

# ESTIMATION DES CARACTERISTIQUES DES REACTIFS SEROLOGIQUES IBR A PARTIR DE LA COMPARAISON DE CINQ REACTIFS MENEES PAR L'AFSSA\*

Etienne Petit <sup>1</sup> et Myriam Perrin <sup>2</sup>

**RESUME :** Afin de comparer plusieurs réactifs sérologiques ELISA détectant des anticorps dirigés contre le virus de la rhinotrachéite infectieuse bovine, une sérothèque a été constituée par l'AFSSA à partir de 700 sérums provenant de différents laboratoires départementaux d'analyses. A partir des différentes combinaisons de résultats obtenus entre les réactifs, les sensibilités et spécificités des réactifs, ainsi que leurs facteurs de dépendance respectifs, ont été évalués en l'absence de technique de référence et en utilisant une méthode basée sur le maximum de vraisemblance. Plusieurs approches ont été utilisées. Les réactifs ont été évalués deux par deux sur la base de deux et trois sous-populations issues de la sérothèque. Enfin, en utilisant un modèle développé pour calculer la probabilité d'obtention d'une combinaison de plusieurs résultats, les sensibilités, spécificités et facteurs de dépendance de cinq réactifs ont été ajustés pour maximiser la vraisemblance des combinaisons obtenues à partir des cinq réactifs. Selon les méthodes utilisées, l'ajustement des modèles aux données peut être très variable, ce qui se traduit par une estimation également variable des paramètres. Le meilleur ajustement est obtenu à partir du modèle complet. Ces estimations permettent d'approcher et de comparer les caractéristiques des réactifs étudiés. Il faut cependant nuancer ces résultats du fait de la sélection des sérums qui ont constitué la sérothèque. En effet, une partie d'entre eux ont été adressés au Laboratoire de référence de l'AFSSA pour avoir présenté des résultats discordants. L'ensemble des résultats est présenté et discuté.

**SUMMARY :** In order to compare several serological tests detecting antibodies against BHV-1, a sera collection was constituted by the AFSSA with 700 sera coming from various departmental laboratories of analyses. From the various combinations of results obtained with the various tests, sensitivities and specificities of each test, like their respective factors of dependence, were evaluated in the absence of a "Gold standard" reference by using a maximum likelihood method. Several approaches were used. Each possible pair of tests was evaluated on the basis of two and three populations resulting from the sera collection. Lastly, by using a model developed to calculate the probability of obtaining a combination of several results, sensitivities, specificities and factors of dependence of the five tests were adjusted to maximize the likelihood of such results combinations obtained with the five tests. According to the methods used, the adjustment of the models to the data can be very variable, which results in a so variable estimation of the parameters. The best adjustment is obtained with the complete model. With these method it's possible to approach and compare characteristics of the studied tests. It is however necessary to moderate these results because of the selection of the sera which constituted the collection. Indeed, a part of them were addressed to the reference Laboratory of the AFSSA because they had unmatched results. The whole results are presented and discussed.



\* Communication présentée lors des Journées AEEMA-AESA, 22-23 mai 2003

<sup>1</sup> F.R.G.D.S. Bourgogne, 42 rue de Mulhouse, 21000 Dijon, France

<sup>2</sup> AFSSA Lyon, 31 Avenue Tony Garnier, 69364 Lyon cedex 7, France

## I – INTRODUCTION - OBJECTIFS

Dans le cadre des qualifications relatives à la rhinotrachéite infectieuse bovine (IBR), le cahier des charges défini par l'Association pour la certification en santé animale (ACERSA) prévoit le contrôle et l'agrément des réactifs utilisés. Les contrôles réalisés imposent aux réactifs un minimum de détectabilité vis-à-vis d'un sérum de référence, ainsi qu'un minimum de spécificité évalué initialement sur une population de 1 000 sérums issus d'une population bovine présumée indemne. Chaque lot de production de chaque réactif est également contrôlé selon un protocole prédéfini. Cependant, ces contrôles ne sauraient imposer aux réactifs de fournir la même réponse en face d'un même sérum et des réactions discordantes sont régulièrement constatées sur le terrain, semant le trouble et l'incompréhension parmi les éleveurs et les vétérinaires. En 2000, à la

demande de l'ACERSA, l'Agence française de sécurité sanitaire des aliments (AFSSA) de Lyon a entrepris une comparaison de plusieurs réactifs agréés sur la base d'une sérothèque constituée à partir de sérums provenant de différents laboratoires départementaux, travaillant dans des zones très variées sur le plan épidémiologique. L'objectif de cette étude est de reprendre les résultats fournis par cette comparaison afin d'estimer la sensibilité et la spécificité de chaque réactif en l'absence d'une méthode de référence. Pour cela la méthode du maximum de vraisemblance décrite par Gardner et Greiner [1999] d'après Hui et Walter a été utilisée.

La finalité de ce travail vise à apprécier les apports et les limites de cette approche pour la comparaison des réactifs en vue de leur utilisation optimale.

## II – MATERIEL ET METHODE

### 1. MATERIEL

Six cent quatre vingt onze sérums ont été collectés par l'AFSSA Lyon à partir de différents laboratoires d'analyses vétérinaires départementaux (départements 01-03-08-10-12-14-15-22-35-38-40-44-50-51-56-58-60-61-82-89). Ces sérums ont été analysés par six réactifs ELISA agréés pour le cahier des charges IBR et désignés par les lettres A, B, C, D, E et F.

La sérothèque a été constituée de façon équilibrée entre sérums positifs et négatifs, auxquels ont été associés des sérums dits « discordants », c'est-à-dire ayant présenté des résultats différents avec différents réactifs dans le laboratoire départemental d'origine.

Parmi les réactifs retenus dans l'étude de l'AFSSA, le lot utilisé du réactif C s'est révélé défectueux et n'a pas été retenu pour la suite de cette étude, qui ne portera que sur les réactifs A, B, D, E et F.

Les résultats de chaque analyse ont été fournis par l'AFSSA sous forme quantitative (densité optique rapportée à celle des sérums témoins) et sous forme qualitative (résultat positif, douteux ou négatif). Seuls les résultats non douteux ont été pris en compte dans cette étude.

### 2. METHODE UTILISEE

Cette méthode est décrite par Gardner et Greiner [1999] d'après Hui et Walter.

Dans une population de prévalence réelle « F », on peut établir les probabilités d'obtenir les différentes combinaisons de résultats avec deux tests connus de sensibilités respectives Se1 et Se2, et de spécificités respectives Sp1 et Sp2. Le tableau I décrit ces probabilités pour les animaux infectés et les animaux non infectés.

Tableau I

## Probabilité d'obtenir une combinaison de résultats pour des animaux infectés ou non

Combinaison des résultats	Probabilité qu'un sérum présente une combinaison de résultats	Echantillon observé		
		Infectés	Non infectés	
T1 pos, T2 pos	$p_1 = F \text{ Se}_1 \text{ Se}_2$	+	$(1 - F) (1 - \text{Sp}_1 - \text{Sp}_2 + \text{Sp}_1 \text{ Sp}_2)$	$n_1$
T1 pos, T2 nég	$p_2 = F \text{ Se}_1 (1 - \text{Se}_2)$	+	$(1 - F) (1 - \text{Sp}_1) \text{ Sp}_2$	$n_2$
T1 nég, T2 pos	$p_3 = F (1 - \text{Se}_1) \text{ Se}_2$	+	$(1 - F) (1 - \text{Sp}_2) \text{ Sp}_1$	$n_3$
T1 nég, T2 nég	$p_4 = F (1 - \text{Se}_1 - \text{Se}_2 + \text{Se}_1 \text{ Se}_2)$	+	$(1 - F) \text{ Sp}_1 \text{ Sp}_2$	$n_4$
Total	$p = F$	+	$1 - F$	$N$

Pour une population étudiée conjointement avec les deux tests, et pour chaque combinaison de résultats observée, on calcule la probabilité d'obtenir le résultat observé en fonction des paramètres du modèle posé (ici,  $\text{Se}_1$ ,  $\text{Se}_2$ ,  $\text{Sp}_1$ ,  $\text{Sp}_2$  et  $F$ ).

La vraisemblance globale correspond au produit de ces probabilités. Dans le cas du tableau I, cette vraisemblance globale est :  $V = p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}$ . L'estimation des paramètres s'obtient en faisant varier leurs valeurs jusqu'à obtenir la valeur maximale de la vraisemblance globale (ou plus précisément son logarithme). Cet ajustement nécessite l'utilisation de programmes spécifiques comme le solveur d'équations du logiciel Microsoft Excel ND ou une routine informatique développée pour cette étude. Cette routine recherche successivement pour chaque paramètre une variation qui accroît la vraisemblance du modèle, tout en respectant les contraintes existantes sur tous les paramètres, jusqu'à obtention d'une vraisemblance invariable.

On conçoit que l'ajustement des paramètres ne peut aboutir que si le nombre de données d'observation indépendantes est supérieur au nombre de paramètres du modèle. Or, ici, seules quatre données indépendantes sont disponibles ( $n_1$ ,  $n_2$ ,  $n_3$  et  $n_4$ ), alors que cinq paramètres doivent

être estimés ( $\text{Se}_1$ ,  $\text{Se}_2$ ,  $\text{Sp}_1$ ,  $\text{Sp}_2$  et  $F$ ). Il est donc nécessaire de disposer de deux populations de prévalences différentes ( $F_a$  et  $F_b$ ), les performances intrinsèques des tests ( $\text{Se}_1$ ,  $\text{Se}_2$ ,  $\text{Sp}_1$  et  $\text{Sp}_2$ ) étant supposées identiques dans ces deux populations. On dispose alors de huit données indépendantes, qui permettent d'estimer les six paramètres ( $\text{Se}_1$ ,  $\text{Se}_2$ ,  $\text{Sp}_1$ ,  $\text{Sp}_2$ ,  $F_a$  et  $F_b$ ).

Dans la pratique, deux tests de diagnostic révèlent souvent des réponses biologiques proches, ou dont les cinétiques d'apparition, chez l'animal infecté, sont corrélées. Il devient alors nécessaire de compléter le modèle précédent en ajoutant des facteurs de dépendance de la sensibilité (noté  $K_{se}$ ) et de la spécificité (noté  $K_{sp}$ ). Les probabilités élémentaires de ce nouveau modèle sont détaillées dans le tableau II. La vraisemblance globale se calcule de la même façon que ci-dessus. L'ajout de deux nouveaux paramètres nécessite de recourir à une troisième population pour l'estimation des paramètres.

Finalement, quel que soit le modèle utilisé, si l'ajustement des paramètres permet de trouver la combinaison de valeurs qui minimise la vraisemblance globale des données observées, il ne permet bien sûr pas toujours de trouver la solution idéale grâce à laquelle le modèle reproduit à l'identique les données observées.

Tableau II

## Probabilité d'obtenir une combinaison de résultats pour des animaux infectés ou non en tenant compte de la dépendance des réactifs

Combinaison des résultats	Infectés	Non infectés
T1 pos, T2 pos	$F \text{ Se}_1 \text{ Se}_2 K_{se}$	$(1 - F) (1 - \text{Sp}_1 - \text{Sp}_2 + K_{sp} \text{ Sp}_1 \text{ Sp}_2)$
T1 pos, T2 nég	$F \text{ Se}_1 (1 - \text{Se}_2 K_{se})$	$(1 - F) (1 - \text{Sp}_1 K_{sp}) \text{ Sp}_2$
T1 nég, T2 pos	$F (1 - \text{Se}_1 K_{se}) \text{ Se}_2$	$(1 - F) (1 - \text{Sp}_2 K_{sp}) \text{ Sp}_1$
T1 nég, T2 nég	$F (1 - \text{Se}_1 - \text{Se}_2 + \text{Se}_1 \text{ Se}_2 K_{se})$	$(1 - F) \text{ Sp}_1 \text{ Sp}_2 K_{sp}$
Total	$F$	$1 - F$

### 3. LES DEUX APPROCHES UTILISEES

Deux approches ont été utilisées pour estimer les paramètres :

- La première est basée sur une analyse « deux à deux » des réactifs et a consisté à fractionner la population générale en deux ou trois sous-populations de prévalences plus ou moins élevées et *a priori* différentes, et à étudier deux à deux chaque combinaison de réactifs. Ces sous-populations ont été constituées en scindant le fichier original, trié dans l'ordre des départements d'origine des prélèvements, en deux ou trois parties d'effectifs équivalents.

Une première estimation a été réalisée dans l'hypothèse d'une indépendance des réactifs sur deux sous-populations, une seconde estimation a intégré les facteurs de dépendance sur trois sous-populations

- La seconde méthode s'est appuyée sur la combinaison simultanée des cinq réactifs étudiés appliquée à la population générale. Les modèles ci-dessus (avec et sans dépendance entre tests) ont été généralisés au cas où cinq tests (au lieu de deux) sont effectués sur les mêmes sérums. Pour ce faire, il est nécessaire de recourir à une modélisation qui permet d'établir la probabilité d'obtention d'une combinaison à partir des sensibilités et spécificités des tests et des seules dépendances deux à deux (Petit, 2002). Ce sont alors 32 ( $= 2^5$ ) combinaisons de résultats qui sont possibles. Le modèle sans dépendance entre tests comporte 11 paramètres (= 5 sensibilités + 5 spécificités + 1 prévalence) et le modèle avec dépendances entre tests 31 paramètres (= 11 + 10 dépendances des sensibilités + 10 dépendances des spécificités). L'estimation des paramètres peut donc être faite dans les deux cas avec une seule population.

## III – RESULTATS

Les tableaux III, IV et V fournissent les résultats obtenus par les différentes méthodes, respectivement pour les sensibilités, les spécificités et les facteurs de dépendance.

Les estimations de prévalence ne sont pas données dans les résultats car elles n'ont pas de signification épidémiologique, du fait de la constitution arbitraire des sous-populations. Leur estimation a varié entre 34 et 39% pour la sous-population 1, et 60 et 69% pour la sous-population 2, lors des comparaisons établies sur deux populations (dans l'hypothèse d'indépendance). Pour les comparaisons établies sur trois sous-populations, les fourchettes d'estimation ont été respectivement de 41 à 50%, 35 à 37% et 65 à 72% pour chaque sous-population.

Dans chaque cellule des tableaux III et IV figure l'estimation de la sensibilité ou de la spécificité du test indiqué en entête de colonne

lorsqu'il a été comparé avec le test indiqué en entête de ligne.

On constate une assez grande variabilité des estimations apportées par les approches deux à deux, mais l'ordre de grandeur et surtout la hiérarchie des réactifs sont globalement respectés entre les deux approches (modèles de comparaison de paires de réactifs et modèle global de comparaison des cinq réactifs).

Les approches « deux à deux » ne fournissent pas des niveaux très importants de dépendance, hormis la dépendance des spécificités des réactifs B et D.

De même, l'estimation portée sur la population générale ne révèle qu'un niveau important de dépendance entre les sensibilités des réactifs E et F.

Tableau III

Estimation des sensibilités des réactifs à partir des analyses « deux à deux »  
et de l'analyse sur la population générale

Comparaison avec	Sensibilités	Test A	Test B	Test D	Test E	Test F
Test A	2 populations (indépendance des tests)		96,67%	95,38%	92,66%	90,96%
	3 populations (dépendance des tests)		98,09%	95,82%	95,36%	90,90%
Test B	2 populations (indépendance des tests)	99,10%		97,92%	94,22%	93,19%
	3 populations (dépendance des tests)	99,10%		98,31%	94,24%	92,86%
Test D	2 populations (indépendance des tests)	99,69%	100,00%		93,96%	92,32%
	3 populations (dépendance des tests)	99,57%	100,00%		93,93%	92,07%
Test E	2 populations (indépendance des tests)	99,67%	98,42%	98,12%		97,08%
	3 populations (dépendance des tests)	99,64%	98,39%	97,92%		96,42%
Test F	2 populations (indépendance des tests)	100,00%	100,00%	100,00%	100,00%	
	3 populations (dépendance des tests)	99,83%	99,68%	99,94%	99,29%	
Moyennes	2 populations (indépendance des tests)	99,61%	98,77%	97,86%	95,21%	93,39%
	3 populations (dépendance des tests)	99,54%	99,04%	98,00%	95,71%	93,06%
Ensemble des tests A,B,D,E F	population générale (indépendance des tests)	99,68%	98,73%	98,38%	97,47%	95,90%
	population générale (dépendance des tests)	99,68%	98,67%	98,31%	97,38%	95,48%

Tableau IV

Estimation des spécificités des réactifs à partir des analyses « deux à deux »  
et de l'analyse sur la population générale

Comparaison avec	Spécificités	Test A	Test B	Test D	Test E	Test F
Test A	2 populations (indépendance des tests)		100,00%	99,71%	100,00%	94,88%
	3 populations (dépendance des tests)		99,35%	99,84%	100,00%	94,78%
Test B	2 populations (indépendance des tests)	99,14%		98,51%	99,45%	94,17%
	3 populations (dépendance des tests)	97,02%		97,44%	99,39%	94,12%
Test D	2 populations (indépendance des tests)	100,00%	100,00%		100,00%	94,08%
	3 populations (dépendance des tests)	99,43%	98,47%		99,92%	94,00%
Test E	2 populations (indépendance des tests)	100,00%	100,00%	100,00%		95,12%
	3 populations (dépendance des tests)	97,07%	99,95%	99,92%		94,77%
Test F	2 populations (indépendance des tests)	100,00%	100,00%	100,00%	100,00%	
	3 populations (dépendance des tests)	99,90%	99,92%	99,94%	99,71%	
Moyennes	2 populations (indépendance des tests)	99,78%	100,00%	99,55%	99,86%	94,56%
	3 populations (dépendance des tests)	98,36%	99,42%	99,29%	99,76%	94,42%
Ensemble des tests A,B,D,E F	population générale (indépendance des tests)	97,25%	99,08%	99,69%	99,70%	95,29%
	population générale (dépendance des tests)	97,48%	99,03%	99,66%	99,99%	95,23%

Tableau V

**Estimation des dépendances des sensibilités et spécificités des réactifs  
à partir des analyses « deux à deux » et de l'analyse sur la population générale**

		3 populations (dépendance des tests)				
		Dépendances des sensibilités				
		A	B	D	E	F
Dépendances des spécificités	A		0,99983	0,99987	1,00035	1,00169
	B	1,00654		1,00005	0,99981	1,00321
	D	1,00004	1,01077		1,00154	1,00246
	E	1,00005	1,00054	1,00078		1,00714
	F	1,00097	1,00076	1,00057	1,00286	

  

		population générale (dépendance des tests)				
		Dépendances des sensibilités				
		A	B	D	E	F
Dépendances des spécificités	K Spécificités					
	A		0,99998	1,00000	1,00000	1,00000
	B	1,00292		1,00688	1,00000	1,00000
	D	1,00000	0,99998		0,99993	1,00000
	E	1,00000	1,00000	1,00000		1,01997
	F	1,00000	1,00000	1,00000	1,00000	

## IV - DISCUSSION

### 1. LA POPULATION

Bien que provenant de différents départements à la situation épidémiologique très variable, la sérothèque constituée par l'AFSSA ne se veut pas représentative d'une situation générale ou moyenne, mais vise plutôt à représenter l'ensemble des situations possibles, en particulier les sérums problématiques qui présentent des résultats discordants. Ainsi, les caractéristiques des réactifs estimées dans cette étude ne sont pas directement applicables à une population normale. Cependant, les estimations révèlent les caractéristiques des réactifs dans les situations les plus délicates et devraient permettre de porter un jugement sur leur aptitude à discriminer les sérums difficiles.

Une nuance peut être cependant apportée à ce propos, notamment pour la méthode portant sur la population générale. En effet, tous les sérums présentant un résultat douteux à un des réactifs étudiés ont été de fait éliminés de la sélection qui ne comptait plus que 628 sérums sur les 691 initiaux. Il est donc possible que l'élimination des douteux ait sélectionné une population plus facile à discriminer.

De même, dans les approches « deux à deux » seuls les sérums douteux à l'un ou l'autre de la paire de réactifs étudiés étaient éliminés. De ce fait, les populations n'ont jamais été parfaitement semblables entre

chaque paire. Ce phénomène a pu concourir à la variabilité des résultats produits dans les analyses « deux à deux ». La constitution arbitraire selon les provenances départementales des sous-populations a pu également créer une hétérogénéité de ces populations, au delà des différences de prévalence. En effet, chaque département peut utiliser des réactifs différents, et la sélection des sérums adressés à l'AFSSA va dépendre fortement des réactifs utilisés localement. Cependant, l'association des départements par numéro minéralogique est complètement indépendante de ce phénomène. Compte tenu de ces éléments, il est possible que l'hypothèse sous-jacente à la méthode, à savoir la constance des caractéristiques des tests (sensibilité, spécificité et facteurs de dépendance) dans les différentes sous-populations, ne soit pas respectée. Pour apprécier cette éventualité, un nouveau calcul aurait pu être entrepris sur une nouvelle partition de sous-populations.

### 2. LA METHODE

La méthode du maximum de vraisemblance est séduisante dans son approche car elle permet de s'affranchir de la présence d'une méthode idéale de référence, qui fait souvent défaut en réalité. Sur le plan pratique, les solutions fournies par les calculs informatiques ne sont pas toujours très concluantes et

peinent parfois à reproduire les données observées. Par manque de temps et de compétence statistique, l'inférence statistique des paramètres estimés n'a pu être étudiée. Des indicateurs ont cependant été utilisés pour apprécier la pertinence des estimations : la distance du Chi<sup>2</sup> entre les données observées et les données prédites avec les paramètres ou plus simplement la somme des distances élevées au carré entre observations et prédictions. Des variations importantes ont pu être constatées entre les différentes estimations, d'un écart quasi nul à des distances carrées dépassant la valeur 120. L'explication des forts écarts peut être imputable à l'incapacité des programmes informatiques utilisés à trouver une solution satisfaisante, ou, plus probablement, à l'absence d'une telle solution.

L'utilisation du modèle permettant de calculer les probabilités d'obtention des différentes combinaisons de résultats offre un apport intéressant car cette approche est globale et prend en compte simultanément tous les paramètres. La modélisation qui permet d'estimer les dépendances d'ordre supérieur à 2 à partir des dépendances « deux à deux » limite considérablement le nombre de nouveaux paramètres à estimer et rend possible le calcul à partir d'une seule population. Le faible écart entre les données observées et prédites (voir ci-dessous) et la relative cohérence avec les estimations « deux à deux » semblent lui conférer une pertinence assez élevée. Ce modèle n'a cependant pas encore été précisément évalué et mériterait une validation.

### 3. LES RESULTATS

Face au nombre important de résultats générés par les études « deux à deux » et leur relative variabilité, il nous a semblé plus intéressant de nous pencher sur les résultats produits par l'estimation réalisée sur la population générale qui offre une vision plus globale. Le tableau VI reproduit les valeurs des différents paramètres estimés dans l'hypothèse de dépendance des réactifs. Les facteurs de dépendance entre deux réactifs se trouvent à l'intersection des lignes et colonnes affectées à chaque réactif, à droite de la diagonale du tableau pour les dépendances entre sensibilités et à gauche pour les dépendances entre spécificités.

On peut noter une assez forte dépendance (Kse = 1,02) entre les sensibilités des réactifs E et F qui sont les deux seuls réactifs recherchant des anticorps totaux, ainsi qu'une dépendance moindre entre les sensibilités des réactifs B et D (Kse = 1,0069). En ce qui concerne la spécificité, la seule dépendance observée est celle des réactifs A et B (Ksp = 1,0029).

Le tableau VII reproduit les différentes combinaisons observées de résultats des cinq réactifs sous forme de séquences du type « nppn » où « n » représente un résultat négatif et « p » un résultat positif, et les prédictions du modèle, qui distingue animaux infectés et non infectés.

Tableau VI

**Estimation des sensibilités, spécificités et facteurs de dépendances des réactifs à partir de la population générale dans l'hypothèse d'une dépendance des réactifs**

	Prévalence <b>49,86%</b>	SeA	SeB	SeD	SeE	SeF
SpA	97,48%		1,0000	1	1	1
SpB	99,03%	1,0029		1,0069	1	1
SpD	99,66%	1	1,0000		0,9999	1
SpE	99,99%	1	1	1,0000		1,0200
SpF	95,23%	1	1	1	1,0000	

Tableau VII

**Ajustement des résultats observés et prédits par l'estimation réalisée  
à partir de la population générale dans l'hypothèse d'une dépendance des réactifs**

	Combinaison ABDEF	Observés	Prédits	dont infectés	dont non infectés	VPP
1	nnnnn	289	289,27	0,00	289,27	0%
2	nnnnp	15	14,49	0,00	14,49	0%
3	nnnpn	0	0,02	0,00	0,02	1%
4	nnppp	0	0,01	0,01	0,00	99%
5	nnpnn	1	1,00	0,00	1,00	0%
6	nnpnp	0	0,05	0,00	0,05	0%
7	nnppn	0	0,00	0,00	0,00	3%
8	nnppp	0	0,00	0,00	0,00	96%
9	npnnn	2	2,00	0,00	2,00	0%
10	npnnp	0	0,10	0,00	0,10	0%
11	npnpn	0	0,00	0,00	0,00	63%
12	npnpp	0	0,01	0,01	0,00	100%
13	npppn	0	0,02	0,02	0,00	100%
14	npppp	0	0,01	0,01	0,00	100%
15	ppppn	0	0,02	0,02	0,00	100%
16	ppppp	1	0,93	0,93	0,00	100%
17	pnnnn	7	6,65	0,04	6,61	1%
18	pnnnp	0	0,35	0,01	0,33	4%
19	pnpnp	1	0,06	0,05	0,00	99%
20	pnpnp	2	2,04	2,04	0,00	100%
21	pnpnn	0	0,06	0,04	0,02	63%
22	pnpnp	0	0,01	0,01	0,00	92%
23	pnpnp	0	0,05	0,05	0,00	100%
24	pnpnp	2	1,90	1,90	0,00	100%
25	ppnnn	1	0,98	0,06	0,92	6%
26	ppnnp	0	0,05	0,00	0,05	0%
27	ppnpn	0	0,08	0,08	0,00	100%
28	ppnpp	3	2,99	2,99	0,00	100%
29	pppnn	6	6,03	6,02	0,00	100%
30	ppppp	2	2,00	2,00	0,00	100%
31	ppppn	7	7,77	7,77	0,00	100%
32	ppppp	289	289,05	289,05	0,00	100%

On constate une bonne adéquation entre données observées et prédites. La somme des distances carrées est en effet égale à 2,11. Seule la combinaison « pnnpn » (ligne 19) reste mal prédite par le modèle. Cette adéquation peut s'expliquer par le nombre de paramètres (31) très proche du nombre de données indépendantes (32) : on a donc un modèle saturé. Une estimation des intervalles de confiance de chaque paramètre serait nécessaire pour vérifier la stabilité des estimations.

Un autre élément remarquable est la valeur prédictive positive (VPP indiquée en dernière colonne du tableau) de chaque résultat calculée à partir des prédictions du modèle. En effet pour chaque combinaison effectivement

observée dans l'échantillon, la VPP est soit proche de 100%, soit proche de 0%. La seule exception est la combinaison ppnnn (ligne 25) qui présente une VPP de 6%. Cette observation permet de classer presque sans appel les sérums analysés, malgré la présence de résultats discordants. Il est possible cependant que l'exclusion des sérums douteux ait favorisé ce phénomène.

Quelle que soit la méthode employée (« deux à deux » ou « population générale ») l'estimation des caractéristiques des réactifs fournit sensiblement le même constat en terme de comparaison. Ainsi, les spécificités des réactifs F, et dans une moindre mesure des réactifs A et B, sont bien en dessous des autres valeurs et surtout en dessous de la



valeur attendue de 99,5% testée lors de la validation des réactifs. Il faut rappeler cependant que la population testée dans cette étude n'est pas représentative d'une population normale. Mais les réactifs présentant de telles valeurs sont sans doute moins indiqués pour lever des doutes sur le statut d'un sérum douteux ou à résultats discordants.

De même, les sensibilités des réactifs semblent assez inégales, le réactif A semble présenter une sensibilité bien supérieure à

celles des autres réactifs, mais il faut garder à l'esprit les remarques précédentes sur la population étudiée.

De cette étude il ressort malgré tout que certains réactifs paraissent plus performants que d'autres. Si l'on peut choisir de privilégier la sensibilité (par exemple le réactif A) ou la spécificité (par exemple le réactif E), il est évident qu'un test comme le réactif F présente peu d'intérêt technique par rapport aux autres.

## V - CONCLUSION

Cette étude visait en premier lieu à explorer la méthode du maximum de vraisemblance pour apprécier les caractéristiques des réactifs sérologiques IBR. Cette méthode peut présenter des limites dans certaines situations comme en atteste la variabilité de la qualité des ajustements dans les analyses deux par deux. Il serait donc nécessaire d'affiner cette approche par une estimation des intervalles de confiance.

L'approche par la combinaison de plusieurs réactifs sur une même population nécessite de faire appel à une modélisation des combinaisons de tests si l'on souhaite tenir compte d'une éventuelle dépendance des réactifs entre eux. La modélisation retenue pour cette étude a le mérite d'offrir des résultats cohérents avec les observations moyennes des analyses « deux à deux », ainsi qu'un ajustement entre données observées et prédites plutôt remarquable.

Malgré ses imperfections, la méthode permet de comparer plusieurs réactifs et permet d'optimiser leur utilisation, notamment dans le cas des sérums à résultats discordants. Certains réactifs se révèlent ainsi particulièrement plus performants que d'autres.

La méthode du maximum de vraisemblance est facilement applicable dans la mesure où plusieurs réactifs sont utilisés sur les mêmes sérums, en l'absence de toute méthode de référence. De nombreux laboratoires réalisent

de telles analyses conjointes sur une sérothèque constituée localement en vue de comparer les réactifs sur le marché. La compilation de leurs essais offrirait une estimation répétée et finalement assez précise des qualités des réactifs testés.

Le problème des sérums douteux reste cependant entier, et nécessiterait sans doute d'être approfondi au travers d'une telle méthode. Dans ce cas particulier, mais aussi d'une façon plus générale, cette approche basée sur des résultats uniquement qualitatifs mériterait d'être associée à une analyse des résultats quantitatifs. Une optimisation des seuils d'interprétation pourrait sans doute en découler.

Une connaissance plus précise des qualités des réactifs, associée à une connaissance des circonstances épidémiologiques des sérums analysés, ne peut qu'améliorer la valeur prédictive des résultats fournis par les différents tests et limiter les erreurs de jugement, souvent préjudiciables à tous les acteurs des garanties sanitaires, de l'éleveur acheteur ou vendeur au fabricant de réactifs, en passant par le vétérinaire, les gestionnaires de la garantie et le laboratoire d'analyses. La méthode décrite dans cette étude offre un outil supplémentaire qu'il convient d'approfondir et de mettre en pratique pour améliorer les connaissances des uns et des autres, ce qui nécessite une mise en commun des observations réalisées.

## BIBLIOGRAPHIE

Gardner I.A., Greiner M. - Advanced Methods for Test Validation and Interpretation in *Veterinary Medicine*, 1999, 78 pages, site [www.berlin-info.de](http://www.berlin-info.de), 44-51.

Petit E. - Etude sur la procédure "résultats aberrants" IBR proposée par l'ACERSA. *Epidémiol. et santé anim.*, 2002, **42**, 133-150.

Pouillot R. - Les tests multiples : effets de la dépendance conditionnelle entre tests sur

l'interprétation. In : Méthodes avancées de validation et d'interprétation des tests de diagnostic, CDrom des Ateliers de l'AEEMA du 16 mai 2001.

Toma B. et coll. - Le dépistage des maladies infectieuses animales. In : Epidémiologie appliquée à la lutte contre les maladies transmissibles majeures. AEEMA (Ed), 2002, 696 p.

