A genetic interpretation of heightened risk of BSE in offspring of affected dams

NEIL M. FERGUSON*, CHRISTL A. DONNELLY,
MARK E. J. WOOLHOUSE AND ROY M. ANDERSON

Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK (neil.ferguson@zoo.ox.ac.uk)

SUMMARY

An analysis is presented of the results of a cohort study designed to test whether or not the aetiological agent of bovine spongiform encephalopathy (BSE) in cattle can be transmitted maternally (vertically) from dam to offspring. Various genetic models are fitted to the data under the assumption that the results could be explained entirely by genetic predisposition to disease (as opposed to maternal transmission) given exposure of offspring of diseased and unaffected dams to contaminated cattle feed. The analyses suggest that the results could be explained by the hypothesis of genetic predisposition, provided a large difference exists in the susceptibility of resistant and susceptible hosts, and explore the ranges of genotypic parameters and frequencies consistent with the limited currently available data. The results presented are broadly robust, even under the scenario that a portion of the observed maternally enhanced risk of BSE is due to a low level of maternal transmission in late incubation stage dams.

1. INTRODUCTION

The epidemic of bovine spongiform encephalopathy (BSE) in British cattle has had a very serious effect both on animal health and the European agriculture industry. The pattern of the epidemic has been documented in detail since the first diagnosis of the disease in November 1986 (Wilesmith *et al.* 1991, 1992; Kimberlin & Wilesmith 1994). Recent research using data on the demography of the cattle population and information on the incubation period of the disease has helped to identify the key epidemiological processes that have shaped the pattern of the epidemic in Great Britain (Anderson *et al.* 1996). However, many uncertainties remain concerning the determinants of transmission, including age-dependent susceptibility to infection with the aetiological agent (believed to be an abnormal form of the prion protein (PrP) (Prusiner *et al.* 1996)), the likelihoods of maternal and horizontal transmission in the absence of exposure to contaminated cattle feed, and the role of cattle genotype in influencing both susceptibility to infection and pathogenesis (the length of the incubation period) after infection (Fraser *et al.* 1992; Curnow *et al.* 1994; Hau & Curnow 1996; Curnow & Hau 1996; Neibergs *et al.* 1994; Hunter *et al.* 1994).

The existence of maternal transmission has been controversial following the release of interim results from a pair-matched cohort study of infection in calves born to dams that did or did not develop clinical signs

of BSE by the Spongiform Encephalopathy Advisory Committee (SEAC) of the Department of Health and Ministry of Agriculture, Fisheries and Food in the UK in July 1996 (SEAC 1996; Anderson *et al.* 1996; Ridley & Baker 1996). Evidence for maternal (or vertical) transmission of other transmissible spongiform encephalopathies (TSEs) such as scrapie and kuru in rodent models, sheep, goats and humans also remains inconclusive at present. However, the issue is one of great importance given recent evidence that the new variant Creutzfeldt-Jakob disease (nvCJD) in humans most probably arose via exposure to the aetiological agent of BSE (Collinge *et al.* 1996).

It is not our intention in this paper to determine the correct genetic model (if any) of BSE susceptibility in cattle; to do so would require analysis of detailed sequence and/or progeny data. Instead, we explore a range of genetic models to determine the genotypic parameter ranges compatible with the cohort study, under the assumption of stationary genotype frequencies. Some progeny analyses have been undertaken (Curnow *et al.* 1994; Hau & Curnow 1996; Curnow & Hau 1996) to examine the possibility of two classes of animals (susceptible and resistant) within five Holstein Friesian pedigree herds. However, the power of these studies was not sufficient, given the relatively small numbers of cases involved, to distinguish between alternative genetic models. Hence, until larger scale studies are performed, the role of analyses of the type presented in this paper is to define better the potential magnitude of genetic heterogeneity in BSE susceptibility within the UK cattle population.

* Author for correspondence.

We examine the results of the cohort study for the detection of maternal transmission in the light of the suggestion that they could have arisen (at least in part) via genetic predisposition to infection as opposed to the maternal transmission of an aetiological agent (Ridley & Baker 1996). Our aim is to ascertain the frequency of any highly susceptible genotype in the UK cattle herd for a range of genetic models, given epidemiological data on BSE cases and the results of the maternal cohort study.

2. GENETIC PREDISPOSITION TO TSEs

It is well known that host genotype has a major influence on susceptibility to TSEs and the typical duration of the incubation period of the disease following infection by a specific route (SEAC 1995; Prusiner *et al.* 1996). For example, the incubation period of BSE and scrapie in mice that have been fed infected material is strongly dependent on the breed and *Sinc* genotype of the host (Fraser *et al.* 1992). Similarly, Hunter *et al.* (1996) have documented disease-linked PrP gene polymorphisms in a closed flock of sheep in which scrapie is endemic, and associations of such polymorphisms (at codons 136, 154 and 171) with susceptibility to infection and the typical duration of the incubation period. The evidence of an important role for host genotype in BSE infection and pathogenesis in cattle is less clear-cut. To date, published experimental studies of the incubation period in cattle do not reveal differences in either susceptibility or incubation period between breeds (Hunter *et al.* 1994). Furthermore, Martin *et al.* (1991) genotyped 103 animals for the PrP gene by length polymorphisms, 16 of which subsequently developed BSE, and failed to find an association between genotype and BSE. However, a recent and more detailed study has revealed polymorphisms in the bovine prion gene. The work of Neibergs *et al.* (1994) suggests that BSE-affected cattle and their relatives are more likely to have a particular homozygous genotype than unrelated non-BSE animals of the same breed or animals of different breeds. The contradictory nature of the few published studies argues strongly for further work on this issue. However, epidemiological analysis of the fate of offspring of affected dams and sires may augment the limited data available, as will more detailed analyses of the maternal cohort study of BSE in calves born to affected and non-affected dams.

3. MATERNAL TRANSMISSION COHORT STUDY

Interim results from a cohort study of the fate of offspring born to affected and non-affected dams were made public by SEAC in mid-1996 (SEAC 1996). The study was begun in July 1989 and has now been completed (SEAC 1997; Wilesmith *et al.* 1997). The offspring of dams that subsequently developed BSE were recruited into the study as a maternally exposed group, and for each animal, a matched control animal was selected which was born in the same herd in the same calving season and whose dam had reached at least six years of age without developing clinical signs of BSE. Animals

recruited into the study were born as early as 1987 and as late as 1989. Almost all of the maternally exposed animals were born within five months of the onset of clinical signs of BSE in the dam. All recruited animals were monitored until they either developed BSE or reached seven years of age. There were 55 cases of BSE among the recruited animals, of which 42 were in the maternally exposed group and 13 in the control group (table 1) among the 301 pairs of animals. The presence of infection in the control group may be interpreted in a variety of ways. First, the calves born to the control group dams could have been fed contaminated feed in their early years of life. This seems likely since recent analyses estimate that some cattle continued to acquire infection via this route at least up to the early part of 1994 (Anderson *et al.* 1996; Ferguson *et al.* 1997). A second explanation is that the dams of these calves would have developed BSE if they had survived or been monitored for more than the six-year period prescribed in the study design. However, independent of cause, the occurrence of BSE in the maternally exposed group is significantly different from that in the control group. If the difference is interpreted as being solely due to maternal transmission of the aetiological agent, then the rate of transmission is 9.6% (95% confidence interval, 5.1–14.2%) over the later stages of the maternal incubation period (including after disease onset). Nothing can be said about the likelihood of transmission for the early stages of the maternal incubation period due to the design of the study.

Because disease occurred in the control group animals (probably due to exposure to contaminated feed), another explanation of the recorded difference between the two groups is possible, namely a selection bias with underlying genetic predisposition to disease.

Detailed statistical analyses of data from the cohort study have now been completed (Donnelly *et al.* 1997b; Curnow *et al.* 1997; Gore *et al.* 1997), and indicate that whilst there is evidence of low level maternal transmission in late incubation stage animals, genetic variability in BSE susceptibility may also play a role. Indeed, the best fitting survival analysis model (Donnelly *et al.* 1997b) provides estimates of 8.2% maternal transmission occurring for the last 88 days of the maternal incubation period (and after disease onset in the dam), together with a 2.71-fold susceptibility difference between the maternally exposed and control group calves. By comparison, the susceptibility ratio between the two groups, assuming no maternal transmission, can be simply calculated to be $42/13 = 3.23$. Hence, whilst for simplicity we assume zero maternal transmission in the analyses presented below, our conclusions are likely to remain broadly robust even when some of the observed maternally enhanced risk of BSE is explained by a low level of maternal transmission.

4. AN INHERITANCE MODEL OF SUSCEPTIBILITY

To start, we ignore the details of any possible genetic basis of susceptibility to infection that leads to BSE, and consider the situation in which a proportion, f , of the cattle herd are highly susceptible and a proportion

Table 1. Final results of the maternal cohort study (SEAC 1997; Wilesmith *et al.* 1997).

	maternally exposed	control
number of cases of BSE	42	13
calves not subsequently developing BSE (up to age 7 years)	259	288
total	301	301

(1 - f) are of low susceptibility (with no intermediates). To keep the analysis general in form we define the susceptibility (the fraction of maternally exposed animals that develop the disease) of the 'high' class as a factor s greater than the susceptibility of the 'low' class (set to equal unity). Furthermore, to continue to retain generality, we define the likelihood that susceptibility is inherited as p_{ij} where this term defines the probability that a dam in susceptibility class j gives birth to a calf in susceptibility class i (where subscripts H and L denote, respectively, 'high' and 'low' susceptibility). The cattle population has stationary genotype frequencies provided

$$\mathbf{P}_{eq} = \mathbf{G}\mathbf{P}_{eq}$$

where $\mathbf{G}$ is a matrix (we call it the generation matrix) of the probabilities of inheritance

$$\mathbf{G} = \begin{pmatrix} p_{H|H} & p_{H|L} \\ 1 - p_{H|H} & 1 - p_{H|L} \end{pmatrix}$$

and $\mathbf{P}_{eq}$ is the eigenvector of $\mathbf{G}$ corresponding to the unit eigenvalue. At equilibrium, the trivial condition

$$f = fp_{H|H} + (1 - f)p_{H|L} \quad (1)$$

must be satisfied, and hence the matrix $\mathbf{G}$ can be redefined in terms of $p_{H|H}$ and the proportion of the herd of high susceptibility f .

The dams selected for the maternal cohort study were from different birth cohorts, and so had differing degrees of exposure to contaminated feed. (As in Anderson *et al.* (1996), annual birth cohorts were defined so that, for example, the 1989 cohort consists of cattle born between 1 July 1988 and 30 June 1989.) The estimated pattern of change in the exposure risk has been documented recently (Anderson *et al.* 1996; Ferguson *et al.* 1997). We define the cumulative infection hazard c_i for the i th cohort of dams in the usual way, such that the proportion of the low susceptibility class remaining susceptible after exposure in cohort i is $x_i = e^{-c_i}$. The proportion of dams in cohort i infected, I_i is therefore given by

$$I_i = \mathbf{S}(x_i)\mathbf{P}_{eq}, \quad (2)$$

where $\mathbf{S}$ is the vector of attack rates for each cohort, corresponding to the high and low susceptibility classes within the population, respectively,

$$\mathbf{S}(x_i) = \begin{pmatrix} 1 - x_i^s \\ 1 - x_i \end{pmatrix}, \quad (3)$$

where s is the ratio of susceptibility defined earlier. We estimate the fraction of the i th yearly birth cohort infected, I_i , from the number of confirmed BSE cases arising from that cohort adjusted for the survival function as described and utilized in Anderson *et al.* (1996), Donnelly *et al.* (1997a) and Ferguson *et al.* (1997). For a given value of s , x_i can thus be estimated from equations (2) and (3).

Turning specifically to the dams of the animals recruited into the 'maternally exposed' and 'control' arms of the cohort study, these arise from a number of different cohorts and can arise from both high and low susceptibility classes. The distribution of high (H) and low (L) susceptibility dams whose offspring were in those recruited to the control arm (C) can be written as the vector

$$\mathbf{P}_C = \frac{\sum_{i=t-20}^{t-2} a_i \mathbf{P}_{eq}^T \mathbf{I}(\mathbf{1} - \mathbf{S}(x_i))}{\sum_{i=t-20}^{t-2} a_i \mathbf{P}_{eq}^T (\mathbf{1} - \mathbf{S}(x_i))} \quad (4)$$

where t is time of recruitment to the study, a_i is the probability that an animal alive at time t is from the i th cohort, $\mathbf{1}$ is a vector of ones and $\mathbf{I}$ is the identity matrix. The limits 2–20 years define possible ages when dams could in principle produce calves. Similarly, the distribution for dams whose offspring were recruited to the maternally exposed (E) arm is given by

$$\mathbf{P}_E = \frac{\sum_{i=t-20}^{t-2} b_i \mathbf{P}_{eq}^T \mathbf{S}(x_i)}{\sum_{i=t-20}^{t-2} b_i \mathbf{P}_{eq}^T (\mathbf{S}(x_i))}, \quad (5)$$

where b_i is the probability that an animal which was a case at time t is from the i th cohort.

The proportions of control and maternally exposed animals in the high susceptibility class (H) depend on the proportions in their dams and the inheritance probabilities. The vector of proportions of H and L in the control group is given by

$$\mathbf{K}_C = \mathbf{G}\mathbf{P}_C, \quad (6)$$

and for the maternally exposed group

$$\mathbf{K}_E = \mathbf{G}\mathbf{P}_E. \quad (7)$$

The prevalence of BSE cases in the control group calves, I_C , is expected to be

$$I_C = \mathbf{S}(y)^T \mathbf{K}_C, \quad (8)$$

and for the maternally exposed group the prevalence, I_E , is expected to be

$$I_E = \mathbf{S}(y)^T \mathbf{K}_E, \quad (9)$$

where, as before, $\mathbf{S}$ is the vector of attack rates for the recruited animals. The attack rate term y can be implicitly solved for in terms of s , f , $p_{H|H}$ and I_C using equation (8). The minimum value of s (the susceptibility ratio) that is consistent with a specified value of f is found as a function of $p_{H|H}$ and subject to the constraint equation (9). Assuming that all maternal infection probabilities I_i equal I_M , the minimum s with respect to f as a function of $p_{H|H}$ and subject to equation (9) can be plotted as a function of I_M and $p_{H|H}$.

As the infection rates in the dams and the calves become increasingly small, the genotype frequencies in

the control arm of the study are expected to approximately equal the population frequencies. Making this approximation, there is a closed-form solution for the relative susceptibility factor s . Mathematically, this limit is obtained when $x_i \rightarrow 1$ and $y \rightarrow 1$. The constraint on the minimization of s is then

$$\frac{\mathbf{S}(y)^T \mathbf{G} \mathbf{P}_E}{\mathbf{S}(y)^T \mathbf{G} \mathbf{P}_C} = R, \quad (10)$$

where R is the observed ratio of prevalences in the maternally exposed and control groups. The ratio R can be shown to equal

$$R = \frac{s^2 f p_{HH} + 2sf(1 - p_{HH}) + (1 - 2f + p_{HH}f)}{(1 + (s - 1)f)^2}.$$

Thus, the relative susceptibility of the high class can be shown to equal

$$s = 1 + \frac{R - 1 + \sqrt{(R - 1)(p_{HH}/f - 1)}}{p_{HH} - Rf}. \quad (11)$$

In the special case where $p_{HH} = 1$ and $p_{HL} = 0$, that is each calf is in the same susceptibility class as its dam, s is independent of f .

5. A SINGLE LOCUS SUSCEPTIBILITY MODEL

(a) Two allele system

Polymorphisms in the sheep PrP gene are closely associated with variation in incubation time in the case of scrapie (Goldmann *et al.* 1991; Hunter *et al.* 1989). Susceptibility is thought to be controlled by a single gene, *Sip*, with two alleles, *sA* (susceptible) and *pA* (resistant) with the susceptibility gene being dominant. For BSE in cattle the existence (let alone the genetic detail of control) of genetic predisposition in susceptibility or incubation period is as yet uncertain. However, if we follow the example of the sheep scrapie model, the model outlined in the previous section can be applied with specificity if we assume a single locus and a two allele framework (H and L for high and low susceptibility). We let s be the ratio of susceptibilities in high divided by low groups with heterozygotes given susceptibility $1 + (s - 1)m$ relative to the low susceptibility group where $0 \leq m \leq 1$. The case $m = 1$ equates with dominant high susceptibility and $m = 0$ with dominant low susceptibility. Following the procedures in the previous section the generation matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{pmatrix} q & q/2 & 0 \\ 1 - q & 1/2 & q \\ 0 & (1 - q)/2 & 1 - q \end{pmatrix}$$

corresponding to the *HH* genotype, the heterozygotes *HL* and the *LL* genotype where q is the frequency of the *H* allele in the population. As above, the assumption of stationary gene frequencies requires that $\mathbf{P}_{eq} = \mathbf{G} \mathbf{P}_{eq}$. Thus,

$$\mathbf{P}_{eq} = \begin{pmatrix} q^2 \\ 2q(1 - q) \\ (1 - q)^2 \end{pmatrix}.$$

As before, given an exposure c_i to the i th cohort of dams in the low susceptibility class, we assume that the proportion x_i remain susceptible where $x_i = e^{-c_i}$. Thus, the proportion of the dams in cohort i that are infected is defined as:

$$I_i = \mathbf{S}(x_i) \mathbf{P}_{eq},$$

where $\mathbf{S}$ is the attack rate vector

$$\mathbf{S}(x_i) = \begin{pmatrix} 1 - x_i^s \\ 1 - x_i^{1+(s-1)m} \\ 1 - x_i \end{pmatrix}$$

corresponding to the *HH* homozygotes, the *HL* heterozygotes and the *LL* homozygotes, respectively.

Turning to the interpretation of the results of the maternal cohort study, the distribution of genotypes (*HH*, *HL*, *LL*) within the control and maternally exposed group are as defined in equations (4) and (5). Equations (6)–(8) may be used to solve for y in terms of s , m , q and I_C . We can then evaluate the minimum value of the average relative susceptibility in animals of genotype *HH* and *HL*, $\bar{s}$, defined by

$$\bar{s} = \frac{sq^2 + 2(1 + (s - 1)m)q(1 - q)}{q(2 - q)},$$

with respect to q (the frequency of the *H* allele in the population) subject to the constraint of equation (9). The minimum value thus obtained is a function of m and the prevalences observed in the dams.

As the infection rates in the dams and the calves become increasingly small, the genotype frequencies in the control arm of the study are expected to approximately equal the population frequencies. Making this approximation, there is a closed-form solution for the relative susceptibility factor s . Mathematically, this limit is obtained when $x_i \rightarrow 1$ and $y \rightarrow 1$ with the minimization constraint given in equation (10). It can be shown that the ratio of expected susceptibilities in the maternally exposed and control groups, R , is given by:

$$R = 1 + \frac{(s - 1)^2 q [q^2 + m(1 - q)(m + 2q)] - q^3 - 4mq^2(1 - q) - 4m^2 q(1 - q)^2}{[1 + (s - 1)q(q + 2m(1 - q))]^2}$$

Hence,

$$s = 1 + \frac{(R - 1)q(q + 2m(1 - q)) + (m + q(1 - 2m))\sqrt{(R - 1)q(1 - q)}}{q^3 + mq(1 - q)(m + 2q) - Rq^2(q + 2m(1 - q))^2}$$

for $s \geq 1$.

(b) Three allele system

Assuming three alleles at a single locus determine susceptibility to infection, we can write a similar model. Let the alleles be denoted *A*, *B* and *C*. The relative susceptibilities of each genotype are denoted s_{ij} for genotype ij with the restriction that one of the genotype susceptibilities is assumed to equal unity. The generation matrix depends on the population frequencies of the alleles (*AA*, *AB*, *AC*, *BB*, *BC*, *CC*) as given by

$$\mathbf{G} = \begin{pmatrix} q & \frac{1}{2}q & \frac{1}{2}q & 0 & 0 & 0 \\ r & \frac{1}{2}(q+r) & \frac{1}{2}r & q & \frac{1}{2}q & 0 \\ (1-q-r) & \frac{1}{2}(1-q-r) & \frac{1}{2}(1-r) & 0 & \frac{1}{2}q & q \\ 0 & \frac{1}{2}r & 0 & r & \frac{1}{2}r & 0 \\ 0 & \frac{1}{2}(1-q-r) & \frac{1}{2}r & (1-q-r) & \frac{1}{2}(1-q) & r \\ 0 & 0 & \frac{1}{2}(1-q-r) & 0 & \frac{1}{2}(1-q-r) & (1-q-r) \end{pmatrix}$$

where q and r are the frequencies of alleles A and B in the population, respectively. In this case, the set of parameter combinations that would be consistent with the data observed in the maternal cohort study is large. This model, thus, yields little insight into the mechanisms underlying the observed maternal cohort data unless additional information is available about genotype frequencies.

6. MODELS WITH ONE RESISTANT CLASS

In the previous sections, we have assumed that s corresponds to the high susceptibility class, and thus $s > 1$. Another interesting limit exists, however, given by $s = 0$. In this case one of the classes of animals is entirely resistant to infection. The results derived above remain valid, but what previously was considered to be a trait for increased susceptibility is now a trait for resistance.

In the case $s = 0$, an analytic solution for f can be obtained for the inheritance of susceptibility model. Let $\bar{I}$ be the weighted mean probability of infection over the cohorts from which the control group dams were drawn:

$$\bar{I} = \sum_{i=t-2}^{t-20} a_i I_i.$$

It can be shown that $h = 1 - f$, the fraction of the population in the susceptible class, is linearly dependent on $p_{L|L}$, being given by:

$$h = \left(\bar{I} + \frac{1 - \bar{I}}{R} \right) p_{L|L}$$

where valid solutions only exist for

$$p_{L|L} > I_E.$$

Here R is the ratio of prevalences between the exposed and control arms, and $p_{L|L}$ is the probability that a susceptible dam has a susceptible calf ($p_{L|L} = 1 - p_{H|L}$).

The single locus susceptibility model must be solved numerically for the general $0 < m < 1$ case, but for $m = 1$ it can be shown that $p = 1 - q$, the fraction of the population with the L allele (H homo- and heterozygotes now being entirely resistant), is given by

$$p = \frac{R - \sqrt{R^2 - 4(R-1)\bar{I}}}{2\bar{I}(R-1)} \quad (12)$$

where the solution is valid for

$$\frac{1 + I_E - \sqrt{(1 - I_E)(5 - I_E)}}{2} < p < \frac{1 + I_E + \sqrt{(1 - I_E)(5 - I_E)}}{2}.$$

This constraint only has effect for $I_E > 5/6$. For $\bar{I} \rightarrow 0$, (12) gives

$$p \rightarrow \frac{1}{R}.$$

For the $m = 0$ case, p is given by the relevant root of the quartic equation:

$$Rp(2-p)[(1-\bar{I})(1+p(1-p)) + (1-p)^2] = (1-\bar{I}p(2-p))(1+P(1-p)).$$

7. APPLICATION OF MODELS

The single locus model can be applied to help interpret the results of the maternal cohort study if all the transmission was feed-borne and the risk difference arose via genetic predisposition to infection. If a single locus, two allele framework applies, we can estimate the minimum value of s that could give rise to the observed prevalences of BSE in maternally exposed and control arms (table 1). Such estimation requires data on the trends of BSE prevalence in different cohorts (obtained from the BSE case database compiled at the Central Veterinary Laboratory (CVL)) and information on the demography of the cohorts (Anderson *et al.* 1996; Donnelly *et al.* 1997a).

We estimate the incidence of BSE in dams of calves recruited to the study and by applying the survival distribution of the dams we estimate the I_i terms of equation (2) from the uncensored BSE case numbers for each cohort of dams (figure 1).

In the study design, the dams of control and maternally exposed animals arise from different sets of yearly birth cohorts. Control animals were required to have a dam not showing clinical signs of BSE by the age of six, but there was no matching with maternally exposed group animals by birth cohort of the dam. As such, the dams of the latter group of maternally exposed animals were not necessarily over six years of age when the calf was born. The a_i terms of equation (4) (the probability that an animal alive at time t is from the i th cohort) are obtained directly from the survival distribution of cohorts with the restriction that dams had to be at least six years of age. The b_i terms of equation (5) (the probability that an animal which was a case at time t is from the i th cohort) are estimated from the

Figure 1. Estimated infection prevalence (adjusted for survival) (I_i) by cohort within herds with at least one confirmed BSE case with the onset of clinical signs in 1988 for cohorts 1982, 1983, 1984 and 1985. The fraction of cases (b_i), born between July 1981 and June 1985, with onset of clinical signs in 1988 arising from each birth cohort.

distribution of cohorts among animals with the onset of BSE in 1988 (figure 1).

The computational burden of the solution of the equation system increases linearly with the number of a_i and b_i terms. Thus, the approximation was made that all dams of the control group calves were assumed to arise from the 1983 birth cohort and all dams of the maternally exposed group calves were assumed to arise from the four birth cohorts with the highest b_i terms (1982, 1983, 1984 and 1985).

(a) Inheritance model

A plot of the minimum value of s (ratio of susceptibilities between the two susceptibility phenotypes), $s_{\min}$, that could lead to the prevalences recorded in the maternal cohort study is given in figure 2a, for various values of $p_{H|H}$ (probability of susceptibility trait being inherited by calves) and I_M (prevalence of BSE infection in dams; i.e. $I_j = I_M$ for all j). For the case when $p_{H|H} = 1$ and for low I_M , $s_{\min}$ is approximately equal to 10, and around 10% of the population is predicted to belong to the high susceptibility class. In other words, the maternal cohort study results could be interpreted as reflecting zero maternal transmission of the infectious agent but a tenfold increased susceptibility in 10% of the population compared with the rest.

As $p_{H|H}$ declines, however, $s_{\min}$ increases dramatically to compensate for the reduction in the susceptibility 'amplification' effect caused by the dam selection criterion in the maternally exposed arm of the study. This is accompanied by a corresponding decrease in the fraction of the population in the high susceptibility class (figure 2b), required to maintain the relatively low incidences seen in the two arms of the study (14.0% and 4.3%).

The dependence of $s_{\min}$ on I_M , (the fractional BSE case incidence seen in the cohorts from which the dams were drawn) is interesting. As I_M rises, the fraction of the control arm dams that is drawn from the high susceptibility class declines, therefore amplifying the selection biases in the study design. The value of $s_{\min}$ therefore declines, as a smaller initial susceptibility ratio is required to produce the observed mean susceptibility differences between the calves in the two study arms. The fraction of animals in the high susceptibility class increases with increasing I_M in order to generate the higher levels of incidence in the dams whilst maintaining the required susceptibility difference between the two susceptibility classes.

The final feature of interest in figures 2a and 2b is that no solutions exist once $p_{H|H}$ drops below a certain critical value $p_{\text{crit}}(I_M)$. This critical value is largely determined by the condition $p_{H|H} > p_{H|L}$, which requires $p_{H|H} > f$ (from equation (1)); i.e. for the mean susceptibility of the maternally exposed animals to be greater than the controls, the probability of a high susceptibility dam having a high susceptibility calf needs to be larger than the probability of a low susceptibility dam having a high susceptibility calf. As $p \rightarrow p_{\text{crit}}(I_M)$, $s_{\min} \rightarrow \infty$, with such solutions clearly having a one-to-one mapping onto the $s = 0$ solution set considered below.

Figure 2c shows the values of $s_{\min}$ and f as a function of $p_{H|H}$ for the incidences actually observed in the maternal cohorts. To a first approximation the curves represent cross-sections of figure 2a,b for $I_M \simeq 0.027$. The minimum possible value of $s_{\min} \simeq 10.5$ occurs when transmission of the susceptibility trait is certain ($p_{H|H} = 1$), with $s_{\min}$ increasing rapidly (approximately as $1/p$) with decreasing $p_{H|H}$. Note that f varies almost linearly with $p_{H|H}$ except $p_{H|H} \rightarrow p_{\text{crit}}(I_M)$. The reason for this is clear when one considers the $I_M \rightarrow 0$ limit given by equation (11). It is trivial to show that the values of f giving minimum values of s are of the form $p_{H|H}/f = F(R)$, where F is a function of R alone. For $p_{H|H}/f \ll 1$, this is effectively equivalent to constraining the ratio of $p_{H|H}/p_{H|L}$ to be constant (from equation (1)).

Finally, figure 2d shows the coefficient of variation (CV) of the susceptibility of individuals ($CV = \text{s.d.}/\text{mean}$) corresponding to the curves shown in figure 2c. The CV has the advantage of being definable for any distribution, whether discrete or continuous, and provides a measure of the variability of some attribute in the population. $CV = 0$ implies no heterogeneity in susceptibility, whilst $CV > 1$ is indicative of major variability. CV effectively tracks the rise in $s_{\min}$ seen with declining $p_{H|H}$, starting at about 1.5 for $m = 1$ and rising to above 4 for $p_{H|H} \rightarrow p_{\text{crit}}(I_M)$.

(b) Single locus, two allele system

Figure 3a shows the minimum value of $\bar{s}$, $\bar{s}_{\min}$, for different values of I_M (the incidence in the maternal cohorts; $I_j = I_M$ for all j), and m , the relative dominance factor ($m = 1$ gives H dominant, $m = 0$ gives H recessive). The global minimum of $\bar{s}_{\min}$, approximately 10, occurs at $I_M = 1$ and $m \simeq 0.1$. However, for more

reasonable values of $I_M < 0.5$, the minimum is approximately 19, at $m = 0$ and $I_M = 0$. Figure 3b shows the values of q , the frequency of the H allele in the general population, which correspond to the values of $\bar{s}_{\min}$ shown in figure 3a.

A number of interesting features emerge upon examination of figures 3a,b. The system exhibits two distinct types of behaviour. For $q > 0.25$, the behaviour can be characterized as H dominant, with $\bar{s}_{\min}$ rapidly increasing from around 20 to infinity at some critical value of I_M , $I_{\text{crit}}(m)$. For $I_M > I_{\text{crit}}(m)$, the model is inconsistent with the maternal cohort study results, and no solution for $\bar{s}$ exists. Underlying this behaviour is a simple process: as I_M increases, q needs to increase so that the fraction of the population with the more susceptible HH and HL genotypes, $q(2-q)$, approximately equals I_M —since it is the high susceptibility genotype that is responsible for most of the prevalence in the dams. However, if q grows too large, the probability that an LL homozygote dam will produce a high susceptibility HH or HL calf becomes too large to generate the incidence difference observed between the control and maternally exposed offspring.

The second, H -recessive, behaviour type occurs for $q < 0.25$. Solutions exist for all I_M , with $\bar{s}_{\min}$ reaching some maximum in I_M (for fixed m) before declining once more as I_M increases further. Here, as I_M increases from zero, q increases as before, since infection in the dams largely occurs in the HH homozygotes. $\bar{s}_{\min}$ correspondingly increases to maintain the incidence difference between the two offspring groups. Beyond a certain point, infection in the HH genotype saturates, so further increases in I_M are made up by increasing infection prevalence in the low susceptibility LL and HL genotypes. The model can still reproduce the maternal cohort study results, however, as, even for large I_M , it is still predominately just the HH homozygotes that experience infection in the offspring groups, rather than both the HH and HL genotypes as in the H -dominant regime. Indeed, $\bar{s}_{\min}$ declines rapidly as I_M increases further, as the genotypic frequencies of the BSE affected dams increasingly resemble the equilibrium frequencies, whilst the BSE free dams become increasingly dominated by LL homozygotes which have zero probability of producing high susceptibility HH homozygotes. The value of I_M for which $\bar{s}_{\min}$ is maximum decreases with decreasing $m < 0.25$ since, as m declines, infection saturation in the HH homozygote dams occurs more rapidly.

For the observed maternal cohort incidences (see figure 1), two regimes of behaviour are still evident (figure 3c), with the crossover occurring at $m \simeq 0.25$. For $m > 0.25$, both $\bar{s}_{\min}$ and q are remarkably constant. Below $m \simeq 0.25$, a dramatic increase in $\bar{s}_{\min}$ is seen (reflecting the underlying increase in s), as the HH homozygotes have to increasingly account for nearly all the infection in both the control and maternally exposed offspring. A simultaneous drop in q is seen—since when minimizing $\bar{s}$, it is beneficial to dramatically increase s whilst decreasing q , rather than to keep q constant and increase s by a smaller amount. The pattern seen with decreasing m of a gradual increase then a sharp decline in q is therefore an artefact of the minimization. How-

ever, the pattern of nearly constant $\bar{s}$ for $m > 0.25$, then a rapid increase in $\bar{s}$ for $m < 0.25$, is still seen even if q is held constant. Figure 3d shows how CV corresponds to the solutions given in figure 3c. As seen for the inheritance model, CV closely matches the behaviour of $\bar{s}_{\min}$, remaining nearly constant at $CV \simeq 2.4$ for $m > 0.25$, but rising rapidly to above 5 as m declines below 0.25.

Turning finally to the $s = 0$ case, when H represents an allele endowing resistance, we see a similar two state behaviour. Figure 3e shows the value of $p = 1 - q$, the frequency of the L allele, for $0 \leq m \leq 1$ (where $m = 0$ now represents dominance of the susceptibility allele, L) consistent with the results of the maternal cohort study, and allowing for the observed incidences in the cohorts of the dams. Here, the L -dominant regime is defined by $m < 0.8$, throughout which p is nearly constant at $p \simeq 0.1$. For $m > 0.8$ —the L -recessive regime— p increases rapidly to around 0.33, as the LL homozygotes have to account for an increasing fraction of all infection.

(c) Single locus, three allele system

The analyses presented so far have lacked genotypic frequency data, and have therefore explored the entire range of genotypic frequencies and susceptibilities consistent with the maternal cohort study. If the genetical processes are more complex than the single locus, two allele system, then in the absence of genotypic data, the parameter space consistent with the maternal cohort study results becomes so large as to be relatively uninformative in helping to discriminate between different hypotheses explaining the study results.

We therefore present an example of an analysis based on a more complex model—the single locus, three allele system—which utilizes data from the sole study (to date) to find any differences in the PrP gene between BSE-affected and BSE-unaffected cows (Neibergs *et al.* 1994). We use maximum likelihood methods to estimate population allelic frequencies, genotypic susceptibilities, and the feed-based exposure of the maternal cohort study animals (eight parameters in total). The model is fitted to the genotypic frequencies observed by Neibergs *et al.* (1994), together with the infection prevalences observed in the two arms of the maternal cohort study (12 data points in total). The data were well fit by the model (goodness-of-fit $\chi^2 = 3.48$, d.f. = 4, p value = 0.48), and the results are presented in Table 2.

The good fit obtained by the model indicates that the Neibergs *et al.* (1994) data and the incidence data from the maternal cohort study are consistent with the model with a single locus, three allele system. We urge caution in the interpretation of the above results, however. The genotypic frequencies given in Neibergs *et al.* (1994) were measured using single-strand conformational polymorphism techniques, and no formal sequencing analysis or pedigree analysis has been performed to date to confirm the results. The observed effect may correspond to a point mutation, or be generated by linkage to some, as yet unidentified, causal gene. Rather, the results are indicative of how mathematical

Figure 2. (a) The minimized value of s as a function of $p_{H|H}$ and I_M assuming $I_i = I_M$ for all i for the inheritance susceptibility model. (b) The value of f corresponding to the values of s shown in (a) as a function of $p_{H|H}$ and I_M . (c) The minimized value of s and the corresponding f value as a function of $p_{H|H}$ for the estimated maternal cohort incidence values for the inheritance susceptibility model. (d) The coefficient of variation ($CV = \text{s.d.}/\text{mean}$) of susceptibility of the population as a function of f .

models enable quantitative estimates to be made of the variation in epidemiological characteristics of the host (i.e. the fitness matrix), given epidemiological data (incidence and/or prevalence) and genotypic data. Such an analysis, whilst complex, is essential to gain any precise understanding of the epidemiological consequences of genotypic variation in the host (in susceptibility and/or incubation period) on the transmission dynamics of the aetiological agent of BSE—especially in the long-term.

8. CONCLUSIONS

The results of the maternal cohort study (SEAC 1997; Wilesmith *et al.* 1997), which was designed to assess whether or not the aetiological agent of BSE could be transmitted vertically, can be explained by a genetic model based on the assumption of genetic predisposition to disease, acting in tandem with exposure to contaminated feed in both control and exposed arms of the study. In the absence of any specific genotypic data on the frequency of 'resistance' and 'susceptibility' alleles in the UK cattle herd, it is only possible to explore the range of frequencies and the type of genetic control (i.e. one locus or many loci, two or more alleles) that are consistent with the recorded prevalence of disease in the offspring of BSE-affected and unaffected dams. For simple models (i.e.

one locus, two alleles), the difference in susceptibility between 'resistant' and 'susceptible' animals must be high to explain the maternal cohort study results. For example, for the one locus, two allele model, at least a 20-fold mean difference in susceptibility between the higher (*HH* and *HL*) and lower (*LL*) susceptibility genotypes is required.

The design of the maternal cohort study makes investigation of a genetic interpretation of the results (i.e. no maternal transmission) extremely difficult. In order to estimate to what extent the segregation of dams by BSE diagnosis status in the two arms of the study ('exposed' and 'unexposed' to infected dam) represents a selection bias—amplifying the frequency of susceptibility in the 'exposed' arm—it is necessary to have detailed information on infection rates in the population from which the dams are drawn and detailed molecular data on candidate susceptibility loci of each animal included in the study. In the absence of such information, our analyses show clearly that the study results could be explained by maternal transmission, genetic predisposition (under a wide variety of genetic models) or a combination of the two factors.

More generally, both genotypic (frequency) and detailed epidemiological (incidence and/or prevalence) data are required to provide a better interpretation of the results. The latter are available on the basis of back-calculation methods working from data on BSE

Figure 3. (a) The value of $\bar{s}$ corresponding to the minimum average susceptibility in those animals not *LL* homozygotes as a function of m and I_M assuming $I_i = I_M$ for all i for the single locus susceptibility model. (b) The value of q corresponding to the values of $\bar{s}$ shown in (a) as a function of m and I_M . (c) The value of $\bar{s}$ (minimized as in (c)) and the corresponding q value as a function of m for the estimated maternal cohort incidence values for the single locus susceptibility model. (d) The coefficient of variation ($CV = \text{s.d.}/\text{mean}$) of susceptibility of the population as a function of m for the single locus susceptibility model. (e) Solution for the $s = 0$ case, where there are two classes: one resistant and one susceptible, showing p (the frequency of the L allele) and the CV of susceptibility in the population as a function of m .

incidence (Anderson *et al.* 1996): the former are not available and are urgently required. It is known that susceptibility to scrapie is determined by both host genotype and pathogen strain. The aetiological agent of BSE may not have significant strain variability as yet due to the novelty of the disease in cattle. The scrapie-sheep system has been evolving for at least 250 years, giving time for selection of host and pathogen types via multiple passaging of the prion agent.

However, the work of Neibergs *et al.* (1994) strongly hints at variation in cattle host susceptibility to infection/disease. More work is urgently required in this area given the applied significance of breeding from a stock of 'resistant' animals.

Finally, further insight into the questions of genetic predisposition and maternal transmission of BSE in cattle may emerge from detailed epidemiological and statistical analysis of the BSE incidence database and

Table 2. *Maximum likelihood estimates MLE and 95% confidence intervals for the model with a single locus with three alleles determining susceptibility to infection.*(Observed genotype and allele frequencies (Neibergs *et al.* 1994) (a) and BSE incidence data from the maternal cohort study (b) were analysed.)

genotypes	susceptibility	confirmed BSE genotype frequencies		Non-BSE-affected genotype frequencies	
	s MLE (95% CI)	observed (n = 56)	MLE (95% CI)	observed (n = 38)	MLE (95% CI)
AA	1.0	0.48	0.46 (0.30,0.59)	0.29	0.28 (0.16,0.43)
AB	4.80 (0.94,12.7)	0.16	0.18 (0.04,0.38)	0.05	0.022 (0.005,0.059)
AC	0.39 (0.11,1.1)	0.30	0.31 (0.11,0.55)	0.45	0.48 (0.42,0.49)
BB	0.0 (0.0,49.6)	0.00	0.00 (0.00,0.035)	0.00	0.00 (0.00,0.0035)
BC	0.0 (0.0,4.1)	0.00	0.00 (0.00,0.11)	0.00	0.019 (0.00,0.053)
CC	0.16 (0.006,0.82)	0.06	0.053 (0.00,0.22)	0.21	0.20 (0.10,0.33)
		allele frequencies		allele frequencies	
allele		observed	MLE	observed	MLE
A		0.71	0.70 (0.58,0.77)	0.54	0.53 (0.40,0.66)
B		0.08	0.09 (0.02,0.19)	0.03	0.022 (0.005,0.06)
(b)					
		maternally exposed group		BSE incidence control group	
		observed (n = 301)	MLE (95% CI)	observed (n = 301)	MLE (95% CI)
		0.140	0.130 (0.075,0.202)	0.043	0.053 (0.029,0.085)

detailed herd breeding records. Attempts at such analyses to date (Curnow *et al.* 1994; Hau & Curnow 1996; Curnow & Hau 1996) have tended to be based on genetic models with somewhat extreme assumptions such as complete resistance in a large fraction of the UK cattle herd. More detailed analyses of the full epidemiological data on the BSE epidemic based on a broader and more realistic range of genetic models should shed further light on the issue, as will the molecular genotypic analysis of biopsy material collected from the animals in the maternal cohort study. Our main conclusion, however, is that the results of the maternal cohort study are consistent with elements of both maternal transmission and genetic predisposition hypotheses.

We are grateful to John W. Wilesmith and colleagues at the Central Veterinary Laboratory for provision of data, and to Dr David Goldstein and Dr Tim Anderson for useful discussions. N. M. F., C. A. D. and R. M. A. thank the Wellcome Trust for research grant support. M. E. J. W. thanks the Royal Society and the BBSRC.

REFERENCES

- Anderson, R. M., Donnelly, C. A., Ferguson, N. M., Woolhouse, M. E. J., Watt, C. J., Udy, H. J., MaWhinney,

- S., Dunstan, S. P., Southwood, T. R. E., Wilesmith, J. W., Ryan, J. B. M., Hoinville, L. J., Hillerton, J. E., Austin, A. R. & Wells, G. A. H. 1996 Transmission dynamics and epidemiology of BSE in British cattle. *Nature* **382**, 779–788.
- Collinge, J., Sidle, K. C. L., Meads, J., Ironside, J. & Hill, A. F. 1996 Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* **383**, 685–690.
- Curnow, R. N. & Hau, C. M. 1996 The incidence of bovine spongiform encephalopathy in the progeny of affected sires and dams. *Vet. Rec.* **138**, 407–408.
- Curnow, R. N., Hodge, A. & Wilesmith, J. W. 1997 Analysis of the bovine spongiform encephalopathy maternal cohort study: the discordant case-control pairs. *Appl. Statist.* **46**, 345–349.
- Curnow, R. N., Wijeratne, W. V. S. & Hau, C. M. 1994 The inheritance of susceptibility to BSE. In *Proceedings of European Commission consultation on transmissible spongiform encephalopathies* (eds R. Bradley & B. Marchant), pp. 109–124. Brussels.
- Donnelly, C. A., Ferguson, N. M., Ghani, A. C., Woolhouse, M. E. J., Watt, C. J., Wilesmith, J. W. & Anderson, R. M. 1997a The epidemiology of BSE in cattle in Great Britain. I. Epidemiological processes, demography of cattle and approaches to control by culling. *Phil. Trans. R. Soc. Lond. B* **352**, 781–801.
- Donnelly, C. A., Ghani, A. C., Ferguson, N. M., Wilesmith, J. W. & Anderson, R. M. 1997b Analysis of the bovine spongiform encephalopathy maternal cohort study: evidence for direct maternal transmission. *Appl. Statist.* **46**, 321–344.

- Ferguson, N. M., Donnelly, C. A., Woolhouse, M. E. J. & Anderson, R. M. 1997 The epidemiology of BSE in cattle herds in Great Britain. II. Model construction and analysis of transmission dynamics. *Phil. Trans. R. Soc. Lond. B* **352**, 803–838.
- Fraser, H., Bruce, M. E., Chree, A., McConnell, I & Wells, G. A. H. 1992 Transmission of bovine spongiform encephalopathy and scrapie to mice. *J. Gen. Virol.* **73**(8), 1891–1897.
- Goldmann, W., Hunter, N., Martin, T., Dawson, M. & Hope, J. 1991 Different forms of the bovine PrP gene have five or six copies of a short, G-C-rich element within the protein-coding exon. *J. Gen. Virol.* **72**(1), 201–204.
- Gore, S. M., Gilks, W. R. & Wilesmith, J. W. 1997 Bovine spongiform encephalopathy maternal cohort study—exploratory analysis. *Appl. Statist.* **46**, 305–320.
- Hau, C. M. & Curnow, R. N. 1996 Separating the environmental and genetic factors that may be causes of bovine spongiform encephalopathy. *Phil. Trans. R. Soc. Lond. B* **351**, 913–920.
- Hunter, N., Foster, J. D., Dickinson, A. G. & Hope, J. 1989 Linkage of the gene for the scrapie-associated fibril protein (PrP) to the *Sip* gene in Cheviot sheep. *Vet. Rec.* **124**, 364–366.
- Hunter, N., Foster, J. D., Goldmann, W., Stear, M. J., Hope, J. & Bostock, C. 1996 Natural scrapie in a close flock of Cheviot sheep occurs only in specific PrP genotypes. *Arch. Virol.* **141**(5), 809–824.
- Hunter, N., Goldmann, W., Smith, G. & Hope, J. 1994 Frequencies of PrP gene variants in healthy cattle with BSE in Scotland. *Vet. Rec.* **135**, 400–403.
- Kimberlin, R. H. & Wilesmith, J. W. 1994 Bovine spongiform encephalopathy: epidemiology, low dose exposure and risks. *Ann. N.Y. Acad. Sci.* **724**, 210–220.
- Martin, T. C., Hunter, N., Goldmann, W., Hughes, S., Broekhuizen, J., Hope, J., Wilesmith, J. W. & Dawson, M. 1991 A study of bovine PrP gene polymorphisms in a BSE-affected herd. *Proceedings of the Second Congress of the European Society for Veterinary Virology*. Uppsala, Sweden, p. 148.
- Neibergs, H. L., Ryan, A. M., Womack, J. E., Spooner, R. L. & Williams, J. L. 1994 Polymorphism analysis of the prion gene in BSE-affected and unaffected cattle. *Anim. Genet.* **25**, 313–317.
- Prusiner, S. B., Telling, G., Cohen, F. E. & DeArmond, S. J. 1996 Prion diseases of human and animals. *Semin. Virol.* **7**, 159–173.
- Ridley, R. M. & Baker, H. F. 1996 No maternal transmission? *Nature* **384**, 17.
- Spongiform Encephalopathy Advisory Committee 1995 *Transmissible Spongiform Encephalopathies: a summary of present knowledge and research*. HMSO: London.
- Spongiform Encephalopathy Advisory Committee 29 July 1996 *Statement*.
- Spongiform Encephalopathy Advisory Committee 16 April 1997 *Statement*.
- Wilesmith, J. W., Ryan, J. B. M. & Atkinson, M. J. 1991 Bovine spongiform encephalopathy: epidemiological studies on the origin. *Vet. Rec.* **128**, 199–203.
- Wilesmith, J. W., Ryan, J. B. M., Hueston, W. D. & Hoinville, L. J. 1992 Bovine spongiform encephalopathy: epidemiological features 1985–1990. *Vet. Rec.* **130**, 90–94.
- Wilesmith, J. W., Wells, G. A. H., Ryan, J. B. M., Gavier-Widen, D. & Simmons, M. M. 1997 A cohort study to examine maternally associated risk factors for bovine spongiform encephalopathy. *Vet. Rec.* (In the press.)

Received 8 May 1997; accepted 3 June 1997

As this paper exceeds the maximum length normally considered for publication in *Proceedings B*, the authors have agreed to make a contribution towards production costs.