Epidémiol. santé anim., 1997, 31-32

DEPENDENCY BETWEEN SENSITIVITY, SPECIFICITY AND PREVALENCE ANALYSED BY MEANS OF GIBBS SAMPLING

Chriel M.[1], Willeberg P.[1]

L'une des suppositions, lors de l'interprétation d'un test diagnostique, est que ce dernier est indépendant de la prévalence de la maladie dans la population à laquelle il est appliqué.
De bonnes estimations de ce type de test peuvent être faites de plusieurs manières, telles que l'estimation par la méthode du maximum de vraisemblance et l'approche Bayesienne. L'estimation Bayesienne utilise l'échantillonnage de Gibbs. L'avantage de cette méthode est lié au fait qu'une seule population est suffisante pour estimer la sensibilité et la spécificité du test en l'absence d'un test standard. La méthode peut être utilisée pour estimer les caractéristiques d'un, deux ou trois tests diagnostiques. Elle nécessite néanmoins une taille de l'échantillon suffisante et des informations préalables en vue de produire une estimation raisonnable. Cet article se base sur des données d'abattoirs après inspection post-mortem des lésions de pleurésie chez les porcs au Danemark. Deux méthodes d'inspection ont été appliquées : la méthode traditionnelle et la méthode visuelle. Trois populations avec des prévalances de pleurésie chronique différentes ont été examinées en vue de comparer l'association entre sensibilité spécificité et prévalence.
Les résultatat suggèrent que l'hypothèse d'indépendance entre la prévalence et la sensibilité lors d'inspection visuelle n'est pas satisfaite. Quand la prévalence de la pleurésie chronique augmente, une baisse de la sensibilité est observée. Il n'y a que peu d'effet de la prévalence sur les autres caractéristiques du test. En conséquence, il faut être prudent lors d'utilisation d'un test diagnostique, notamment sur la prévalence de la maladie dans la population étudiée avant toute conclusion.

INTRODUCTION

Diagnostic tests are assumed to provide the user with the true state of nature. Furthermore, the test characteristics (sensitivity and specificity) of a given test are assumed to be independent of the population to which it's applied, in contrary to the dependency between the prevalence of the given disease in a population and the estimates of the predictive values.
Reasons for developing a new diagnostic test are varying from insufficiency of the already developed test, the price of the test, to the need for a more rapid method of determining the true state of disease of the animal. In the process of developing new diagnostic tests, an evaluation and comparison with the standard (or old) test ought to take place before applying the new test in the field. However, the basis of these evaluations has often been performed in only one population for which reason it is not possible to evaluate the test under the varying prevalences and field conditions.
A large number of frequentist-based methods have been developed for evaluating the sensitivity and specificity of a test in the absence of a gold standard (*Walter and Irwig, 1988*). These methods were designed to compare and evaluate one, two or three diagnostic tests applied to one or two populations with different prevalence of the disease. This class of models are referred to as "latent" class models due to the unknown true classification of the individual. The techniques are all based on the maximum likelihood estimation method, with constraining one or more parameters or replicate testings.
A different method for estimation is based on Bayes Theorem (*Joseph et al, 1995*). This methodology avoids the problems of the above-mentioned and allows the uncertainty of the true value of all the parameters of interest to vary.
The method has been developed for comparison of one, two or three diagnostic tests in one population and can be viewed as a generalization of the frequentist-based methods. An iterative Markov-Chain Monte Carlo (MCMC)- technique, using the Gibbs sampler, is used to approximate the marginal posterior densities of the parameters of interest in the absence of a gold standard.
The purpose of this study was to evaluate the assumption of independency between prevalence and test characteristics such as sensitivity and specificity in three different populations by means of the Gibbs sampling method.

MATERIAL AND METHODS

In 1993 a study was carried out in a Danish slaughterhouse in order to compare the detection of lesion rates from a visual inspection procedure (VMI) and traditional meat inspection (TMI) procedure (*Willeberg et al, 1997*). A total of 183,383 fattening pigs was inspected. The pigs were delivered from farms with three different health status with varying levels of chronic parietal pleuritis. The material was divided into the following three categories:
a. pigs derived from conventional fattening units (N=97,129; apparent prevalence: TMI=20.76%, VMI=17.30%) ;

[1] Division of Ethology and Health, The Royal Veterinary and Agricultural University, Bülowsvej 13, DK-1870 Frederiksberg C, Copenhagen, Denmark.

b. pigs form mycoplasma (only) infected (MS) herds (N = 7258; apparent prevalence: TMI = 18.30%, VMI = 14.95%) ;
c. pigs from specific pathogen free (SPF) herds (N = 2271; apparent prevalence: TMI=10.88%, VMI = 9.20%).

The unobserved latent variables, Y_1, Y_2, Y_3, and Y_4, represent the true number of classified subjects out of the observed cell values u, v, w, and x respectively in the 2 X 2 data in table I.

Table I
Observed data from two diagnostic tests in the absence of a gold standard. The latent variables (Y_1, Y_2, Y_3, and Y_4) are stated in parentheses.

		TEST 2		
		Positive	Negative	
TEST 1	Positive	u (Y_1)	v (Y_2)	u+v
	Negative	w (Y_3)	x (Y_4)	w+x
		u+w	v+x	N

The beta(α,β) prior densities of the prevalences, sensitivities and specificities were set identically for all three populations (Table II) with an equally tailed 95% probability range in order not to influence the posterior distributions differently for the three populations.
The MCMC-method is based on the bayesian approach and carries out a number of numerical integrations using simulations. Instead of calculating exact or approximate estimates, this technique generates a stream of simulated values, using the Gibbs Sampler, for each quantity of interest (*Gelfand and Smith, 1990*). The densities were obtained by smoothing the output from the Gibbs sampler with a normal kernel.

Table II
The beta prior densities for the two meat inspection procedures.

	Range (%)	Beta coefficients	
		α	β
Prevalence	5-45%	4.44	13.31
Sensitivity	35-100%	4.1	1.76
Specificity	90-100%	71.25	3.75

Expert based informative priors (table II) have been specified in order to draw useful inferences on the posterior distributions (*Joseph et al, 1997*). The Gibbs sampler is used to construct the marginal posterior densities of all the parameters of interest.
A total of 5500 iterations were performed by the Gibbs sampler. Inferences were based on the last 5000 Gibbs samples in all three populations.

RESULTS
The result of the analyses demonstrated that the normal assumption of independency between the test characteristics and the prevalence of the population was not fulfilled in this material. By increasing prevalence of chronic parietal pleuritis in the populations a decreasing sensitivity and specificity were found especially for the sensitivity of the VMI procedure (table III).

Table III
The posterior Bayesian estimates (95%Credible Intervals) of the prevalences of chronic pleuritis in slaughter pigs and the sensitivities and specificities of the two meat inspection procedures.

		Traditional meat inspection		Visual meat inspection	
POPULATION	PREVALENCE	SENSITIVITY	SPECIFICITY	SENSITIVITY	SPECIFICITY
CONVENTIONAL	20.05 (17.43;21.72)	90.98 (85.82;95.57)	96.59 (94.58;98.24)	75.61 (70.54;83.25)	96.99 (96.13;98.11)
MS	16.72 (13.63;19.93)	91.09 (83.64;98.11)	96.38 (93.62;98.65)	78.69 (69.83;93.01)	97.84 (96.66;99.26)
SPF	9.53 (7.75;11.80)	91.38 (82.82;98.84)	97.48 (96.10;99.09)	85.86 (72.29;97.38)	98.69 (97.77;99.31)

Furthermore, the results demonstrated that the VMI procedure tends to underestimate the prevalence in all three populations.
Allowing uncertainty on all the estimated parameters also allow positive findings to be classified as false positives. Therefore, the posterior estimates of the prevalences were shrinked toward the average of the two meat inspection procedures.

DISCUSSION

There are conventional and deeply rooted beliefs that the sensitivity and specificity of a diagnostic test remains constant when the same test is applied in different populations, and that only the predictive values are related to disease prevalence.
Sensitivity and specificity are constant only when the diagnosis are made by an absolute reliable reference test or gold standard and when the diseases have no grading in severity.
As demonstrated in the above example this is not always the case. Especially the VMI procedure demonstrated a decrease in sensitivity for increasing prevalence. This might be caused by an association between the condition of interest, here the chronic parietal lesions, and the applied test. In this situation one would expect, that the size of the lesions would have a greater impact on the detection rate when the VMI procedure was applied than the TMI procedure due to the more intensive control procedure in the latter.

Furthermore, one might suspect that the persons involved in the meat inspection procedures intended to prove a higher efficiency (reporting bias) in detecting lesions with the TMI in comparison with the new procedure resulting in a higher apparent prevalence in all three populations for the TMI procedure than for the VMI procedure.
When a test-disease association exists, like in this case with different levels and degrees of severity of chronic parietal pleuritis, the test performance is only valid in the particular population for which it has been evaluated. This situation has also been observed for many diagnostic tests used in human health screenings (*Choi, 1997*).
Using the Gibbs sampler for estimating the posterior distributions requires a careful assessment of the prior distributions. Uniform prior density for the prevalence will lead to a uniform posterior density even if the sample size increases (*Joseph, 1997*). The Bayesian method allows the investigator to estimate test characteristics when only one population is available in contrast to the frequentist-based methods (*Hui and Walter 1980*). This allows one to draw inferences of test characteristics and thereby focus on the consequences of the predictive values when a given test is applied on populations of different prevalences.

CONCLUSION

In the absence of a gold standard this method using Gibbs sampling in order to estimate the test characteristics has shown to be applicable. However, this method requires caution when the priors are to be specified in order not to influence the posterior estimates. Combinations of expert knowledge and earlier reported results are strongly advisable. An increase of the sample size does not alleviate the cautions that should be taken in specifying the priors (*Joseph, 1997; Andersen, 1997*). The applied method has demonstrated that not only the predictive values are influences by the prevalence of the disease of concern, but also the sensitivities and specificities of a given test. Reported sensitivity and specificity ought to be interpreted as average probabilities of a positive or negative test result, respectively, in a particular population.
Therefore, it should be recommended to the producers of diagnostic tests to evaluate the performances of the tests in different population with a varying prevalence in order to determine the influence of the above mentioned on the test characteristics.

REFERENCES

Andersen, S., 1997 :Re: "Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Test in the Absence of a Gold Standard". Am.J. of Epidemiol, vol. 145:290.

Choi, B.C.K., 1997: Causal Modelling to Estimate Sensitivity and Specificity of a Test when Prevalence Changes. Epidemiology, vol 8:80-86.

Hui, S.L., and Walter, S.D., 1980: Estimating the Error Rates of Diagnostic Tests. Biometrics, 36:167-171.

Joseph, L. 1997. The first author reply. Am.J. of Epidemiol, vol. 145:291.

Joseph, L., Gyorkos, T.W., Coupal, L., 1995: Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Test in the Absence of a Gold Standard. Am.J. of Epidemiol, vol. 141:263-272..

Walter, S.D., Irwig, L.M., 1988: Estimation of Test Error Rates , Disease Prevalence and Relative Risk from Misclassified Data: A Review. J. Clin. Epidemiol, vol. 41:923-937.

Willeberg, P., Wedam, J.M., Gardner, I.A., Holmes, J.C., Mousing, J., Kyrval, J., 1997, Enøe, C., Andersen, S., Leontides, L.: A comparative study of visual and traditional post-mortem inspection of slaughter pigs: Estimation of sensitivity, specificity and differences in non-detection rates. Proc. 8 ISVEE, Paris, France.