PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Separating the Environmental and Genetic Factors that may be Causes of Bovine Spongiform Encephalopathy

C. M. Hau and R. N. Curnow

Phil. Trans. R. Soc. Lond. B 1996 **351**, 913-920
doi: 10.1098/rstb.1996.0084

Separating the environmental and genetic factors that may be causes of bovine spongiform encephalopathy

C. M. Hau[1] and R. N. Curnow[2]

[1]*Department of Epidemiology and Public Health, Ninewells Hospital and Medical School, Dundee DD1 9SY, U.K.*
[2]*Department of Applied Statistics, The University of Reading, P.O. Box 240, Earley Gate, Reading RG6 6FN, U.K.*

SUMMARY

The initial cause of the bovine spongiform encephalopathy (BSE) epidemic is generally accepted to have been the feeding of infected animal protein to cattle. The proportion of animals affected in any year in a particular herd has generally been low. This suggests either considerable variation in the extent of challenge of the individual animals or variation in their susceptibility to challenge or both. There is known to be genetic variation in susceptibility in other spongiform encephalopathies, such as scrapie in sheep. However, earlier indications that there may be associations between the incidence of BSE in cattle and polymorphisms and mutations in the PrP gene have not been confirmed (Hunter *et al.* 1994). Here, we attempt to model the likely extent of challenge of the individual animals in five Holstein Friesian pedigree herds and also the distribution of incubation times to the date of clinical onset. By studying the incidence of the disease in related animals we first found that single locus genetic models fitted the data much better than a non-genetic model. This was the first statistical evidence found of genetic variation in susceptibility to BSE. A check on the model in which individual animals were randomly allocated to 'parents' showed that the result was due to the lack of allowance in the non-genetic model for those animals insufficiently challenged or, for non-genetic reasons, resistant to their level of challenge. Thus there is still no evidence, molecular or statistical, for genetic variation in susceptibility. The importance of checking the attribution of genetic effects in complex models by the random allocation of progeny to parents is clear.

1. INTRODUCTION

There is general, but not complete, agreement that the initial cause of the epidemic of bovine spongiform encephalopathy (BSE) was the consumption by cattle of sheep tissues affected by scrapie, following changes in the procedures for the production of concentrate feeds. A ban on the feeding of such protein to ruminants was issued by the Ministry of Agriculture, Fisheries and Food in July 1988 (Wilesmith *et al.* 1992*a*,*b*) and this has produced an appreciable reduction in the incidence of BSE. There are, however, sufficient cases in animals born since the ban to suggest that some other form of infection may now be operating (Hoinville 1994; Hoinville *et al.* 1995). Other forms of infection are horizontal or vertical transmission from animal to animal. Experiments and studies to date have not been successful in identifying the occurrence of transmissions of this kind (Hoinville *et al.* 1995). The statistical study described in this paper is based on an analysis of data on animals born before the feed ban. We shall assume that, for these animals, any infection was caused by the consumption of contaminated animal protein. Thus no allowance will be made for the possibility of horizontal or vertical transmission. For a recent review of the literature on BSE, see Schreuder (1994).

The main purpose of this study was to investigate whether the susceptibility of an animal to infection by contaminated feed was determined by the animal's genotype. Although there are certainly other transmissible encephalopathies in which genetic factors determine susceptibility or resistance (Prusiner 1992), earlier indications of associations between the incidence of BSE in cattle and polymorphisms and mutations in the PrP gene have not been confirmed (Hunter *et al.* 1994). Our statistical investigations of the occurrences of the disease in related animals were planned to highlight any evidence of genetic factors that could then be investigated at the molecular level. Many genetic models of susceptibility are possible. Only a single locus may be involved or many loci may play a part, in which case the effects at different loci could combine additively or non-additively. The occurrence of alleles at different loci may be associated and their effects may interact. Distinguishing between different genetic models is extremely difficult when, as with BSE, environmental factors, such as the extent of challenge, could be involved. Often, all that can be said is that there is some form of inheritance. Due to this lack of discrimination, this paper assumes that if there is genetic involvement then a single locus is involved. We shall assume that there are two different alleles segregating at this single locus. S is the allele for susceptibility and R is the allele for resistance. We studied two genetic models. In one susceptibility is

Phil. Trans. R. Soc. Lond. B (1996) **351**, 913–920
Printed in Great Britain

dominant so that individuals with genotypes SS and SR are susceptible and, in the other, susceptibility is recessive so that only individuals with genotype SS are susceptible. We assumed that mating is at random with respect to this locus and so the population frequencies of the three genotypes SS, SR and RR are q^2, $2pq$ and p^2 respectively, where q is the frequency of the allele for susceptibility, and $p = 1-q$ is the frequency of the allele for resistance. The basis for resistance may be resistance to challenge or may be a lengthening of incubation time to such an extent that it exceeds the lifetimes of the infected animals. We cannot distinguish between these two possible bases of resistance because there is, to date, no method of determining whether or not a cow is incubating the disease. The implications of the two mechanisms are, of course, very different, particularly if the incubating animal is, in any sense, or at any time, a possible source of infection.

We cannot assume that, in an affected herd, all the susceptible animals were sufficiently challenged by the infective agent that they will start incubating the disease. Further, the incubation time, that is the time from infection to clinical onset, of BSE can be as much as six or seven years and is certainly very variable (Wilesmith *et al.* 1992*a*,*b*; Wooldridge 1995). We cannot assume that animals which were last exposed to challenge many years ago are not now incubating the disease. These two factors, extent of challenge and length of incubation time, result in incomplete penetrance of the genotype and therefore need to be modelled if any genetic contribution is to be revealed and estimated. An alternative approach to investigating a possible genetic factor would be to compare the incidences of the disease in groups of related animals. This has, so far, failed to identify any genetic involvement in susceptibility to BSE. (Wijeratne & Curnow 1990; Wijeratne *et al.* 1991; Curnow *et al.* 1994; Curnow & Hau 1995).

Here, we construct a model for the extent of challenge and the distribution of incubation times. The model provided, for each animal, the probabilities of diagnosis of BSE in terms of its exposure to the contaminated feed and the time interval from first exposure to the feed to its date of last record. Data from five pedigree Holstein Friesian herds were used to estimate the parameters of the model and the parameters representing the importance of the genotype of the individual animal. All the animals included in the analysis were born before the feed ban in July 1988. Cases reported to us by the Central Veterinary Laboratory (CVL) up to the end of August 1993 are included in the analysis.

2. MODELLING THE EXTENT OF EXPOSURE AND THE DISTRIBUTION OF INCUBATION TIMES

The age of a cow at first exposure to the infected feed is not thought to be important in terms of extent of challenge or of likely incubation time (J. W. Wilesmith, personal communication). Assuming that infection occurs as a Poisson process at instantaneous rate λ over the whole period of exposure, t_E, we can measure the effective incubation time either as the expected time since first infection:

$$IT_1 = t - \frac{1}{\lambda} + \frac{t_E \exp(-\lambda t_E)}{1-\exp(-\lambda t_E)}$$

or the expected accumulated times initiated by all infections,

$$IT_2 = \lambda t_E (t - t_E/2).$$

The probability of diagnosis by time t after first exposure will be assumed to be a logistic function of one of these measures of incubation times, IT_1 or IT_2, so that Prob(diseased and diagnosed by time t) $= 0$ if $IT_i \leqslant k$, $(i = 1, 2)$,

$$= (IT_i - k)^\gamma / [m^\gamma + (IT_i - k)^\gamma] \quad \text{if } IT_i > k, \ (i = 1, 2),$$

where k is the value of IT below which disease cannot be diagnosed, m is the median value of $(IT-k)$ for infected animals, and γ determines the rate at which the probability of disease increases as IT moves to values above k. The slope at $IT = k$ is zero providing $\gamma > 1$. The probability tends to 1 as IT becomes very large as all infected animals will, if they survive long enough, be diagnosed BSE.

The time since first exposure and the length of exposure were used to calculate the probability that each animal would, by the time of its last record, be diagnosed as having BSE. The disease statuses of all the animals at last record then provided the likelihood of the data as a function of the four parameters in the model λ, k, m and γ. We analysed the data from each herd separately. The alternative was to pool the data over herds and then include terms in the model for differences between herds in incidence and, possibly, in allele frequency and measures of exposure. We preferred to fit and examine the fit of the models separately for each herd, looking for consistency over herds of the best genetic model and the estimated allele frequencies as indicators of the reliability of the conclusions.

We still have to take into account the genetic component of the model and this requires considering the familial relationships among the animals in each herd.

3. GENETIC MODEL AND LIKELIHOOD ANALYSIS

The additional parameter brought in by the genetic models is the frequency of the S allele, q. The non-genetic model is the special case where $q = 1$, which makes all animals susceptible. The genetic models divide the animals into two groups, one group is resistant and so will never have the disease. The parameters defined above λ, k, m and γ, therefore refer only to the susceptible animals. In the non-genetic model the parameters refer to all the animals because all are susceptible.

The likelihood analysis combines the environmental information and familial relationships for each animal

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

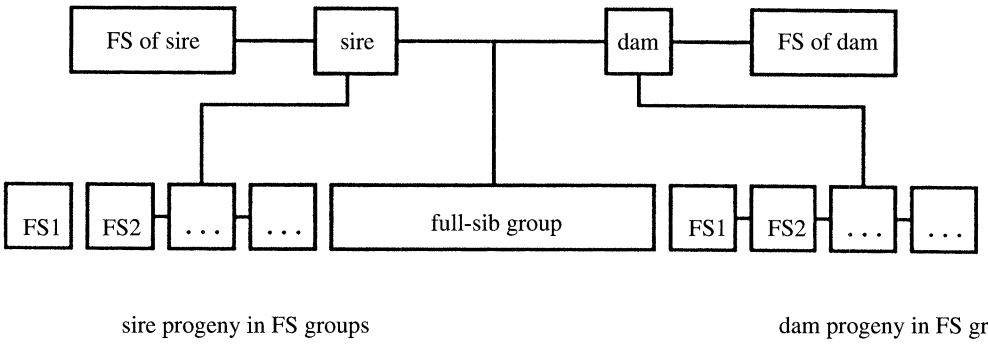

Figure 1. The related animals providing data on each full-sib group. FS indicates full-sib or full-sibs.

and requires evaluation of the likelihood of the disease status of each of the full-sib groups in the herd. The related individual animals providing data on each full-sib group are shown in figure 1.

For each full-sib group the required relatives were identified from data supplied by the Holstein Friesian Society of Great Britain and Ireland.

We first used, for each full-sib progeny group, the disease statuses and exposure histories of the sire, the dam, the half-sibs, and the full-sibs of the sire and dam, to update, using Bayes' theorem, the probabilities of each of the nine combinations of the three genotypes SS, SR and RR for the sire and dam. In the following derivation the probabilities for the disease status of an animal was always calculated using either the first or the cumulative infection model and the animal's exposure history. Writing $G_{S\times D}$ and G_R for the possible genotypes of the sire and dam and the other relatives, respectively, and similarly $B_{S\times D}$ and B_R for their disease statuses, we calculated the probabilities of each of the nine sire and dam genotype combinations given the information on their disease statuses and those of their relatives

$$P(G_{S\times D}\,|\,B_{S\times D}, B_R).$$

The disease statuses of related animals are independent given their genotypes and so we can use Bayes' theorem to write

$$P(G_{S\times D}\,|\,B_{S\times D}, B_R) = \frac{P(G_{S\times D})P(B_R\,|\,G_{S\times D})P(B_{S\times D}\,|\,G_{S\times D})}{P(B_{S\times D}, B_R)}. \quad (1)$$

In this equation, $P(G_{S\times D})$ is the probability, derived from the random mating assumption, of the genotypes of the sire and dam before anything is known of their disease status. $P(B_R\,|\,G_{S\times D})$ in (1) needs to be written in relation to the probabilities of all the possible genotypes for the relatives given the genotypes of the sire and dam,

$$P(B_R\,|\,G_{S\times D}) = \sum_{G_R} P(B_R\,|\,G_R)P(G_R\,|\,G_{S\times D}).$$

All the terms in (1) are calculable from the familial relationships of the animals, $P(G_R\,|\,G_{S\times D})$, and the model for exposure and incubation time, $P(B\,|\,G)$. If the susceptibility allele is dominant, for example, the latter term for an animal with BSE will be:

$$P(B\,|\,G = SS \text{ or } SR) = 0$$

if $IT_i \leqslant k (i = 1,2)$

$$= \frac{(IT_i - k)^\gamma}{m^\gamma + (IT_i - k)^\gamma}$$

if $IT_i > k (i = 1,2)$

$$P(B\,|\,G = RR) = 0$$

The denominator in (1) can best be calculated by summing the numerators for the nine combinations of sire and dam genotypes.

For each of the nine combinations of sire and dam genotypes, the probabilities of the possible genotypes of the full-sib group, $P(G_{FS}\,|\,G_{S\times D})$, are calculated. The likelihood of the disease status of the full-sib group is then calculated as

$$\sum_{G_{FS}G_{S\times D}} P(B_{FS}\,|\,G_{FS})\,P(G_{FS}\,|\,G_{S\times D})\,P(G_{S\times D}\,|\,B_{S\times D}, B_R). \quad (2)$$

There was one affected bull in the five data sets analysed, the probabilities of the three genotypes for the sire $P(G_s\,|\,B_s)$ were allocated as follows:

	RR	SR	SS
recessive model	0	0	1
dominance model	0	1/2	1/2
non-genetic model	1/3	1/3	1/3

Because their exposure histories were not known, the fact that a sire was not BSE affected at last record was not used to estimate the probabilities of its own genotype.

The likelihoods (2) for each full-sib group in the herd were multiplied together to provide an overall likelihood for the herd.

We investigated the likelihood surface by adaptive searches over grids of the values of the four parameters, λ, q, k and γ. For each parameter set we estimated m by the median value of $(IT_i - k)$, $(i = 1,2)$ for the infected animals in the herd. This may provide slight underestimates of m because animals with long incubation times may not yet have been diagnosed. We found γ to

be a relatively unimportant parameter in terms of its effect on the estimates of the other more important parameters.

If there is evidence of inheritance of susceptibility, a crucial concern will be the estimated values of λ and q. Are we dealing with a high infection rate and low susceptibility or with a low infection rate and high susceptibility?

4. RESULTS

Table 1 shows, for each of the five herds studied, the numbers of animals: male, female, calved, affected, affected but not calved, affected but not exposed and the total number in the analysis. Unaffected females that had not calved were excluded from the analysis because they were generally young animals and even if incubating the disease, unlikely to have reached the stage of diagnosis. The small number of affected but presumed not exposed animals and one identified affected animal in herd 4 with an unknown date of disease onset were omitted from the analysis. The Med t_E column gives, for each herd, the estimated median length of time that the animals in the analyses were exposed to the contaminated feed. Nearly 90% of the animals had no full-sibs. The distribution of the remaining animals to full-sib progeny groups of various sizes is shown in the final three columns of table 1.

Tables 2*a*, *b* show the best fitting models for the cumulative and first infection models respectively. On the first line, for each herd and model, LR is the ratio of the maximized likelihood under the genetic model to the maximized likelihood under the non-genetic model. The following rows show the estimated values of the parameters λ, γ, and q. The final row shows the proportions of susceptible animals, $2q-q^2$, q^2 and 1 for the dominant, recessive and non-genetic models, respectively.

The likelihood ratios, LR, are only approximate because many cows will appear more than once in the analysis, once in the proband full-sib group and many times, perhaps, as sire or dam of another full-sib group or relatives of these sires or dams. Also the comparison is with the non-genetic model which corresponds to a point on the edge of the parameter space, $q = 1$. However, the LR are consistently larger than 1, often appreciably so. There is no proper way to compare the likelihoods for the two infection models because one is not nested within the other. However, the maximized likelihoods for the first infection model always exceed those for the cumulative infection model. There is no consistency in the relative values of the likelihoods for the dominant and recessive models.

The likelihood surfaces showed, as expected, that separating the extent of challenge, λ, and the frequency of susceptible animals, depending on q, is difficult. This may explain why the values of λ and q vary so substantially from herd-to-herd, often, as expected, with high values of λ being associated with low values of q and vice versa.

These analyses based on modelling the exposure times and incubation times are the only analyses we

Table 1. *Summary of information available about the five herds*

herd number	number of males 1	number of females 2	number calved 3	numbers of affected & not calved 4	not exposed 5	total in analysis 1+3 +4−5	number of affected	number with onset known	med te (years)	number of families with 2 fullsibs	number of families with 3 fullsibs	number of families with 4 fullsibs
1	13	1003	611	1	0	625	69 (64 identified	64	3.79	37	3	0
2	0	664	407	0	0	407	38 (33 identified)	33	2.41	4	0	0
3	5	690	469	6	2	478	99 (74 identified)	74	1.79	21	0	0
4	3	551	322	6	0	331	39 (36 identified)	35	2.34	38	6	2
5	0	647	367	5	3	369	59 (56 identified)	56	0.92	16	0	0

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 2*a*. *Likelihoods for the five herds*: *cumulative infection model*

(In the cases where the parameter is given as > x no maximum value could be found, the likelihood still increasing at *x*.)

		genetic model		
herd		dominant	recessive	non-genetic
1	*LR*	3.50×10^{4}	1.83×10^{2}	1
	λ	0.15	0.1	0.05
	γ	1.5	2.2	1.8
	q	0.27	0.8	1
	freq susc	0.47	0.64	1
2	*LR*	1.09×10^{17}	3.33×10^{20}	1
	λ	> 10	> 10	0.03
	γ	1.08	1.2	1.4
	q	0.14	0.44	1
	freq susc	0.26	0.19	1
3	*LR*	2.56×10^{24}	5.43×10^{17}	1
	λ	0.25	0.35	0.05
	γ	1.8	2.0	2.1
	q	0.04	0.6	1
	freq susc	0.20	0.36	1
4	*LR*	1.53×10^{7}	5.75×10^{11}	1
	λ	0.2	20	0.03
	γ	1.2	1.25	1.05
	q	0.17	0.29	1
	freq susc	0.31	0.08	1
5	*LR*	2.73×10^{25}	3.49×10^{36}	1
	λ	> 20	> 20	0.04
	γ	1.01	1.4	1.5
	q	0.02	0.51	1
	freq susc	0.04	0.26	1

Table 2*b*. *Likelihoods for the five herds*: *genetic model* (*first infection model*)

(In the cases where the parameter is given as > *x* no maximum value could be found, the likelihood still increasing at *x*.)

		genetic model		
herd		dominant	recessive	non-genetic
1	*LR*	1.19×10^{3}	5.11×10^{1}	1
	λ	0.15	0.12	0.06
	γ	2.0	2.0	2.2
	q	0.3	0.8	1
	freq susc	0.51	0.64	1
2	*LR*	7.31×10^{7}	1.17×10^{11}	1
	λ	0.25	0.55	0.04
	γ	2.5	2.0	1.8
	q	0.12	0.44	1
	freq susc	0.23	0.19	1
3	*LR*	2.33×10^{19}	5.95×10^{23}	1
	λ	0.2	3.0	0.06
	γ	3.0	3.5	3.0
	q	0.06	0.43	1
	freq susc	0.12	0.19	1
4	*LR*	1.75×10^{7}	5.21×10^{4}	1
	λ	0.3	0.35	0.04
	γ	2.0	2.8	2.0
	q	0.16	0.38	1
	freq susc	0.29	0.14	1
5	*LR*	5.55×10^{16}	8.52×10^{28}	1
	λ	1.0	> 30	0.05
	γ	1.6	3.0	4.5
	q	0.03	0.46	1
	freq susc	0.06	0.22	1

have attempted that have suggested genetic differences in susceptibility (Wijeratne & Curnow 1990; Wijeratne *et al*. 1991; Curnow *et al*. 1994; Curnow & Hau 1995). The program to calculate the likelihoods is extremely complex. It has been checked for internal inconsistencies and tested extensively on small but wide-ranging artificial data sets. There remains a possibility that the higher likelihoods are artefacts, being a consequence of inadequacies in the modelling of the extent of challenge and the length of incubation period. These inadequacies may generate apparent linkages between progeny and parents that are then erroneously attributed to the inheritance of susceptibility. In the next section we describe a check on our results by a random labelling technique of the kind commonly used to test the statistical significance of spatial associations between different sets of objects, for example pollution sources and cases of a disease.

5. RANDOM LABELLING ANALYSIS OF FULL-SIB PROGENY GROUPS

We formed, within each herd, new progeny groups by randomly allocating to each sire and dam combination the correct number of 'progeny' where the 'progeny' are matched with the real progeny as closely as possible on the basis of year of birth and year of last record. They were chosen to have no close genetic relationship to the sire or dam. The likelihoods for each of the three models – dominant, recessive, non-genetic – were then calculated for this allocation using the first infection model and the parameters estimated from the real data. If there is inheritance of susceptibility and/or maternal transmission then this likelihood should be less than with the real data where the groupings and calculations do associate progeny with their real parents. The calculations were repeated for 25 random allocations. To avoid excessive computations individual animals in each random allocation could be allocated to more than one progeny group. Understandably therefore, the likelihoods are influenced very much by the number of affected animals included in each allocation. Figure 2 shows, for each of the three genetic models, the 25 changes in log likelihood from that with the real data for herd 1, plotted against the number of affected animals allocated in each of the simulations. The effect of the number of affected animals allocated is clear. There were 64 affected animals in herd 1 and, for all three models, there is virtually no difference between the random labelling and real data log likelihoods when the random labelling simulations contained about 64 affected animals. The log likelihoods for the simulated data would be further increased if new estimates were calculated for each simulation. This pattern was repeated for three of the other four herds. One herd showed a different pattern of relationship to the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

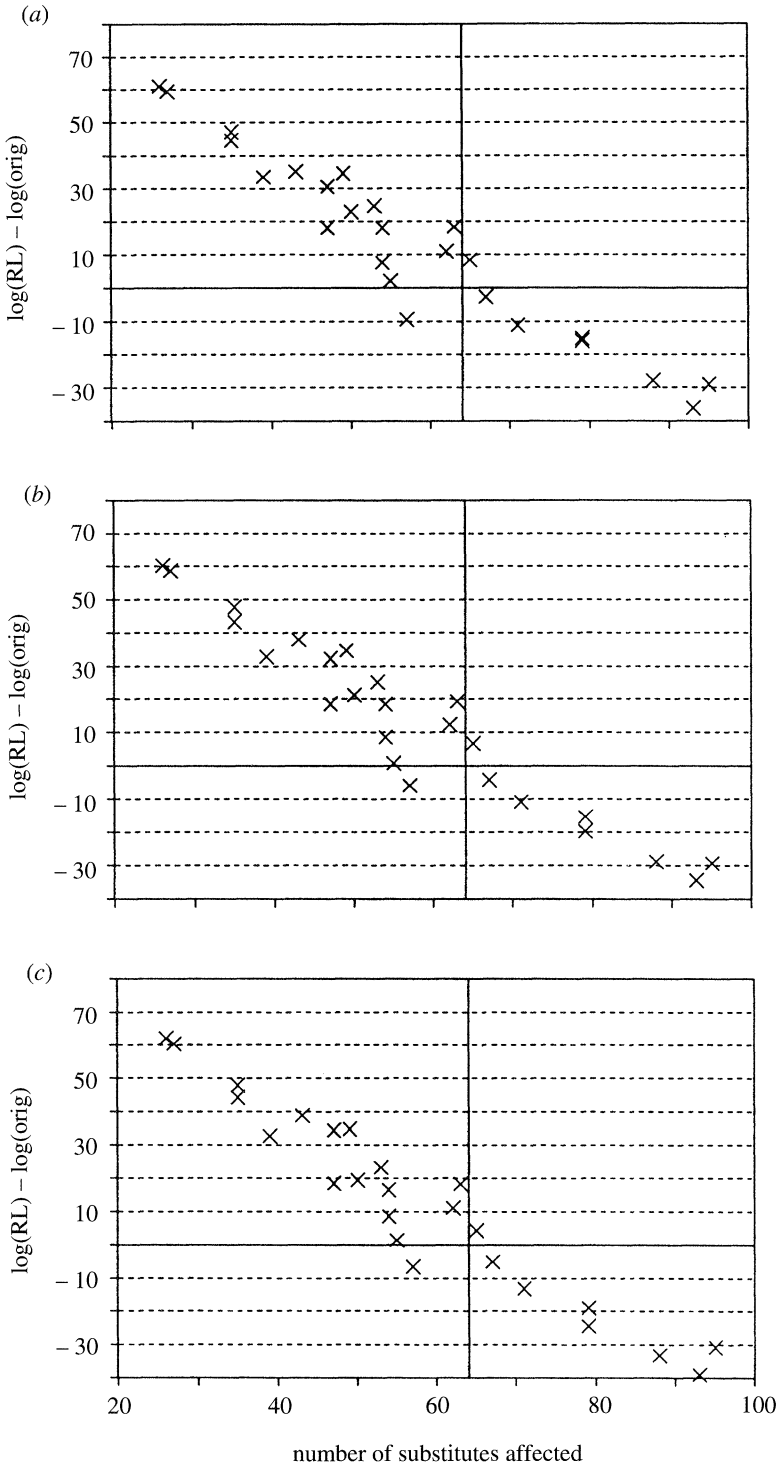

Figure 2. Random-labelling (RL) tests for herd 1. Plots of [log likelihood (RL) – log likelihood (real data)] against number of substitutes that were affected. Solid vertical lines indicate the number of affected animals in the real dataset. (*a*) Dominant model; (*b*) recessive model; and (*c*) non-genetic model.

numbers affected in the random samples but, again, no evidence of lower likelihoods than for the real data. Clearly the randomly labelled data sets have similar likelihoods to the real data sets when account is taken of the number of affected animals allocated in the artificial data sets.

The higher likelihoods associated with the genetic models when the full-sib progeny groups are allocated randomly to the parents could be due to the matching of the allocated 'progeny' to the real progeny according to year of birth and year of last record. This would be an artefact resulting from inadequacies in the modelling of exposure and incubation times. To check this we have randomly allocated the progeny without matching according to year of birth or year of last record but have avoided allocating progeny to real parents. The necessary randomizations are now much more expensive in computer time and so we studied

Table 3. *Likelihood values*, L, *for herd 3 with random allocation of progeny to parents*

	genetic model					
	dominant		recessive		non-genetic	
run	no. affected allocated	L	no. affected allocated	L	no. affected allocated	L
1	18	0.208×10^{-53}	17	0.472×10^{-34}	17	0.133×10^{-70}
2	24	0.109×10^{-59}	16	0.589×10^{-30}	4	0.194×10^{-31}
3	14	0.903×10^{-40}	17	0.307×10^{-33}	23	0.426×10^{-91}
4	13	0.396×10^{-35}	17	0.222×10^{-33}	21	0.107×10^{-78}
5	21	0.339×10^{-60}	23	0.179×10^{-42}	17	0.106×10^{-69}
real data	18	0.727×10^{-46}	18	0.453×10^{-35}	18	0.658×10^{-74}

118 of the 478 animals in one herd only, herd 3, and carried out five instead of 25 simulation runs for each type of model. The likelihood values obtained are given in table 3. We see that the likelihoods are generally larger for the genetic models and so the anomaly cannot be explained by the matching process used in the earlier simulations. For the same number of affected animals, the likelihoods are similar to those for the real data.

The explanation of the better fit of the genetic models is almost certainly that the genetic models allow a proportion of the cows to be resistant to challenge whereas the non-genetic model assumes that all exposed animals will, if they live long enough, be diagnosed BSE. Thus an animal with a long exposure time, t_E, and still alive a long time after exposure will, with the genetic model, be classified with high probability as genetically resistant to challenge. With the non-genetic model, in which all exposed animals are assumed infected, this animal and all similar animals, will contribute a small value to the likelihood. As the majority of animals are not diagnosed BSE and many of them have long exposure times and live several years after first exposure, the likelihood for a whole herd is therefore much reduced when a non-genetic model is fitted.

6. AN IMPROVED NON-GENETIC MODEL

The animals we have studied were in 'exposed' herds. However, individual animals may not have been sufficiently challenged by the contaminated feed to start incubating the disease. The extent of challenge required to initiate incubation in an individual animal may depend on the state of the animal at the time of challenge. The non-genetic model for the probability of disease now to be studied allows, by introducing a factor P, for the sufficiency, or otherwise, of the extent of challenge and for any non-genetically determined factors representing the susceptibility of the animal at any times of challenge.

$$\text{Prob (diseased by time } t) = P(t-t_0)^\gamma/[m^\gamma + (t-t_0)^\gamma].$$

In this model we have made the further simplification, based on previous work, that the actual extent of exposure is unimportant in relation to incubation time. As before, γ is a parameter representing the rate of increase in incidence as t exceeds the threshold value t_0. We have estimated t_0 as the actual observed minimum time to the disease after exposure and we have then omitted this animal from the analysis. This approximation avoids the estimation of the minimum observed incubation time minus the minimum possible incubation time. The value of m, the median value of $(t-t_0)$ for susceptible and challenged animals, was estimated as a multiple of the median value of $(t-t_0)$ for the animals observed to be affected in each of the herds. The log likelihood is

$$L = d\ln m + \ln P - 2\sum_{d} \ln(M+T) + \sum_{nd} \ln[1 - PT/(M+T)],$$

where d is the number of diseased animals minus 1, the animal that provided the t_0 value; $\sum_{d}$ *and* $\sum_{nd}$ indicate summation over the diseased and not diseased animals respectively, $T = (t-t_0)^\gamma$ and $M = m^\gamma$.

The likelihoods of the data for all five herds were found to be maximized by setting $\gamma = 1$. This corresponds to a rapid increase in disease rate as time since first exposure exceeds the minimum value, t_0.

Table 4 shows the maximizing values of m as a fraction of the median time since first exposure of the observed affected animals and the maximizing values of P, the proportion of animals sufficiently challenged to start incubating the disease. The optimization was over a coarse grid of values of m/median and P. The main conclusion is that, if variability in the effect of extent of exposure is ignored, about 30 or 40% of

Table 4. *Estimates of parameters for improved non-genetic model with* $\gamma = 1$

(P is the proportion of animals sufficiently challenged to start incubating the disease.)

		estimated values	
herd	median value $(t-t_0)$ (months)	m/median	P
1	61	1.2	.4
2	61	1.4	.3
3	56	1.4	.4
4	61	1.6	.3
5	55	1.0	.3

animals were sufficiently challenged to start incubating the disease and that the median time to diagnosis of these animals varied from about 55 months (herd 5) to 61 months (herd 4). The larger median times are almost certainly overestimates with the likelihoods being very insensitive to the choice of values of m/median.

The split of the 30 or 40 % between the effects of the extent of challenge and any environmentally determined susceptibility at the time of challenge is difficult to estimate. The relationship of incidence to length of time exposed is weak, partly because of difficulties in assessing time exposed but also because it may not be the relevant measure of challenge. For this reason, we can only report that, for the non-genetic model and ignoring exposure times, our best estimate of the proportion of animals sufficiently challenged to be incubating the disease is, for all five herds, about 30 or 40 %.

7. CONCLUSIONS

We conclude that the non-genetic models in tables 2*a*, *b* do not fit the data from the five herds as well as the genetic models not because susceptibility is inherited but because the non-genetic models do not include a term representing the proportion of animals sufficiently challenged to be incubating the disease. This proportion may be of the order of 30 or 40 %. The modelling described in this paper provides no evidence that the susceptibility of an animal to infection by contaminated feed is influenced by the animal's genotype and therefore confirms previous work by the authors. We emphasise the importance in complex modelling of checking, by the use of randomly allocated data sets, the attribution of genetic causes to familial correlations.

This work formed part of MAFF Project CSA 2428. We thank Mr J. W. Wilesmith (CVL) for his cooperation in discussing this project at various stages and informing us of recent findings relevant to our study; Miss J. Ryan (CVL) for providing information on BSE-affected animals; Mr M. Coffey and his staff (Holstein Friesian Society of Great Britain and Ireland) for supplying pedigrees, fact sheets and other useful information; Ms A. Crew and other staff (National Milk Records, Milk Marketing Board) for calving and pedigree information on milk recorded Holstein Friesian cows; Mrs M. Collins and Mr K. H. Freeman for computing assistance and Mrs M. Black for secretarial assistance.

REFERENCES

Curnow, R. N., Wijeratne, W. V. S. & Hau, C. M. 1993 The inheritance of susceptibility to BSE. *Proceedings of European commission consultation on transmissible spongiform encephalopathies* (ed. R. Bradley & B. Marchant), pp. 109–124. Brussels.

Curnow, R. N. & Hau, C. M. 1995 The incidence of bovine spongiform encephalopathy in the progeny of affected sires and dams. *Vet. Rec.* (In the press.)

Hoinville, L. J. 1994 Decline in the incidence of BSE in cattle born after the introduction of the 'feed ban'. *Vet. Rec.* **134**, 274–5.

Hoinville, L. J., Wilesmith, J. R. & Richards, M. S. 1995 An investigation of risk factors for cases of bovine spongiform encephalopathy born after the introduction of the 'feed ban'. *Vet. Rec.* **136**, 312–318.

Hunter, N., Goldmann, W., Smith, G. & Hope, J. 1994 Frequencies of PrP gene variants in healthy cattle and cattle with B.S.E. in Scotland. *Vet. Rec.* **135**, 400–403.

Prusiner, S. B. 1992 Molecular biology and genetics of neurodegenerative diseases caused by prions. *Adv. Virus Res.* **41**, 241–280.

Schreuder, B. E. C. 1994 Animal spongiform encephalopathies – An update Part II. Bovine spongiform encelphalopathy (BSE). *Vet. Quart.* **3**, 18–192.

Wijeratne, W. V. S. & Curnow, R. N. 1990 A study of the inheritance of susceptibility to bovine spongiform encephalopathy. *Vet. Rec.* **126**, 5–8.

Wijeratne, W. V. S., Curnow, R. N. & Hudson, C. M. 1991 A study of the population genetics of bovine spongiform encephalopathy (BSE). Sub-acute Spongiform Encephalopathies (ed. R. Bradley *et al.*), pp. 221–223. Brussels.

Wilesmith, J. W., Ryan, J. B. M., Hueston, W. D. & Hoinville, L. J. 1992*a* Bovine spongiform encephalopathy: epidemiological features 1985 to 1990. *Vet. Rec.* **130**, 90–94.

Wilesmith, J. W., Hoinville, L. J., Ryan, J. B. M. & Sayers, A. R. 1992*b* Bovine spongiform encephalopathy: aspects of the clinical picture and analyses of possible changes 1986–1990. *Vet. Rec.* **130**, 197–201.

Wooldridge, M. J. A. 1995 A study of the incubation period, or age at onset, of the transmissible spongiform encephalopathies/prion diseases. Ph.D. thesis, University of London.

Received 7 December 1995; accepted 17 January 1996

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES