

## Technical Appendix

### Basic Model

The method used to map the relative risk for BSE is based on a hierarchical Bayesian model (1–3), which enables to limit the overdispersion of the relative risk (4) using a smoothing. This method was introduced in disease mapping by Clayton and Kaldor (5) through a spatial prior named Conditional Autoregressive component, based on Gaussian distributions.

The delivery areas were labeled  $i = 1 \dots 943$  and the observed number of BSE cases in each area followed a Poisson distribution. The basic model for the disease mapping, without covariate (model 0), was:

$$y_i \sim P(\lambda_i) = e^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!} \text{ with } \lambda_i = e_i \cdot r_i \quad (1)$$
$$\ln \lambda_i = \ln e_i + u_i$$

The distribution parameter  $\lambda_i$  is the product of the expected number of BSE cases,  $e_i$ , and the relative risk for BSE,  $r_i$ , in a given area. The expected number of BSE, computed on the basis of the overall French incidence of BSE cases, took into account the main demographic structure of the bovine population: the dairy versus beef cattle (6–8). So, the expected number of BSE cases was:

$$e_i = p_{dairy} \cdot DAIRY_i + p_{beef} \cdot BEEF_i \quad (2)$$

$DAIRY_i$  and  $BEEF_i$  were the numbers of adult cattle in each area provided by the Agricultural Census 2000 and  $p_{dairy}$  and  $p_{beef}$  were the overall probabilities of infection assessed from the data. The relative risk  $r_i$  was the variation of the risk compared to a standard risk evaluated on the whole French territory. The spatial prior  $u_i$  was based on the spatial contiguities between areas. These components followed a normal distribution:

$$u_i \sim N(\bar{u}_{\partial i}, \tau_{u_i}) \quad (3)$$

$\bar{u}_{\partial i}$  was the mean of the spatial components in the set  $\partial i$  of areas adjacent to area  $i$  (neighbouring) and  $\tau_{u_i}$  was the variance inversely weighted by the number of neighbours of area  $i$ . This component is also called the ‘clustering effect’ (9).

## Estimation Step

Hierarchical models can be fitted by Markov Chain Monte Carlo (MCMC) method as implemented in LinBUGS, (Bayesian inference using Gibbs Sampling, <http://mathstat.helsinki.fi/openbugs/Home.html>). Gibbs Sampling is an adaptation of the general Metropolis algorithm (10). It consists in visiting each parameter in turn, and simulating a new value for this parameter from its full posterior conditional distribution, given the current values for the remaining parameters (11). The 50,000 first cycles of the Markov chain were discarded from computations. The effective chain had a size of 250,000 cycles. The parameters were estimated with their posterior mean computed from a sampling of 20,000 values (1 value every 5 cycles was used to reduce the autocorrelation). The stability of the chains was verified with the Heidelberger-Welch convergence diagnostic (12). The tests of conformity about the parameters were made from the 95% prediction interval given by the empirical quantiles of the chain. The spatial prior, named Conditional Autoregressive Model, was implemented in the geographical extension of LinBUGS: GeoBUGS (Department of Epidemiology and Public Health of the Imperial College at St Mary's Hospital, London).

## Test of covariates in the model

We added linearly the covariates use of MBM (*MBM*), animal fat (*FAT*) and animal DCP (*DCP*) as prior distributions (13):

$$\ln \lambda_i = \ln e_i + u_i + \beta \cdot X_i \quad (5)$$

In this equation,  $X_i$  was the vector of covariates for the area  $i$  and  $\beta$  was the vector of the regression coefficients. The covariates were incorporated one at a time and then together, so the dimensions of  $X_i$  and  $\beta$  varied according to the number of covariates tested. The first component of  $\beta$ , the baseline risk  $\beta_0$ , was the average risk. We assumed that the prior distributions of all regression parameters were uniform.

The effect of the covariates on the relative risk were assessed by the Deviance Information Criterion (DIC) used to compare models with variable complexity (number of parameters and hierarchical levels), fitted with the MCMC method (14). This method is a generalisation of the Akaike Information Criterion that is not appropriate for hierarchical models. DIC is calculated by adding the effective number of parameters (complexity) to the posterior

mean deviance (adequacy) of a model. The effective number of parameters is estimated by the difference between the posterior mean of the deviance and the deviance at the posterior estimates of the parameters of interest. The ‘best fit’ model is the one with the smallest DIC value. This criterion was assessed at the same time as the MCMC simulation. When covariates were chosen with the DIC, we performed a conformity test  $H_0 : \beta = 0$  on the regression coefficient based on the 95% prediction interval given by the empirical quantiles of MCMC simulations. An estimation of the p-value for this test was computed with the  $n$  simulated values of the Markov chain of the parameter  $\beta$  by  $p = \min\left(1 - \frac{m}{n}, \frac{m}{n}\right)$  where  $m$  was the number of values less than 0 (15). Finally, the estimated regression parameters  $\hat{\beta}$  can be interpreted as an odds ratio computed as  $\exp(\hat{\beta})$ . The odds ratio measured how much the relative risk for BSE (in a delivery area) was increased for each unit of the covariate (in this case a 100% increase of the proportion of factories using a given byproduct in the area).

## References

1. Banerjee S, Carlin BP, Gelfand AE. Hierarchical modeling and analysis for spatial data. London: Chapman & Hall; 2004. p. 99–128.
2. MacNab YC. Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics*. 2003;59:305–16. [Medline](#)
3. Richardson S, Best N. Bayesian hierarchical models in ecological studies of health-environment effects. *Environmetrics*. 2003;14:129–47.
4. Militino AF, Ugarte MD, Dean CB. The use of mixture models for identifying high risks in disease mapping. *Stat Med*. 2001;20:2035–49. [Medline](#)
5. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 1987;43:671–81. [Medline](#)
6. Ducrot C, Roy P, Morignat E, Baron T, Calavas D. How the surveillance system may bias the results of analytical epidemiological studies on BSE: prevalence among dairy versus beef suckler cattle breeds in France. *Vet Res*. 2003;34:185–92. [Medline](#)

7. Morignat E, Ducrot C, Roy P, Baron T, Vianard JL, Biacabe AG, et al. Targeted surveillance to assess of BSE in high risk populations in western France and the associated risk factors. *Vet Rec.* 2002;151:73–7. [Medline](#)
8. Wilesmith JW, Wells GAH, Cranwell MP, Ryan JBM. Bovine spongiform encephalopathy: epidemiological studies. *Vet Rec.* 1988;123:638–44. [Medline](#)
9. Richardson S, Monfort C, Green M, Draper G, Muirhead C. Spatial variation of natural radiation and childhood leukaemia incidence in Great Britain. *Stat Med.* 1995;14:2487–501. [Medline](#)
10. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov chain Monte Carlo in practice.* London: Chapman & Hall; 1996.
11. Mollié A. Bayesian and Empirical Bayes approaches to disease mapping. In: *Disease mapping and risk assessment.* Aberdeen: Wiley; 1999: p.15-29.
12. Heidelberger P, Welch PD. Simulation run length control in the presence of an initial transient. *Oper Res.* 1983;31:1109–44.
13. Lawson AB, Browne WJ, Vidal Rodeiro CL. *Disease mapping with WinBUGS and MLwiN.* Aberdeen: Wiley; 2003. p. 17–28.
14. Spiegelhalter DJB, Carlin BP, van der Linder A. Bayesian measures of model complexity and fit. *J Royal Stat Soc B.* 2002;64:583–639.
15. Davison AC, Hinkley DV, Gill R, Ripley BD, Ross S, Silverman BW, et al. *Bootstrap methods and their application.* Cambridge: Cambridge University Press; 1997.