Statistical Analysis of Bovine Spongiform Encephalopathy Data in Belgium.

Cortiñas Abrahantes, J.[1], Wanyama, S.S.[1], Mintiens, K.[2], Geys, H.[1,7], van Everbroeck, B.[3], Cras,P.[4], Berkvens, D.[5], Saegerman, C. [6], Roels, S.[2] and Aerts, M.[1].

[1] Center for Statistics, Hasselt University; [2] Veterinary and Agrochemical Research Centre; [3] Biobank, Institute Born Bunge, University of Antwerp; [4] Department of Neurology, University Hospital of Antwerp, University of Antwerp; [5] Department of Animal Health, Institute of Tropical Medicine; [6] Department of Infectious and Parasitic Diseases, Epidemiology and Risk analysis, Faculty of Veterinary Medicine, University of Liege; 7 Johnson & Johnson, PRD.

Abstract

Bovine spongiform encephalopathy (BSE), widely known as "mad cow disease", is a chronic, degenerative disease affecting the central nervous system of cattle. Worldwide there have been more than 180,000 cases since the disease was first diagnosed in 1986 in Great Britain. The first case of BSE on the Belgian territory was confirmed on 31st October 1997. From 1990 BSE has been monitored by a passive surveillance scheme, where only animals with clinical symptoms were examined with very few cases being diagnosed (19 cases). However in 2001, an active surveillance scheme (based on EU regulation 999/2001) was implemented and since then many more cases have been detected. The data were obtained from the Belgian Federal Agency for the Safety of the Food Chain. We focus on the year 2001, which includes information from animals that were either slaughtered for consumption or under emergency, and animals which died on the farms due to some illness and were disposed off after sample analysis (rendering plan). Several statistical techniques were used to analyze the data in order to examine relationships between rapid and confirmatory tests and several covariates.

Introduction

Bovine Spongiform Encephalopathy (BSE) was first described as a fatal pathological entity, in the United Kingdom, in November 1986 by the Central Veterinary Laboratory, Weysbridge. The advent of this new disease in the United Kingdom in the 1980s was followed by a substantial epidemic which reached its peak in the UK in 1992. Since then, it has been recognized in many other countries around the world. The first case of BSE on the Belgian territory was confirmed on 31st October 1997. This was in Havenge in a province of Belgium in an animal which was born in 1992. In Belgium, BSE has been monitored since 1990 by a passive surveillance scheme (1990 to 2000), detecting only 19 cases. Concern raise that BSE is a potential threat to human health through food products derived from affected animals. This was heightened with the discovery in Britain of a new human transmissible spongiform encephalopathies (TSE), a variant form of Creutzfeldt – Jakob disease (vCJD), in the 1990s that had characteristics consistent with BSE. In 2001, in Belgium, an active surveillance scheme based on EU regulation 999/2001 was introduced (Roels *et. al.,* 2003). It was undertaken with the assistance of the Belgian Federal Agency for Food and the safety of the Food Chain. Since then many more cases have been detected. We will focus on the year 2001, accounting for the fact that the testing and practice procedures were adapted in the course of that year.

The objective of this research is to establish factors that may well be used to predict the probability of an animal testing BSE positive in a confirmatory test, given that it was detected as a positive animal in a first screening testing.

Materials and Methods

The data consist of 363866 animals (cows) which were either slaughtered normally or under emergency, and some that were found dead on the farms. All were sampled and analyzed for BSE in the year 2001. Most of the animals were born in Belgium itself though there are a few imported cases. The results show that 99 animals were positive for BSE at the first screening. For the regional incidence: - Flanders had 70 cases with the majority of these from Oriental Flanders (31 cases), Walloonia had 29 cases with the highest number of cases from Liege (9 cases), while Brussels region had no case of BSE.

We will concentrate our research on the animals tested positive in the first screening (99 cows). From the animals tested positive in the first screening one collected the date of birth, date of detection, the age (in months), commune and province (Antwerp (ANV), Brabant Flanders (BRF), Brabant Wallonie (BRW), Occidental Flanders (FOC), Oriental Flanders (FOR), Hainaut (HAI), Liege (LIE), Limburg (LIM), Luxemburg (LUX) and Namur (NAM)) at the time it was detected, the different streams (clinical suspected, cohort, rendering and slaughter), coat color (white, whiteblack, whiteblue, whitered and others), which can be associated with the race and type of animal and finally also depending on the date of detection an extra variable was constructed identifying different periods (1: januari to march; 2: april to august and 3: september to december) in which the threshold or the test itself were adapted in this particular year. Several confirmatory tests were carried out in a Central Lab. Here we will report only the results related to the Elisa test.

In this article we focus on two statistical approaches in order to identify factors that can be associated to animals retesting positive. The first one, classical logistic regression, was used to model the proportion of BSE retested positive animals in function of the variables before mentioned and a stepwise procedure was used to select the final model. The second approach is based upon classification trees (Breiman *et. al.*, 1984). This is a method where, following specific splitting rules (in our case we use the improvement in deviance), disjoint subsets of the data are constructed. This partitioning process results in a saturated tree. The saturated binary tree is then pruned to an optimal size tree. The final step is the selection process, which determines the final tree. A 10-fold crossvalidation procedure is used in order to correct the classification error. We will study the possibility of obtaining a tree which results in a 100 % sensitivity, given the importance of classifying well the true positives and as high specificity as possible. For that, different loss matrices ($\begin{pmatrix} 0 & 1.001 - i/100 \\ i/100 & 0 \end{pmatrix} \text{where } i = 1, \ldots, 100$) will be considered and we will monitor the total error and the sensitivity.

Results

First the dataset was explored and some tables and graphs of the data were created. For the variable color it can be highlighted that if color was white or whiteblue the number of false positives were 6 or 8 times the number of true positives, while for whiteblack or whitered the number of false positives were 0.5 or 0.25 times the number of true positives. For the other colors the numbers of true and false positives were the same. For the variable stream, there were no false positives for clinical suspected, cohort or rendering animals, while 0.625 of the slaughtered animals were false positive.

Table 1 shows the number of false positives, true positives and the total number of animals per province. From this table Oriental Flanders appears to be the province with the most BSE cases. Figure 1 shows the proportion of true positives in 13 age groups indicating a possible quadratic age effect.

Province	False Positive	True Positive	Total
ANV	4	8	45135
BRF	2	1	21396
BRW	2	1	5611
FOC	8	10	84854
FOR	15	16	83582
HAI	5	3	36897
LIE	4	5	29213
LIM	3	3	22744
LUX	7	0	17231
NAM	0	2	17203

Table 1. Number of false and true positive animals per province.

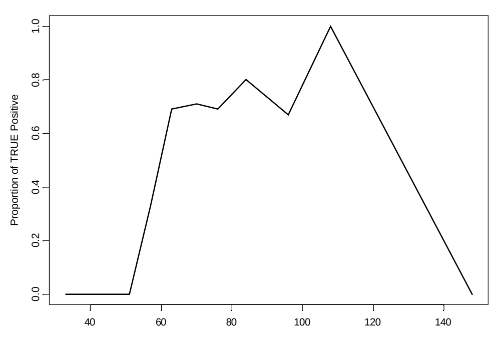

Fig.1 Proportion of true positives in 13 age groups.

In order to further study the possible influence of the different factors on retesting positive, the two approaches previously mentioned were applied. The stepwise logistic regression approach ends up with a model that contains the variable period (p-value<0.0001), age (0.3113, p-value=0.027) and the square of age (-0.00159, p-value=0.059). Note that already in this simple logistic regression, the odds ratio of the second vs first period and third vs first period were 109.06 (CI: 10.407; ∞) and 252.83 (CI: 16.02; ∞). The final model has sensitivity 87.8 and specificity 96.

In addition to the logistic regression procedure, a classification tree approach was used. The total error and sensitivity of the 10-fold crossvalidation for the different loss matrices are shown in figure 2.

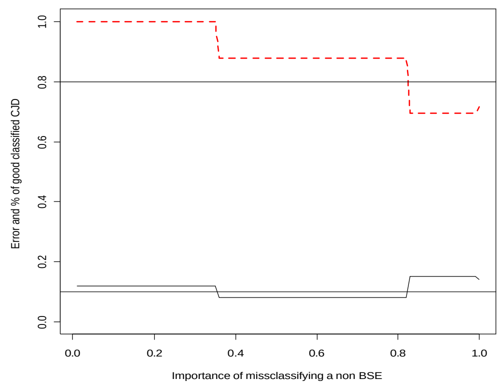

Fig.2 Estimated total error (solid line) and sensitivity (dashed line) for different loss matrices.

It can be seen that the estimated overall error rate can reach values below 10 %, but the sensitivity is then below 90%. However, we were interesting on keeping 100% sensitivity, thus we will compare the results obtained when no loss was considered (x axis =0.5) and when a loss matrix is used to assure 100% sensitivity (loss matrix when i=1).

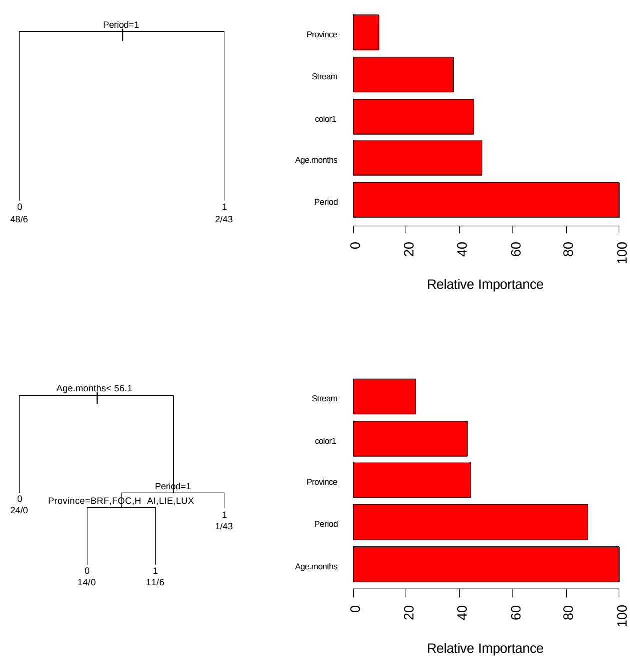

Fig.3 Classification trees and relative importance when NO loss (upper panel) and 100% sensitivity is assured (lower panel).

The classification trees and the relative importance are displayed in figure 3. It is important to highlight that the relative importance are different for the case when NO loss is considered than when a loss matrix is considered. For the latter, it can be seen that age of the animals plays an important role to retest positive which confirms the exploratory analysis and the results obtained with the previous approach.

Discussion and Recommendation

The clinical diagnosis of cattle suspected of having BSE poses a major challenge to the veterinarian. However, based on clinical findings alone, it is not always possible to clearly identify each case as BSE-positive or BSE-negative. The uncertainty that surrounds the diagnosis of BSE is likely to continue as long as there is no reliable test for the antemortem diagnosis of BSE. There are cases in which this disease cannot be ruled out with certainty, not even after repeated tests and especially for younger animals. Here we try to establish possible factors that can be associated to an animal retesting BSE positive. Both statistical approaches produce similar conclusions, highlighting the importance of the period, note that it is in fact due to changing test sensitivity and have thus no epidemiological value and the age of the animal tested to estimate the conditional probabilities of positivity at the confirmatory tests given that the screening test was positive. Saegerman *et. al.* (2004) already pointed out the effect of age, the results obtained here confirm their findings. The classification tree approach offers in general the flexibility of choosing different loss functions to be minimized and the possibility to evaluate variable importance for the different loss functions. The percentage of retesting positive in the Flanders region (54.3%) is much higher than for the Wallonia region (37,9%). Another important remark is that the proportion of true positives for animals between 60 and 114 months was higher than 0.65.
A recommendation can be made taking into account the data analysis. First, careful consideration needs to be made on tests used. There should also be strict follow up on cohort animals from which positives are identified.

References

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone C.J. (1984). *Classification and regression trees.* New York, Chapman and Hall.

Roels, S., De Bosschere, H., Saegerman, C., Dechamps, P., and Vanopdenbosch, E. (2004). BSE Surveillance and testing in Belgium. *New Food,* **1**: 36-40.

Saegerman, C., Speybroeck, N., Roels, S., Vanopdenbosch, E., Thiry, E. & Berkvens, D. (2004). Decision support tools in clinical diagnosis in cows with suspected bovine spongiform encephalopathy. *Journal of Clinical Microbiology*, **42**: 172-178.