1
Paramètres a et b des différentes lois bêta (Be (a/b)) utilisées comme valeurs a priori dans les quatre modèles

Paramètres	I	II	III	IV (NI)
Se _{FC}	Be (207,6/420,5) a	Be (207,6/420,5)	Be (1/1)	Be (1/1)
Sp _{FC}	Be (298,1/1) a	Be (298,1/1)	Be (100/2) e	Be (1/1)
Se _{ELISA}	Be (7,4/7,9) a,b	Be (12,44/4,2) d	Be (1/1)	Be (1/1)
Sp _{ELISA}	Be (58,7/1,4) b,c	Be (58,7/1,4)	Be (100/5,5) e	Be (1/1)
p ₁	Be (6,2/99,7)	Be (6,2/99,7)	Be (6,2/99,7)	Be (0,1/0,9)
p ₂	Be (14,5/77,7)	Be (14,5/77,7)	Be (14,5/77,7)	Be (0,9/0,1)

a : Whitlock *et al.*, 2000 ; b : Robbe-Austerman *et al.*, 2006 ; c : Wintz, 2006 ;
d : Gumber *et al.*, 2006 ; e : Kostoulas *et al.*, 2006

III – RESULTATS

1. RESULTATS DES TESTS

Sur les prélèvements des 600 chèvres, les résultats de l'ELISA et de la culture fécale ont été obtenus pour 532 animaux (tableau 2).

2. METHODE FREQUENTIELLE

L'application du logiciel TAGS aux données récoltées permet d'estimer la sensibilité et la spécificité des deux tests, ainsi que la prévalence dans les deux populations (tableau 3).

Pour la culture fécale, la sensibilité et la spécificité ont été estimées à respectivement 52 et 99%, alors que pour le test ELISA, la sensibilité et la spécificité ont été estimées à 51 et 100%. Les prévalences dans les deux populations ont également été estimées à 2% pour la population 1 (faible prévalence) et 24% pour la population 2 (forte prévalence).

L'intervalle de confiance à 95% n'a pu être défini pour la spécificité de l'ELISA, en raison de la taille de l'échantillon.

Tableau 2
Résultats de la culture fécale et du test ELISA dans les deux sous-populations

Sous-population 1	Culture fécale (FC)			Sous-population 2	Culture fécale (FC)			
	+	-	Total		+	-	Total	
ELISA	+	2	1	3	+	13	13	26
	-	6	311	317	-	15	171	186
Total	8	312	320	Total	28	184	212	

Tableau 3
Estimations des différents paramètres à partir du logiciel TAGS (IC à 95%)
en pourcentages

Paramètres	ML* (IC à 95%)
Sensibilité de la culture fécale	52 (34-69)
Spécificité de la culture fécale	99 (95-100)
Sensibilité de l'ELISA	51 (21-72)
Spécificité de l'ELISA	100 (ND-100)
Prévalence dans la population 1	2 (0,5-6)
Prévalence dans la population 2	24 (16-34)

* ML : Maximum Likelihood : maximum de vraisemblance

3. METHODE BAYESIENNE

L'estimation du paramètre est obtenue à partir de la moyenne, de la médiane ou du mode de la distribution du « posterior » [Branscum *et al.*, 2005]. En fonction du choix des « priors » différentes estimations sont obtenues pour la sensibilité et la spécificité des deux tests ainsi que pour la prévalence dans les deux populations (tableau 4). Les différents modèles ont permis de mettre en évidence l'influence du choix des « priors » sur les estimations des différents paramètres (« posteriors »).

Les prévalences estimées sont de l'ordre de 3% pour la population 1 et de 20% pour la population 2. Les estimations de la spécificité

des deux tests sont comparables, quel que soit le modèle, et sont proches de 100%.

En revanche, les estimations de la sensibilité varient selon le modèle utilisé. Ainsi, l'estimation de la sensibilité de la culture fécale est comprise entre 42 et 53%. L'estimation de la sensibilité de l'ELISA varie encore plus, de 34 à 61%. La modification de la sensibilité de l'ELISA dans le modèle II a pour effet de modifier la sensibilité de la culture fécale. Dans le modèle III, la sensibilité de l'ELISA devient supérieure à la sensibilité de la culture fécale. En revanche, dans le modèle IV, les estimations de la sensibilité de la culture fécale et de l'ELISA sont comparables.

Tableau 4
Valeurs *a posteriori* des différents paramètres pour les quatre modèles utilisés
avec le logiciel WinBUGS : médiane (intervalle de probabilité à 95%) en pourcentages

Paramètres	I	II	III	IV (NI)
Se _{FC}	42 (30-57)	53 (38-70)	50 (33-70)	53 (34-72)
Sp _{FC}	100 (99-100)	100 (99-100)	98 (97-99)	98 (96-100)
Se _{ELISA}	34 (31-38)	34 (31-38)	61 (42-83)	53 (33-78)
Sp _{ELISA}	99 (98-100)	99 (98-100)	99 (97-100)	99 (98-100)
p ₁	4 (2-7)	3 (1-6)	3 (1-5)	1 (0-4)
p ₂	23 (17-30)	22 (16-28)	19 (13-25)	23 (14-34)

IV- DISCUSSION

1. VERIFICATION DU PARADIGME DE HUI ET WALTER

1.1. INDEPENDANCE DES TESTS

Les deux tests utilisés reposent sur des bases biologiques différentes (antigène pour l'un, anticorps pour l'autre) et peuvent donc être considérés comme indépendants [Branscum *et al.*, 2005]. De plus, différents auteurs [Kostoulas *et al.*, 2006, Van Schaik *et al.*, 2006] ont testé l'hypothèse d'une dépendance entre les deux tests et ont démontré que cette hypothèse ne pouvait pas être retenue.

Il existe cependant une méthode développée par Vacek [1985] pour appliquer l'approche bayésienne quand les tests sont dépendants.

1.2. DIFFERENCE DE PREVALENCE ENTRE LES DEUX POPULATIONS

La constitution des deux sous-populations (sur la base des manifestations cliniques) a été réalisée indépendamment des résultats des tests et indépendamment de critères biologiques (âge, sexe). Les données récoltées montrent que les prévalences sont différentes. Cette hypothèse est donc vérifiée.

Il n'est pas nécessaire que la différence de prévalence entre les deux populations soit importante [Enøe *et al.*, 2000].

Cependant, selon différents auteurs [Nielsen *et al.*, 2002 ; Toft *et al.*, 2005] il vaudrait mieux utiliser des critères géographiques pour constituer les deux sous-populations.

1.3. CONSTANCE DES PARAMETRES

C'est l'hypothèse la plus difficile à vérifier. En effet, selon van Schaik *et al.* [2006], les sensibilités sont différentes dans les deux populations car l'excrétion est moins intense dans les troupeaux sans signes cliniques.

2. CHOIX DES PRIORS

Les priors doivent être définis indépendamment des données de l'étude [Branscum *et al.*, 2005]. Idéalement, ils

devraient être définis à partir d'études comparables [Enøe *et al.*, 2000]. Or il existe très peu d'études concernant les caprins. Les estimations *a priori* ont donc été faites sur la base de résultats obtenus pour d'autres espèces, principalement les bovins. Par ailleurs, différents tests ELISA sont commercialisés pour le diagnostic de la paratuberculose et leurs évaluations ont été réalisées par rapport à différents tests (culture fécale, lésions histologiques, signes cliniques,...). Il est donc très difficile de proposer des valeurs *a priori* pour les différents paramètres.

Le fait de tester différents « priors » permet de vérifier la stabilité de l'analyse bayésienne [Enøe *et al.*, 2000].

3. CHOIX D'UNE APPROCHE

Les méthodes probabilistes sont particulièrement adaptées aux maladies chroniques ou à infection persistante telles que la paratuberculose car les antigènes et les anticorps sont détectables pendant l'essentiel du processus infectieux [Branscum *et al.*, 2005].

Les estimations obtenues avec le modèle bayésien non informatif (modèle IV) sont très proches de celles obtenues avec l'approche fréquentielle. L'avantage de l'approche bayésienne, est de pouvoir traiter des échantillons de petite taille, avec des valeurs nulles dans les tableaux de contingence, contrairement à l'approche fréquentielle [Enøe *et al.*, 2000]. Un intervalle de probabilité peut être défini dans tous les cas.

La précision de l'estimation est déterminée par l'amplitude de l'intervalle de probabilité : plus l'intervalle est réduit, meilleure est la précision [Branscum *et al.*, 2005]. Par ailleurs, la taille de l'intervalle de probabilité dépend de la différence de prévalence entre les deux populations [Enøe *et al.*, 2000] : un écart faible entre les deux prévalences induit un intervalle de probabilité large et donc augmente l'imprécision. Les modèles I et II sont les modèles les plus précis.

V - CONCLUSION

L'approche bayésienne permet donc d'estimer les caractéristiques de deux tests pour le diagnostic de l'infection paratuberculeuse. En raison du choix des valeurs *a priori* et de la précision des estimations, c'est le modèle II

qui a été retenu. La sensibilité de la culture fécale est supérieure à celle du test ELISA (53% contre 34%). En revanche, les spécificités des deux tests sont comparables.

BIBLIOGRAPHIE

- Branscum A.J., Gardner I.A., Johnson W.O. - Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. *Prev. Vet. Med.*, 2005, **68**, 145-163.
- Enøe C., Georgiadis M.P., Johnson W.O. - Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.*, 2000, **45**, 61-81.
- Gumber S., Eamens G., Whittington R.J. - Evaluation of a Pourquier ELISA kit in relation to agar gel immunodiffusion (AGID) test for assessment of the humoral immune response in sheep and goats with and without Mycobacterium paratuberculosis infection. *Vet. Microbiol.*, 2006, **115**, 91-101.
- Johnson W.O., Gastwirth J.L., Pearson L.M. - Screening without a «gold standard»: the Hui-Walter paradigm revisited. *Am. J. of Epidem.*, 2001, **153** (9), 921-924.
- Kostoulas P., Leontides L., Enøe C., Billinis C., Florou M., Sofia M. - Bayesian estimation of sensitivity and specificity of serum ELISA and faecal culture for diagnosis of paratuberculosis in greek dairy sheep and goats. *Prev. Vet. Med.*, 2006, **76** (1-2), 56-73.
- Nielsen S.S., Grønbaek C., Agger J.F., Houe H. - Maximum-likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of paratuberculosis. *Prev. Vet. Med.*, 2002, **53**, 191-204.
- Pouillot R., Gerbier G., Gardner I.A. - «TAGS», a program for the evaluation of test accuracy in the absence of a gold standard. *Prev. Vet. Med.*, 2002, **53**, 67-81.
- Praet N., Dorny P., Saegerman C., Marcotty T., Berkvens D. - Estimation de la prévalence d'une maladie et des caractéristiques des tests diagnostiques par une approche bayésienne. *Epidémiol. et santé anim.*, 2006, **49**, 113-130.
- Robbe-Austerman S, Gardner I.A., Thomsen B.V., Morrill D.G., Martin B.M., Palmer M.V., Thoen C.O., Ewing C. - Sensitivity and specificity of the agar-gel-immunodiffusion test, ELISA and the skin test for the detection of paratuberculosis in United States Midwest sheep populations. *Vet. Res.*, 2006, **37**, 553-564.
- Spiegelhalter D., Thomas A., Best N., Gilks W. - BUGS: Bayesian Inference Using Gibbs Sampling, version 0.50. MRC Biostatistics Unit, Cambridge, 1996.
- Toft N., Jørgensen E., Højsgaard S. - Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Prev. Vet. Med.*, 2005, **68**, 19-33.
- Vacek P.M. - The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*, 1985, **41**, 959-968.
- Van Schaik G., Haro F., Mella A., Kruze J. - Bayesian analysis to validate a commercial ELISA to detect paratuberculosis in dairy herds of southern Chile. *Prev. Vet. Med.*, 2007, **79** (1), 59-69.

Whitlock R.H., Wells S.J., Sweeney R.W., Van
Tiem J. - ELISA and faecal culture for
paratuberculosis : sensitivity and
specificity of each method. *Vet. Microbiol.*,
2000, **77**, 387-398.

Wintz E. - Etude comparative de différentes
méthodes de diagnostic de la
paratuberculose caprine. Thèse de
doctorat vétérinaire, Alfort, 2006.

