



## **Etude d'un modèle statistique bayésien intégrant les résultats de tests multiples pour le diagnostic de la cysticercose porcine**

Praet Nicolas



Crédit National Geographic 2002

Thèse présentée pour l'obtention du grade de Master of Science (M.Sc.)  
en Santé Animale Tropicale

Institut de Médecine Tropicale Prince Léopold,  
Département de production et Santé Animales Tropicales,  
Antwerpen, Belgique

Thèse de Master of Science (M.Sc.) en Santé Animale Tropicale

Présentée et défendue le 27 mars 2006

A l'Institut de Médecine Tropicale (IMT) Prince Léopold

Anvers, Belgique

Composition du jury :

Prof. Dr. Pierre DORNY, IMT, Président  
Prof. Dr. Claude SAEGERMAN, Ulg, Membre  
Prof. Dr. Marleen BOELAERT, IMT, Membre  
Dr. Tanguy MARCOTTY, IMT, Membre  
Prof. Dr. Emmanuel LESAFFRE, KUL, Promoteur  
Prof. Dr. Dirk BERKVENs, IMT, Promoteur

## **Remerciements**

Je tiens à remercier chaleureusement mes promoteurs, les Professeurs Dirk Berkvens et Emmanuel Lesaffre pour leur encadrement et le temps qu'ils ont consacré à la réalisation de ce travail. Merci pour la confiance qui s'est installée entre nous et les bons moments déjà passés ensemble.

Mes remerciements s'adressent également au Professeur Pierre Dorny qui s'est fortement investi dans la collection des opinions d'expert, primordiales pour l'application des techniques bayésiennes. Merci pour cette collaboration de jadis et future.

Merci aussi à tous les membres du département de Santé Animale de l'Institut qui me font le plaisir d'être mes nouveaux collègues, ainsi qu'à tous les étudiants du CIPSAT 2001-2002 pour leur enthousiasme, leur expérience et leurs cultures.

Merci bien sûr à mes parents et à mes frères qui ont toujours crû en moi et m'ont soutenu pendant toutes ces années. Merci à Björn, Marie, Alexandra, Dévi, Vincent, Hélène, Danielle, Dirk, Stéphanie, Ana,...pour ce que vous représentez pour moi.

## Table des matières

1. INTRODUCTION.....	1
1.1. La cysticercose .....	1
1.2. Tests diagnostiques. ....	3
1.2.1. L’inspection de la langue. ....	3
1.2.2. L’inspection visuelle de la carcasse. ....	3
1.2.3. L’imagerie médicale.....	3
1.2.4. Tests sérologiques. ....	3
1.2.4.1. Méthodes de détection d’anticorps (méthodes indirectes). ....	3
1.2.4.2. Méthode de détection d’antigènes (méthode directe). ....	4
1.2.4.3. Validation de ces méthodes diagnostiques.....	4
1.3. Etude épidémiologique en Zambie.....	5
1.3.1. Contexte .....	5
1.3.2. Les résultats.....	5
1.4. L’approche bayésienne.....	7
1.4.1. Définition .....	7
1.4.2. Le théorème de Bayes .....	7
1.4.3. Dépendance conditionnelle et re-paramétrisation .....	9
1.5. Les critères de validation.....	13
1.5.1. Le modèle multinomial .....	13
1.5.2. Bayesian p-value .....	14
1.5.3. Le DIC ou « Deviance Information Criterion » : .....	16
1.5.4. $p_D$ ou « nombre de paramètres réellement estimés ».....	18
1.5.5. La convergence .....	21
1.5.6. Les logiciels.....	21
1.5.6.1. WinBUGS .....	21
1.5.6.2. R .....	21
1.5.6.3. BOA .....	21
2. OBJECTIFS .....	23
3. PREMIERE ETUDE : Simulation pour l’étude du comportement des critères de validation .....	24
3.1. Résultats .....	26
3.2. Conclusions .....	26
4. DEUXIEME ETUDE : Etude du modèle à 4 tests.....	29

5. TROISIEME ETUDE : Etude d'un modèle à 3 tests .....	34
5.1. Résultats et conclusions .....	37
6. QUATRIEME ETUDE : Etude des paramètres d'un modèle à 2 tests .....	44
6.1. Résultats .....	45
6.2. Conclusions .....	46
7. DISCUSSION .....	47
8. PERSPECTIVES.....	50
9. REFERENCES BIBLIOGRAPHIQUES .....	51

## Liste des tableaux

Tableau 1 : Résultats de 4 tests diagnostiques appliqués sur une population de 868 porcs zambiens traditionnels. ....	6
Tableau 2 : Table de contingences reprenant les différents résultats possibles d'un test diagnostique sous forme de probabilités. ....	7
Tableau 3 : Probabilités conditionnelles pour un modèle intégrant 4 tests diagnostiques. ....	11
Tableau 4 : Caractéristiques des 4 tests diagnostiques dans le modèle re-paramétrisé. ....	12
Tableau 5: Table de contingences reprenant les différents résultats possibles d'un test diagnostique. ....	18
Tableau 6 : Table de contingences reprenant les différents résultats possibles d'un test diagnostique sous forme de probabilités. ....	19
Tableau 7: Nombre maximal de paramètres estimables, nombre de paramètres à estimer sous dépendance conditionnelle et nombre de paramètres à estimer sous indépendance conditionnelle pour 1, 2, 3, 4, 5, ..., h tests. ....	20
Tableau 8 : Résultats de l'analyse de la simulation décrite en annexe 1. ....	26
Tableau 9: Marche à suivre pour la validation d'un modèle statistique bayésien intégrant plusieurs tests diagnostiques. ....	28
Tableau 10: Informations <i>a priori</i> appliquées au modèle à 4 tests diagnostiques. ....	29
Tableau 11 : Valeurs de $p_D$ et de DIC calculées à partir des probabilités $p_r$ de la distribution multinomiale en utilisant les informations <i>a priori</i> non informatives. ....	30
Tableau 12 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse du modèle à 4 tests intégrant les informations <i>a priori</i> . ....	30
Tableau 13 : Critères de validation du modèle à 4 tests intégrant les informations <i>a priori</i> . ..	30
Tableau 14 : Informations <i>a priori</i> appliquées au modèle à 4 tests diagnostiques issues du screening réalisé sur base des paramètres $p_D$ et DIC. ....	31
Tableau 15 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse du modèle à 4 tests intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC. ....	32
Tableau 16 : Critères de validation du modèle à 4 tests intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC. ....	32
Tableau 17 : Moyennes postérieures des paramètres $p$ et contraintes strictes imposées au modèle à 4 tests diagnostiques. ....	33
Tableau 18 : Combinaisons de 3 tests possibles à partir des 4 tests diagnostiques. ....	34

Tableau 19 : Informations <i>a priori</i> appliquées aux 24 modèles à 3 tests diagnostiques (Communications personnelles, Dorny, 2004).....	35
Tableau 20 : Informations <i>a priori</i> appliquées aux 24 modèles à 3 tests diagnostiques. Celles-ci sont issues du screening réalisé sur base des paramètres $p_D$ et DIC. ....	36
Tableau 21 : Valeurs de $p_D$ et de DIC calculées à partir des probabilités $pr$ de la distribution multinomiale en utilisant les informations <i>a priori</i> non informatives. ....	37
Tableau 22 : Caractéristiques des tests et prévalence issues de l'analyse des modèles à 3 tests intégrant les informations <i>a priori</i> . ....	38
Tableau 23 : Critères de validation des modèles à 3 tests diagnostiques intégrant les informations <i>a priori</i> . ....	39
Tableau 24 : Caractéristiques des tests et prévalence issues de l'analyse des modèles à 3 tests diagnostiques intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC.....	40
Tableau 25 : Critères de validation des modèles à 3 tests diagnostiques intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC.. ....	41
Tableau 26 : Résultats de l'analyse de la <i>Bayesp</i> sans et avec contraintes strictes des modèles à 3 tests diagnostiques intégrant les informations <i>a priori</i> .....	42
Tableau 27 : Résultats de l'analyse de la <i>Bayesp</i> avec contraintes strictes des modèles à 3 tests diagnostiques intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC.....	43
Tableau 28 : Informations <i>a priori</i> appliquées au modèle à 2 tests diagnostiques. ....	44
Tableau 29 : Valeurs de $p_D$ et de DIC calculées à partir des probabilités $pr$ de la distribution multinomiale en utilisant les informations <i>a priori</i> non informatives. ....	45
Tableau 30 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse des modèles à 2 tests diagnostiques intégrant les informations <i>a priori</i> . ....	45
Tableau 31 : Critères de validation des modèles à 2 tests diagnostiques intégrant les informations <i>a priori</i> . ....	45
Tableau 32 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse des modèles à 2 tests diagnostiques intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC.....	46
Tableau 33 : Critères de validation des modèles à 2 tests diagnostiques intégrant les informations <i>a priori</i> issues du screening réalisé sur base des paramètres $p_D$ et DIC. ....	46

## Liste des figures

Figure 1 : Distribution géographique de la cysticerose dans le monde (WHO, 2002).....	1
Figure 2: Cycle de la cysticerose porcine (Centers for Disease Control and prevention : <a href="http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/A-F/Cysticercosis/Cysticercosis_LifeCycle.gif">http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/A-F/Cysticercosis/Cysticercosis_LifeCycle.gif</a> ). ....	2
Figure 3 : Graphique de la valeur du paramètre $p_D$ en fonction du nombre de contraintes imposées. ....	27



## Résumé

Une étude épidémiologique récente utilisant les techniques statistiques bayésiennes menée sur une population de porcs domestiques zambiens, a permis de fournir une estimation de la prévalence réelle de la cysticercose porcine au sein de cette population. Les résultats de 4 tests diagnostiques, la palpation de la langue, l'inspection visuelle de la carcasse, ainsi que deux tests sérologiques, un test ELISA pour la recherche des antigènes et un test ELISA pour la recherche des anticorps, ont été intégrés dans un modèle statistique bayésien et analysés. La présente étude a pour but de déterminer les critères de validation de cette analyse et de vérifier si la diminution du nombre de tests utilisés influence ceux-ci. Après avoir établi les différents critères de validation à l'aide d'une simulation théorique, à savoir, la correspondance des DIC et  $p_D$  calculés à partir des paramètres  $p$  du modèle et à partir des probabilités  $pr$  de la distribution multinomiale, et la tendance de la Bayesian p-value à tendre vers 0 lorsque l'on impose des contraintes strictes autour des valeurs vraies des différents paramètres, le modèle à 4 tests a été validé et des modèles à 3 puis à 2 tests élaborés et également validés. L'application de l'inspection visuelle de la carcasse ou de la palpation de la langue, suivis, si ceux-ci sont négatifs, d'un test ELISA pour la recherche d'antigène fournit une estimation acceptable de la prévalence de la maladie (environ 60%) et des caractéristiques de ces tests diagnostiques.

Mots clés : Statistiques bayésiennes, cysticercose porcine, *Taenia solium*, tests diagnostiques, WinBUGS, R, critères de validation, prévalence, Bayesian p-value, DIC,  $p_D$ .

## Summary

A recent epidemiological study of village pigs in Zambia using a Bayesian approach gave an estimate of the prevalence of porcine cysticercosis in this population. Four diagnostic tests were used: tongue palpation, visual carcass inspection and two ELISAs for the detection of circulating antigen and specific antibodies respectively. The data were integrated in a Bayesian statistical model and analysed. The aim of the present project is to establish validation criteria for this analysis and apply them on models using only 3 and 2 tests. A theoretical simulation gave 2 criteria: the concordance of DIC and  $p_D$  values evaluated in the multinomial probabilities  $pr$  and in the parameters  $p$  of the model and the fact that the Bayesian p-value tends to zero when severe constraints situated around the posterior nodes are applied. The models integrating 4, 3 and 2 tests were validated. The application of the carcass or tongue inspection followed by the ELISA for detection of antigen if the inspection is negative seems to give an acceptable estimation of the prevalence (around 60%) and of the diagnosis test characteristics.

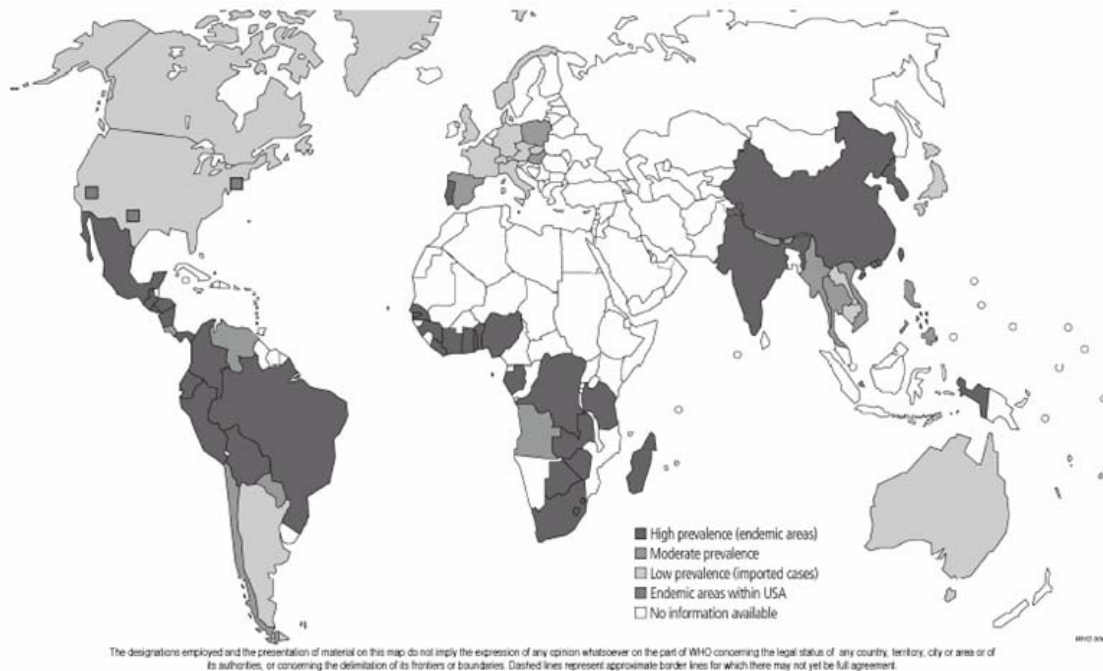
Keywords : Bayesian statistics, cysticercosis, diagnostic tests, WinBUGS, R, validation criteria, prevalence, Bayesian p-value, DIC,  $p_D$ .

# 1. INTRODUCTION

## 1.1. La cysticercose

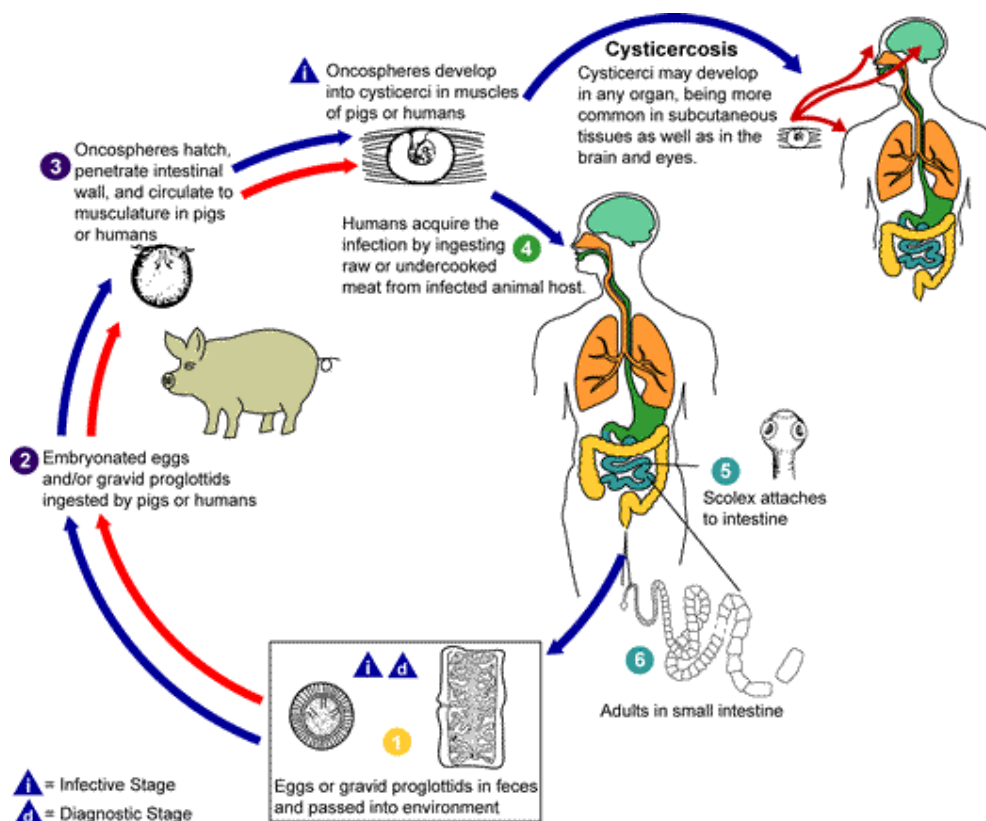
La cysticercose porcine est une helminthose due à *Taenia solium*. En plus d'affecter le porc, cette maladie est une zoonose endémique dans beaucoup de pays en développement en Afrique, Asie et Amérique du Sud (Zoly *et al.*, 2003) (cf. figure 1).

Figure 1 : Distribution géographique de la cysticercose dans le monde (WHO, 2002).



Elle est responsable de fortes pertes économiques (baisse de production et élimination de l'entièreté de la carcasse massivement infestée) et représente un risque important pour la Santé Publique (Garcia & Del Bruto, 2000). L'ingestion de viande de porc infestée de cysticerques entraîne la taeniose chez l'homme (cf. figure 2). Celui-ci est l'hôte définitif du parasite ; le vers adulte se développe dans ses intestins et produit des œufs évacués avec les matières fécales. Le porc s'infeste par l'ingestion de ces œufs. L'homme peut également directement s'infester par ingestion des œufs et développer la cysticercose humaine. Le parasite passe alors dans la circulation sanguine et les cysticerques se développent dans les tissus comme le cerveau (neurocysticercose), les muscles ou le tissu sous-cutané (Nash, 2003).

Figure 2: Cycle de la cysticerose porcine (Centers for Disease Control and prevention : [http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/A-F/Cysticercosis/Cysticercosis\\_LifeCycle.gif](http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/A-F/Cysticercosis/Cysticercosis_LifeCycle.gif)).



Le stade kystique ne provoque généralement pas de réponse immunitaire importante ; mais, une fois mort, le cysticerque est reconnu comme « non soi » et une importante réaction immunitaire entraîne alors des pathologies telles que myosites, myocardites, hydrocéphalie, cécité, méningites, encéphalites et même la mort (Delaram *et al.*, 2004). Les signes cliniques de la maladie sont très variables. Les crises épileptiformes sont fréquentes en cas d'atteinte du système nerveux central (Diop *et al.*, 2003). La Ligue Internationale Contre l'Epileptie indique cette maladie comme la première cause de crises épileptiformes dans les pays en développement (Majofane *et al.*, 2003).

L'augmentation de la production et de la consommation de viande de porc dans le sud et l'est du continent africain a entraîné, ces dernières années, une hausse du nombre de cas de cysticerose porcine et humaine dans ces régions. Les hôpitaux zambiens de Lusaka indiquent une augmentation du nombre de patients présentant des nodules sous-cutanés ou atteints de

crises épileptiformes. Les résultats de tests diagnostiques systématiquement réalisés en abattoir rapportent des cas de cysticercose porcine dans six des neuf provinces zambiennes, évaluant ainsi la prévalence à 20% (Phiri *et al.*, 2003). Ces tests sont la palpation de la base de la langue et l'inspection de la carcasse reconnus comme tests peu sensibles et fournissant, très probablement, une prévalence apparente moindre par rapport à la prévalence réelle de la maladie. D'autres tests diagnostiques plus sensibles sont disponibles.

## **1.2. Tests diagnostiques.**

### **1.2.1. L'inspection de la langue.**

L'inspection de la langue est une méthode diagnostique *in vivo* consistant en l'ouverture de la bouche et l'inspection, la palpation de la base de la langue à la recherche de cysticerques (Gonzalez *et al.*, 1990).

### **1.2.2. L'inspection visuelle de la carcasse.**

L'inspection visuelle est la recherche de cysticerques au niveau des « endroits de prédilection » sur la carcasse, à savoir, les masséters, les triceps brachiaux, la langue, le cœur,...(Boa *et al.*, 2002).

Ces deux techniques sont largement utilisées sur le terrain mais possèdent une sensibilité très faible.

### **1.2.3. L'imagerie médicale.**

En médecine humaine, les techniques d'imagerie médicale permettent de mettre en évidence la présence de cysticerques dans le cerveau. Le diagnostic final doit, malgré tout, faire l'objet d'une confirmation car le diagnostic différentiel comporte également d'autres pathologies (Dorny *et al.*, 2003).

### **1.2.4. Tests sérologiques.**

#### **1.2.4.1. Méthodes de détection d'anticorps (méthodes indirectes).**

Ces méthodes permettent la mise en évidence d'une réaction immunitaire de l'hôte envers le parasite. Plusieurs ont été décrites : le test de fixation du complément, le test d'hémagglutination, le radioimmunoassay, l'ELISA (Enzyme-Linked Immunosorbant Assay), le dipstick-ELISA, la méthode d'agglutination au latex et d'immunoblotting. Les antigènes

utilisés dans ces tests sont issus de liquide de cysticerques ou d'antigènes du parasite *Taenia crassiceps* qui peuvent être produit sur animaux de laboratoire (Nunes *et al.*, 2000). La technique la plus spécifique est l'EITB (Enzyme-linked ImmunoElectroTransfer Blot) utilisant plusieurs glycoprotéines purifiées de cysticerques de *Taenia solium* (Tsang *et al.*, 1989).

Dans les pays en développement, l'ELISA est la méthode de choix vu sa réalisation aisée à bas prix (Dorny *et al.*, 2003).

#### **1.2.4.2. Méthode de détection d'antigènes (méthode directe).**

La méthode directe permettant la détection d'antigènes circulants du parasite est l'ELISA antigène. Celle-ci nécessite la production d'anticorps monoclonaux spécifiques. Le test peut être réalisé sur sérum ou sur liquide céphalo-rachidien (Dorny *et al.*, 2000). Cette technique permet de détecter la présence de cysticerques viables dans l'organisme et se présente donc comme une alternative intéressante aux méthodes indirectes qui fournissent uniquement une information sur l'exposition de l'individu au pathogène et ne permettent pas de prendre de décision thérapeutique. Des réactions croisées avec *Taenia hydatigena* existent chez le porc (Dorny *et al.*, 2003).

#### **1.2.4.3. Validation de ces méthodes diagnostiques.**

Peu de ces techniques sont standardisées et validées. Le seul « gold standard » disponible pour le diagnostic de la cysticercose porcine est la dissection complète de la carcasse du porc à la recherche de cysticerques. Cette technique est évidemment coûteuse et difficilement réalisable à grande échelle (Dorny *et al.*, 2003).

### **1.3. Etude épidémiologique en Zambie**

#### **1.3.1. Contexte**

Une étude épidémiologique récente utilisant les techniques statistiques bayésiennes (Dorny *et al.*, 2004) menée sur une population de porcs domestiques zambiens, a permis de fournir une estimation de la prévalence réelle de la maladie au sein de cette population. Les tests diagnostiques utilisés dans cette étude sont au nombre de quatre : l'inspection de la langue (TONG ou T), l'inspection visuelle de la carcasse (VISUAL ou V), ainsi que deux tests sérologiques plus sensibles; un test ELISA pour la recherche des antigènes (ELISA AG ou AG) et un test ELISA pour la recherche des anticorps (ELISA AB ou AB) (Dorny *et al.*, 2003).

L'approche bayésienne utilisée pour analyser ces données sera discutée au point 1.4.

#### **1.3.2. Les résultats**

Ce travail est basé sur le diagnostic de la cysticercose porcine à l'aide des quatre tests appliqués en parallèle sur une population de 868 porcs traditionnels domestiques zambiens (Dorny *et al.*, 2004). Ces porcs sont issus de l'abattoir de Chibolya et furent examinés et échantillonnés entre les mois de mars et de juin 2000.

Les résultats de ces tests appliqués sur le terrain sont repris dans le tableau 1.

Tableau 1 : Résultats de 4 tests diagnostiques appliqués sur une population de 868 porcs zambiens traditionnels.

TONG	VISUAL	ELISA AG	ELISA AB	Nombre de porcs
0	0	0	0	326
0	0	0	1	42
0	0	1	0	281
0	0	1	1	95
0	1	0	0	0
0	1	0	1	0
0	1	1	0	5
0	1	1	1	4
1	0	0	0	1
1	0	0	1	0
1	0	1	0	2
1	0	1	1	0
1	1	0	0	2
1	1	0	1	1
1	1	1	0	35
1	1	1	1	74

0 = résultat négatif au test, 1 = résultat positif au test ; TONG = la palpation de la langue, VISUAL = l'inspection visuelle de la carcasse, ELISA AG = ELISA Antigène, ELISA AB = ELISA anticorps

## 1.4. L'approche bayésienne

### 1.4.1. Définition

L'approche bayésienne est une technique d'analyse statistique permettant l'intégration de données récoltées sur le terrain et d'opinions d'experts dans un modèle de probabilités (Gelman *et al.*, 2004 ; Branscum *et al.*, 2005 ; Enoe *et al.*, 2000).

### 1.4.2. Le théorème de Bayes

Les statistiques bayésiennes sont basées sur le théorème de Bayes (Bayes, 1763) :

$$\Pr(A|B) = \frac{\Pr(B|A) * \Pr(A)}{\Pr(B|A) * \Pr(A) + \Pr(B|\bar{A}) * \Pr(\bar{A})}$$

avec, par exemple,

Événement A = individu infecté

Événement  $\bar{A}$  = individu indemne

Événement B = résultat positif au test.

Si l'on tient compte de la table de contingences (cf. tableau 2),

Tableau 2 : Table de contingences reprenant les différents résultats possibles d'un test diagnostique sous forme de probabilités.

	Test +	Test -	Total
infectés	$p * se$	$p * (1 - se)$	$p$
indemnes	$(1 - p) * (1 - sp)$	$(1 - p) * sp$	$1 - p$
Total	$p * se + (1 - p) * (1 - sp)$	$p * (1 - se) + (1 - p) * sp$	1

$se$  = sensibilité du test ;  $sp$  = spécificité du test ;  $p$  = prévalence de la maladie.



ce théorème correspond à la formule, plus connue, de la valeur prédictive d'un résultat positif d'un test (Toma *et al.*, 1991) ou :

$$VP_{+} = \frac{p * se}{p * se + (1 - p) * (1 - sp)}$$

avec :

$se$ , la sensibilité du test

$sp$ , la spécificité du test

$p$ , la prévalence

$Pr(B|A)$  est la probabilité d'obtenir un résultat positif au test si l'animal est réellement infecté, soit la sensibilité ( $se$ )

$Pr(A)$  est la probabilité qu'un animal soit infecté, soit la prévalence ( $p$ )

$Pr(B|\bar{A})$  est la probabilité de d'obtenir un résultat positif au test si l'animal est réellement indemne,

soit « 1-la spécificité » ( $1 - sp$ )

$Pr(\bar{A})$  est la probabilité qu'un animal soit indemne, soit « 1-la prévalence » ( $1 - p$ ).

La probabilité qu'un animal soit infecté s'il a obtenu un résultat positif au test est donc proportionnelle à la probabilité d'obtenir un résultat positif au test si l'animal est infecté multiplié par la probabilité qu'un animal soit infecté :

$$Pr(A|B) \propto Pr(B|A) * Pr(A)$$

La généralisation de cette formule à plusieurs tests et donc à un nombre plus élevé de paramètres à estimer (cf.§1.5.4.) permet l'intégration des « résultats de plusieurs tests » obtenus sur le terrain et d'« opinions d'experts » ou « informations *a priori* » pour fournir une estimation des différentes caractéristiques des tests et de la prévalence réelle, aussi appelées « informations *a posteriori* ». Ceci pouvant se résumer par la formule suivante :

$$Pr(\Theta | Y) \sim Pr(Y | \Theta) * Pr(\Theta)$$

avec :

$\Pr(\Theta | Y)$ , informations *a posteriori* des paramètres étant donné la vraisemblance des données de terrain,

$\Pr(Y | \Theta)$ , vraisemblance des observations étant donné la valeur  $\Theta$  des paramètres,

$\Pr(\Theta)$ , informations *a priori*, opinions d'experts ou distribution de probabilité des valeurs  $\Theta$  des paramètres. Nous utilisons une distribution uniforme tronquée dont le domaine s'étend de  $a$  à  $b$ .

### 1.4.3. Dépendance conditionnelle et re-paramétrisation

Si plusieurs tests diagnostiques sont utilisés, il peut y avoir une dépendance conditionnelle entre ceux-ci. Deux tests sont conditionnellement indépendants lorsque la sensibilité et la spécificité du second test ne dépend pas des résultats du premier, au sein d'un groupe d'animaux infectés ou indemnes respectivement (Gardner *et al.*, 2000 ; Dendukuri & Joseph, 2001) :

$$\Pr(T_2^+ | T_1^+ \cap \text{infecté}) = \Pr(T_2^+ | T_1^- \cap \text{infecté}) = \Pr(T_2^+ | \text{infecté})$$

$$\Pr(T_2^- | T_1^- \cap \text{indemne}) = \Pr(T_2^- | T_1^+ \cap \text{indemne}) = \Pr(T_2^- | \text{indemne})$$

La dépendance conditionnelle entre les sensibilités et spécificités de deux tests est une covariance que nous noterons  $\gamma$  et  $\gamma'$  respectivement :

$$\gamma = \text{cov}(T_1, T_2 | D^+) = \frac{\sum_{i=1}^N (T_{1i} - Se_1) \cdot (T_{2i} - Se_2)}{N}$$

$$\gamma' = \text{cov}(T_1, T_2 | D^-) = \frac{\sum_{i=1}^N [(1 - T_{1i}) - Sp_1] \cdot [(1 - T_{2i}) - Sp_2]}{N}$$

avec,

$i$ , l'individu

$T_{1i}$ , le test 1 (0 = résultat négatif, 1 = résultat positif)

$T_{2i}$ , le test 2 (0 = résultat négatif, 1 = résultat positif)

$Se_1$ , la sensibilité du test 1

$Se_2$ , la sensibilité du test 2

$Sp_1$ , la spécificité du test 1

$Sp_2$ , la spécificité du test 2

$N$ , l'effectif total

$D^+$ , un animal infecté

$D^-$ , un animal indemne

Une opinion claire et précise de la sensibilité et de la spécificité d'un test ainsi que de la dépendance conditionnelle éventuelle entre les tests étant difficile à fournir par des experts, le modèle utilisé est une re-paramétrisation du modèle classique pour tests multiples établie par Adel & Berkvens (2002) (cf. tableau 3). Cette re-paramétrisation permet également d'écarter le problème de limites des covariances évoqués par ces derniers. Le tableau 4 reprend les équations permettant le calcul des différentes caractéristiques des tests et de la prévalence à partir de la re-paramétrisation.

Tableau 3 : Probabilités conditionnelles pour un modèle intégrant 4 tests diagnostiques.

Prevalence	$Pr(D^+)$	$p_1$
$Se_I$	$Pr(T_1^+   D^+)$	$p_2$
$Sp_I$	$Pr(T_1^-   D^-)$	$p_3$
	$Pr(T_2^+   D^+ \cap T_1^+)$	$p_4$
	$Pr(T_2^+   D^+ \cap T_1^-)$	$p_5$
	$Pr(T_2^-   D^- \cap T_1^-)$	$p_6$
	$Pr(T_2^-   D^- \cap T_1^+)$	$p_7$
	$Pr(T_3^+   D^+ \cap T_1^+ \cap T_2^+)$	$p_8$
	$Pr(T_3^+   D^+ \cap T_1^+ \cap T_2^-)$	$p_9$
	$Pr(T_3^+   D^+ \cap T_1^- \cap T_2^+)$	$p_{10}$
	$Pr(T_3^+   D^+ \cap T_1^- \cap T_2^-)$	$p_{11}$
	$Pr(T_3^-   D^- \cap T_1^- \cap T_2^-)$	$p_{12}$
	$Pr(T_3^-   D^- \cap T_1^- \cap T_2^+)$	$p_{13}$
	$Pr(T_3^-   D^- \cap T_1^+ \cap T_2^-)$	$p_{14}$
	$Pr(T_3^-   D^- \cap T_1^+ \cap T_2^+)$	$p_{15}$
	$Pr(T_4^+   D^+ \cap T_1^+ \cap T_2^+ \cap T_3^+)$	$p_{16}$
	$Pr(T_4^+   D^+ \cap T_1^+ \cap T_2^+ \cap T_3^-)$	$p_{17}$
	$Pr(T_4^+   D^+ \cap T_1^+ \cap T_2^- \cap T_3^+)$	$p_{18}$
	$Pr(T_4^+   D^+ \cap T_1^+ \cap T_2^- \cap T_3^-)$	$p_{19}$
	$Pr(T_4^+   D^+ \cap T_1^- \cap T_2^+ \cap T_3^+)$	$p_{20}$
	$Pr(T_4^+   D^+ \cap T_1^- \cap T_2^+ \cap T_3^-)$	$p_{21}$
	$Pr(T_4^+   D^+ \cap T_1^- \cap T_2^- \cap T_3^+)$	$p_{22}$
	$Pr(T_4^+   D^+ \cap T_1^- \cap T_2^- \cap T_3^-)$	$p_{23}$
	$Pr(T_4^-   D^- \cap T_1^- \cap T_2^- \cap T_3^-)$	$p_{24}$
	$Pr(T_4^-   D^- \cap T_1^- \cap T_2^- \cap T_3^+)$	$p_{25}$
	$Pr(T_4^-   D^- \cap T_1^- \cap T_2^+ \cap T_3^-)$	$p_{26}$
	$Pr(T_4^-   D^- \cap T_1^- \cap T_2^+ \cap T_3^+)$	$p_{27}$
	$Pr(T_4^-   D^- \cap T_1^+ \cap T_2^- \cap T_3^-)$	$p_{28}$
	$Pr(T_4^-   D^- \cap T_1^+ \cap T_2^- \cap T_3^+)$	$p_{29}$
	$Pr(T_4^-   D^- \cap T_1^+ \cap T_2^+ \cap T_3^-)$	$p_{30}$
	$Pr(T_4^-   D^- \cap T_1^+ \cap T_2^+ \cap T_3^+)$	$p_{31}$

Tableau 4 : Caractéristiques des 4 tests diagnostiques dans le modèle re-paramétrisé.

$$p = p_1$$

$$Se1 = p_2$$

$$Sp_1 = p_3$$

$$Se_2 = p_2 * p_4 + (1 - p_2) * p_5$$

$$Sp_2 = p_3 * p_6 + (1 - p_3) * p_7$$

$$Se_3 = p_2 * (p_4 * p_8 + (1 - p_4) * p_9) + (1 - p_2) * (p_5 * p_{10} + (1 - p_5) * p_{11})$$

$$Sp_3 = p_3 * (p_6 * p_{12} + (1 - p_6) * p_{13}) + (1 - p_3) * (p_7 * p_{14} + (1 - p_7) * p_{15})$$

$$Se_4 = p_2 * (p_4 * (p_8 * p_{16} + (1 - p_8) * p_{17}) + (1 - p_4) * (p_9 * p_{18} + (1 - p_9) * p_{19})) \\ + (1 - p_2) * (p_5 * (p_{10} * p_{20} + (1 - p_{10}) * p_{21}) + (1 - p_5) * (p_{11} * p_{22} + (1 - p_{11}) * p_{23}))$$

$$Sp_4 = p_3 * (p_6 * (p_{12} * p_{24} + (1 - p_{12}) * p_{25}) + (1 - p_6) * (p_{13} * p_{26} + (1 - p_{13}) * p_{27})) \\ + (1 - p_3) * (p_7 * (p_{14} * p_{28} + (1 - p_{14}) * p_{29}) + (1 - p_7) * (p_{15} * p_{30} + (1 - p_{15}) * p_{31}))$$

## 1.5. Les critères de validation

Afin de valider l'analyse bayésienne et de vérifier si les informations *a priori* sont en adéquation avec les données de terrain, plusieurs critères de validation sont disponibles : La « Bayesian p-value », le « Deviance Information Criterion », le paramètre «  $p_D$  » (Spiegelhalter *et al.*, 2003 ; Berkvens *et al.*, 2005) et « l'Analyse de la convergence ».

### 1.5.1. Le modèle multinomial

Les résultats  $r_i$  des tests diagnostiques suivent une distribution multinomiale à  $k$  classes :

$$r_i \sim \text{dmulti}(\pi_1, \dots, \pi_k, N)$$

avec

$N$ , le nombre de résultats

$\pi_i$ , les probabilités pour chaque classe

$k$ , le nombre de classe ( $k = 2^h$ , avec  $h$  le nombre de tests utilisés)

Si nous tenons compte d'un diagnostic basé sur 4 tests diagnostiques en intégrant la dépendance conditionnelle éventuelle décrite ci-dessus, les probabilités de la distribution binomiales seront les suivantes (Adel & Berkvens, 2002) :

$$\begin{aligned} Pr_{0000} &= p_1 * (1-p_2) * (1-p_3) * (1-p_{11}) * (1-p_{23}) + (1-p_1) * p_3 * p_6 * p_{12} * p_{24} \\ Pr_{0001} &= p_1 * (1-p_2) * (1-p_3) * (1-p_{11}) * p_{23} + (1-p_1) * p_3 * p_6 * p_{12} * (1-p_{24}) \\ Pr_{0010} &= p_1 * (1-p_2) * (1-p_3) * p_{11} * (1-p_{22}) + (1-p_1) * p_3 * p_6 * (1-p_{12}) * p_{25} \\ Pr_{0011} &= p_1 * (1-p_2) * (1-p_3) * p_{11} * p_{22} + (1-p_1) * p_3 * p_6 * (1-p_{12}) * (1-p_{25}) \\ Pr_{0100} &= p_1 * (1-p_2) * p_5 * (1-p_{10}) * (1-p_{21}) + (1-p_1) * p_3 * (1-p_6) * p_{13} * p_{26} \\ Pr_{0101} &= p_1 * (1-p_2) * p_5 * (1-p_{10}) * p_{21} + (1-p_1) * p_3 * (1-p_6) * p_{13} * (1-p_{26}) \\ Pr_{0110} &= p_1 * (1-p_2) * p_5 * p_{10} * (1-p_{20}) + (1-p_1) * p_3 * (1-p_6) * (1-p_{13}) * p_{27} \\ Pr_{0111} &= p_1 * (1-p_2) * p_5 * p_{10} * p_{20} + (1-p_1) * p_3 * (1-p_6) * (1-p_{13}) * (1-p_{27}) \\ Pr_{1000} &= p_1 * p_2 * (1-p_4) * (1-p_9) * (1-p_{19}) + (1-p_1) * (1-p_3) * p_7 * p_{14} * p_{28} \\ Pr_{1001} &= p_1 * p_2 * (1-p_4) * (1-p_9) * p_{19} + (1-p_1) * (1-p_3) * p_7 * p_{14} * (1-p_{28}) \\ Pr_{1010} &= p_1 * p_2 * (1-p_4) * p_9 * (1-p_{18}) + (1-p_1) * (1-p_3) * p_7 * (1-p_{14}) * p_{29} \\ Pr_{1011} &= p_1 * p_2 * (1-p_4) * p_9 * p_{18} + (1-p_1) * (1-p_3) * p_7 * (1-p_{14}) * (1-p_{29}) \\ Pr_{1100} &= p_1 * p_2 * p_4 * (1-p_8) * (1-p_{17}) + (1-p_1) * (1-p_3) * (1-p_7) * p_{15} * p_{30} \\ Pr_{1101} &= p_1 * p_2 * p_4 * (1-p_8) * p_{17} + (1-p_1) * (1-p_3) * (1-p_7) * p_{15} * (1-p_{30}) \\ Pr_{1110} &= p_1 * p_2 * p_4 * p_8 * (1-p_{16}) + (1-p_1) * (1-p_3) * (1-p_7) * (1-p_{15}) * p_{31} \\ Pr_{1111} &= p_1 * p_2 * p_4 * p_8 * p_{16} + (1-p_1) * (1-p_3) * (1-p_7) * (1-p_{15}) * (1-p_{31}) \end{aligned}$$

### 1.5.2. Bayesian p-value

La Bayesian p-value est un premier paramètre permettant d'indiquer si les opinions d'experts ne sont pas en contradiction avec les données de terrain. Elle tient compte de la vraisemblance des résultats et des valeurs échantillonnées à partir de la distribution multinomiale du modèle. Elle correspond à la différence entre la déviance des valeurs observées et la déviance des valeurs échantillonnées, soit :

$$p - value = 2 \sum_i r_i \ln \frac{r_i}{\hat{r}_i} - 2 \sum_i \tilde{r}_i \ln \frac{\tilde{r}_i}{\hat{r}_i}$$

La déviance des valeurs observées sur le terrain étant :

$$Dev_{obs} = 2 \sum_i r_i \ln \frac{r_i}{\hat{r}_i} = -2 \ln L$$

avec

$r_i$ , les valeurs observées

$\hat{r}_i$ , les valeurs calculées à partir de la distribution multinomiale du modèle

$L$ , la vraisemblance

La déviance des valeurs échantillonnées étant :

$$Dev_{échant} = 2 \sum_i \tilde{r}_i \ln \frac{\tilde{r}_i}{\hat{r}_i}$$

avec

$\tilde{r}_i$ , les valeurs échantillonnées à partir de la distribution multinomiale du modèle à chaque itération

$\hat{r}_i$ , les valeurs calculées à partir de la distribution multinomiale du modèle.

Nous considérons alors une « *step function* » de la *p-value* telle que :

$$\textit{Step function} = 1 \text{ si } p\text{-value} \geq 0$$

$$\textit{Step function} = 0 \text{ si } p\text{-value} < 0$$

et ce pour chaque itération.

Nous calculons ensuite, la moyenne postérieure de cette « *step function* », la *Bayesp*; celle-ci se situant entre 0 et 1.

Un bon modèle trouvera sa *Bayesp*, de préférence, proche de 0.5. En effet, dans 50% des itérations, la déviance des observations sera plus grande que la déviance des valeurs échantillonnées ; la « *step function* » prendra alors la valeur de 1. Dans les 50% des itérations restantes, la déviance des observations sera plus petite que la déviance des valeurs échantillonnées et la « *step function* » prendra la valeur de 0.



### 1.5.3. Le DIC ou « Deviance Information Criterion » :

Le deuxième paramètre permettant de vérifier l'adéquation entre les opinions d'expert et les résultats expérimentaux est le DIC ou « Deviance Information Criterion » (Spiegelhalter *et al.*, 2003). Il se base également sur la vraisemblance des observations. Le DIC est une généralisation de l'« Akaiques Information Criterion (AIC) » pour un modèle multinomial (Berkvens *et al.*, 2006) :

$$AIC = -2 \sum_{i=1}^k r_i \ln(\hat{\pi}_i) + 2p$$

avec

$r_i$ , les valeurs observées

$\hat{\pi}_i$ , les probabilités pour chaque classe estimées par le modèle

$p$ , le nombre de paramètre réellement estimés

Le  $p$  est ici remplacé par son équivalent bayésien «  $p_D$  » (cf. paragraphe 1.5.4.) pour obtenir l'équation suivante :

$$DIC = -4 \sum_{i=1}^k r_i \ln(\pi_i) + 2 \sum_{i=1}^k r_i \ln(\hat{\pi}_i)$$

avec

$r_i$ , les valeurs observées

$\hat{\pi}_i$ , les probabilités estimées par le modèle pour chaque classe

$\pi_i$ , les probabilités observées pour chaque classe

En se basant sur la formule de l'AIC, le DIC peut être exprimé comme,

$$DIC = -2 \ln L + 2p_D$$

avec

$L$ , la vraisemblance

$p_D$ , le nombre de paramètres réellement estimés.

Si la vraisemblance augmente et/ou le nombre de paramètres à estimer diminue, le DIC diminue.

Si la vraisemblance diminue et/ou le nombre de paramètres à estimer augmente, le DIC augmente.

Le modèle présentant la meilleure adéquation entre les données de terrain et les informations *a priori* sera celui qui possède le DIC le plus faible et positif. En d'autres termes, la vraisemblance doit être la plus élevée possible et sa log-concavité respectée (Spiegelhalter *et al.*, 2002).

#### 1.5.4. $p_D$ ou « nombre de paramètres réellement estimés »

Ce paramètre tient également compte de la vraisemblance des observations et se calcule comme suit :

$$p_D = -2 \sum_{i=1}^k r_i \ln(\pi_i) + 2 \sum_{i=1}^k r_i \ln(\hat{\pi}_i)$$

Le  $p_D$  ou « nombre de paramètres réellement estimés » doit être positif mais inférieur ou égal au nombre de paramètres estimables.

#### Nombre de paramètres estimables :

Prenons la table de contingences pour l'application d'UN test diagnostic (cf. tableau 5):

Tableau 5: Table de contingences reprenant les différents résultats possibles d'un test diagnostique.

	Test +	Test -	<i>total</i>
Infectés	$n_{T^+ D^+}$	$n_{T^- D^+}$	$n_{D^+}$
Indemnes	$n_{T^+ D^-}$	$n_{T^- D^-}$	$n_{D^-}$
<i>Total</i>	$n_{T^+}$	$n_{T^-}$	<b>n</b>

$n_{T^+|D^+}$  = vrais positifs ;  $n_{T^-|D^+}$  = faux positifs ;

$n_{T^-|D^+}$  = faux négatifs ;  $n_{T^-|D^-}$  = vrais négatifs.

Avec la sensibilité et la spécificité respectivement définies par :

$$Se = \frac{n_{T^+|D^+}}{n_{D^+}} \quad Sp = \frac{n_{T^-|D^-}}{n_{D^-}}$$

Cette même table sous forme de probabilités (cf.tableau 6),

Tableau 6 : Table de contingences reprenant les différents résultats possibles d'un test diagnostique sous forme de probabilités.

	Test +	Test -	Total
infectés	$p * se$	$p * (1 - se)$	$p$
indemnes	$(1 - p) * (1 - sp)$	$(1 - p) * sp$	$1 - p$
Total	$p * se + (1 - p) * (1 - sp)$	$p * (1 - se) + (1 - p) * sp$	1

$se$  = sensibilité du test ;  $sp$  = spécificité du test ;  $p$  = prévalence de la maladie.

nous permet d'énoncer les 2 équations suivantes :

$$pr_1 = p.se + (1 - p).(1 - sp)$$

$$pr_0 = p.(1 - se) + (1 - p).sp$$

avec :

$pr_1$ , la probabilité pour un animal d'obtenir un résultat positif au test

$pr_0$ , la probabilité pour un animal d'obtenir un résultat négatif au test

Il y a donc 3 paramètres à estimer : la sensibilité  $se$ , la spécificité  $sp$  et la prévalence  $p$

Le nombre de paramètres estimables est de 1 car il y a une seule équation indépendante.

Si nous utilisons maintenant DEUX tests diagnostiques, il existe alors 4 résultats de tests possibles et donc 4 équations (remarquons que dans un but didactique, ce raisonnement n'est pas réalisé à partir du modèle re-paramétrisé) :

$$Pr_{11} = (se_1 * se_2 + \gamma_{12}) * p + ((1 - sp_1) * (1 - sp_2) + \gamma'_{12}) * (1 - p)$$

$$Pr_{10} = (se_1 * (1 - se_2) - \gamma_{12}) * p + ((1 - sp_1) * sp_2 - \gamma'_{12}) * (1 - p)$$

$$Pr_{01} = ((1 - se_1) * se_2 - \gamma_{12}) * p + (sp_1 * (1 - sp_2) - \gamma'_{12}) * (1 - p)$$

$$P_{00} = ((1 - se_1) * (1 - se_2) + \gamma_{12}) * p + (sp_1 * sp_2 + \gamma'_{12}) * (1 - p)$$

Avec  $se_1$ , la sensibilité du premier test ;  $se_2$ , la sensibilité du deuxième test ;  $sp_1$ , la spécificité de premier test ;  $sp_2$ , la spécificité du deuxième test ;  $\gamma_{12}$ , la dépendance conditionnelle entre les sensibilités des deux tests ;  $\gamma'_{12}$ , la dépendance conditionnelle entre les spécificités des deux tests.

Le nombre de paramètres à estimer est de 7 si nous tenons compte des dépendances conditionnelles éventuelles entre les tests et de 5 dans le cas contraire.

Le nombre de paramètres estimables est de 3 car il y a 3 équations indépendantes.

Le tableau 7 ci-dessous, permet d'élargir ce principe à 3, 4, 5, ..., h tests

Tableau 7: Nombre maximal de paramètres estimables, nombre de paramètres à estimer sous dépendance conditionnelle et nombre de paramètres à estimer sous indépendance conditionnelle pour 1, 2, 3, 4, 5, ..., h tests.

Nombre de tests	Nombre maximal de paramètres estimables	Nombre de paramètres à estimer sous dépendance conditionnelle	Nombre de paramètres à estimer sous indépendance conditionnelle
1	1	3	3
2	3	7	5
3	7	15	7
4	15	31	9
5	31	63	11
h	$2^h - 1$	$2^{h+1} - 1$	$2h + 1$

Nous disposons ainsi du nombre de paramètres à estimer et du nombre maximal de paramètres estimables. Le  $p_D$  est le nombre de paramètres réellement estimés et doit être inférieur ou égal au nombre maximal de paramètres estimables. Le nombre de paramètres à estimer étant généralement supérieur au nombre maximal de paramètres estimables, nous devons utiliser les informations *a priori* pour les réduire.

### 1.5.5. La convergence

Cf. paragraphe 1.5.6.3.

### 1.5.6. Les logiciels

Les analyses statistiques sont réalisées dans différents logiciels

#### 1.5.6.1. WinBUGS

BUGS correspond à « Bayesian inference Using Gibbs Sampling ». Ce logiciel utilise la méthode de Gibbs sampling du Markov Chain Monte Carlo pour fournir une estimation de la prévalence et des différents paramètres des tests diagnostiques à partir des données de terrain et des opinions d'expert insérées dans le modèle multinomial (Spiegelhalter *et al.*, 2003). Le Markov Chain Monte Carlo est un algorithme d'échantillonnage à partir d'une distribution de probabilité 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.

#### 1.5.6.2. R

R est un logiciel destiné à la manipulation, au calcul et à la représentation de données (The R Foundation for Statistical Computing, 2004). Nous utilisons ce programme à deux fins :

- analyser la convergence à travers le « package BOA »
- calculer les paramètres DIC et  $p_D$  de manière alternative à celle utilisée en WinBUGS.

Le logiciel WinBUGS utilise, en effet, les moyennes postérieures des différents paramètres  $p$  du modèle pour calculer les deux critères (nous les nommerons DIC\_ $p$  et  $p_D$ \_ $p$ ) alors que nous utilisons les moyennes postérieures des probabilités  $pr$  du modèle multinomial (nous les nommerons DIC\_ $pr$  et  $p_D$ \_ $pr$ ).

#### 1.5.6.3. BOA

BOA ou « Bayesian Output Analysis program » est un package utilisable en R. Il permet l'analyse de la convergence du modèle en utilisant les CODA's (Smith, 2004).

Les CODA's sont les recueils des valeurs des paramètres pour chaque itération issues de l'analyse des modèles dans le logiciel WinBUGS, et ce, pour chaque chaîne d'itérations. BOA permet de vérifier si ces chaînes d'itérations sont convergentes. Les chaînes sont convergentes si l'estimation du paramètre étudié ne dépend pas de leur point d'amorçage (Brooks & Gelman, 1998).

## 2. OBJECTIFS

Les objectifs de ce travail sont de déterminer les critères de validation (décrits au paragraphe 1.5.) du modèle statistique bayésien utilisé. Nous étudierons d'abord le comportement de ces paramètres à travers une simulation sur un modèle comportant deux tests diagnostiques.

Nous envisagerons alors l'étude et la validation du modèle à 4 tests proposé par Dorny *et al.* (2004).

Nous vérifierons ensuite, à partir de ce modèle, s'il est possible de réduire le nombre de tests diagnostiques utilisés à 3 et ensuite à 2 tests en maintenant une prévalence, des spécificités et sensibilités des tests identiques à celles obtenues par l'analyse du modèle à 4 tests et en validant ces modèles. Ceci est réalisé dans le but de diminuer le nombre de faux négatifs, issus de l'utilisation exclusive d'un test peu sensible (inspection de la langue ou de la carcasse), tout en minimisant les coûts nécessaires au diagnostic de la maladie.



### 3. PREMIERE ETUDE : Simulation pour l'étude du comportement des critères de validation

Cette simulation permet d'étudier le comportement des différents critères de validation développés ci-dessus tout en connaissant la prévalence réelle de la maladie X. Les conditions d'échantillonnage et les valeurs de sensibilité et de spécificité fictives choisies sont proches des conditions retrouvées dans l'étude sur la cysticercose porcine en Zambie décrites auparavant mais seulement 2 tests diagnostiques sont considérés :

Imaginons un échantillonnage de 1000 individus ( $N = 1000$ ) réalisé aléatoirement au sein d'une population.

Imaginons une maladie X dont la prévalence réelle est de 0.50.

Imaginons deux tests diagnostiques (T1 et T2) dont la spécificité ( $Sp$ ) est de 1 ; il n'y a donc pas de faux positif possible. Les sensibilités ( $Se$ ) des tests sont connues et sont respectivement de 0.6 et de 0.7. Il existe une dépendance conditionnelle entre les tests. Il faut donc considérer 7 paramètres à estimer dont les probabilités sont :

$$p_1 = 0.50$$

$$p_2 = 0.60$$

$$p_3 = 1$$

$$p_4 = 0.90$$

$$p_5 = 0.40$$

$$p_6 = 1$$

$$p_7 = 1$$

$$Se_1 = p_2 = 0.6$$

$$Sp_1 = p_3 = 1$$

$$Se_2 = p_2 * p_4 + (1 - p_2) * p_5 = 0.7$$

$$Sp_2 = p_3 * p_6 + (1 - p_3) * p_7 = 1$$

Les probabilités de la distribution binomiale peuvent alors être calculées :

$$pr_1 = pr_{00} = p_1*(1-p_2)*(1-p_5) + (1-p_1)*p_3*p_6 = 0.62$$

$$pr_2 = pr_{01} = p_1*(1-p_2)*p_5 + (1-p_1)*p_3*(1-p_6) = 0.08$$

$$pr_3 = pr_{10} = p_1*p_2*(1-p_4) + (1-p_1)*(1-p_3)*p_7 = 0.03$$

$$pr_4 = pr_{11} = p_1*p_2*p_4 + (1-p_1)*(1-p_3)*(1-p_7) = 0.27$$

avec « 0 » correspondant à un résultat négatif au test et « 1 » à un résultat positif.

Les observations de terrain théoriques des tests sont calculées à partir des informations *a posteriori* décrites ci-dessus:

$$r(00) = 620$$

$$r(01) = 80$$

$$r(10) = 30$$

$$r(11) = 270$$

Plusieurs modèles sont envisagés ; ils diffèrent quant aux contraintes imposées (cf. annexe 1):

- Modèle 1 (M1), aucune contrainte (informations *a priori* non informatives)
- Modèle 2 (M2), les spécificités des deux tests sont égales à 1
- Modèle 3 (M3), les spécificités des deux tests sont égales à 1 ; des contraintes sont appliquées aux sensibilités des deux tests uniformément autour des probabilités des paramètres  $p$  décrits ci-dessus (cf.annexe 1)
- Modèle 4 (M4), les spécificités des deux tests sont égales à 1 ; des contraintes strictes sont appliquées aux sensibilités des deux tests uniformément autour des probabilités des paramètres  $p$  décrits ci-dessus (cf.annexe 1)
- Modèle 5 (M5), les spécificités des deux tests sont égales à 1 ; des contraintes strictes sont appliquées aux sensibilités des deux tests et à la prévalence uniformément autour des probabilités des paramètres  $p$  décrits ci-dessus (cf.annexe 1)
- Modèle 6 (M6), les spécificités des deux tests sont égales à 1 ; des contraintes erronées sont appliquées à la sensibilité du test 1 (cf.annexe 1)
- Modèle 7 (M7), les spécificités des deux tests sont égales à 1 ; des contraintes erronées sont appliquées à la sensibilité du test 1 et à la probabilité conditionnelle d'obtenir un résultat positif au deuxième test si l'animal est infecté et a obtenu un résultat positif au premier test ( $p_4$ ) (cf.annexe 1).

### 3.1. Résultats

Tableau 8 : Résultats de l'analyse de la simulation décrite en annexe 1.

Modèle			$p$		$pr$		$Test 1$			$Test 2$	
	<i>Conv.</i>	<i>Bayesp</i>	DIC <sub><math>p</math></sub>	$P_D$ <sub><math>p</math></sub>	DIC <sub><math>pr</math></sub>	$P_D$ <sub><math>pr</math></sub>	<i>Prev</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>
1	non	0.4916	-90.177	-111.609	24.283	2.936	0.5253	0.3229	1	0.3957	1
2	oui	0.4930	8.303	-13.065	24.342	2.952	0.5688	0.5732	1	0.6680	1
3	oui	0.4796	24.176	2.873	24.279	2.917	0.5058	0.5956	1	0.6939	1
4	oui	0.1852	20.407	0.990	20.471	1.021	0.5006	0.6000	1	0.7000	1
5	oui	0.0004	18.426	0.000	18.445	0.007	0.5000	0.6000	1	0.7000	1
6	oui	0.7000	25.524	2.351	25.468	2.338	0.3850	0.8106	1	0.9049	1
7	oui	1.0000	355.977	1.406	355.899	1.382	0.3848	0.8103	1	0.5006	1

*Conv* = convergence ; *Bayesp* = Bayesian p-value ;  $p$  indique que les paramètres DIC et  $p_D$  sont calculés à partir des paramètres  $p$  du modèle ;  $pr$  indique que les paramètres DIC et  $p_D$  sont calculés à partir des probabilités  $pr$  de la distribution multinomiale.

### 3.2. Conclusions

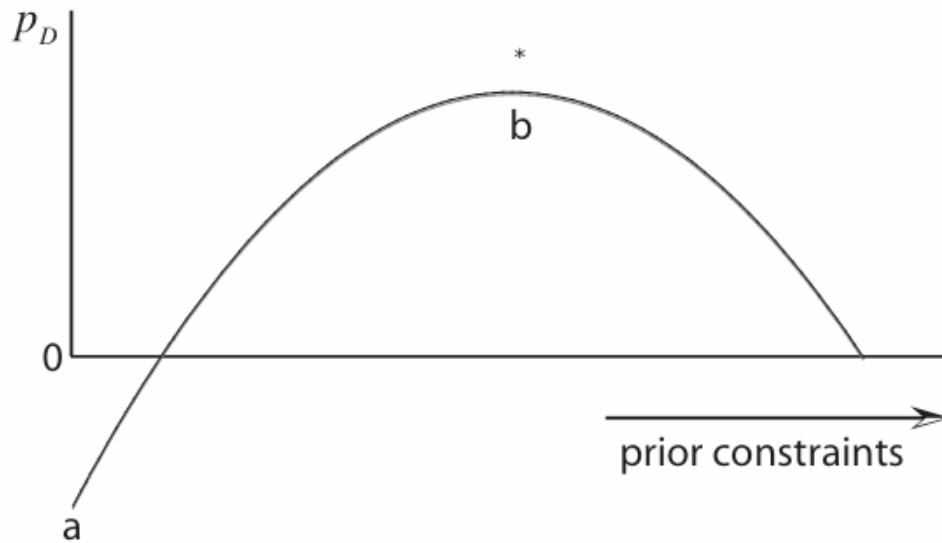
Le tableau 8 montre que :

- lorsqu'il n'y a pas de contraintes (informations *a priori* non informatives, modèle 1), le modèle est « sous spécifié » et non convergent ; les DIC et  $p_D$  calculés à partir des paramètres  $p$  du modèle prennent des valeurs négatives. Ceci s'explique par le fait que la vraisemblance n'est pas log-concave (Spiegelhalter *et al.*, 2002).
- plus on ajoute de contraintes autour des différents paramètres  $p$ , plus les DIC et  $p_D$  calculés à partir des paramètres  $p$  du modèle se rapprochent des valeurs des DIC et  $p_D$  calculés à partir des probabilités  $pr$  de la distribution multinomiale (modèle 2).
- dans le modèle 3, les valeurs des DIC et  $p_D$  calculés des deux manières différentes sont pratiquement égales; la prévalence et les caractéristiques des tests ont atteint leur valeur espérée. Les informations *a priori* semblent fournir juste assez de contraintes au modèle pour réduire le nombre de paramètres à estimer au nombre de paramètres estimables. Le modèle est correctement « spécifié ».
- les résultats de l'analyse du modèle 4 montrent une chute des valeurs de DIC et de  $p_D$ . Ceci indique que le modèle est « surspécifié », c'est-à-dire que « trop » de contraintes sont imposées autour des différents paramètres  $p$ . Les informations *a priori* fournissent plus d'informations que nécessaire au modèle pour réduire le nombre de

paramètres à estimer au nombre de paramètres estimables. La *Bayesp* tend vers 0 et indique dès lors que le modèle s'ajuste aux données de terrain ; la différence entre la déviance des observations ( $Dev_{obs}$ ) et la déviance des valeurs échantillonnées à partir de la distribution multinomiale ( $Dev_{échant}$ ) est négative dans la plupart des itérations.

- lorsque l'on impose des contraintes strictes autour de tous les paramètres  $p$  (modèle 5), les valeurs de DIC et  $p_D$  chutent encore davantage. La prévalence et les caractéristiques des tests gardent toujours leur valeur espérée. La *Bayesp* est pratiquement égale à 0 ; le modèle est parfaitement ajusté aux données de terrain.
- la valeur du paramètre  $p_D$  calculée à partir des paramètres  $p$  du modèle semble, dès lors, suivre une courbe quasi continue (même si une discontinuité ne peut pas être exclue) dont l'asymptote correspond à la valeur maximale et la plus proche du nombre de paramètres estimables, soit  $2^h - 1$  avec  $h$  le nombre de tests (cf. fig.3).

Figure 3 : Graphique de la valeur du paramètre  $p_D$  en fonction du nombre de contraintes imposées.



a : la valeur minimale de  $p_D$  calculée à partir des paramètres  $p$  du modèle et résultant de l'analyse du modèle intégrant les informations *a priori* non informatives ; b : la valeur de  $p_D$  résultant de l'analyse du modèle correctement spécifié (la valeur de  $p_D$  calculée à partir des probabilités  $pr$  est égale à la valeur de  $p_D$  calculée à partir des paramètres  $p$  du modèle; \*: nombre de paramètres estimables, soit  $2^h - 1$  avec  $h$  le nombre de tests

- si on impose des contraintes erronées (ne se situant pas autour des paramètres  $p$  du modèle), les valeurs de DIC et de  $p_D$  ne correspondent plus aux valeurs calculées à partir des probabilités  $pr$  de la distribution multinomiale sur base des informations a

priori non informatives et la *Bayesp* tend vers 1. La prévalence ne correspond plus à sa valeur espérée.

- L'analyse de la convergence ne semble pas permettre la validation du modèle. En effet, les modèles convergent avant que la prévalence et les caractéristiques des tests n'atteignent leur valeur espérée.

La marche à suivre pour valider un modèle statistique bayésien se résume donc dans le tableau 9.

Tableau 9: Marche à suivre pour la validation d'un modèle statistique bayésien intégrant plusieurs tests diagnostiques.

1.	L'analyse du modèle avec les informations <i>a priori</i> non informatives pour le calcul des valeurs des $p_D$ et DIC à partir des probabilités $pr$ de la distribution multinomiale.
2.	L'intégration des opinions d'experts, l'analyse du modèle et le calcul des différents critères de validation. La <i>Bayesp</i> sera de préférence proche de 0,5. La correspondance des valeurs des DIC et $p_D$ calculés selon les deux méthodes décrites ci-dessus est utilisée comme test de screening pour sélectionner la combinaison des informations <i>a priori</i> qui offrent la meilleure adéquation avec les données de terrain. L'utilisation des densités de probabilités des paramètres représente également un outil intéressant pour le screening. Cette étape est réalisée interactivement avec l'expert.
3.	La tendance de la <i>Bayesp</i> à tendre vers 0 lorsque l'on impose des contraintes strictes uniformément autour des moyennes postérieures (informations <i>a posteriori</i> ) des paramètres $p$ .

Si ces critères sont respectés, nous qualifierons le modèle de « **CONFORME** ».

## 4. DEUXIEME ETUDE : Etude du modèle à 4 tests

Après avoir déterminé les critères de validation, nous pouvons appliquer la marche à suivre établie aux données récoltées en Zambie. Dans cette étude, les informations *a priori* sont fournies par le Professeur Pierre Dorny (cf. tableau 10) et les résultats des tests sont ceux des 868 porcs zambiens décrits ci-dessus (Dorny *et al.*, 2004). Le modèle utilisé tient compte de la dépendance conditionnelle éventuelle entre les tests et comporte donc 31 paramètres à estimer (cf. annexe 2).

Tableau 10: Informations *a priori* appliquées au modèle à 4 tests diagnostiques.

Paramètres à estimer	Contraintes brutes
$p_1$	0 - 1
$p_2$	0 - 1
$p_3$	1
$p_4$	0.9 - 1
$p_5$	0 - 0.1
$p_6$	1
$p_7$	-----*
$p_8$	0 - 1
$p_9$	0 - 1
$p_{10}$	0 - 1
$p_{11}$	0 - 1
$p_{12}$	0.9 - 1
$p_{13}$	-----*
$p_{14}$	-----*
$p_{15}$	-----*
$p_{16}$	0 - 1
$p_{17}$	0 - 1
$p_{18}$	0 - 1
$p_{19}$	0 - 1
$p_{20}$	0 - 1
$p_{21}$	0 - 1
$p_{22}$	0 - 1
$p_{23}$	0 - 1
$p_{24}$	0.9 - 1
$p_{25}$	0 - 1
$p_{26}$	-----*
$p_{27}$	-----*
$p_{28}$	-----*
$p_{29}$	-----*
$p_{30}$	-----*
$p_{31}$	-----*

\*ces paramètres ne  
doivent pas être  
estimés car  $p_3 = 1$

**Etape 1:** Analyse du modèle avec les informations *a priori* non informatives

Tableau 11 : Valeurs de  $p_D$  et de DIC calculées à partir des probabilités  $pr$  de la distribution multinomiale en utilisant les informations *a priori* non informatives.

$p_D_{pr}$	DIC $_{pr}$
9.648238	71.44624

**Etape 2:** Analyse du modèle intégrant les informations *a priori* et calcul des critères de validation

Tableau 12 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse du modèle à 4 tests intégrant les informations *a priori*.

		<i>TONG</i>		<i>VISUAL</i>		<i>ELISA AG</i>		<i>ELISA AB</i>	
Combinaison	<i>Prev</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>
T-V-Ag-Ab	0.642 (0.54-0.91)	0.210 (0.14-0.26)	1.000	0.221 (0.15-0.27)	1.000	0.867 (0.62-0.98)	0.947 (0.90-0.99)	0.358 (0.26-0.41)	0.917 (0.85-0.99)

Tableau 13 : Critères de validation du modèle à 4 tests intégrant les informations *a priori*.

Combinaisons	<i>Bayesp</i>	$p_D_{pr}$	$p_D_p$	DIC $_{pr}$	DIC $_p$	<i>Conv.</i>
T-V-Ag-Ab	0.4949	9.925	8.307	70.296	68.678	oui

Le tableau 13 indique une *Bayesp* proche de 0.5. Les valeurs de DIC et de  $p_D$  ne sont pas identiques selon la méthode de calcul utilisée. Ce critère est alors utilisé pour sélectionner la combinaison des informations *a priori* qui offrent la meilleure adéquation avec les données de terrain (cf. tableau 14).

Les résultats de l'analyse et les valeurs de  $p_D$  et DIC obtenues sont reprises dans le tableau 15 et 16.

Tableau 14 : Informations *a priori* appliquées au modèle à 4 tests diagnostiques issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Paramètres à estimer	Contraintes ajustées
$p_1$	0 - 1
$p_2$	0 - 1
$p_3$	1
$p_4$	0 - 1
$p_5$	0 - 0.1
$p_6$	1
$p_7$	-----
$p_8$	0.8 - 1
$p_9$	0.8 - 1
$p_{10}$	0.8 - 1
$p_{11}$	0.8 - 1
$p_{12}$	0.9 - 1
$p_{13}$	-----
$p_{14}$	-----
$p_{15}$	-----
$p_{16}$	0 - 1
$p_{17}$	0 - 1
$p_{18}$	0 - 1
$p_{19}$	0 - 1
$p_{20}$	0 - 1
$p_{21}$	0 - 1
$p_{22}$	0 - 1
$p_{23}$	0 - 1
$p_{24}$	0.9 - 1
$p_{25}$	0.9 - 1
$p_{26}$	-----
$p_{27}$	-----
$p_{28}$	-----
$p_{29}$	-----
$p_{30}$	-----
$p_{31}$	-----



Tableau 15 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse du modèle à 4 tests intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

		<i>TONG</i>		<i>VISUAL</i>		<i>ELISA AG</i>		<i>ELISA AB</i>	
Combinaison	Prev	Se	Sp	Se	Sp	Se	Sp	Se	Sp
T-V-Ag-Ab	0.603 (0.54-0.67)	0.221 (0.18-0.27)	1.000	0.233 (0.19-0.28)	1.000	0.911 (0.84-0.98)	0.947 (0.90-0.99)	0.373 (0.32-0.43)	0.939 (0.90-0.99)

Tableau 16 : Critères de validation du modèle à 4 tests intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaison	<i>Bayesp</i>	$p_D\text{-}pr$	$p_D\text{-}p$	DIC <sub>pr</sub>	DIC <sub>p</sub>	conv.
T-V-Ag-Ab	0.5092	9.416	9.456	69.609	69.649	oui

Nous remarquons que les valeurs de DIC et  $p_D$  sont maintenant pratiquement égales dans les deux méthodes de calcul utilisées.

**Etape 3 :** Application de contraintes strictes autour des moyennes postérieures des différents paramètres  $p$

Des contraintes strictes sont imposées uniformément autour des moyennes postérieures des différents paramètres  $p$  (cf. tableau 17). Nous obtenons une *Bayesp* de 0,0146 qui tend vers 0 ; le modèle est donc « conforme ».

Tableau 17 : Moyennes postérieures des paramètres  $p$  et contraintes strictes imposées au modèle à 4 tests diagnostiques.

Paramètres à estimer	Moyennes postérieures des paramètres $p$	Contraintes strictes
$p_1$	0.60	0.59 - 0.61
$p_2$	0.22	0.21 - 0.23
$p_3$	1	1
$p_4$	0.97	0.96 - 0.98
$p_5$	0.02	0.01 - 0.03
$p_6$	1	1
$p_7$	-----	-----
$p_8$	0.96	0.95 - 0.97
$p_9$	0.87	0.86 - 0.88
$p_{10}$	0.93	0.92 - 0.94
$p_{11}$	0.90	0.89 - 0.91
$p_{12}$	0.95	0.94 - 0.96
$p_{13}$	-----	-----
$p_{14}$	-----	-----
$p_{15}$	-----	-----
$p_{16}$	0.68	0.67 - 0.69
$p_{17}$	0.40	0.39 - 0.41
$p_{18}$	0.25	0.24 - 0.26
$p_{19}$	0.34	0.33 - 0.35
$p_{20}$	0.45	0.44 - 0.46
$p_{21}$	0.51	0.50 - 0.52
$p_{22}$	0.26	0.25 - 0.27
$p_{23}$	0.54	0.53 - 0.55
$p_{24}$	0.94	0.93 - 0.95
$p_{25}$	0.95	0.94 - 0.96
$p_{26}$	-----	-----
$p_{27}$	-----	-----
$p_{28}$	-----	-----
$p_{29}$	-----	-----
$p_{30}$	-----	-----
$p_{31}$	-----	-----

## 5. TROISIEME ETUDE : Etude d'un modèle à 3 tests

Cette étude est réalisée dans le but de savoir si l'on peut se limiter à l'utilisation de 3 tests diagnostiques, au lieu des 4 décrits ci-dessus, tout en respectant les critères de validation, afin de diminuer les coûts du diagnostic à grande échelle.

Afin de déterminer quel(s) modèle(s) de combinaisons à 3 tests est(sont) conforme(s), nous devons considérer et analyser toutes les combinaisons de 3 tests possibles à partir des 4 tests (Tong, Visual, ELISA AG, ELISA Ab). Cette analyse nous permet une étude approfondie du modèle à 3 tests et de l'importance éventuelle de l'ordre des tests au sein de ce modèle.

Il existe 24 combinaisons de 3 tests possibles à partir des 4 tests diagnostiques. Celles-ci sont reprises dans le tableau 18 :

Tableau 18 : Combinaisons de 3 tests possibles à partir des 4 tests diagnostiques.

Combinaisons	
T-V-AG	V-AG-AB
T-AG-V	V-AB-AG
V-AG-T	AB-AG-V
V-T-AG	AG-AB-V
AG-V-T	AB-V-AG
AG-T-V	AG-V-AB
T-V-AB	T-AG-AB
T-AB-V	T-AB-AG
V-AB-T	AB-AG-T
V-T-AB	AG-AB-T
AB-T-V	AG-T-AB
AB-V-T	AB-T-AG

Un modèle à 3 tests est élaboré et adapté pour chacune de ces combinaisons (cf. annexe 3).

Les opinions de l'expert (Communications personnelles Dorny, 2004) sont récoltées dans le tableau 19.

Tableau 19 : Informations *a priori* appliquées aux 24 modèles à 3 tests diagnostiques (Communications personnelles, Dorny, 2004).

Combinaisons	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$	$p_9$	$p_{10}$	$p_{11}$	$p_{12}$	$p_{13}$	$p_{14}$
<b>T-V-AG</b>	0 - 0.5	1	0.9 - 1	0 - 0.1	1	-----	0.95 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	-----	-----
<b>T-AG-V</b>	0 - 0.5	1	0.95 - 1	0 - 1	0.9 - 1	-----	0.95 - 1	0 - 1	0.5 - 1	0 - 1	1	1	-----
<b>V-AG-T</b>	0 - 1	1	0 - 1	0 - 1	0.9 - 1	-----	0.5 - 1	0 - 0.1	0 - 0.1	0 - 1	1	1	-----
<b>V-T-AG</b>	0 - 1	1	0 - 1	0 - 0.1	1	-----	0.95 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	-----	-----
<b>AG-V-T</b>	0.9 - 1	0.9 - 1	0.2 - 1	0 - 0.5	1	1	0 - 1	0 - 1	0 - 0.5	0 - 0.1	1	-----	1
<b>AG-T-V</b>	0.9 - 1	0.9 - 1	0.2 - 1	0 - 0.1	1	1	0.9 - 1	0 - 1	0.5 - 1	0.5 - 1	1	-----	1
<b>T-V-AB</b>	0 - 0.5	1	0.9 - 1	0 - 0.1	1	-----	0.7 - 1	0.7 - 1	0.7 - 1	0 - 1	0.9 - 1	-----	-----
<b>T-AB-V</b>	0 - 0.5	1	0.7 - 1	0 - 1	0.9 - 1	-----	0.9 - 1	0 - 1	0 - 0.5	0 - 1	1	1	-----
<b>V-AB-T</b>	0 - 1	1	0 - 1	0 - 1	0.9 - 1	-----	0.1 - 1	0 - 1	0 - 1	0 - 1	1	1	-----
<b>V-T-AB</b>	0 - 1	1	0 - 1	0 - 0.1	1	-----	0.7 - 1	0.7 - 1	0.7 - 1	0 - 1	0.9 - 1	-----	-----
<b>AB-T-V</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	1	1	0.9 - 1	0 - 1	0.9 - 1	0 - 1	1	-----	1
<b>AB-V-T</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	1	1	0 - 1	0 - 1	0 - 1	0 - 1	1	-----	1
<b>V-AG-AB</b>	0 - 1	1	0 - 1	0 - 1	0.9 - 1	-----	0.7 - 1	0 - 1	0.5 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>V-AB-AG</b>	0 - 1	1	0 - 1	0 - 1	0.9 - 1	-----	0 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>AB-AG-V</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	0.5 - 1	0 - 1	0.5 - 1	0 - 1	1	1	1
<b>AG-AB-V</b>	0.9 - 1	0.9 - 1	0.5 - 1	0 - 1	0.9 - 1	0.9 - 1	0.5 - 1	0.5 - 1	0 - 1	0 - 1	1	1	1
<b>AB-V-AG</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	1	1	0.9 - 1	0.9 - 1	0.9 - 1	0.7 - 1	0.9 - 1	-----	0.9 - 1
<b>AG-V-AB</b>	0.9 - 1	0.9 - 1	0.2 - 1	0 - 0.5	1	1	0.5 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	-----	0.9 - 1
<b>T-AG-AB</b>	0 - 0.5	1	0.95 - 1	0 - 1	0.9 - 1	-----	0.7 - 1	0 - 1	0.7 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>T-AB-AG</b>	0 - 1	1	0 - 1	0 - 1	0.9 - 1	-----	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	0.9 - 1	0.9 - 1	-----
<b>AB-AG-T</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0 - 1	0 - 1	1	1	1
<b>AG-AB-T</b>	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0 - 1	0 - 1	1	1	1
<b>AG-T-AB</b>	0.9 - 1	0.9 - 1	0.2 - 1	0 - 0.1	1	1	0.7 - 1	0.7 - 1	0 - 1	0 - 1	0.9 - 1	-----	0.9 - 1
<b>AB-T-AG</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	1	1	0.9 - 1	0.9 - 1	0.7 - 1	0 - 1	0.9 - 1	-----	0.9 - 1

La marche à suivre instituée dans la première étude est appliquée :

Les modèles sont d'abord analysés en utilisant les informations *a priori* non informatives (étape 1), puis en intégrant les informations *a priori* (cf. tableau 19). La correspondance des valeurs des DIC et  $p_D$  calculés selon les deux méthodes décrites ci-dessus est utilisée comme test de screening pour sélectionner la combinaison des informations *a priori* qui offrent la meilleure adéquation avec les données de terrain (cf. tableau 20) (étape 2).

Des contraintes strictes, autour des informations *a posteriori* issues de l'analyse du modèle intégrant les contraintes ajustées, sont ensuite imposées et la *Bayesp* est calculée (étape 3). Afin d'illustrer l'importance et la puissance de cette étape, des contraintes strictes sont également appliquées autour des informations *a posteriori* issues de l'analyse du modèle intégrant les contraintes brutes.

Tableau 20 : Informations *a priori* appliquées aux 24 modèles à 3 tests diagnostiques. Celles-ci sont issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaisons	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$	$p_9$	$p_{10}$	$p_{11}$	$p_{12}$	$p_{13}$	$p_{14}$
<b>T-V-AG</b>	0 - 1	1	0 - 1	0 - 0.1	1	-----	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	-----	-----
<b>T-AG-V</b>	0 - 0.5	1	0.95 - 1	0 - 1	0.9 - 1	-----	0.95 - 1	0 - 1	0.5 - 1	0 - 1	1	1	-----
<b>V-AG-T</b>	0 - 1	1	0.8 - 1	0.8 - 1	0.8 - 1	-----	0.5 - 1	0 - 1	0 - 0.1	0 - 0.1	1	1	-----
<b>V-T-AG</b>	0 - 1	1	0 - 1	0 - 0.1	1	-----	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	0.9 - 1	-----	-----
<b>AG-V-T</b>	0.8 - 1	0.8 - 1	0.2 - 1	0 - 0.5	1	1	0 - 1	0 - 1	0 - 1	0 - 0.1	1	-----	1
<b>AG-T-V</b>	0.8 - 1	0.8 - 1	0 - 1	0 - 1	1	1	0.9 - 1	0 - 1	0 - 1	0.5 - 1	1	-----	1
<b>T-V-AB</b>	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>T-AB-V</b>	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>V-AB-T</b>	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>V-T-AB</b>	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>AB-T-V</b>	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>AB-V-T</b>	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>V-AG-AB</b>	0 - 1	1	0.8 - 1	0.8 - 1	0.8 - 1	-----	0 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>V-AB-AG</b>	0 - 1	1	0 - 1	0 - 1	0.9 - 1	-----	0.8 - 1	0.8 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>AB-AG-V</b>	0 - 1	0.9 - 1	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	0 - 1	0 - 0.1	0 - 1	0 - 0.1	1	1	1
<b>AG-AB-V</b>	0.8 - 1	0.8 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0 - 0.1	0 - 0.1	1	1	1
<b>AB-V-AG</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	1	1	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	0.8 - 1	-----	0.8 - 1
<b>AG-V-AB</b>	0.8 - 1	0.8 - 1	0.2 - 1	0 - 0.1	1	1	0.5 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	-----	0.9 - 1
<b>T-AG-AB</b>	0 - 0.5	1	0.95 - 1	0 - 1	0.8 - 1	-----	0.7 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>T-AB-AG</b>	0 - 0.5	1	0.5 - 1	0 - 1	0.5 - 1	-----	0.5 - 1	0.5 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	-----
<b>AB-AG-T</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0 - 1	0 - 1	1	1	1
<b>AG-AB-T</b>	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0.9 - 1	0.9 - 1	0 - 1	0 - 1	0 - 1	0 - 1	1	1	1
<b>AG-T-AB</b>	0.9 - 1	0.9 - 1	0.2 - 1	0 - 0.1	1	1	0.7 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	-----	0.9 - 1
<b>AB-T-AG</b>	0 - 1	0.9 - 1	0 - 1	0 - 1	1	1	0.9 - 1	0 - 1	0 - 1	0 - 1	0.9 - 1	-----	0.9 - 1

## 5.1. Résultats et conclusions

### Etape 1 :

Tableau 21 : Valeurs de  $p_D$  et de DIC calculées à partir des probabilités  $pr$  de la distribution multinomiale en utilisant les informations *a priori* non informatives.

Combinaisons	$p_D_{pr}$	DIC $_{pr}$
T-V-AG	4.801	38.510
T-AG-V	4.855634	38.83363
V-AG-T	4.895368	38.77237
V-T-AG	4.894125	38.60013
AG-V-T	4.849424	38.68442
AG-T-V	4.932984	38.83698
T-V-AB	5.387	41.322
T-AB-V	5.180049	42.04105
V-AB-T	5.32866	42.21066
V-T-AB	5.3323	41.2213
AB-T-V	5.481708	42.36571
AB-V-T	5.358716	42.24872
V-AG-AB	6.107	47.746
V-AB-AG	5.95118	48.54918
AB-AG-V	5.902142	48.67714
AG-AB-V	6.072613	48.70761
AB-V-AG	5.888445	48.63744
AG-V-AB	6.08677	47.82477
T-AG-AB	6.088	48.212
T-AB-AG	5.987694	48.82369
AB-AG-T	6.071	49.091
AG-AB-T	6.072607	48.92261
AG-T-AB	6.072646	48.16165
AB-T-AG	6.210327	49.17533

Lorsque nous utilisons les informations *a priori* non informatives, l'ordre des tests ne semble pas influencer les valeurs des paramètres  $p_D$  et DIC calculées à partir des probabilités  $pr$  de la distribution multinomiale.

## Etape 2 :

Tableau 22 : Caractéristiques des tests et prévalence issues de l'analyse des modèles à 3 tests intégrant les informations *a priori*.

Combinaisons	$p_1$	$SeT$	$SeV$	$SeAg$	$SeAb$	$SpT$	$SpV$	$SpAg$	$SpAb$
<b>T-V-AG</b>	0.7054	0.1939	0.204	0.804	-	1	1	0.9478	-
<b>T-AG-V</b>	0.5697	0.2369	0.2489	0.9167	-	1	1	0.8849	-
<b>V-AG-T</b>	0.6738	0.201	0.2118	0.8341	-	1	1	0.9476	-
<b>V-T-AG</b>	0.6906	0.198	0.2081	0.8186	-	1	1	0.9475	-
<b>AG-V-T</b>	0.5735	0.2322	0.2457	0.9551	-	1	1	0.947	-
<b>AG-T-V</b>	0.5323	0.2518	0.2643	0.9662	-	1	1	0.8841	-
<b>T-V-AB</b>	0.3828	0.3683	0.3867	-	0.5567	1	1	-	0.9424
<b>T-AB-V</b>	0.4023	0.3507	0.3655	-	0.458	1	1	-	0.8241
<b>V-AB-T</b>	0.3592	0.3901	0.4128	-	0.6055	1	1	-	0.9391
<b>V-T-AB</b>	0.3478	0.4141	0.4354	-	0.6368	1	1	-	0.9405
<b>AB-T-V</b>	0.4056	0.3546	0.369	-	0.5575	1	1	-	0.942
<b>AB-V-T</b>	0.1997	0.7288	0.7688	-	0.5797	1	1	-	0.819
<b>V-AG-AB</b>	0.6054	-	0.2353	0.8718	0.3764	-	1	0.8845	0.9471
<b>V-AB-AG</b>	0.7284	-	0.1962	0.7752	0.3272	-	1	0.9472	0.9485
<b>AB-AG-V</b>	0.6138	-	0.2333	0.9019	0.3635	-	1	0.9462	0.9343
<b>AG-AB-V</b>	0.5698	-	0.2483	0.9603	0.372	-	1	0.9466	0.9226
<b>AB-V-AG</b>	0.6899	-	0.0.2102	0.7787	0.343	-	1	0.8679	0.9449
<b>AG-V-AB</b>	0.5596	-	0.2531	0.93	0.3699	-	1	0.8876	0.9105
<b>T-AG-AB</b>	0.6048	0.2232	-	0.8738	0.3766	1	-	0.8866	0.9423
<b>T-AB-AG</b>	0.7231	0.1888	-	0.7829	0.3301	1	-	0.9477	0.9468
<b>AB-AG-T</b>	0.6111	0.2227	-	0.9039	0.3643	1	-	0.946	0.934
<b>AG-AB-T</b>	0.5729	0.235	-	0.9558	0.3712	1	-	0.9468	0.9235
<b>AG-T-AB</b>	0.5584	0.2416	-	0.929	0.39	1	-	0.8854	0.9328
<b>AB-T-AG</b>	0.7162	0.191	-	0.7879	0.3316	1	-	0.9476	0.9465

Tableau 23 : Critères de validation des modèles à 3 tests diagnostiques intégrant les informations *a priori*.

Combinaisons	<i>Bayesp</i>	$p_{D\_pr}$	$p_{D\_p}$	DIC <sub>pr</sub>	DIC <sub>p</sub>	conv
<b>T-V-AG</b>	0,4901	4,9997	3,9720	37,216	36,163	oui
<b>T-AG-V</b>	0,4868	5,0375	-4,5190	37,541	26,925	oui
<b>V-AG-T</b>	0,4894	5,1263	4,3600	37,554	36,789	oui
<b>V-T-AG</b>	0,4895	4,9379	4,1100	37,158	36,329	oui
<b>AG-V-T</b>	0,4916	5,0260	4,6680	37,147	36,789	oui
<b>AG-T-V</b>	0,4999	5,1088	5,0320	37,437	37,360	oui
<b>T-V-AB</b>	0,4851	5,5372	-8,6850	40,550	26,328	oui
<b>T-AB-V</b>	0,4327	5,4738	-16,8870	41,201	18,840	oui
<b>V-AB-T</b>	0,4856	5,6018	1,0890	40,545	36,031	oui
<b>V-T-AB</b>	0,4738	5,5698	3,3090	40,707	38,446	oui
<b>AB-T-V</b>	0,4826	5,1677	2,3370	40,346	37,515	oui
<b>AB-V-T</b>	0,4732	5,6144	-5,4030	40,802	29,789	oui
<b>V-AG-AB</b>	0,4566	6,3274	3,9050	47,842	45,421	oui
<b>V-AB-AG</b>	0,4904	6,1110	5,3880	48,041	47,318	oui
<b>AB-AG-V</b>	0,4839	6,1339	-2,3730	47,963	39,456	oui
<b>AG-AB-V</b>	0,4786	6,2617	3,4250	48,038	45,201	oui
<b>AB-V-AG</b>	0,4945	5,5670	5,5680	47,552	47,552	oui
<b>AG-V-AB</b>	0,4525	6,2157	-0,2680	47,648	41,164	oui
<b>T-AG-AB</b>	0,4963	5,2804	3,1560	47,581	45,457	oui
<b>T-AB-AG</b>	0,4932	6,3981	2,4220	48,621	44,645	oui
<b>AB-AG-T</b>	0,4902	6,3925	-0,9160	48,560	41,252	oui
<b>AG-AB-T</b>	0,4864	6,2181	3,6320	48,309	45,723	oui
<b>AG-T-AB</b>	0,4692	5,1073	5,1410	47,702	47,736	oui
<b>AB-T-AG</b>	0,4892	6,3050	5,3780	48,475	47,547	oui

Les combinaisons T-Ag-V, T-V-Ab, T-Ab-V, Ab-V-T, Ab-Ag-V, Ag-V-Ab et Ab-Ag-T présentent des valeurs négatives de  $p_D$  calculées à partir des paramètres  $p$  des modèles.

Nous observons que les DIC et  $p_D$  de la combinaison Ab-V-Ag sont identiques dans les deux méthodes utilisées (en rouge). Certaines combinaisons semblent donc être plus faciles à spécifier que d'autres. Les combinaisons optimales d'informations *a priori* sont sélectionnées à l'aide du screening réalisé à partir des paramètres  $p_D$  et DIC. Les résultats de ces analyses sont décrits dans les tableaux 24 et 25.



Tableau 24 : Caractéristiques des tests et prévalence issues de l'analyse des modèles à 3 tests diagnostiques intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaisons	$p_1$	$SeT$	$SeV$	$SeAg$	$SeAb$	$SpT$	$SpV$	$SpAg$	$SpAb$
<b>T-V-AG</b>	0.5999	0.2229	0.2346	0.9167	-	1	1	0.9477	-
<b>T-AG-V</b>	0.5827	0.2293	0.241	0.9398	-	1	1	0.9466	-
<b>V-AG-T</b>	0.5961	0.2229	0.2351	0.9212	-	1	1	0.9474	-
<b>V-T-AG</b>	0.6004	0.2226	0.2342	0.916	-	1	1	0.9477	-
<b>AG-V-T</b>	0.6103	0.2299	0.2187	0.9019	-	1	1	0.9475	-
<b>AG-T-V</b>	0.6178	0.2149	0.2274	0.8932	-	1	1	0.9483	-
<b>T-V-AB</b>	-	-	-	-	-	-	-	-	-
<b>T-AB-V</b>	-	-	-	-	-	-	-	-	-
<b>V-AB-T</b>	-	-	-	-	-	-	-	-	-
<b>V-T-AB</b>	-	-	-	-	-	-	-	-	-
<b>AB-T-V</b>	-	-	-	-	-	-	-	-	-
<b>AB-V-T</b>	-	-	-	-	-	-	-	-	-
<b>V-AG-AB</b>	0.6017	-	0.2332	0.9129	0.3719	-	1	0.9475	0.9411
<b>V-AB-AG</b>	0.6041	-	0.2321	0.9095	0.3601	-	1	0.9472	0.9236
<b>AB-AG-V</b>	0.6252	-	0.2254	0.8807	0.3628	-	1	0.9444	0.941
<b>AG-AB-V</b>	0.6262	-	0.2248	0.8814	0.3644	-	1	0.9473	0.9434
<b>AB-V-AG</b>	0.6027	-	0.2336	0.9107	0.3609	-	1	0.9463	0.9243
<b>AG-V-AB</b>	0.6145	-	0.2278	0.8953	0.3713	-	1	0.9464	0.9445
<b>T-AG-AB</b>	0.6055	0.2201	-	0.9077	0.371	1	-	0.9471	0.9397
<b>T-AB-AG</b>	0.6062	0.2204	-	0.906	0.3599	1	-	0.9459	0.9247
<b>AB-AG-T</b>	0.6321	0.2111	-	0.872	0.3598	1	-	0.9443	0.9415
<b>AG-AB-T</b>	0.6325	0.2111	-	0.8732	0.3599	1	-	0.9472	0.9419
<b>AG-T-AB</b>	0.6208	0.2143	-	0.8877	0.3689	1	-	0.9469	0.9452
<b>AB-T-AG</b>	0.6072	0.2208	-	0.9057	0.3596	1	-	0.947	0.9243

Tableau 25 : Critères de validation des modèles à 3 tests diagnostiques intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaisons	$Bayesp$	$p_{D\_pr}$	$p_{D\_p}$	DIC <sub>pr</sub>	DIC <sub>p</sub>	conv
<b>T-V-AG</b>	0.5442	4.79198	4.749	37.3672	37.324	oui
<b>T-AG-V</b>	0.514	5.022743	4.663	37.26774	36.908	oui
<b>V-AG-T</b>	0.5182	5.147474	4.990	37.57347	37.416	oui
<b>V-T-AG</b>	0.5525	4.792294	4.855	37.49029	37.553	oui
<b>AG-V-T</b>	0.4993	5.057101	4.832	37.2791	37.054	oui
<b>AG-T-V</b>	0.5308	5.195142	4.859	37.94114	37.605	oui
<b>T-V-AB</b>	0.5116	5.65773	3.451	40.63073	38.425	oui
<b>T-AB-V</b>	-	-	-	-	-	oui
<b>V-AB-T</b>	-	-	-	-	-	oui
<b>V-T-AB</b>	-	-	-	-	-	oui
<b>AB-T-V</b>	-	-	-	-	-	oui
<b>AB-V-T</b>	-	-	-	-	-	oui
<b>V-AG-AB</b>	0.464	6.318029	6.320	47.80903	47.811	oui
<b>V-AB-AG</b>	0.4726	5.77225	5.853	47.48125	47.561	oui
<b>AB-AG-V</b>	0.414	6.061337	5.947	46.94834	46.833	oui
<b>AG-AB-V</b>	0.4284	6.285076	6.152	47.46608	47.333	oui
<b>AB-V-AG</b>	0.4772	5.869478	5.951	47.72948	47.811	oui
<b>AG-V-AB</b>	0.4072	5.942395	5.918	46.85839	46.834	oui
<b>T-AG-AB</b>	0.4607	6.347439	6.449	48.28244	48.384	oui
<b>T-AB-AG</b>	0.4756	6.080623	5.967	48.03462	47.921	oui
<b>AB-AG-T</b>	0.4148	5.817279	5.938	47.00628	47.127	oui
<b>AG-AB-T</b>	0.4306	6.356913	6.182	47.81691	47.642	oui
<b>AG-T-AB</b>	0.4496	6.383764	6.267	48.05876	47.942	oui
<b>AB-T-AG</b>	0.4714	5.997656	5.911	47.89666	47.810	oui

Le screening ne permet pas l'obtention de combinaisons d'informations *a priori* optimales pour les combinaisons composées exclusivement des tests T, V et Ab. Ceci est dû au manque d'informations *a priori* disponibles afin d'appliquer des contraintes suffisantes pour réduire le nombre de paramètres à estimer au nombre de paramètres estimables. Les autres combinaisons sont ajustables et permettent une estimation assez précise de la prévalence et des paramètres des tests.

### Etape 3

Les valeurs de la *Bayesp* calculées sans et avec contraintes strictes autour des moyennes postérieures des paramètres  $p$  issues des modèles intégrant les informations *a priori* brutes se trouvent dans le tableau 26 :

Tableau 26 : Résultats de l'analyse de la *Bayesp* sans et avec contraintes strictes des modèles à 3 tests diagnostiques intégrant les informations *a priori*.

Combinaisons	<i>Bayesp</i> sans contrainte stricte	<i>Bayesp</i> avec contraintes strictes
<b>T-V-AG</b>	0.4901	0.171
<b>T-AG-V</b>	> 0.9	-
<b>V-AG-T</b>	> 0.9	-
<b>V-T-AG</b>	0.4895	0,1708
<b>AG-V-T</b>	> 0.9	-
<b>AG-T-V</b>	> 0.9	-
<b>T-V-AB</b>	> 0.9	-
<b>T-AB-V</b>	0.6891	0.8346
<b>V-AB-T</b>	0.5608	0.6036
<b>V-T-AB</b>	> 0.9	-
<b>AB-T-V</b>	0.5174	0.2952
<b>AB-V-T</b>	0.5618	0.2728
<b>V-AG-AB</b>	1	-
<b>V-AB-AG</b>	0.4904	0.0216
<b>AB-AG-V</b>	> 0.9	-
<b>AG-AB-V</b>	1	-
<b>AB-V-AG</b>	0.4945	0.0248
<b>AG-V-AB</b>	0.4434	0.0224
<b>T-AG-AB</b>	1	-
<b>T-AB-AG</b>	1	-
<b>AB-AG-T</b>	0.4902	0.6504
<b>AG-AB-T</b>	0.4864	0.1026
<b>AG-T-AB</b>	1	-
<b>AB-T-AG</b>	0.5353	0.0574

Lorsque nous n'utilisons pas de contrainte stricte, nous observons que certaines combinaisons (en rouge) atteignent directement une *Bayesp* située proche de 0.5 quand d'autres tendent vers 1. Lorsque nous les appliquons, la *Bayesp* des mêmes combinaisons tend vers 0 tandis que les autres tendent vers 1. Cette analyse préliminaire nous donne déjà une idée des modèles correctement et mal spécifiés.

L'application de contraintes strictes sur les modèles intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC (cf. tableau 27) montre un rapprochement sensible de la *Bayesp* vers 0 pour toutes les combinaisons.

Tableau 27 : Résultats de l'analyse de la *Bayesp* avec contraintes strictes des modèles à 3 tests diagnostiques intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaisons	<i>Bayesp</i>	Combinaisons	<i>Bayesp</i>
V-T-AG	0.1686	V-AG-AB	0.0046
AG-T-V	0.1324	AG-AB-V	0.0032
T-AG-V	0.1282	AG-AB-T	0.0032
AG-V-T	0.1152	AB-AG-V	0.003
V-AG-T	0.1038	AG-V-AB	0.0028
T-V-AG	0.0966	AB-AG-T	0.002
T-AB-AG	0.0264	T-V-AB	-
AB-T-AG	0.026	T-AB-V	-
V-AB-AG	0.0186	V-AB-T	-
AB-V-AG	0.0116	V-T-AB	-
AG-T-AB	0.0074	AB-T-V	-
T-AG-AB	0.005	AB-V-T	-

L'ordre des tests dans le modèle ne semble pas avoir d'influence sur leur analyse statistique. Par contre, le « poids » des contraintes appliquées sur les paramètres à estimer semble varier selon l'ordre des tests et sont, dès lors, plus ou moins « puissantes » selon leur place dans le modèle.

## 6. QUATRIEME ETUDE : Etude des paramètres d'un modèle à 2 tests

Dans cette partie, nous allons étudier la possibilité de réduire le nombre de tests à deux, soit, premièrement et systématiquement l'inspection visuelle de la carcasse ou la palpation de la base de langue, suivis, si ceux-ci sont négatifs, d'un test ELISA Antigène.

En effet, l'inspection de la carcasse et la palpation de la langue possèdent une spécificité élevée mais une sensibilité assez faible. Nous ne pouvons dès lors pas nous satisfaire de ces tests pour affirmer que les animaux négatifs soient de vrais négatifs. Nous devons donc joindre un test possédant une sensibilité plus élevée ; ce qui est le cas du test ELISA Antigène. Les combinaisons étudiées sont « V, et Ag si V- » et « T, et Ag si T- ».

A nouveau, la diminution du nombre de tests permettra une diminution du coût du diagnostic. Un modèle à deux tests est élaboré (cf. annexe 4) ; celui-ci tient compte du fait qu'un animal fournissant un résultat positif au premier test ne sera pas testé en ELISA. La marche à suivre instituée dans la première étude est appliquée.

Les modèles intégrant les informations *a priori* non informatives sont, tout d'abord, analysés (**Etape 1**). Ils sont ensuite testés en intégrant les informations *a priori* (**Etape 2**) (Communications personnelles Dorny, 2004 ; cf. tableau 28); la correspondance des valeurs des DIC et  $p_D$  calculés selon les deux méthodes décrites ci-dessus est utilisée comme test de screening pour sélectionner la combinaison des informations *a priori* qui offrent la meilleure adéquation avec les données de terrain. Les contraintes strictes sont, enfin, appliquées (**Etape 3**) (cf. tableau 28).

Tableau 28 : Informations *a priori* appliquées au modèle à 2 tests diagnostiques.

Paramètres à estimer	Informations <i>a priori</i>		Informations <i>a priori</i> issues du screening		Contraintes strictes	
	V et Ag si V-	T et Ag si T-	V et Ag si V-	T et Ag si T-	V et Ag si V-	T et Ag si T-
$p_1$	0 - 1	0 - 1	0 - 1	0 - 1	0.59 - 0.61	0.6 - 0.61
$p_2$	0 - 1	0 - 0.5	0 - 1	0 - 0.5	0.22 - 0.24	0.21 - 0.23
$p_3$	1	1	1	1	1	1
$p_4$	0.9 - 1	0.9 - 1	0.8 - 1	0.8 - 1	0.89 - 0.91	0.89 - 0.91
$p_5$	0 - 1	0 - 1	0.8 - 1	0.8 - 1	0.89 - 0.91	0.89 - 0.91
$p_6$	0.9 - 1	0.9 - 1	0.9 - 1	0.9 - 1	0.94 - 0.96	0.94 - 0.96

## 6.1. Résultats

### Etape 1 :

Tableau 29 : Valeurs de  $p_D$  et de DIC calculées à partir des probabilités  $pr$  de la distribution multinomiale en utilisant les informations *a priori* non informatives.

Combinaisons	$p_D_{pr}$	DIC $_{pr}$
V et Ag si V-	1.912	17.382
T et Ag si T-	1.924	17.367

### Etape 2 :

Tableau 30 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse des modèles à 2 tests diagnostiques intégrant les informations *a priori*.

		<i>TONG</i>		<i>VISUAL</i>		<i>ELISA AG</i>		<i>ELISA AB</i>	
Combinaisons	<i>Prev</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>	<i>Se</i>	<i>Sp</i>
V et Ag si V-	0.723 (0.54-0.98)	-	-	0.200 (0.14-0.27)	1.000	0.784 (0.57-0.98)	0.948 (0.90-1)	-	-
T et Ag si T-	0.720 (0.54-0.98)	0.189 (0.13-0.26)	1.000	-	-	0.790 (0.58-0.98)	0.947 (0.90-1)	-	-

Tableau 31 : Critères de validation des modèles à 2 tests diagnostiques intégrant les informations *a priori*.

Combinaisons	<i>Bayesp</i>	$p_D_{pr}$	$p_D_p$	DIC $_{pr}$	DIC $_p$	<i>conv</i>
V et Ag si V-	0.497	1.983	0.897	17.528	16.443	oui
T et Ag si T-	0.497	2.167	0.932	17.689	16.454	oui

Tableau 32 : Caractéristiques des tests, prévalence et intervalles de crédibilité (95%) issus de l'analyse des modèles à 2 tests diagnostiques intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaisons	Prev	TONG		VISUAL		ELISA AG		ELISA AB	
		Se	Sp	Se	Sp	Se	Sp	Se	Sp
V et Ag si V-	0.603 (0.53-0.68)	-	-	0.233 (0.19-0.28)	1.000	0.898 (0.82-0.98)	0.948 (0.90-1)	-	-
T et Ag si T-	0.603 (0.53-0.68)	0.22 (0.18-0.28)	1.000	-	-	0.898 (0.82-0.98)	0.948 (0.90-1)	-	-

Tableau 33 : Critères de validation des modèles à 2 tests diagnostiques intégrant les informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC.

Combinaisons	Bayesp	$p_{D\_pr}$	$p_{D\_p}$	DIC <sub>pr</sub>	DIC <sub>p</sub>	Conv.
V et Ag si V-	0.497	1.999	1.992	17.555	17.547	oui
T et Ag si T-	0.498	1.994	1.993	17.512	17.513	oui

## 6.2. Conclusions

Lorsque nous intégrons les informations *a priori*, la *Bayesp* se situe bien, pour les deux combinaisons, autour de 0.5 mais les valeurs de DIC et de  $p_D$  ne sont pas identiques selon la méthode de calcul utilisée. En revanche, l'application des informations *a priori* issues du screening réalisé sur base des paramètres  $p_D$  et DIC permet cette égalité.

L'application de contraintes strictes fournit une valeur de *Bayesp* de 0.1062 pour la combinaison comprenant le test visuel et 0.1098 pour la combinaison comprenant la palpation de la langue.

Ces deux modèles sont donc conformes.

## 7. DISCUSSION

Ce travail nous a permis d'établir les critères de validation d'un modèle statistique bayésien intégrant les résultats de plusieurs tests diagnostiques, soit :

- la correspondance entre les paramètres DIC et  $p_D$  calculés à partir des probabilités  $pr$  de la distribution multinomiale et à partir des paramètres  $p$  du modèle. Le calcul des valeurs des DIC et  $p_D$  calculés à partir des probabilités  $pr$  de la distribution multinomiale sans utiliser de contraintes (utilisation des informations *a priori* non informatives) nous fournit directement les valeurs attendues de DIC et de  $p_D$ . La comparaison de ces valeurs avec celles obtenues à partir des paramètres  $p$  du modèle permet la validation de l'analyse en cas d'égalité. Celles-ci ne fournissent, en effet, les valeurs correctes de DIC et  $p_D$  que quand le modèle est conforme.
- la tendance de la *Bayesp* à tendre vers 0 lorsque l'on impose des contraintes strictes autour des valeurs vraies des différents paramètres, signifiant ainsi le bon ajustement du modèle aux données de terrain.

Ces deux critères réunis permettent d'établir la « conformité » du modèle étudié et d'estimer avec précision la prévalence et les caractéristiques des tests diagnostiques utilisés.

L'analyse de la convergence n'a pu fournir aucune information quant à la « conformité » des modèles. En effet, les modèles convergent avant que la prévalence et les caractéristiques des tests n'atteignent leur valeur espérée. Ce critère n'est donc que peu utilisable.

L'analyse des combinaisons de 4, 3 et 2 tests fournissent des valeurs de prévalence et de caractéristiques des tests équivalentes. La prévalence se situe autour de 0.6, soit trois fois plus que la prévalence apparente estimée au préalable dans cette région (Phiri *et al.*, 2003) ; les sensibilités des tests T, V, ELISA AG et AB autour de 0.2, 0.2, 0.9 et 0.35 respectivement ; les spécificités des tests ELISA AG et AB autour de 0.9 et 0.95 respectivement. Ces résultats sont en accord avec une étude réalisée en parallèle sur 65 porcs issus de la même population. Ces porcs ont été disséqués à la recherche de cysticerques et soumis au 4 tests diagnostiques étudiés ici (Dorny *et al.*, 2004).

L'établissement d'un diagnostic de la cysticercose porcine par l'application de deux tests diagnostiques, « V et Ag si V - » ou « T et Ag si T - », semble donc fournir une estimation



acceptable de la prévalence de la maladie et des différents paramètres de ces derniers. Ces séries de tests sont toutes deux intéressantes; la première, si les animaux sont systématiquement abattus en abattoirs, la deuxième si l'on désire pratiquer un diagnostic ante-mortem.

Nous démontrons ici que la limitation du diagnostic de la cysticercose aux seuls tests d'inspection de la langue ou de la carcasse ne suffit pas pour obtenir une estimation acceptable de la prévalence réelle de la maladie, et que quatre tests ne sont pas nécessaires. L'approche bayésienne nous a permis d'obtenir un diagnostic de qualité tout en minimisant ses coûts.

L'étude des combinaisons de 3 tests montre l'importance des tests intégrés dans le modèle. Seuls les modèles intégrant le test ELISA AG, l'inspection de la carcasse et/ou la palpation de la langue sont identifiables. L'ELISA AG possède, en effet, une sensibilité et une spécificité élevées tandis que T et V ont une spécificité de 100%. Ces valeurs élevées permettent d'imposer des contraintes suffisantes et nécessaires à la « conformité ». Les modèles ne comportant pas l'ELISA AG et V et/ou T sont systématiquement non-conformes ; ceci pouvant s'expliquer par le manque de contraintes disponibles en utilisant le test ELISA AB. L'ordre des tests dans le modèle ne semble pas influencer leur analyse statistique. Néanmoins, le réarrangement de la place des tests a permis :

- de présenter les questions posées à l'expert de manière alternative et a permis à celui-ci de fournir des informations *a priori* très précises.
- de constater que la valeur des contraintes appliquées sur les paramètres à estimer varie selon l'ordre des tests et seront dès lors plus ou moins « puissante » selon leur place dans le modèle. Cette hypothèse nécessiterait une confirmation par la réalisation de simulations.

De plus, la synthèse des informations *a priori* nous a confronté à l'importance d'établir les conditions précises de reconnaissance d'un animal comme positif ou négatif. La découverte de cysticerques calcifiés entraîne-t-elle la positivité de l'animal? Un animal possédant des anticorps dirigés contre le parasite est-il considéré comme positif si les autres tests sont négatifs? Le test ELISA AB peut-il être utilisé sans son homologue permettant la détection d'antigènes circulants lorsqu'il s'agit d'estimer la prévalence de la maladie au sein d'une population ?

Les données récoltées sur le terrain exigent des informations *a priori* précises et non contradictoires avec celles-ci. Elles déterminent la « conformité » du modèle bayésien intégrant les opinions d'expert. La bonne réalisation des tests en laboratoire est ainsi primordiale pour ne pas insérer un biais dû à la manipulation.

L'approche bayésienne se présente comme une technique fiable de modélisation de données récoltées sur le terrain. Elle permet de valider cette modélisation afin d'obtenir des informations *a posteriori* sûres car soumises à des critères très strictes. Celles-ci sont une synthèse entre les résultats de test diagnostiques récoltés sur le terrain et les opinions d'experts acquises jusqu'au moment de l'analyse et représentent donc une nouvelle information générée et utilisable par la suite. Elle permet de faire évoluer, d'affiner les connaissances en les confrontant à de nouvelles données de terrain.

## 8. PERSPECTIVES

A l'issue de ce travail, nous démontrons que la limitation du diagnostic à un ou deux tests peu sensibles (soit, la palpation de la langue et/ou l'inspection de la carcasse) fournit un grand nombre de faux négatifs. L'ajout d'un test plus sensible comme « l'ELISA antigène » paraît donc primordial pour estimer la prévalence réelle de la maladie au sein de la population.

Cette modélisation s'adaptant à tout type de test, son utilisation peut s'étendre à toutes les maladies soumises à un diagnostic systématique. Dans un premier temps, il serait intéressant de pouvoir l'appliquer au diagnostic de la cysticercose porcine et bovine dans d'autres régions et pays du monde. Cette généralisation demandera une normalisation des différents critères de validation ainsi qu'une étude théorique préliminaire du comportement de l'analyse de bases de données plus conséquentes. Il est à noter que les informations *a priori* utilisables dans une étude donnée ne sont pas automatiquement applicables dans une autre et qu'elles peuvent même fortement varier. Par exemple, le test ELISA AG utilisé pour le diagnostic de la cysticercose due à *Taenia solium* montre des réactions croisées avec une infestation par *Cysticercus tenuicollis* chez le porc (Dorny *et al.*, 2003). Les informations *a priori* disponibles dépendent donc de la présence de ce dernier dans la population étudiée.

L'importance de la précision des informations *a priori* a été soulignée dans ce travail. Une manière de s'en rapprocher serait l'intégration de l'opinion de plusieurs experts dans un même modèle.

Ce type d'approche nous a permis de réduire le nombre de test pour obtenir une information identique à celle fournie par l'utilisation d'un nombre plus élevé de tests et de diminuer ainsi les coûts nécessaires au diagnostic de la maladie. Il semble également qu'elle puisse donner une information sur le nombre minimum d'échantillons à récolter au sein d'une population afin d'estimer la prévalence d'une infection tout en se limitant aux frais d'échantillonnage strictement nécessaires. La vérification de ce concept nécessiterait une étude supplémentaire tant empirique que mathématique.

## 9. REFERENCES BIBLIOGRAPHIQUES

Adel A. and Berkvens D. 2002. Extension of the classic multi-test model, involving covariance between the tests, to a three- and four-test situation. Internal Technical Report, ITM, 16.04.2002.

Bayes T. 1763. An essay towards solving a problem in the doctrine of chances. MD Computing 8(3): 157-171.

Berkvens D., Speybroek N., Praet N., Adel A. and Lesaffre E. 2006. Estimating disease prevalence in a Bayesian framework using probabilistic constraints. Epidemiology 17(2), 145-153.

Boa M.E., Kassuku A.A., Willingham A.L., Keyyu J.D., Phiri I.K. and Nansen P. 2002. Distribution and density of cysticerci of *Taenia solium* by muscle groups and organs in naturally infected local finished pigs in Tanzania. Veterinary Parasitology 106, 155-164.

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

Brooks S.P. and Gelman A. 1998. General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics, 7, 434-455.

Delaram G. and Burns M. 2004. Cysticercosis. Emedecine,  
<http://www.emedicine.com/ped/topic537.htm>

Dendukuri N. and Joseph L. 2001. Bayesian approaches to modelling the conditional dependence between multiple diagnostics tests. Biometrics 57, 208-217.

Diop A.G., de Boer H.M., Mandlhate C., Prilipko L. and Meinardi H. 2003. The global campaign against epilepsy in Africa. Acta Tropica 87(1) : 149-159.

Dorny P., Vercammen F., Brandt J., Vansteenkiste W., Berkvens D. and Geerts S. 2000. Sero-epidemiological study of *Taenia saginata* cysticercosis in Belgian cattle. *Veterinary Parasitology* 88, 43-49.

Dorny P., Brandt J., Zoli A and Geerts S. 2003. Immunodiagnostic tools for human and porcine cysticercosis. *Acta Tropica* 87(1), 79-86.

Dorny P., Phiri I.K., Vercruysse J., Gabriel S., Willingham A.L. III, Brandt J., Victor B., Speybroeck N. and Berkvens D. 2004. A Bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. *International Journal for Parasitology* 34(5), 569-76.

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

Garcia H.H. and Del Bruto O.H. 2000. *Taenia solium* cysticercosis. *Infectious Disease Clinics of North America* 14, 97-119.

Gardner I.A., Stryhn H., Lind P. and Collins M.T. 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Preventive Veterinary Medicine* 45, 107-122.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. 2004. *Bayesian Data Analysis*. Second Edition, Chapman & Hall/CRC.

Gonzalez A.E., Cama V., Gilman R.H., Tsang V.C.W., Pilcher J.B., Chavera A., Castro M., Montenegro T., Verastegui M., Miranda E. and Bazalar H. 1990. Prevalence and comparison of serological assays, necropsy and tongue examination for the diagnosis of porcine cysticercosis in Peru. *Am. J. Trop. Med. Hyg.* 43, 194-199.

Majofane N.A., Appleton C.C., Krecek R.C., Michael L.M., Willingham A.L. 2003. The current status of neurocysticercosis in Eastern and Southern Africa. *Acta Tropica* 87(1), 25-33.

Nash T.E. 2003. Human case management and treatment of cysticercosis. *Acta Tropica* 87(1), 61-69

Nunes C.M., Biondi G.F., Heinemann M.B., Richtzenhain L.J. 2000. Comparative evaluation of an indirect ELISA test for diagnosis of swine cysticercosis employing antigen from *Taenia solium* and *Taenia crassiceps* metacestodes. *Veterinary Parasitology* 93, 135-140.

Phiri I.K., Ngowni H., Afonso S., Matenga E., Boa M., Mukaratirwa S., Githigia S, Saimo M., Sikasunge C., Maingi N., Lubega G.W., Kassuku A., Michael L., Siziya S., Krecek R.C., Noormahomed E., Vilhena M., Dorny P.& Willingham A.L. 2003. The emergence of *Taenia solium* cysticercosis in Eastern and Southern Africa as a serious agricultural problem and public health risk. *Acta Tropica* 87(1), 13-23.

Smith B.J. 2004. Bayesian Output Analysis Program (BOA) Version 1.1.2 for i386, mingw32.

Spiegelhalter D.J., Best N.G., Carlin B.P. and van der Linde A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B* 64, 583-640.

Spiegelhalter D.J., Thomas A., Best N.G. and Lunn D. 2003. WinBUGS Version 1.4 User.

The R Foundation for Statistical Computing. 2004. R Version 1.9.1 , ISBN 3-900051-00-3.

Toma B., Benet J-J., Dufour B., Eloit M., Moutou F., Sanaa M. 1991. Glossaire d'épidémiologie animales. Editions du point vétérinaire, Maison Alfort.

Tsang V.C.W., Brand J.A., Boyer A.E.1989. An enzyme-linked immunoelectrotransfer blot assay and glycoprotein antigens for diagnosing human cysticercosis (*Taenia solium*). *Journal of Infectious Diseases* 159, 50-59.

World Health Organization. 2004. Control of neurocysticercosis. Fifty-fifth World Health Assembly report.

Zoli A., Shey-Njila O, Assana E., Nguekam J-P., Dorny P., Brandt J. and Geerts S. 2003. Regional status, epidemiology and impact of *Taenia solium* cysticercosis in Western and Central Africa. *Acta Tropica* 87(1), 35-42.

## Annexe 1

### M1: Informations *a priori* non informatives

```
model
{
  r[1:4] ~ dmulti( pr[1:4], N )

  pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])*p[3]*p[6]
  pr[2] <- p[1]*(1-p[2])*p[5]+(1-p[1])*p[3]*(1-p[6])
  pr[3] <- p[1]*p[2]*(1-p[4])+(1-p[1])*(1-p[3])*p[7]
  pr[4] <- p[1]*p[2]*p[4]+(1-p[1])*(1-p[3])*(1-p[7])

  p[1] ~ dbeta(1,1)
  p[2] ~ dbeta(1,1)
  p[3] ~ dbeta(1,1)
  p[4] ~ dbeta(1,1)
  p[5] ~ dbeta(1,1)
  p[6] ~ dbeta(1,1)
  p[7] ~ dbeta(1,1)

  r2[1:4] ~ dmulti( pr[1:4], N)
  for (i in 1:4)
  {
    z1[i] <- equals(0,pr[i])
    y1[i] <- max(z1[i],pr[i])
    x1[i] <- max(r[i],1)
    d[i] <- r[i]*log(x1[i]/(y1[i]*N))
    z2[i] <- equals(0,pr[i])
    y2[i] <- max(z2[i],pr[i])
    x2[i] <- max(r2[i],1)
    d2[i] <- r2[i]*log(x2[i]/(y2[i]*N))
  }
  G0 <- 2 * sum(d[])
  Gt <- 2 * sum(d2[])
  bayesp <- step(G0 - Gt)

  prev <- p[1]
  se1 <- p[2]
  sp1 <- p[3]
  se2 <- p[2]*p[4]+(1-p[2])*p[5]
  sp2 <- p[3]*p[6]+(1-p[3])*p[7]
}

list(r=c(620,80,30,270), N=1000)
```

### M2: spécificité du test 1 == 1, spécificité du test 2 ==1

```
pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])
pr[2] <- p[1]*(1-p[2])*p[5]
pr[3] <- p[1]*p[2]*(1-p[4])
pr[4] <- p[1]*p[2]*p[4]

p[1] ~ dbeta(1,1)
p[2] ~ dbeta(1,1)
p[4] ~ dbeta(1,1)
p[5] ~ dbeta(1,1)
```



**M3: spécificité du test 1 == 1, spécificité du test 2 ==1, contraintes sur sensibilité du test 1, contraintes sur sensibilité du test 2**

```
pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])
pr[2] <- p[1]*(1-p[2])*p[5]
pr[3] <- p[1]*p[2]*(1-p[4])
pr[4] <- p[1]*p[2]*p[4]
```

```
p[1] ~ dbeta(1,1)
p[2] ~ dbeta(1,1)I(0.5,0.7)
p[4] ~ dbeta(1,1)I(0.8,1)
p[5] ~ dbeta(1,1)I(0.3,0.5)
```

**M4: spécificité du test 1 == 1, spécificité du test 2 ==1, contraintes strictes sur sensibilité du test 1, contraintes strictes sur sensibilité du test 2**

```
pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])
pr[2] <- p[1]*(1-p[2])*p[5]
pr[3] <- p[1]*p[2]*(1-p[4])
pr[4] <- p[1]*p[2]*p[4]
```

```
p[1] ~ dbeta(1,1)
p[2] ~ dbeta(1,1)I(0.5999,0.6001)
p[4] ~ dbeta(1,1)I(0.8999,0.9001)
p[5] ~ dbeta(1,1)I(0.3999,0.4001)
```

**M5: spécificité du test 1 == 1, spécificité du test 2 ==1, contraintes strictes sur sensibilité du test 1, contraintes strictes sur sensibilité du test 2, contraintes strictes sur la prévalence**

```
pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])
pr[2] <- p[1]*(1-p[2])*p[5]
pr[3] <- p[1]*p[2]*(1-p[4])
pr[4] <- p[1]*p[2]*p[4]
```

```
p[1] ~ dbeta(1,1)I(0.4999,0.5001)
p[2] ~ dbeta(1,1)I(0.5999,0.6001)
p[4] ~ dbeta(1,1)I(0.8999,0.9001)
p[5] ~ dbeta(1,1)I(0.3999,0.4001)
```

**M6: spécificité du test 1 == 1, spécificité du test 2 ==1, contraintes erronées sur sensibilité du test 1**

```
pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])
pr[2] <- p[1]*(1-p[2])*p[5]
pr[3] <- p[1]*p[2]*(1-p[4])
pr[4] <- p[1]*p[2]*p[4]
```

```
p[1] ~ dbeta(1,1)
p[2] ~ dbeta(1,1)I(0.8,1)
p[4] ~ dbeta(1,1)
p[5] ~ dbeta(1,1)
```

**M7: spécificité du test 1 == 1, spécificité du test 2 ==1, contraintes erronées sur sensibilité du test 1, contraintes erronées sur probabilité conditionnelle de tester positif au deuxième test si l'animal est infecté et a obtenu un résultat positif au premier test**

```
pr[1] <- p[1]*(1-p[2])*(1-p[5])+(1-p[1])
pr[2] <- p[1]*(1-p[2])*p[5]
pr[3] <- p[1]*p[2]*(1-p[4])
pr[4] <- p[1]*p[2]*p[4]
```

```
p[1] ~ dbeta(1,1)
p[2] ~ dbeta(1,1)I(0.8,1)
p[4] ~ dbeta(1,1)I(0.2,0.4)
p[5] ~ dbeta(1,1)
```

## Annexe 2

```

model
{
  r[1:16] ~ dmulti( pr[1:16], n )
  pr[1] <- p1*(1-p2)*(1-p5)*(1-p11)*(1-p23)+(1-p1)*p3*p6*p12*p24
  pr[2] <- p1*(1-p2)*(1-p5)*(1-p11)*p23+(1-p1)*p3*p6*p12*(1-p24)
  pr[3] <- p1*(1-p2)*(1-p5)*p11*(1-p22)+(1-p1)*p3*p6*(1-p12)*p25
  pr[4] <- p1*(1-p2)*(1-p5)*p11*p22+(1-p1)*p3*p6*(1-p12)*(1-p25)
  pr[5] <- p1*(1-p2)*p5*(1-p10)*(1-p21)+(1-p1)*p3*(1-p6)*p13*p26
  pr[6] <- p1*(1-p2)*p5*(1-p10)*p21+(1-p1)*p3*(1-p6)*p13*(1-p26)
  pr[7] <- p1*(1-p2)*p5*p10*(1-p20)+(1-p1)*p3*(1-p6)*(1-p13)*p27
  pr[8] <- p1*(1-p2)*p5*p10*p20+(1-p1)*p3*(1-p6)*(1-p13)*(1-p27)
  pr[9] <- p1*p2*(1-p4)*(1-p9)*(1-p19)+(1-p1)*(1-p3)*p7*p14*p28
  pr[10] <- p1*p2*(1-p4)*(1-p9)*p19+(1-p1)*(1-p3)*p7*p14*(1-p28)
  pr[11] <- p1*p2*(1-p4)*p9*(1-p18)+(1-p1)*(1-p3)*p7*(1-p14)*p29
  pr[12] <- p1*p2*(1-p4)*p9*p18+(1-p1)*(1-p3)*p7*(1-p14)*(1-p29)
  pr[13] <- p1*p2*p4*(1-p8)*(1-p17)+(1-p1)*(1-p3)*(1-p7)*p15*p30
  pr[14] <- p1*p2*p4*(1-p8)*p17+(1-p1)*(1-p3)*(1-p7)*p15*(1-p30)
  pr[15] <- p1*p2*p4*p8*(1-p16)+(1-p1)*(1-p3)*(1-p7)*(1-p15)*p31
  pr[16] <- p1*p2*p4*p8*p16+(1-p1)*(1-p3)*(1-p7)*(1-p15)*(1-p31)
  for (i in 1:16)
  {
    d[i] <- r[i]*log(max(r[i],1)/(pr[i]*n))
  }
  G0 <- 2 * sum(d[])
  r2[1:16] ~ dmulti(pr[1:16], n)
  for (i in 1:16)
  {
    d2[i] <- r2[i]*log(max(r2[i],1)/(pr[i]*n))
  }
  Gt <- 2 * sum(d2[])
  bayesp <- step(G0 - Gt)
  p1 ~ dbeta(1,1)
  p2 ~ dbeta(1,1)
  p3 ~ dbeta(1,1)
  p4 ~ dbeta(1,1)
  p5 ~ dbeta(1,1)
  p6 ~ dbeta(1,1)
  p7 ~ dbeta(1,1)
  p8 ~ dbeta(1,1)
  p9 ~ dbeta(1,1)
  p10 ~ dbeta(1,1)
  p11 ~ dbeta(1,1)
  p12 ~ dbeta(1,1)
  p13 ~ dbeta(1,1)
  p14 ~ dbeta(1,1)
  p15 ~ dbeta(1,1)
  p16 ~ dbeta(1,1)
  p17 ~ dbeta(1,1)
  p18 ~ dbeta(1,1)
  p19 ~ dbeta(1,1)
  p20 ~ dbeta(1,1)
  p21 ~ dbeta(1,1)
  p22 ~ dbeta(1,1)
  p23 ~ dbeta(1,1)
  p24 ~ dbeta(1,1)
  p25 ~ dbeta(1,1)
  p26 ~ dbeta(1,1)
  p27 ~ dbeta(1,1)
  p28 ~ dbeta(1,1)
  p29 ~ dbeta(1,1)
  p30 ~ dbeta(1,1)
  p31 ~ dbeta(1,1)
}
list(r=c(326,42,281,95,0,0,5,4,1,0,2,0,2,1,35,74), n=868)

```

### Annexe 3

```
model
{
  r[1:8] ~ dmulti( pr[1:8], n )
  pr[1] <- p1*(1-p2)*(1-p5)*(1-p11)+(1-p1)*p3*p6*p12
  pr[2] <- p1*(1-p2)*(1-p5)*p11+(1-p1)*p3*p6*(1-p12)
  pr[3] <- p1*(1-p2)*p5*(1-p10)+(1-p1)*p3*(1-p6)*p13
  pr[4] <- p1*(1-p2)*p5*p10+(1-p1)*p3*(1-p6)*(1-p13)
  pr[5] <- p1*p2*(1-p4)*(1-p9)+(1-p1)*(1-p3)*p7*p14
  pr[6] <- p1*p2*(1-p4)*p9+(1-p1)*(1-p3)*p7*(1-p14)
  pr[7] <- p1*p2*p4*(1-p8)+(1-p1)*(1-p3)*(1-p7)*p15
  pr[8] <- p1*p2*p4*p8+(1-p1)*(1-p3)*(1-p7)*(1-p15)
  for (i in 1:8)
  {
    d[i] <- r[i]*log(max(r[i],1)/(pr[i]*n))
  }
  G0 <- 2 * sum(d[])
  r2[1:8] ~ dmulti(pr[1:8], n)
  for (i in 1:8)
  {
    d2[i] <- r2[i]*log(max(r2[i],1)/(pr[i]*n))
  }
  Gt <- 2 * sum(d2[])
  bayesp <- step(G0 - Gt)
  p1 ~ dbeta(1,1)
  p2 ~ dbeta(1,1)
  p3 ~ dbeta(1,1)
  p4 ~ dbeta(1,1)
  p5 ~ dbeta(1,1)
  p6 ~ dbeta(1,1)
  p7 ~ dbeta(1,1)
  p8 ~ dbeta(1,1)
  p9 ~ dbeta(1,1)
  p10 ~ dbeta(1,1)
  p11 ~ dbeta(1,1)
  p12 ~ dbeta(1,1)
  p13 ~ dbeta(1,1)
  p14 ~ dbeta(1,1)
  p15 ~ dbeta(1,1)

  se1 <- p2
  se2 <- p2*p4+(1-p2)*p5
  se3 <- p2*(p4*p8+(1-p4)*p9)+(1-p2)*(p5*p10+(1-p5)*p11)
  sp1 <- p3
  sp2 <- p3*p6+(1-p3)*p7
  sp3 <- p3*(p6*p12+(1-p6)*p13)+(1-p3)*(p7*p14+(1-p7)*p15)
}
list(r=c(r[1], ..., r[i]), n=868)
```

#### Annexe 4.

Model (V et Ag si V-)

```
{
r[1:3] ~ dmulti( pr[1:3], n )
pr[1] <- p1*(1-p2)*(1-p5) + (1-p1)*p6
pr[2] <- p1*(1-p2)*p5 + (1-p1)*(1-p6)
pr[3] <- p1*p2
for (i in 1:3)
{
    d[i] <- r[i]*log(max(r[i],1)/(pr[i]*n))
}
G0 <- 2 * sum(d[])
r2[1:3] ~ dmulti(pr[1:3], n)
for (i in 1:3)
{
    d2[i] <- r2[i]*log(max(r2[i],1)/(pr[i]*n))
}
Gt <- 2 * sum(d2[])
bayesp <- step(G0 - Gt)
p1 ~ dbeta(1,1)
p2 ~ dbeta(1,1)
p4 ~ dbeta(1,1)
p5 ~ dbeta(1,1)
p6 ~ dbeta(1,1)

se1 <- p2
se2 <- p2*p4 + (1-p2)*p5
sp2 <- p6
}
list(r=c(369, 378, 121), n = 868)
```

```

Model (T et Ag si T -)
{
r[1:3] ~ dmulti( pr[1:3], n )
pr[1] <- p1*(1-p2)*(1-p5) + (1-p1)*p6
pr[2] <- p1*(1-p2)*p5 + (1-p1)*(1-p6)
pr[3] <- p1*p2
for (i in 1:3)
{
    d[i] <- r[i]*log(max(r[i],1)/(pr[i]*n))
}
G0 <- 2 * sum(d[])
r2[1:3] ~ dmulti(pr[1:3], n)
for (i in 1:3)
{
    d2[i] <- r2[i]*log(max(r2[i],1)/(pr[i]*n))
}
Gt <- 2 * sum(d2[])
bayesp <- step(G0 - Gt)
p1 ~ dbeta(1,1)
p2 ~ dbeta(1,1)
p4 ~ dbeta(1,1)
p5 ~ dbeta(1,1)
p6 ~ dbeta(1,1)

se1 <- p2
se2 <- p2*p4 + (1-p2)*p5
sp2 <- p6
}
list(r=c(368, 385, 115), n = 868)

```