PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The epidemiology of BSE in cattle herds in Great Britain. II. Model construction and analysis of transmission dynamics

N. M. Ferguson, C. A. Donnelly, M. E. J. Woolhouse and R. M. Anderson

Phil. Trans. R. Soc. Lond. B 1997 **352**, 803-838
doi: 10.1098/rstb.1997.0063

The epidemiology of BSE in cattle herds in Great Britain. II. Model construction and analysis of transmission dynamics

N. M. FERGUSON*, C. A. DONNELLY, M. E. J. WOOLHOUSE AND R. 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)

CONTENTS

SUMMARY

Mathematical models that describe the key processes determining the pattern of the bovine spongiform encephalopathy (BSE) epidemic in British cattle are derived that allow for infection from feed as well as maternal and direct horizontal transmission. Heterogeneous susceptibility classes are also incorporated into the analysis. Maximum likelihood methods are used to estimate parameters and to obtain confidence intervals from available experimental and epidemiological data. A comprehensive sensitivity analysis of all model parameters and distributional assumptions is presented. Additional validation is provided by fitting the model to independent data collected in Northern Ireland. Model estimates and predictions based on BSE case data for Great Britain and Northern Ireland, together with their implications, are reviewed, and future research priorities discussed.

1. INTRODUCTION

The first bovine spongiform encephalopathy (BSE) case was diagnosed in Great Britain (GB) in November 1986 (Wells *et al.* 1987). By the end of May 1997, over 167 000 cases had been confirmed. The effect of the epidemic on animal health and the European agricultural industry has been severe. From both an economic and public health perspective, this impact has been heightened by recent evidence supporting a link between BSE and the emergence of a new variant of Creutzfeldt–Jacob disease (CJD) in humans (Collinge *et al.* 1996). It is therefore of great importance that we gain the fullest possible understanding of both the biology and epidemiology of the disease BSE and its aetiological agent. Only through such insight is it possible to quantify the risk to humans from past exposure to contaminated meat and meat products, predict future trends in infection and disease incidence, and devise control strategies to hasten the end of the epidemic.

In this paper we present the methodology and results of an analysis of the data on confirmed BSE cases collated by the Central Veterinary Laboratory (CVL) of the Ministry of Agriculture, Fisheries and Food (MAFF) at Weybridge, Surrey. We derive a model for the probability that an individual animal is observed to experience the onset of the clinical signs of BSE at a particular time, allowing for a variety of possible transmission routes (feed, maternal and non-feed-related horizontal transmission). We also consider the effects of heterogeneous susceptibility classes and under-reporting of cases. This work refines and generalizes the analyses presented in Anderson *et al.* (1996). Using an extensive range of sensitivity analyses, we show that the key results presented in that paper remain valid, and are robust to changes in a wide variety of model assumptions and parameter values.

Evidence to date suggests that BSE in cattle arose from supplementary feed containing meat and bone meal (MBM) contaminated by a scrapie-like agent derived from sheep or cattle (Wilesmith *et al.* 1991, 1992). The epidemic was then fuelled by recycling of

* Author for correspondence.

infective tissues from BSE-infected cattle back into cattle feed, with the growth of the epidemic of infections only being stopped after ruminant-derived protein was banned in the production of cattle feed in July 1988 (Statutory Instrument 1988). To date, no ante-mortem diagnostic test for pre-symptomatic BSE infection exists.

There is evidence that calves of BSE-affected dams are more likely to become BSE cases than herd- and age-matched calves of unaffected dams (Wilesmith *et al.* 1997; Donnelly *et al.* 1997*a*, *c*; Gore *et al.* 1997; Curnow *et al.* 1997). One explanation of this result is the existence of a low level of maternal transmission of BSE. However, in the absence of any known biological mechanism for such a transmission route, it is also important to consider the alternative hypothesis; namely, that the maternal risk factor is inherited, implying the existence of genetically determined variability in susceptibility or incubation period—though no work to date has found evidence for such variability (Wijeratne & Curnow 1990; Curnow *et al.* 1994; Curnow & Hau 1996; Hau & Curnow 1996). There is also no evidence to date for direct (contact- or pasture-based) horizontal transmission.

Given the indirectly horizontal nature of the feed-based transmission route, one approach to developing mathematical models of the epidemic would be to use a traditional deterministic compartmental model (Anderson & May 1991) to explicitly describe the recycling of infection through feed that gave rise to the epidemic. However, since the movements of individuals between disease states in such models are typically governed by Poisson processes, they would be unable to accurately mimic the incubation period distribution seen in the BSE epidemic. The recycling governing the epidemic is also complex in form, involving both a time delay—caused by the time taken from the slaughter of an animal to the consumption of contaminated material from this animal by another via feed—and a time-varying transmission probability—dependent on the nature of the rendering processes used, the level of implementation of regulations, and the degree of cross-contamination in feed mills. Given the very limited information available on these processes, this casts doubt on whether any significant additional information would be gained from models explicitly incorporating recycling, at least at the population level. Indeed, given current epidemiological knowledge and the lack of an ante-mortem diagnostic test for BSE infection, the primary role for simple models is to accurately quantify the magnitude and epidemiological characteristics of the epidemic of *infections* of the aetiological agent of BSE—given knowledge of the epidemic of reported *cases* of the disease induced by the agent.

With these requirements in mind, and given the many parameters that need to be estimated in any model (given the lack of biological and epidemiological information), it is therefore sensible to adapt the backcalculation models used in the study of the AIDS epidemic to derive estimates of the time-varying incidence of HIV infection (Brookmeyer & Gail 1986, 1988; Gail & Rosenberg 1992; Bacchetti *et al.* 1993) for application to the BSE epidemic in cattle. These techniques have the advantage of being explicitly statistical in form, thus allowing a variety of parametrizations and transmission mechanisms to be robustly tested against the case data. Modelling then consists of fitting a set of parameters describing the feed-risk profile, incubation period distribution, and age-at-infection distribution to the age- and time-stratified reported case frequencies.

While the models described in this paper only utilize two levels of case stratification (age and time)—thereby fitting to the 'mean field' epidemic in Great Britain as a whole—it is also possible to further stratify the data by county, parish or even herd. Incorporating spatial and between-herd variation in exposure will certainly be vital to gain a further understanding of how contact/dispersion processes operating at the individual or herd level generated the observed spatio-temporal epidemic. However, such complex models can only be sensibly developed given an understanding of the relevant epidemiological and biological factors provided by both detailed analyses of the case database—as presented in the companion paper to this work (Donnelly *et al.* 1997*b*)—and simpler modelling exercises.

In § 2, we present detailed (and necessarily mathematical) descriptions of the model design and maximum likelihood methods used in its implementation. Emphasis is placed on incorporation of a range of potential transmission routes and heterogeneous susceptibility classes into the model, and methods for predicting future infection trends are discussed. Important biases in the original case report data are discussed in § 3, together with a description of debiasing methods. Section 4 then presents the results of a detailed and comprehensive set of sensitivity analyses. These explore the robustness of model estimates and predictions of numbers infected and cases of disease to changes in nearly all model assumptions and parameters. Factors explored include: the incubation period and age-dependent susceptibility distributions; the mean of the incubation period distribution; maternal and horizontal transmission; genetically variable susceptibility; under-reporting trends; feed-risk estimation through time; and sensitivity to demographic parameters. Section 5 then discusses the results and implications from the best fit model, with emphasis on methods of evaluating the (relative) risk of human exposure to contaminated meat and meat products in the past. We conclude with a discussion on future research priorities in the epidemiological study of BSE.

2. METHODS

(*a*) *Backcalculation*

Backcalculation methods reconstruct the past pattern of infection from disease-incidence data and estimates of the incubation period. This approach was used by Brookmeyer & Gail (1986, 1988) in order to obtain short-term projections of AIDS cases and to estimate HIV incidence

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

in the previous time periods. Backcalculation methodology has been reviewed by Gail & Rosenberg (1992) and Bacchetti *et al.* (1993), and has been applied in a veterinary context by Haydon *et al.* (1997).

Let the function $I(u)$ represent an infection process in chronological time such that the probability of infection by time s is $\int_{-\infty}^{s} I(u)\,\mathrm{d}u$. The basic convolution utilized relates the probability of disease onset in a time interval to the infection process and the incubation period probability density function (PDF), $f(t)$, such that $\int_{T_1}^{T_2} f(t)\,\mathrm{d}t$ is the probability that disease onset occurs between T_1 and T_2 years after infection. The probability of an individual becoming a case in the time period $[T_1, T_2]$ is

$$\int_{T_1}^{T_2}\int_{0}^{t} I(s)f(t-s)\,\mathrm{d}s\mathrm{d}t. \tag{1}$$

It should be noted that backcalculation of this type provides no information about future infection rates and little information about recent infection rates.

In early work the incubation period distribution was assumed to be stationary and known. The accuracy of the information obtained about the infection rates over time depended on the accuracy of the incidence data and the validity of the assumed incubation period distribution. The backcalculation methodology has been widely studied and generalized in a number of aspects. Brookmeyer & Liao (1990) allowed the incubation period distribution to change over time (i.e. $f(t-s)$ was replaced by $f(t-s, s)$ in (1)).

In the case of BSE in cattle we will consider birth cohorts individually, and therefore express the problem in terms of the age of a particular birth cohort rather than in terms of explicit calendar time.

Let $Q(t, a)$ be the instantaneous infection rate at time t, among susceptibles of age a, and $p_{\mathrm{I}}(t_0, a)$ be the probability that an animal born at time t_0 is infected by age a, in the absence of mortality. Thus,

$$\frac{\partial p_{\mathrm{I}}(t_0, a)}{\partial a} = Q(t_0 + a, a)(1 - p_{\mathrm{I}}(t_0, a)).$$

Hence $Q(t, a)$ is a hazard function:

$$p_{\mathrm{I}}(t_0, A) = \int_0^A Q(t_0 + a, a) \times \exp\left(-\int_0^a Q(t_0 + a', a')\,\mathrm{d}a'\right)\mathrm{d}a \tag{2}$$

$$= 1 - \exp\left(-\int_0^A Q(t_0 + a', a')\,\mathrm{d}a'\right)$$

$$\simeq \int_0^A Q(t_0 + a', a')\,\mathrm{d}a', \tag{3}$$

where (3) holds for $p_{\mathrm{I}}(t_0, A) \ll 1$.

The infection process can be factorized into the sum of the product of univariate functions in t and a: one representing a time-dependent risk of infection and one representing an age-dependent susceptibility/exposure distribution, such that

$$Q(t, a) = \sum\nolimits_j r_j(t) g_j(a), \tag{4}$$

where the jth time-dependent risk of infection is denoted $r_j(t)$, t is chronological time, while the function $g_j(a)$ describes how exposure and/or susceptibility to infection varies with age, a. Of course, this factorization assumes that the age-dependent patterns of exposure and susceptibility are constant over time.

While $g_j(a)$ may in principle take any form, it proves useful to normalize (over the lifespan of an animal—$0 \leqslant a \leqslant 18$) the function to be a PDF, since it then represents the distribution of ages at infection given a constant infinitesimal risk of infection ($r_j(t) = \delta\ \forall t$). When performing sensitivity analyses, this has the benefit of allowing fitted $r_j(t)$ profiles to be directly compared even when different forms of $g_j(a)$ are being used.

Thus, in the absence of mortality, the PDF that an individual born at time t_0 becomes infected at age a and a case at age u is

$$\rho_{\mathrm{C}}(t_0, a, u) = \sum\nolimits_j r_j(t_0 + a) g_j(a) \times \exp\left(-\int_0^a Q(t_0 + a', a')\,\mathrm{d}a'\right) f_j(u - a), \tag{5}$$

where $f_j(t)$ is the incubation period PDF for the jth time-dependent risk of infection. The probability that an individual born at time t_0 becomes a case by age A is therefore

$$\int_0^A \int_0^u \rho_{\mathrm{C}}(t_0, a, u)\,\mathrm{d}a\mathrm{d}u.$$

It is particularly important to include the survival probability with age if the incubation period of disease or mean age at infection is substantial relative to mean life expectancy. Let $S(u, t_0)$ represent the probability that an animal born at t_0 survives until age u. Thus, including mortality, the probability of an individual born at time t_0 becoming a case by age A is

$$p_{\mathrm{C}}(t_0, A) = \int_0^A S(u, t_0) \int_0^u \rho_{\mathrm{C}}(t_0, a, u)\,\mathrm{d}a\mathrm{d}u. \tag{6}$$

(*b*) *Inclusion of different transmission routes*

We consider three possible transmission routes: contaminated feed (F); maternal transmission (M); and non-feed-related horizontal transmission (H). Two of these represent continuous infection hazards throughout the lifetime of an animal, while maternal transmission is described by a point risk of infection at birth. Thus,

$$Q(t, a) = r_{\mathrm{F}}(t) g_{\mathrm{F}}(a) + r_{\mathrm{H}}(t) g_{\mathrm{H}}(a).$$

and

$$p_{\mathrm{I}}(t_0, A) = R_{\mathrm{M}}(t_0) + (1 - R_{\mathrm{M}}(t_0)) \times \int_0^A Q(t_0 + a, a) \exp\left(-\int_0^a Q(t_0 + a', a')\,\mathrm{d}a'\right)\mathrm{d}a, \tag{7}$$

where $R_{\mathrm{M}}(t)$ is the probability of maternal transmission for an animal born at time t.

Hence,

$$\rho_{\mathrm{C}}(t_0, a, u) = R_{\mathrm{M}}(t_0)\delta(a) f_{\mathrm{M}}(u) + (1 - R_{\mathrm{M}}(t_0)) \times \sum\nolimits_{j=\mathrm{F,H}} r_j(t_0 + a) g_j(a) \times \exp\left(-\int_0^a Q(t_0 + a', a')\,\mathrm{d}a'\right) f_j(u - a), \tag{8}$$

where $\delta(a)$ is the Dirac δ function.

This formulation is not quite identical to representing maternal transmission as an age-dependent infection hazard $Q_{\mathrm{M}}(t, a) = r_{\mathrm{M}}(t) g_{\mathrm{M}}(a)$, with $g_{\mathrm{M}}(a) = \delta(a)$, but the latter causes $r_{\mathrm{M}}(t)$ to have a complex and non-intuitive form. We will therefore use the form (7) when explicitly considering maternal transmission, but use the general form (4) when mathematical clarity is required.

We need the probability that an animal of age a (in the absence of mortality) is infectious through transmission route k at time t, $y_k(t, a)$. Let us define $\Omega_k(v)$ to be

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the fractional infectiousness through transmission route k (i.e. $0 \leqslant \Omega_k \leqslant 1$) of an infected animal at time v before clinical onset ($v < 0$ corresponding to times after clinical onset). For instance, if animals are only infectious through transmission route k for time T_{I} before onset, then $\Omega_k(v) = 1$ for $v \leqslant T_{\mathrm{I}}$, and $\Omega_k(v) = 0$ for $v > T_{\mathrm{I}}$. It can then be shown that:

$$y_k(t,a) = \int_0^\infty \Omega_k(v) \int_0^a \rho_{\mathrm{C}}(t-a, a', v+a)\,\mathrm{d}a'\mathrm{d}v. \qquad (9)$$

This equation is merely expressing the fact that an animal aged a at time t had to be born at time $t-a$, has been infected via feed at an age a' (where $0 < a' < a$), and is time v away from disease onset, making its incubation period $v + a - a'$.

Note that the feed-based transmission route is here treated as being characterized by an independent function $r_{\mathrm{F}}(t)$. However, given the feed recycling process, it could also be described by a time-delayed horizontal route with time-varying transmission coefficient—representing the varying effectiveness of controls on the use of feed containing MBM over time.

The maternal transmission probability, $R_{\mathrm{M}}(t)$, is then obtained by calculating the proportion of dams at time t (defined to be all cattle over two years of age) which are infectious, and multiplying by the reproduction rate of cows (assumed to be one calf per year), and the probability of maternal transmission from an infectious dam to its calf, ϵ:

$$R_{\mathrm{M}}(t) = \epsilon \int_0^\infty \sigma_{\mathrm{M}}(t,a) y_{\mathrm{M}}(t,a)\,\mathrm{d}a, \qquad (10)$$

where $\sigma_{\mathrm{M}}(t,a)$ is the PDF for the fraction of dams at time t that are of age a, defined thus:

$$\sigma_{\mathrm{M}}(t,a) = \begin{cases} 0, & a \leqslant 2, \\ \dfrac{B(t-a)S(a,t-a)}{\displaystyle\int_2^\infty B(t-a)S(a,t-a)\,\mathrm{d}a}, & a > 2. \end{cases} \qquad (11)$$

Here, $B(t)$ is the recruitment (birth) rate of cattle at time t, and we are assuming that all cattle over two years of age are female, since the great majority of males are slaughtered before that age.

Similarly, the horizontal transmission rate (force of infection) is given by the product of the proportion of cattle alive at time t which are infectious and a transmission coefficient, β:

$$r_{\mathrm{H}}(t) = \beta \int_0^\infty \sigma(t,a) y_{\mathrm{H}}(t,a)\,\mathrm{d}a, \qquad (12)$$

where the PDF for the fraction of cattle alive at time t which are of age a, $\sigma(t,a)$, is given by:

$$\sigma(t,a) = \frac{B(t-a)S(a,t-a)}{\displaystyle\int_0^\infty B(t-a)S(a,t-a)\,\mathrm{d}a}. \qquad (13)$$

Note that we have neglected the effect on BSE-induced mortality on the adult and total herd sizes. While this is acceptable when considering the entire GB cattle population as a whole, it would not be were individual high-incidence herds to be modelled.

Putting the last four equations together gives a pair of integral equations for r_{H} and R_{M}:

$$\begin{aligned} R_{\mathrm{M}}(t) = \epsilon \int_0^\infty \sigma_{\mathrm{M}}(t,a) \int_0^\infty \Big[& R_{\mathrm{M}}(t-a) f_{\mathrm{M}}(v+a) \\ & + (1 - R_{\mathrm{M}}(t-a)) \int_0^a \sum\nolimits_{j=\mathrm{F},\mathrm{H}} r_j(t-a+u) g_j(u) \\ & \times \exp\left(- \int_0^u Q(t-a+u', u')\,\mathrm{d}u' \right) \\ & \times f_j(v+a-u)\,\mathrm{d}u \Big] \Omega_{\mathrm{M}}(v)\,\mathrm{d}v\mathrm{d}a, \end{aligned} \qquad (14)$$

$$\begin{aligned} r_{\mathrm{H}}(t) = \beta \int_0^\infty \sigma(t,a) \int_0^\infty \Big[& R_{\mathrm{M}}(t-a) f_{\mathrm{M}}(v+a) \\ & + (1 - R_{\mathrm{M}}(t-a)) \int_0^a \sum\nolimits_{j=\mathrm{F},\mathrm{H}} r_j(t-a+u) g_j(u) \\ & \times \exp\left(- \int_0^u Q(t-a+u', u')\,\mathrm{d}u' \right) \\ & \times f_j(v+a-u)\,\mathrm{d}u \Big] \Omega_{\mathrm{H}}(v)\,\mathrm{d}v\mathrm{d}a. \end{aligned} \qquad (15)$$

This can be represented symbolically in terms of integral operators. For instance, for $\beta = 0$, we can define $\mathcal{F}$ and $\mathcal{G}$ (Anderson *et al.* 1996) thus

$$\begin{aligned} [\mathcal{F} \cdot \boldsymbol{y}](t) = \int_0^\infty \sigma_{\mathrm{M}}(t,a) \int_0^\infty & \boldsymbol{y}(t-a) \\ & \times f_{\mathrm{M}}(v+a) \Omega_{\mathrm{M}}(v)\,\mathrm{d}v\mathrm{d}a, \end{aligned} \qquad (16)$$

$$\begin{aligned} [\mathcal{G} \cdot \boldsymbol{y}](t) = \int_0^\infty \sigma_{\mathrm{M}}(t,a) \int_0^\infty & \boldsymbol{y}(t-a) \\ & \times \Big[\int_0^a r_{\mathrm{F}}(t-a+u) g_{\mathrm{F}}(u) \\ & \times \exp\left(- \int_0^u Q(t-a+u', u')\,\mathrm{d}u' \right) \\ & \times f_{\mathrm{F}}(v+a-u)\,\mathrm{d}u \Big] \Omega_{\mathrm{M}}(v)\,\mathrm{d}v\mathrm{d}a, \end{aligned}$$

where $\boldsymbol{y}(t)$ is an arbitrary function of t. Note that $\mathcal{F}$ and $\mathcal{G}$ are effectively generation operators: $\mathcal{F}$ calculates the fraction maternally infected in one generation by dams who themselves were infected maternally, while $\mathcal{G}$ calculates those maternally infected by dams who were infected via feed.

This allows (14) to be rewritten as

$$R_{\mathrm{M}} = \epsilon((\mathcal{F} - \mathcal{G}) \cdot R_{\mathrm{M}} + \mathcal{G} \cdot 1). \qquad (17)$$

The solution to (17) can then be generated iteratively, stepping through the generations

$$R_{\mathrm{M}}^{(n+1)} = \epsilon((\mathcal{F} - \mathcal{G}) \cdot R_{\mathrm{M}}^{(n)} + \mathcal{G} \cdot 1), \qquad (18)$$

where $R_{\mathrm{M}}^{(0)}(t) = 0$. It can then be shown that

$$R_{\mathrm{M}}^{(n)} = \left[\sum\nolimits_{i=1}^{n} \epsilon^{i-1} (\mathcal{F} - \mathcal{G})^{i-1} \right] \cdot \epsilon\mathcal{G} \cdot 1, \qquad (19)$$

where $R_{\mathrm{M}}^{(n)}$ converges to R_{M} as $n \to \infty$ provided $\epsilon < 1$.

When $\beta > 0$, the convergence criterion $\epsilon < 1$, becomes $\epsilon + \beta < 1$, though, if high infectiousness is restricted to the end of the incubation period, this is overly conservative due to the low probability of an animal surviving to late-stage incubation.

Finally, in the case of horizontal transmission, expressions for the basic reproduction ratio via that route, $\boldsymbol{R}_0^{(\mathrm{H})}$, and the number of secondary *infections* generated

by one primary *case* occurring at time t, $\boldsymbol{I}^{(\mathrm{H})}$, are useful:

$$\boldsymbol{R}_0^{(\mathrm{H})}(t) = \beta \int_0^t \int_0^\infty \int_0^{t-t_0} \int_{t-(t_0+a)}^t \int_0^{t-v} q(t_\mathrm{p}, v, t) \times \Omega_\mathrm{H}(t-(t_0+a))\sigma(t_0+u+a, u+a) \times g_\mathrm{H}(a) f_\mathrm{H}(u)\, \mathrm{d}t_\mathrm{p}\mathrm{d}v\mathrm{d}a\mathrm{d}u\mathrm{d}t_0, \quad (20)$$

$$\boldsymbol{I}^{(\mathrm{H})}(t) = \beta \int_0^t \int_0^{t-t_0} \int_{t-(t_0+a)}^t \int_0^{t-v} q(t_\mathrm{p}, v, t) \times \Omega_\mathrm{H}(t-(t_0+a)) \times \sigma(t_0+a, a) g_\mathrm{H}(a)\, \mathrm{d}t_\mathrm{p}\mathrm{d}v\mathrm{d}a\mathrm{d}t_0. \quad (21)$$

Here $q(t_\mathrm{p}, v, t)$ is the joint PDF for the time of birth, t_p, and the incubation period for the primary case, v, conditional on it experiencing disease onset at time t, and under the assumption that the population experiences an infinitesimal force of infection that is constant through time:

$$q(t_\mathrm{p}, v, t) = \frac{\sigma(t, t-t_\mathrm{p}) g_\mathrm{F}(t-v-t_\mathrm{p}) f_\mathrm{F}(v)}{\int_0^t \int_0^{t-v} \sigma(t, t-t_\mathrm{p}) g_\mathrm{F}(t-v-t_\mathrm{p}) f_\mathrm{F}(v)\, \mathrm{d}t_\mathrm{p}\mathrm{d}v}. \quad (22)$$

These equations are necessarily complex, as they are effectively averaging over the distributions of possible case ages and incubation times for both the primary case and secondary cases.

(c) *Heterogeneity in susceptibility*

The model can be further generalized to allow for variable susceptibility to infection of classes within the population. Let s_i denote the proportion of the total population in class i and let w_i denote the relative susceptibility to infection of class i to class 1 for $i > 1$. Thus, assuming the relative susceptibilities are independent of infection risk, $Q^i(t, a)$, the infection hazard for class i, is defined by:

$$Q^i(t, a) = w_i Q(t, a),$$

where $w_1 = 1$. We therefore obtain different joint PDFs for the ages at infection and disease onset for each susceptibility class:

$$\rho_\mathrm{C}^i(t_0, a, u) = w_i \sum\nolimits_j r_j(t_0+a) g_j(a) \times \exp\left(-\int_0^a Q^i(t_0+a', a')\, \mathrm{d}a' \right) f_j(u-a). \quad (23)$$

Similarly, each susceptibility class has its own probability that an animal is infectious through transmission route k, given as

$$y_k^i(t, a) = \int_0^\infty \Omega_k(v) \int_0^a \rho_\mathrm{C}^i(t-a, a', v+a)\, \mathrm{d}a'\mathrm{d}v. \quad (24)$$

The feed-related risk rate is as given previously; however, the horizontal transmission rate becomes

$$r_\mathrm{H}(t) = \beta \sum\nolimits_i s_i \int_0^\infty \sigma(t, a) y_\mathrm{H}^i(t, a)\, \mathrm{d}a. \quad (25)$$

We do not consider the effect of susceptibility classes on the maternal transmission rate since it requires the inheritance probabilities to be specified for each dam–calf class combination.

The probability of an animal born at time t_0 becoming a case by age A is:

$$p_\mathrm{C}(t_0, A) = \int_0^A S(u, t_0) \left(\sum\nolimits_i s_i \int_0^u \rho_\mathrm{C}^i(t_0, a, u)\, \mathrm{d}a \right) \mathrm{d}u. \quad (26)$$

(d) *Under-reporting*

The available data are on confirmed cases reported as suspects to MAFF. Thus, any under-reporting will need to be accounted for in the probability model. Let $\Lambda(t)$ be the probability that a case that onset at time t was reported. Thus, the probability of an individual born at time t_0 becoming a case by age A and being subsequently reported to MAFF is:

$$p_\mathrm{RC}(t_0, A) = \int_0^A \Lambda(t_0+u) S(u, t_0) \times \left(\sum\nolimits_i s_i \int_0^u \rho_\mathrm{C}^i(t_0, a, u)\mathrm{d}a \right) \mathrm{d}u. \quad (27)$$

(e) *Maximum likelihood methods*

The data arise from a multinomial distribution and the likelihood can be written in terms of the probability $p_\mathrm{RC}(t_0, A)$. Let N_{t_0} be the number of calves born in the time interval t_0 to $t_0 + \Delta_t$, so that $N_{t_0} = \int_{t_0}^{t_0+\Delta_t} B(t)\, \mathrm{d}t$. Let X_{i,t_0} be the number of cases among calves born in the time interval t_0 to $t_0 + \Delta_t$ with onset between ages A_i and A_{i-1} from $A_0 = 0$ up to the maximum possible age of onset $A_{i_{\max(t_0)}}$. The maximum age $A_{i_{\max(t_0)}}$ varies with t_0 since cohorts have been observed for variable amounts of time. Thus,

$$N_{t_0} - \sum\nolimits_{i=1}^{i_{\max(t_0)}} X_{i,t_0}$$

is the number of calves born in the time interval t_0 to $t_0 + \Delta_t$ which were not observed to experience disease onset by age $A_{i_{\max,t_0}}$. Ignoring additive constants, the complete data log likelihood (l) is written as a sum over cohorts, t_0:

$$l = \sum\nolimits_{t_0} \left(N_{t_0} - \sum\nolimits_{i=1}^{i_{\max(t_0)}} X_{i,t_0} \right) \times \ln\left(1 - \frac{\int_{t_0}^{t_0+\Delta_t} B(t) p_\mathrm{RC}(t, A_{i_{\max(t_0)}})\, \mathrm{d}t}{\int_{t_0}^{t_0+\Delta_t} B(t)\, \mathrm{d}t} \right) + \sum\nolimits_{i=1}^{i_{\max(t_0)}} X_{i,t_0} \times \ln\left(\frac{\int_{t_0}^{t_0+\Delta_t} B(t) p_\mathrm{RC}(t, A_i)\, \mathrm{d}t}{\int_{t_0}^{t_0+\Delta_t} B(t)\, \mathrm{d}t} - \frac{\int_{t_0}^{t_0+\Delta_t} B(t) p_\mathrm{RC}(t, A_{i-1})\, \mathrm{d}t}{\int_{t_0}^{t_0+\Delta_t} B(t)\, \mathrm{d}t} \right).$$

The following distributions/functions needed to be parametrized or estimated from independent data: (i) the time-dependent risk of infection from feed, ($r_\mathrm{F}(t)$) (specific forms were presented above for maternal and horizontal transmission); (ii) the incubation period distributions, ($f_j(t)$)—which may be transmission route spe-

cific; (iii) the age-dependent susceptibility/exposure distributions to various risks, $(g_j(a))$; (iv) the infectiousness distributions, $(\Omega_k(v))$; (v) the survival distributions $(S(a,t_0))$; (vi) the birth rate through time, $(B(t))$; (vii) the proportions in susceptibility classes, (s_i), and their relative susceptibilities, (w_i); and (viii) the probability that a case is reported, $(\Lambda(t))$.

Independent demographic information provides estimates of the birth rates and the survival distribution (Donnelly *et al.* 1997*b*). The maternal cohort study provides evidence that calves of BSE-affected dams are more likely to become cases than calves of unaffected dams (Wilesmith *et al.* 1997; Donnelly *et al.* 1997*c*; Gore *et al.* 1997; Curnow *et al.* 1997). This increased risk could be due to maternal transmission or differential genetic susceptibility to the BSE aetiological agent, or both. The maternal cohort study and the oral dosing study (Anderson *et al.* 1996) both provide some information on the incubation period distribution.

Parameters requiring estimation were fitted by maximizing the likelihood using direction set (Jacobs 1977; Press *et al.* 1992) and Latin hypercube sampling (McKay *et al.* 1979) methods. Some use was also made of simulated annealing techniques (Kirkpatrick *et al.* 1983; Kirkpatrick 1984; Vanderbilt & Louie 1984).

(*f*) *Feed-risk estimation and prediction of future trends*

In order to predict future infections and cases, it is necessary to predict future trends in feed risk. A parametric form may be assumed for the time-dependent risk of infection from feed throughout the epidemic. In this case, the prediction of future feed risk is straightforward. If, however, the feed risk is fit non-parameterically (as is preferable), then the future feed risk is more difficult to predict. Assumptions must be made about the form of future trends because no information is available to allow non-parametric fitting. This is also true of feed-risk levels in the recent past, where the probability that any cases resulting from infection in this period will yet have been observed is low. For this reason, feed risk can only be reliably fitted up to approximately two years before the most recent case data.

We use model estimates of the recent trends in feed risk (from mid-1991 to mid-1993 approximately) to extrapolate, using simple linear regression, to mid-1996. A further constraint is also imposed: namely that feed risk beyond mid-1996 is assumed to be zero. The justification for this is that all mammalian MBM was banned from use in livestock feed from 29 March 1996, and from 1 August 1996 it became an offence to possess feed containing mammalian MBM.

The purpose of these constraints is to give a better indication of uncertainty in past infection rates (about which nothing can now be done), and how this will affect the future of the epidemic, rather than to effectively speculate about future compliance with the MBM ban. Exploring the complete range of future risk without restriction is fairly meaningless, since it is obvious that if MBM feed restrictions were relaxed, any number of future infections might occur.

(*g*) *Confidence and prediction intervals*

The goodness-of-fit of the model can be judged by comparing the maximum model likelihood with the saturated data likelihood using a χ^2 distribution. For the BSE epidemic, using maximum likelihood (ML) estimation is preferable to using simple least-squares methods on numbers of cases stratified by age and time of onset: the observed number of cases with ages of onset greater than 10 years is too small to permit minimization of the standardized residuals for individual age-of-onset years. The loss of information incurred by amalgamating all such cases can have the effect of producing artefactually good fits for models that, in reality, cannot fit to the tail of the age-at-onset distributions of (in particular) earlier cohorts.

The likelihood analyses do assume conditional independence of the observations. McCullagh & Nelder (1989) indicate that the nominal standard errors for multinomial (generalized linear) models are underestimates if overdispersion has been ignored. Thus the standard errors need to be inflated and the goodness-of-fit statistic needs to be correspondingly deflated. This is particularly relevant in the case of the BSE epidemic, due to the observed within-holding clustering of cases seen (Donnelly *et al.* 1997*b*). This phenomenon results from the reduction in the effective sample size caused by correlation and would affect our standard error and goodness-of-fit measures as well. Conversely, however, when one samples the distribution of χ^2 goodness-of-fit statistics directly from the model using bootstrap techniques, the effective degrees-of-freedom appear to be consistently fewer than the number obtained by subtracting the number of model parameters and constraints from the number of multinomial outcomes. This appears to arise because of the relatively large number of outcomes with low probabilities. This effect will tend to counterbalance the former, in that it will result in over-optimistic estimates of the goodness-of-fit. However, it will not affect judgements of differences in goodness-of-fit between different model variants or estimates of parameter confidence limits.

The simultaneous likelihood ratio confidence region for all of the ML estimated parameters contains all combinations of parameters which provide a similar goodness-of-fit to the observed data, as measured by the likelihood ratio χ^2 statistic. Thus, the likelihood ratio 95% confidence intervals are defined by the multidimensional region containing only the combinations of parameters corresponding to a log likelihood within $\frac{1}{2}\chi^2_{P,0.95}$ of the maximum log likelihood where P is the number of parameters estimated using maximum likelihood methods.

The corresponding confidence interval for any resulting quantity is obtained from the range of values arising from the combinations of parameters contained in the confidence region. Prediction intervals can be similarly defined to bound the predictions resulting from all combinations of parameters providing a similar goodness-of-fit to the observed data.

The high dimensionality and complex geometry of the parameter space makes the characterization of this region highly computationally intensive. We therefore restrict our search to determining the boundaries of the region that intersect with the individual parameter axes through the best-fit point, i.e. the univariate confidence bounds. This inevitably means that the widths of confidence limits quoted below are lower bounds, especially in the case of highly correlated parameters (e.g. those defining the age-dependent susceptibility and incubation period distributions). However, many of the sensitivity analyses presented below are explicitly aimed at identifying correlated parameters. Moreover, through the use of many different functional forms for all parametric distributions, we begin to explore the space of possible models, as well as the parameter space for individual models.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

3. CORRECTION OF DATA BIASES

The model utilizes two main data types: BSE case reports, stratified by birth cohort and age at onset; and demographic data on the British cattle herd. Both are discussed in the companion paper to this (Donnelly *et al.* 1997*b*), so here we concentrate on a discussion of the data biases present in the case database and methods for their elimination.

All model fits described in this paper use case data stratified by year of age-at-onset and birth cohort year. There are several problems associated with using finer age or time stratifications. Most importantly, there are serious biases in the case database, which pose increasingly serious problems for models trying to fit to more finely stratified case data. These are described below.

Beyond data biases, a further impediment to using more finely stratified cohorts is the requirement for extremely reliable birth rate data to be available at the required resolution. While we have reasonable estimates of trends in birth seasonality from 1988 to 1995, this data cannot be considered to be an accurate enough reflection of birth rates in the British herd as a whole to allow reliable estimates of recruitment to be made.

(*a*) *Age-at-onset reporting biases*

A significant proportion of reported BSE cases do not have known dates of birth. Instead the owners have reported an estimated age (in months) at onset, which is significantly more likely than expected to be an integer number of years. Figure 1 (and the discussion in the companion paper to this (Donnelly *et al.* 1997*b*)) illustrates the nature of the problem. In the 1981–1988 cohorts between 25 and 50% of case reports are missing date-of-birth information; this translates into around 25% of cases which arose at the peak of the case epidemic in 1992–1993. However, the reported age at onset is only significantly biased for cases with onset dates before 1991, with a roughly constant 60–70% of such cases arising from the 1981–1988 birth cohorts having their estimated age at onset reported in whole years.

This bias distorts the age-at-onset structure of cases in the affected cohort—contributing to (and perhaps being the sole cause of) an apparent excess of cases in each first half year of age at onset. It even has an effect when yearly stratification is used, since it is thought that farmers without precise birth records for an animal typically underestimate (round down) its age—resulting in a consequent bias in the estimate of the date of birth. Given the form of the survivorship curve, this can result in slight—but significant—underestimation of infection numbers in the early years of the epidemic due to animals being mistakenly attributed to the incorrect birth cohort.

Removing this bias can be addressed using resampling methods. The procedure adopted here was to randomize the month of the reported age at onset for cases with onset dates up to the end of 1990 and for which no dates of birth were given. Due to the bias towards integer ages of onset in the raw data, this has the effect of increasing the mean age at onset for the cases affected. The procedure affects approximately 5000 cases, making up 20% of all cases with onset dates before 1991 and 3% of all reported BSE cases. No resampling of cases after 1990 was performed since the proportion of cases with estimated ages at onset in whole years falls too close to that expected by chance (8.7%) after that time. Figure 1 illustrates the effectiveness of this relatively crude method at removing the observed bias.

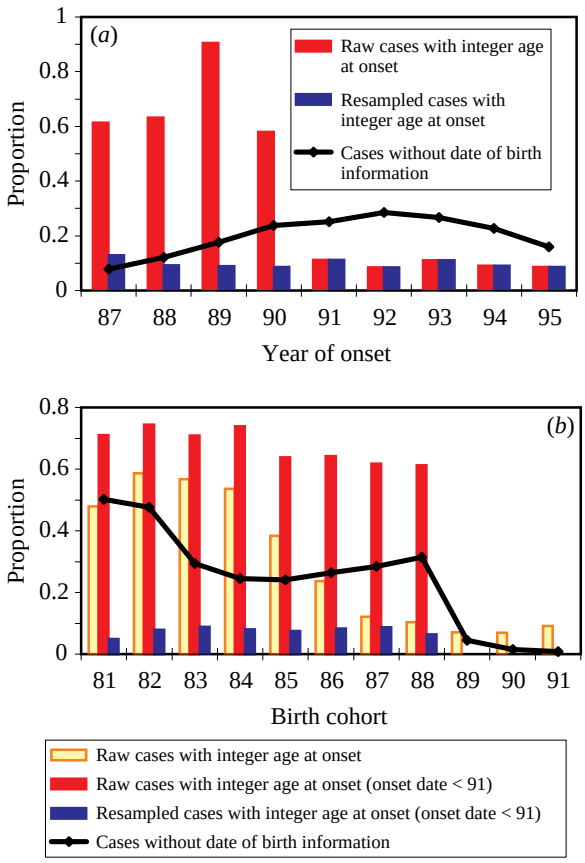


Figure 1. (*a*) The proportion of cases with integer age at onset for the un-resampled and resampled (data, together with the proportion of cases without reported dates of birth, each stratified by onset year. (*b*) The proportion of cases with integer age at onset for the un-resampled data unresampled data with onset dates before 1991 and resampled with onset dates before 1991, together with the proportion of cases without reported dates of birth, each stratified by birth cohort.

If case stratification at temporal resolutions of less than a year were needed, a more sophisticated procedure would be required. An obvious candidate would be to resample the cases with missing date-of-birth information according to the birth distribution seen for cases where dates of birth are known. This is problematical, however, since *a priori* we might expect the frequency with which precise dates of birth were not given to be greater for older animals or for animals which onset early in the epidemic. Hence the age-at-onset distributions (by onset date or cohort) might be expected to be different for animals with

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

and without reported dates of birth. Resolving these complexities presents a considerable hurdle to those wishing to examine the structure of the early BSE epidemic at a fine temporal resolution. Study of the latter stages of the epidemic poses few difficulties; cases born after the introduction of the ruminant feed ban in July 1988 have nearly (greater than 96%) complete date of birth records.

As the debiasing procedure adopted was stochastic, it was necessary to generate a number of resampled datasets and fit the model to each one. The results of this procedure are described in § 4 *b*. However, the spread of model fits was sufficiently narrow to permit sensitivity analysis to be performed using a single representative sample dataset.

(*b*) *Onset seasonality*

There is significant seasonality in month of onset in the case database—with a distinct deficit in January (following Christmas), and excess in September–October. This is probably related to the degree to which cattle are under close observation by farmers, but again is non-trivial to correct. This is discussed more completely in Donnelly *et al.* (1997*b*).

4. SENSITIVITY ANALYSES

This section explores the sensitivity of the model results to changes in: (i) the functional form of the incubation period and age-dependent susceptibility distributions; (ii) resampling to address the biases in reported age-at-onset data; (iii) the mean of the incubation period distribution; (iv) the level of maternal transmission and the duration of maternal infectiousness; (v) the level of horizontal transmission and the duration of horizontal infectiousness; (vi) genetically determined susceptibility levels; (vii) under-reporting of cases in the early stages of the epidemic; (viii) the form (parametric or non-parametric) of the feed-risk profile through time; (ix) demographic factors—birth rate, birth seasonality and survivorship.

We do not explore every combination of factors; to do so proved prohibitively computationally intensive. This is important to note when considering the univariate likelihood profiles presented below; these should be taken as a guide to the comparative goodness-of-fit of the model variants explored, but it should be borne in mind that certain model parameters are inevitably correlated (e.g. rates of maternal and horizontal transmission and the corresponding infectious periods), though wherever possible such correlations are noted. Moreover, in fitting to the case data summed over herds, we are ignoring the significant within-herd and geographical case clustering seen in the BSE epidemic. Models that reproduce such higher moment behaviour will therefore be vital in making a thorough assessment of whether the epidemiological pattern seen in the BSE epidemic are consistent with, in particular, direct horizontal transmission or extreme genetic variability in susceptibility.

That said, we show that the basic results (estimates of numbers of infected animals entering the food chain, predictions of future BSE case numbers) of the model are largely robust to changes in parametrizations of key distributions, and change in a predictable way as a function of other factors (e.g. the level of maternal transmission).

As our baseline, we use the model with no maternal or horizontal transmission, a single susceptibility class (i.e. no genetically variable susceptibility), a non-parametric feed-risk profile with 20 independently fitted knots, and a fitted under-reporting profile $\Lambda(t)$ before July 1988. We also assume that $f_{\mathrm{F}}(u) = f_{\mathrm{M}}(u) = f_{\mathrm{H}}(u) = f(u)$; i.e. that all infected animals show the same incubation period distribution. Note that there is no *a priori* reason for this assumption to be valid, though limited data from the maternal cohort study (Donnelly *et al.* 1997*c*) tends to indicate broadly similar incubation period distributions for the feed-borne and maternal transmission routes. However, by assuming identical incubation period distributions for different routes of transmission, our estimates of the level of maternal or horizontal transmission compatible with the case data are likely to be conservative; were the incubation period distributions substantially different, the goodness-of-fit would be likely to worsen more rapidly with increasing maternal/horizontal transmission due to the growing bimodality of the model age-at-onset distribution.

For the reasons discussed in § 3, we stratified the case data by cohort of birth and age at onset in yearly steps. The cohort of birth is defined on the half-year boundaries, as agricultural censuses are taken in July of each year. Thus the 1987 cohort contains animals born between July 1986 and June 1987. We describe results obtained using the resampled case data (see § 3 *a*) throughout most of this work, but note if these are changed when the raw case data are used. Finally, due to the biases caused by reporting and confirmation delays, the model was fitted to case data with onset dates up to the end of 1995.

(*a*) *Incubation period and age-dependent susceptibility distributions*

For future reference, we define a representative range of the incubation period and age-dependent susceptibility/exposure distributions examined. These are listed in tables 1 and 2.

Of the incubation period distributions, C is worthy of note. This distribution can be derived from a mechanistic model of disease pathogenesis—first explored by Medley & Short (1996)—that assumes that animals are infected with some initial dose d_0, that is distributed as $h(d_0)$, and that prion densities then grow exponentially with time thus:

$$d(t) = d_0 \mathrm{e}^{\gamma_1 t}.$$

Disease onset occurs when the prion density reaches some critical level (arbitrarily set to unity here), at time $t = u$. The incubation period is therefore given

Table 1. *Incubation period distributions, $f(u)$*

(Parameters are α_1, α_2 and α_3, and all are restricted to be positive. Functional forms given are un-normalized.)

code	functional form	description
A	$0, \quad u \leqslant (1-\alpha_1)\alpha_2$ $(u-(1-\alpha_1)\alpha_2)^{\alpha_1^2\alpha_2^2/\alpha_3-1}$ $\times \exp[-(u-(1-\alpha_1)\alpha_2)\alpha_1\alpha_2/\alpha_3], \quad u > (1-\alpha_1)\alpha_2$	gamma distribution with time delay $(1-\alpha_1)\alpha_2$ and mean α_2
B	$0, \quad u \leqslant (1-\alpha_1)\alpha_2$ $(u-(1-\alpha_1)\alpha_2)^{\alpha_3-1}$ $\times \exp[-((u-(1-\alpha_1)\alpha_2)\Gamma(1+1/\alpha_3)/\alpha_1\alpha_2)^{\alpha_3}], \quad u > (1-\alpha_1)\alpha_2$	Weibull distribution with time delay $(1-\alpha_1)\alpha_2$ and mean α_2
C	$(\alpha_2 e^{-u/\alpha_1}/\alpha_3)^{\alpha_2^2/\alpha_3} \exp[-\alpha_2 e^{-u/\alpha_1}/\alpha_3]$	mechanistic incubation period distribution with gamma distributed initial dose

Table 2. *Age-dependent susceptibility/exposure distributions, $g(a)$*

(Parameters are γ_1, γ_2, γ_3 and γ_4, and all are restricted to be positive. Functional forms given are un-normalized. CDF is cumulative density function.)

code	functional form	description
1	$\text{PDF} = e^{-a/\gamma_1}$	exponentially decaying susceptibility (constant exposure assumed)
2	$\text{PDF} = e^{-a/\gamma_1} + \gamma_2$	exponentially decaying susceptibility with constant background (constant exposure)
3	$\text{PDF} = a^{\gamma_2-1}e^{-a/\gamma_1}$	gamma distributed susceptibility (constant exposure)
4	$\text{PDF} = e^{-a/\gamma_1} + \gamma_2, \quad a \leqslant 2$ $= 2(e^{-a/\gamma_1} + \gamma_2), \quad a > 2$	as 2, but with step function exposure, doubling at two years of age
5	$\text{PDF} = 1, \quad a \leqslant \gamma_2$ $= e^{-a/\gamma_1} + \gamma_3, \quad a > \gamma_3$	constant susceptibility to age γ_2, then exponentially decaying with constant background (constant exposure)
6	$\text{PDF} = 1$	constant exposure and susceptibility
7	$\text{CDF} = (1-\exp[-(\gamma_1 a)^{\gamma_2}])(1-\exp[-(\gamma_3 a)^{\gamma_2+\gamma_4}])$	empirically derived, extremely flexible distribution (Anderson *et al.* 1996) (constant exposure)
8	complex (see description)	as 7, but with step function exposure, doubling at two years of age

by:

$$u = -\frac{\log d_0}{\gamma_1},$$

and is distributed as

$$f(u) = h(d_0(u))\frac{\mathrm{d}d_0}{\mathrm{d}u} = h(e^{-\gamma_1 u})\gamma_1 e^{-\gamma_1 u}.$$

The primary reason for the success of this distribution is that it can better reproduce the observed time-lag of two years before any cases are seen than simpler forms. This is because the best-fit initial exposure distribution peaks at doses far below the critical dose, thereby always requiring a considerable period of exponential growth before disease onset is reached. Moreover, the exact form of $h(d_0)$ is not important; while we use the gamma distribution here, nearly identical results are obtained using the Beta or Weibull distributions.

A and B represent more usually adopted incubation period distributions for infectious agents. However, in order to properly reproduce the observed initial delay before any disease onset is seen, it is necessary to explicitly include a time delay into their functional form. Their complex parametrization in table 1 is to allow the mean incubation period, μ_I, to be one of the parameters—enabling exploration of the likelihood profile of the model as μ_I is varied. Such a parametrization was not possible for form C, making it computationally unfeasible to fix μ_I when using that distribution.

There is little biological or epidemiological information on which to base a choice of the age-

Table 3. *Sensitivity analysis of model results to choice of $f(u)$ and $g(a)$, using case data resampled to remove biases described in § 3 a*

(The goodness-of-fit χ^2 and its degrees of freedom (d.f.) are given for each fit, together with parameter values, I_T (total number of infections 1974–1995), and C_F (the predicted number of cases from 1997–2001). Zero maternal and horizontal transmission was assumed for all model runs. Under-reporting before July 1988 is also fitted (see § 4 *e*).)

age-dependent susceptibility distribution, $g(a)$		incubation period distribution, $f(u)$ A	B	C
1	χ^2 (d.f.)	1731 (221)	2238 (221)	1310 (221)
	I_T	945 000	905 000	1 024 000
	C_F	5060	4390	6960
	parameter values	$\alpha_1 = 0.570$, $\alpha_2 = 5.28$ $\alpha_3 = 1.47$, $\gamma_1 = 0.907$	$\alpha_1 = 0.463$, $\alpha_2 = 5.14$ $\alpha_3 = 2.36$, $\gamma_1 = 0.961$	$\alpha_1 = 0.863$, $\alpha_2 = 0.00405$ $\alpha_3 = 2.52 \times 10^{-5}$, $\gamma_1 = 0.811$
2	χ^2 (d.f.)	1026 (220)	1612 (220)	800 (220)
	I_T	955 000	920 000	1 117 000
	C_F	5580	4720	18 900
	parameter values	$\alpha_1 = 0.643$, $\alpha_2 = 5.37$ $\alpha_3 = 1.41$, $\gamma_1 = 0.731$ $\gamma_2 = 0.00393$	$\alpha_1 = 0.471$, $\alpha_2 = 5.21$ $\alpha_3 = 2.42$, $\gamma_1 = 0.801$ $\gamma_2 = 0.00389$	$\alpha_1 = 1.038$, $\alpha_2 = 0.00772$ $\alpha_3 = 7.19 \times 10^{-5}$, $\gamma_1 = 0.573$ $\gamma_2 = 0.00321$
3	χ^2 (d.f.)	469 (220)	1296 (220)	350 (220)
	I_T	1 046 000	1 050 000	1 039 000
	C_F	18 200	26 200	11 000
	parameter values	$\alpha_1 = 0.690$, $\alpha_2 = 5.68$ $\alpha_3 = 1.46$, $\gamma_1 = 2.99$ $\gamma_2 = 0.126$	$\alpha_1 = 0.502$, $\alpha_2 = 5.49$ $\alpha_3 = 2.84$, $\gamma_1 = 2.63$ $\gamma_2 = 0.198$	$\alpha_1 = 1.305$, $\alpha_2 = 0.0176$ $\alpha_3 = 2.34 \times 10^{-4}$, $\gamma_1 = 3.02$ $\gamma_2 = 0.103$
4	χ^2 (d.f.)	1053 (220)	1645 (220)	748 (220)
	I_T	1 023 000	928 000	1 110 000
	C_F	13 600	5200	19 000
	parameter values	$\alpha_1 = 0.658$, $\alpha_2 = 5.56$ $\alpha_3 = 1.50$, $\gamma_1 = 0.552$ $\gamma_2 = 0.00187$	$\alpha_1 = 0.477$, $\alpha_2 = 5.28$ $\alpha_3 = 2.49$, $\gamma_1 = 0.640$ $\gamma_2 = 0.00179$	$\alpha_1 = 1.076$, $\alpha_2 = 0.00887$ $\alpha_3 = 8.85 \times 10^{-5}$, $\gamma_1 = 0.516$ $\gamma_2 = 0.00144$
5	χ^2 (d.f.)	1027 (219)	1546 (219)	800 (219)
	I_T	955 000	924 000	1 117 000
	C_F	5580	4860	18 900
	parameter values	$\alpha_1 = 0.643$, $\alpha_2 = 5.37$ $\alpha_3 = 1.41$, $\gamma_1 = 0.731$ $\gamma_2 = 0.0$, $\gamma_3 = 0.00393$	$\alpha_1 = 0.467$, $\alpha_2 = 5.19$ $\alpha_3 = 2.30$, $\gamma_1 = 0.726$ $\gamma_2 = 0.599$, $\gamma_3 = 0.00770$	$\alpha_1 = 1.038$, $\alpha_2 = 0.00772$ $\alpha_3 = 7.19 \times 10^{-5}$, $\gamma_1 = 0.573$ $\gamma_2 = 0.0$, $\gamma_3 = 0.00321$
6	χ^2 (d.f.)	74 116 (222)	72 990 (222)	74 685 (222)
	I_T	11 144 000	11 103 000	11 129 000
	C_F	101	114	93
	parameter values	$\alpha_1 = 1.0$, $\alpha_2 = 3.92$ $\alpha_3 = 0.157$	$\alpha_1 = 0.847$, $\alpha_2 = 3.96$ $\alpha_3 = 10.0$	$\alpha_1 = 1.674$, $\alpha_2 = 0.100$ $\alpha_3 = 6.25 \times 10^{-4}$
7	χ^2 (d.f.)	377 (218)	710 (218)	308 (218)
	I_T	926 000	931 000	954 000
	C_F	9550	16 200	9340
	parameter values	$\alpha_1 = 0.675$, $\alpha_2 = 4.87$ $\alpha_3 = 1.66$, $\gamma_1 = 1.35$ $\gamma_2 = 0.655$, $\gamma_3 = 0.823$ $\gamma_4 = 3.44$	$\alpha_1 = 0.509$, $\alpha_2 = 4.75$ $\alpha_3 = 2.14$, $\gamma_1 = 1.31$ $\gamma_2 = 0.651$, $\gamma_3 = 0.777$ $\gamma_4 = 2.94$	$\alpha_1 = 1.103$, $\alpha_2 = 0.0196$ $\alpha_3 = 4.06 \times 10^{-4}$, $\gamma_1 = 1.45$ $\gamma_2 = 0.645$, $\gamma_3 = 0.826$ $\gamma_4 = 3.77$
8	χ^2 (d.f.)	383 (218)	724 (218)	313 (218)
	I_T	927 000	926 000	953 000
	C_F	9500	15 600	9390
	parameter values	$\alpha_1 = 0.676$, $\alpha_2 = 4.87$ $\alpha_3 = 1.66$, $\gamma_1 = 2.78$ $\gamma_2 = 0.533$, $\gamma_3 = 0.807$ $\gamma_4 = 3.67$	$\alpha_1 = 0.509$, $\alpha_2 = 4.74$ $\alpha_3 = 2.14$, $\gamma_1 = 2.64$ $\gamma_2 = 0.532$, $\gamma_3 = 0.774$ $\gamma_4 = 3.23$	$\alpha_1 = 1.119$, $\alpha_2 = 0.0210$ $\alpha_3 = 4.51 \times 10^{-4}$, $\gamma_1 = 2.95$ $\gamma_2 = 0.530$, $\gamma_3 = 0.800$ $\gamma_4 = 4.15$

Table 4. *Sensitivity analysis of model results to choice of $f(u)$ and $g(a)$, using raw (uncorrected) case data*

(The goodness-of-fit χ^2 and its degrees of freedom (d.f.) are given for each fit, together with parameter values, I_T (total number of infections 1974–1995), and C_F (prediction of total number of cases from 1997–2001). Zero maternal and horizontal transmission was assumed for all model runs. Under-reporting before July 1988 is also fitted (see § 4 *e*).)

age-dependent susceptibility distribution, $g(a)$		incubation period distribution, $f(u)$ A	B	C
1	χ^2 (d.f.)	1918 (221)	2420 (221)	1527 (221)
	I_T	932 000	886 000	1 017 000
	C_F	5500	4380	7520
	parameter values	$\alpha_1 = 0.574$, $\alpha_2 = 5.29$ $\alpha_3 = 1.46$, $\gamma_1 = 0.914$	$\alpha_1 = 0.463$, $\alpha_2 = 5.14$ $\alpha_3 = 2.37$, $\gamma_1 = 0.967$	$\alpha_1 = 0.877$, $\alpha_2 = 0.00430$ $\alpha_3 = 2.77 \times 10^{-5}$, $\gamma_1 = 0.818$
2	χ^2 (d.f.)	1250 (220)	1786 (220)	917 (220)
	I_T	1 023 000	914 000	1 119 000
	C_F	14 300	5200	20 000
	parameter values	$\alpha_1 = 0.655$, $\alpha_2 = 5.53$ $\alpha_3 = 1.51$, $\gamma_1 = 0.635$ $\gamma_2 = 0.00450$	$\alpha_1 = 0.473$, $\alpha_2 = 5.24$ $\alpha_3 = 2.45$, $\gamma_1 = 0.799$ $\gamma_2 = 0.00403$	$\alpha_1 = 1.047$, $\alpha_2 = 0.00795$ $\alpha_3 = 7.55 \times 10^{-5}$, $\gamma_1 = 0.571$ $\gamma_2 = 0.00343$
3	χ^2 (d.f.)	581 (220)	1424 (220)	447 (220)
	I_T	1 050 000	953 000	1 043 000
	C_F	19 300	8870	12 200
	parameter values	$\alpha_1 = 0.689$, $\alpha_2 = 5.70$ $\alpha_3 = 1.49$, $\gamma_1 = 2.87$ $\gamma_2 = 0.160$	$\alpha_1 = 0.502$, $\alpha_2 = 5.49$ $\alpha_3 = 2.84$, $\gamma_1 = 2.48$ $\gamma_2 = 0.238$	$\alpha_1 = 1.298$, $\alpha_2 = 0.0171$ $\alpha_3 = 2.26 \times 10^{-4}$, $\gamma_1 = 2.82$ $\gamma_2 = 0.148$
4	χ^2 (d.f.)	1161 (220)	1813 (220)	866 (220)
	I_T	1 022 000	925 000	1 113 000
	C_F	14 100	5820	20 300
	parameter values	$\alpha_1 = 0.659$, $\alpha_2 = 5.57$ $\alpha_3 = 1.51$, $\gamma_1 = 0.551$ $\gamma_2 = 0.00194$	$\alpha_1 = 0.480$, $\alpha_2 = 5.30$ $\alpha_3 = 2.52$, $\gamma_1 = 0.639$ $\gamma_2 = 0.00186$	$\alpha_1 = 1.084$, $\alpha_2 = 0.00910$ $\alpha_3 = 9.22 \times 10^{-5}$, $\gamma_1 = 0.516$ $\gamma_2 = 0.00153$
5	χ^2 (d.f.)	1250 (219)	1710 (219)	917 (219)
	I_T	1 023 000	914 000	1 119 000
	C_F	14 300	5080	20 000
	parameter values	$\alpha_1 = 0.655$, $\alpha_2 = 5.53$ $\alpha_3 = 1.51$, $\gamma_1 = 0.635$ $\gamma_2 = 0.0$, $\gamma_3 = 0.00450$	$\alpha_1 = 0.467$, $\alpha_2 = 5.19$ $\alpha_3 = 2.30$, $\gamma_1 = 0.726$ $\gamma_2 = 0.621$, $\gamma_3 = 0.00799$	$\alpha_1 = 1.047$, $\alpha_2 = 0.00795$ $\alpha_3 = 7.55 \times 10^{-5}$, $\gamma_1 = 0.571$ $\gamma_2 = 0.00$, $\gamma_3 = 0.00343$
6	χ^2 (d.f.)	74 201 (222)	73 082 (222)	74 802 (222)
	I_T	11 145 000	11 110 000	11 114 000
	C_F	107	121	102
	parameter values	$\alpha_1 = 1.0$, $\alpha_2 = 3.92$ $\alpha_3 = 0.152$	$\alpha_1 = 0.827$, $\alpha_2 = 3.95$ $\alpha_3 = 10.0$	$\alpha_1 = 1.673$, $\alpha_2 = 0.100$ $\alpha_3 = 6.03 \times 10^{-4}$
7	χ^2 (d.f.)	455 (218)	794 (218)	388 (218)
	I_T	903 000	924 000	930 000
	C_F	9460	17 400	9450
	parameter values	$\alpha_1 = 0.696$, $\alpha_2 = 4.80$ $\alpha_3 = 1.66$, $\gamma_1 = 1.26$ $\gamma_2 = 0.669$, $\gamma_3 = 0.766$ $\gamma_4 = 3.87$	$\alpha_1 = 0.512$, $\alpha_2 = 4.73$ $\alpha_3 = 2.14$, $\gamma_1 = 1.29$ $\gamma_2 = 0.652$, $\gamma_3 = 0.759$ $\gamma_4 = 3.01$	$\alpha_1 = 1.146$, $\alpha_2 = 0.0241$ $\alpha_3 = 5.71 \times 10^{-4}$, $\gamma_1 = 1.29$ $\gamma_2 = 0.672$, $\gamma_3 = 0.771$ $\gamma_4 = 4.64$
8	χ^2 (d.f.)	462 (218)	813 (218)	392 (218)
	I_T	900 000	919 000	915 000
	C_F	9290	16 600	9060
	parameter values	$\alpha_1 = 0.699$, $\alpha_2 = 4.79$ $\alpha_3 = 1.66 = 5$, $\gamma_1 = 2.51$ $\gamma_2 = 0.548$, $\gamma_3 = 0.750$ $\gamma_4 = 4.18$	$\alpha_1 = 0.511$, $\alpha_2 = 4.73$ $\alpha_3 = 2.15$, $\gamma_1 = 2.58$ $\gamma_2 = 0.534$, $\gamma_3 = 0.762$ $\gamma_4 = 3.30$	$\alpha_1 = 1.158$, $\alpha_2 = 0.0262$ $\alpha_3 = 6.57 \times 10^{-4}$, $\gamma_1 = 2.44$ $\gamma_2 = 0.562$, $\gamma_3 = 0.730$ $\gamma_4 = 5.34$

dependent susceptibility/exposure distribution. It should be noted, however, that the two most extreme forms of age-dependent susceptibility, namely no age variation in susceptibility, and no susceptibility except in the first few weeks of life, can be rejected. Were $g(a)$ uniform (form 6), and at the level seen in young animals, much higher incidence levels would have been expected, and the distinctive age distribution of BSE cases seen (majority of BSE cases occurring in cattle between four and seven years of age) would be impossible to explain. Similarly, older cattle must have some low level of susceptibility, since BSE cases have been seen in cattle born as long ago as 1974.

We therefore explore a wide range of potential forms of $g(a)$ (table 2). All forms other than 4 and 8 assume constant exposure levels with age, while the latter assume exposure doubles once animals move into the dairy herd at two years of age. Some justification can be gained for this hypothesis from limited data collected on animal feeding practices in study herds (Donnelly *et al.* 1997*b*). Forms 7 and 8 are unique in being able to take on a very wide range of shapes, including highly asymmetric distributions that rise to a very sharp peak in the first year of life, but have a significant long tail.

Tables 3 and 4 show the results of model fitting with all combinations of the three forms of $f(u)$ and eight forms of $g(a)$. In each case, all parameters were fitted. A model with complete flexibility in the forms of both $f(u)$ and $g(a)$ could not in general be expected to independently estimate both an incubation-period distribution and age-dependent susceptibility distribution from one set of age-structured case data, but the restricted range of shapes of many of the $g(a)$ forms makes such an exercise informative here. Table 3 gives results obtained using the resampled case data (see § 3 *a*), while table 4 shows results obtained using the raw cases data. In general, all models fit better to the resampled case data, though care must obviously be taken in interpreting improvements in model fit associated with such relatively crude resampling techniques. It is more relevant to note that the ranking of model fits is nearly entirely consistent for both tables. Indeed, for all the sensitivity analyses presented in this paper, the qualitative effect of parameter changes on goodness-of-fit was identical for both the resampled and raw datasets.

Firstly, note that models A6, B6 and C6, which assume $g(a)$ is uniform, completely fail to qualitatively fit the observed data. Indeed, they only perform as well as they do by predicting very high feed infectivity in the period 1981–1983, producing 40%+ rates of incidence in those years, and then relying on the under-reporting profile $\Lambda(t)$ to begin to match the observed temporal pattern of case numbers.

Otherwise, the consistent pattern is that incubation period distribution C is always the best-fit form out of the three considered, followed by A, while B always fits significantly worse. Similarly, the best fitting forms of $g(a)$ are always 7 and 8, followed by 3. Forms 7 and 8 have the unique property of being able to peak at non-zero a while maintaining a low variance but long tail. A7 is the model design used in Anderson *et al.* (1996); it achieves a better fit here than in the earlier work due to use of an updated and more complete version of the case database, incorporation of more detailed data on birth seasonality into the model, and use of a more generalized under-reporting profile $\Lambda(t)$ (see § 4 *e*).

Figure 2*a* shows $f(u)$ for the models A3, B3 and C3 (from table 3). While the incubation period distributions are very similar in form, the better fitting ones always have increased probability density in their tails. This trend is reflected in figure 2*c*, which shows $f(u)$ for the models A7, B7 and C7. The corresponding plots of $g(a)$ for the form 3 and form 7 models are shown in figures 2*b*, *d*. The key difference between these two forms of $g(a)$ is the location of the maximum: for the best fitting form, 7, the peak occurs at one year of age, not at birth.

Figures 3*a*, *b* give an indication of the quality of fit for models C7 ($\chi^2_{218} = 308$) and B7 ($\chi^2_{218} = 710$), respectively. These plots show the observed and expected numbers of cases by year of age for the birth cohorts 1981–1992. Most of the 402 difference in χ^2 is accounted for by the 1985, 1988, 1989 and 1991 cohorts. In the cases of the 1985 and 1989 cohorts, the difference in goodness-of-fit is obvious; the reason for the difference in fit to the 1988 and 1991 cohorts is not visible, however, being caused by the fact that those two cohorts have the only two known cases of BSE occurring in one-year-old cattle. Incubation period distribution B relies on an initial time delay to reproduce the observed very low probability of cases in animals less than two years old—in which region, the probability density is very low (less than 10^{-15}). Form C generates a much higher, but still small, density in that region, and can therefore 'fit' rare cases in one-year-olds much better. However, even excluding this effect, C7 still fits better than B7 by a χ^2 difference of over 250.

It is worth noting the consistency in the estimates of the mean incubation period from table 3; all (excluding models A6, B6 and C6) lie in the range 4.7–5.3 years. Similarly, estimates of the total number of animals infected all lie in the range 900 000–1 130 000. The main reason for the latter variation is that models with a low mean age at infection will inevitably predict higher numbers of total infections since younger animals have a lower probability of surviving to the typical age at onset of disease.

Predictions of the number of cases between 1997 and 2001 vary significantly, however. One underlying cause for this is the variation in the rate of decline of the feed-risk profile caused by the complex interplay and varying significance of the tails of the incubation period and age-at-infection distributions. For example, for the forms of $g(a)$ that peak at $a = 0$, a longer tail generates a less rapid decline in the feed-risk profile after 1989, since an increasing proportion of older cases are attributed to having been infected late in life, rather than having incubated for longer than usual.

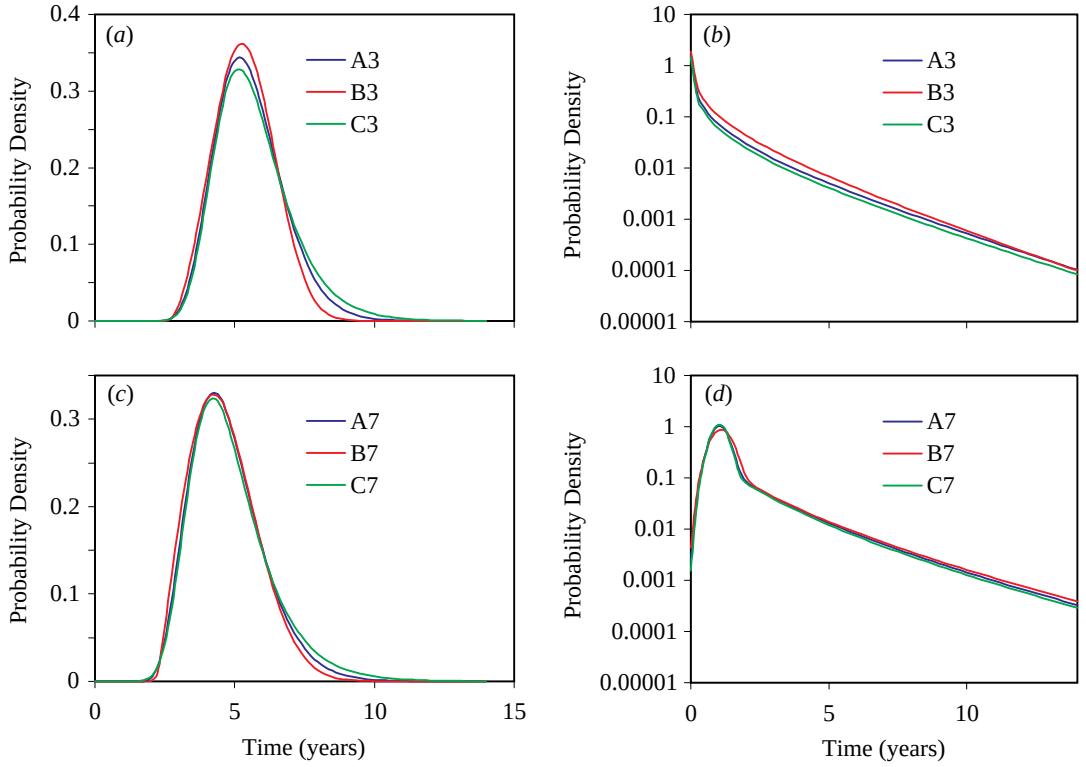


Figure 2. The maximum likelihood estimated PDF corresponding to the incubation-period distribution, $f(u)$, and to the age-at-infection distribution, $g(a)$. (*a*) The maximum likelihood incubation distributions, $f(u)$, of the forms A, B, and C assuming a gamma age-at-infection distribution (form 3), $g(a)$. (*b*) The maximum likelihood gamma age-at-infection distributions, $g(a)$, assuming forms A, B, and C for the incubation-period distribution, $f(u)$. (*c*) The maximum likelihood incubation distributions, $f(u)$, of the forms A, B, and C assuming form 7 (Anderson *et al.* 1996) for the age-at-infection distributions, $g(a)$. (*d*) The maximum likelihood age-at-infection distributions of form 7, $g(a)$, assuming forms A, B, and C for the incubation-period distribution, $f(u)$.

A further confounding factor—responsible for many of the large case number predictions—is that confidence in estimates of feed risk close to the present (the last feed-risk time point fitted was at the end of 1992) is inevitably low, with spuriously high values being sometimes fitted to generally poor fitting models. It should also be noted that models incorporating maternal transmission (§ 4 *c*) generally allow much better estimates of future case numbers to be made, as the effect of maternal transmission is to explain much of the residual tail of the epidemic (of infections), thereby driving feed risk to zero earlier in time—at a point where more confidence can be had in feed-risk estimates.

Finally, it has been postulated, based on serial passaging experiments of transmissable spongiform encephalopathies (TSEs) in rodents (Weissmann 1991; Kimberlin 1993), that the incubation period of BSE might have shortened as the epidemic progressed. However, the case data gives little evidence of a dramatic change—unless it was accompanied by a simultaneous increase in the mean age at infection, and fitting a slight decline is problematic when underreporting is also being fitted. This is discussed further in § 4 *g*.

(*b*) *Fit variability due to date-of-birth estimation debiasing*

As has been noted earlier, a random resampling procedure was used to estimate dates of birth for those cases for which only estimated ages at onset were given (see § 3 *a*). It is therefore necessary to explore how model fits vary for a range of resampled datasets. Figure 4*a* illustrates the relatively tight (mean = 308, s.d. = 5.67) distribution of χ^2 goodness-of-fit values obtained by fitting model C7 to each of 100 resampled datasets. Figure 4*b* shows the corresponding distribution of the estimated numbers of animals infected between 1974 and 1995, I_{T} (mean = 953 800, s.d. = 943), where the variability in sample values is approximately 0.1%. Similarly, figure 4*c* shows the approximately 0.25% variability in C_{F}, the numbers of cases predicted from 1997–2001 (mean = 9364, s.d. = 22.4). The mean incubation period, μ_{I}, also shows less than 0.1% variability (figure 4*d*—mean = 5.0085, s.d. = 0.0032) across fits to 100 resampled datasets. However, the resampled datasets do give model fits with I_{T} and C_{F} some 3–4% larger than obtained with the raw data (see table 4), due to the fact that the best fit value of μ_{I} is some 2% smaller for the raw data. Hence use of the resampled data is conservative, in the sense of examining the worst-case scenario. As has already

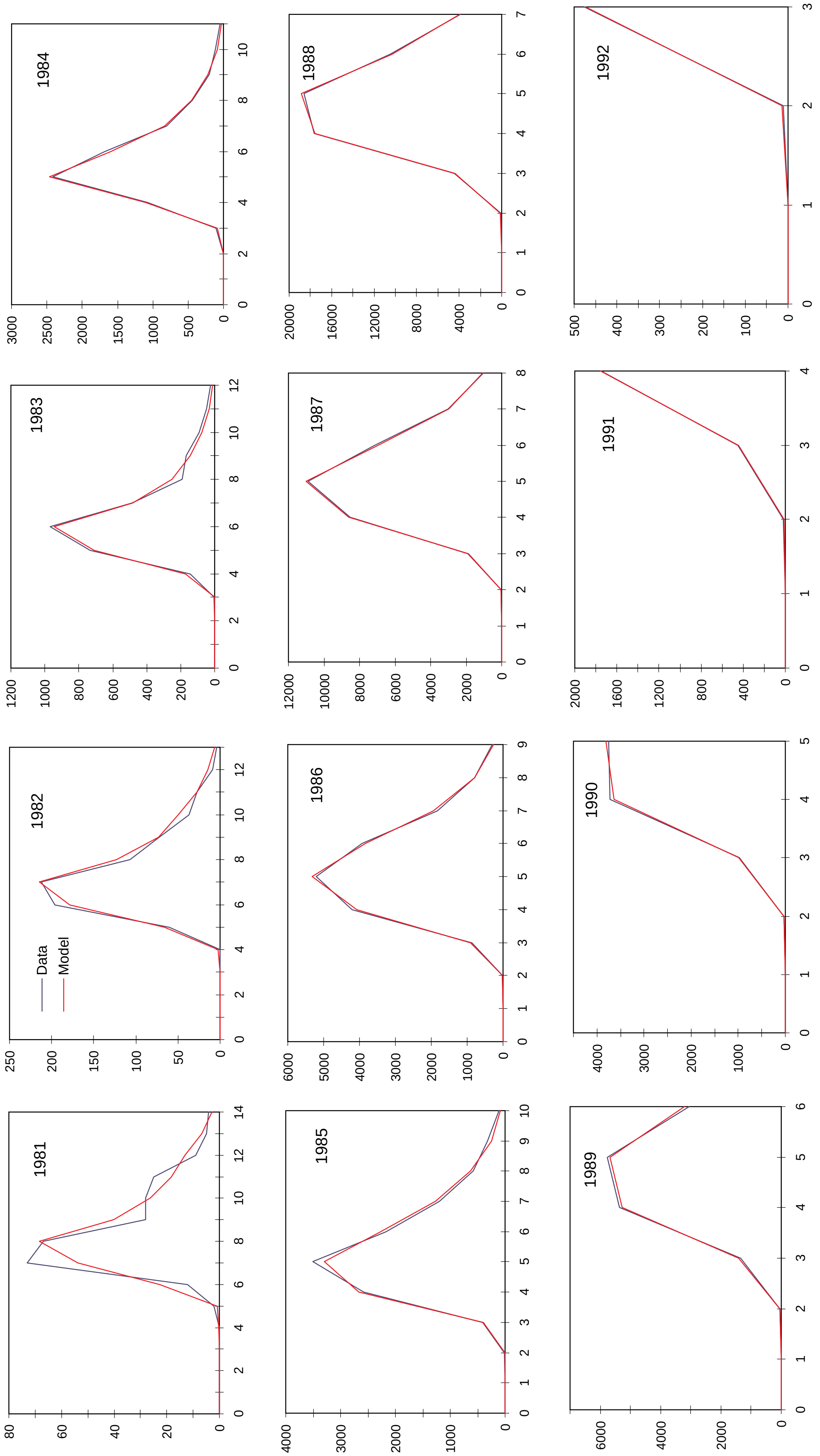


Figure
1992 as

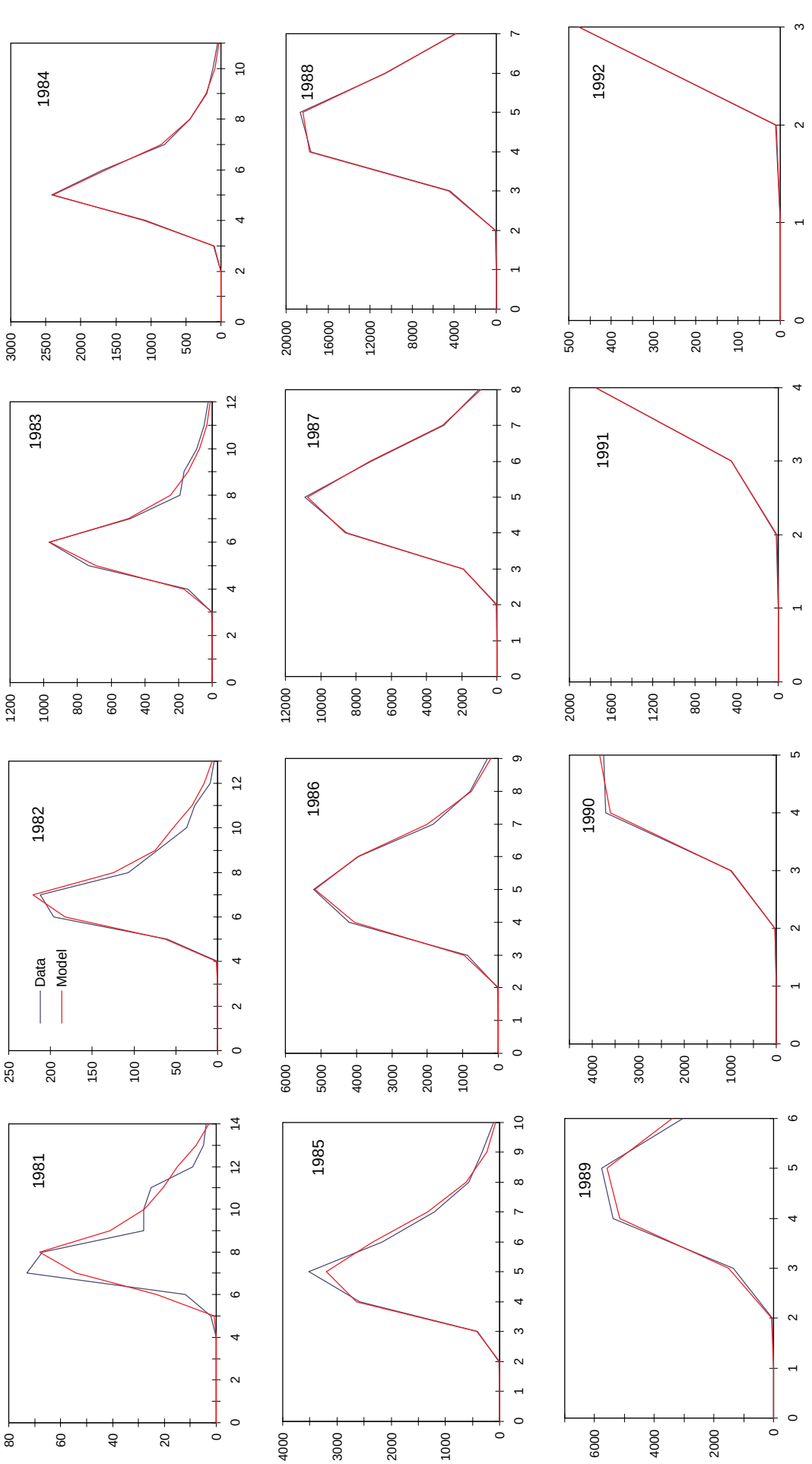


Figure 3. 1981 to 19

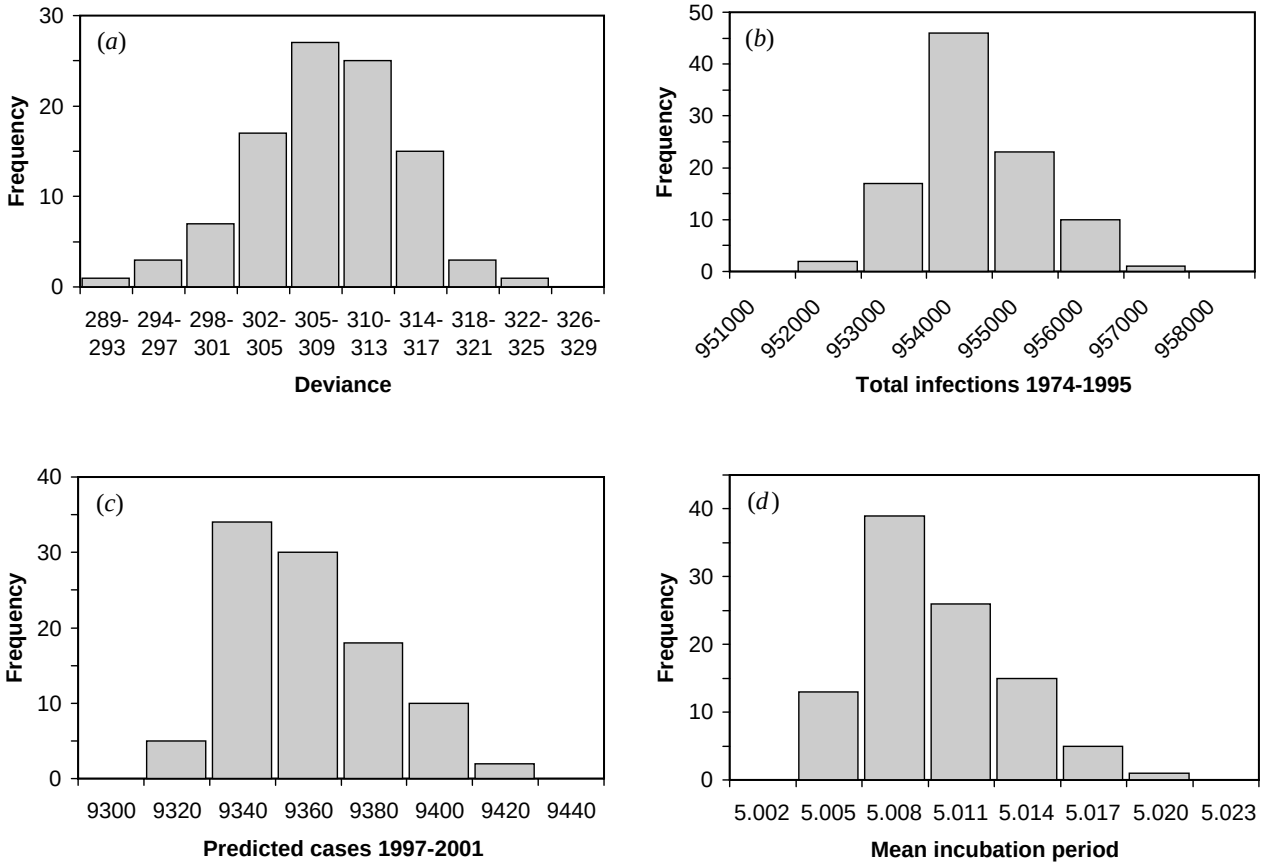


Figure 4. The model was fitted to 100 resampled datasets in order to explore the variability in model results due to the resampling procedure. The frequency distributions of the resulting goodness-of-fit χ^2 values (*a*); the estimated total number of cattle infected between 1974 and 1995 (*b*); the number of cases predicted to arise 1997–2001 (*c*); and the estimated mean of the incubation-period distribution are displayed (*d*).

been noted in § 4 *a*, the model fits to the resampled datasets were always substantially better than those to the unresampled data. This was due to the (relatively minor) changes to the assignment of cases to cohorts: resampling has the effect of moving a few cases to the birth cohort below that to which it might naively be assigned when the date of birth is estimated by subtracting the estimated age at onset from the date of onset. This has the effect of somewhat ameliorating the rapid drop in the mean age at onset seen in the cohorts affected at the start of the epidemic. The improvement in goodness-of-fit therefore reflects the relative difficulty the model has in fitting this rapidly changing age structure of cases well (see § 4 *g*). It should be noted that the goodness-of-fit is sensitive to the resampling algorithm used: algorithms that resample many fewer cases result in χ^2 values more comparable to that obtained with the raw data, while crude resampling of many more cases (e.g. applying the algorithm used here to cases with onset after 1990) can also result in a considerably worsened fit.

(*c*) *Mean incubation period*

The results above already give some indication of the relationship between mean incubation period, μ_{I}, mean age at infection and goodness-of-fit. For incubation period distributions A and B, however, we can fix, rather than fit, the mean incubation period. This allows us to examine the likelihood profile and epidemiological properties of the model as a function of incubation period. In what follows below, we use the best fit model out of those with the A or B incubation-period distributions, namely A7. However, the basic trends are unaffected if B7 is used.

Figure 5*a* shows the χ^2_{218} profile and sum of the mean incubation period and age at infections for $4 \leqslant \mu_{\mathrm{I}} \leqslant 6$. The χ^2 profile is almost archetypal in shape, with the 95% confidence interval for μ_{I} being within the range 4.75–5.00. It is interesting to note that within this region, the mean age at infection, μ_a, is fitted so that $\mu_{\mathrm{I}} + \mu_a$ is held constant at a value of 6.45 years. Even outside this region, it is only when $\mu_{\mathrm{I}} + \mu_a$ can no longer be fitted close to this value that the model fit dramatically worsens. It is for this reason that we adopted form 7 for $g(a)$. Other forms fare much worse at being able to maintain a good fit for a variety of values of μ_{I}, while form 7 encompasses a wider range of distribution shapes.

Figure 5*b* illustrates the trend for the estimated total number of infections (from 1974 to 1995), I_{T}, to increase with μ_{I}; as commented upon in § 4 *a*, this is because for longer incubation periods, animals must have been infected earlier and so have a lower probability of surviving to onset—due to the form of the

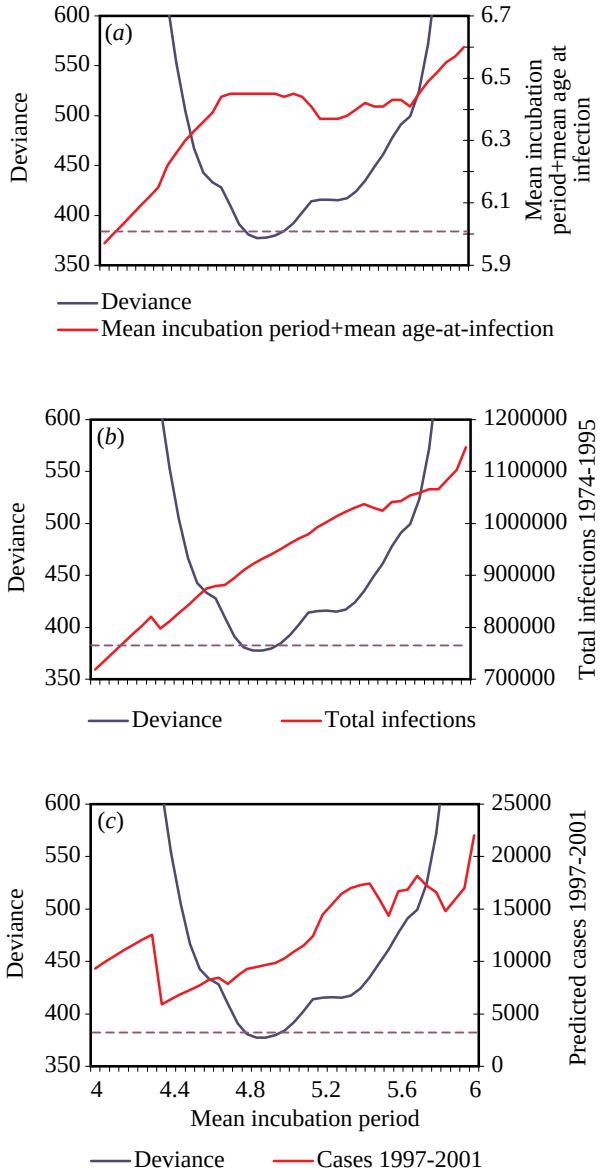


Figure 5. The sensitivity of the model results to the mean of the incubation-period distribution is displayed for (*a*) the sum of the mean incubation period and the mean age at infection, (*b*) the estimated total number of cattle infected between 1974 and 1995 and (*c*) the number of cases predicted to arise 1997–2001. The goodness-of-fit χ^2 value is given for each value of the mean of the incubation-period distribution. The mauve dashed line indicates the 95% confidence level for 1 d.f., relative to the minimum of the χ^2 profile.

survivorship curve (Donnelly *et al.* 1997*b*). A larger number of animals therefore need to be infected to produce the observed number of cases.

Figure 5*c* illustrates how the predicted number of cases from 1997–2001, C_{F}, varies with μ_{I}. Here, for the reasons discussed earlier, the fluctuations in C_{F} seen in the large χ^2 regions should be ignored. Across the 95% (and 90%) confidence interval for μ_{I}, a steady increase in μ_{I} with C_{F} is seen, however.

(*d*) *Maternal transmission*

The results presented here are based on the best fit model C7 from § 4 *a*. Using alternative models (e.g. A7 or C5) does not substantially change the conclusions. Maternal transmission rates are calculated accurately up to third order in ϵ. We use a step function infectiousness distribution $\Omega_{\mathrm{M}}(v)$, i.e. animals have a probability, ϵ, of transmitting BSE to their unborn offspring if they are within time, ω_{M}, of disease onset, otherwise there is zero probability of maternal transmission. We then explore how the model results change as ϵ and ω_{M} are varied.

The key epidemiological concept to bear in mind in the following discussion is that incidence of maternal infections tracks case incidence in the cattle population. For this reason, maternal transmission only reaches significant levels towards 1992—some four years after the feed-based infection peak. Hence maternal transmission only has a significant effect on the tail of the epidemic.

Figure 6 summarizes these results. Figure 6*a* plots the χ^2 deviance against ϵ for three values of ω_{M}, 0.25, 0.5 and 0.75 years. Note that the three curves on each graph in figure 6 are nearly identical, bar a linear rescaling along the ϵ axis. For the $\omega_{\mathrm{M}} = 0.25$ case the 95% confidence interval (for 1 d.f.) for $\omega_{\mathrm{M}} = 0.25$ is contained in the range $[0, 0.16]$, for $\omega_{\mathrm{M}} = 0.5$, $[0, 0.08]$, while for $\omega_{\mathrm{M}} = 0.75$ the range is $[0, 0.04]$. However, it should be noted that it would be naive to expect to be able to estimate the rate of maternal transmission from any model that does not explicitly use dam–calf pair data. That said, the best estimates of the rate of maternal transmission for BSE in cattle obtained from detailed analyses of the results of the maternal cohort study (Donnelly *et al.* 1997*c*) give $\epsilon = 0.082$ and $\omega_{\mathrm{M}} = 0.24$—very close to the best fitting values of ϵ and ω_{M} seen here.

More interestingly, figures 6*b*, *c* show how I_{T} and C_{F} vary with ϵ. The effect of increasing ϵ above zero is to initially decrease the estimated total number of infections over the period 1974–1995. In this regime, maternal transmission is better at explaining the form of the tail of the epidemic—driving the feed risk in the latter stages of the epidemic to disproportionately lower values than might be expected just from the total numbers of maternal infections (figure 6*d*). As ϵ is increased further, however, μ_{I} and μ_a (figures 6*f*, *g*) tend to increase and decrease, respectively, to prevent the increasing numbers of animals being maternally infected at birth from disrupting the age structure of the epidemic. This causes (due to the form of the survivorship function) a gradual increase in I_{T} (figure 6*b*). Of course, this increase is not reflected in the predicted case numbers, C_{F} (figure 6*c*), where a gradual increase is only seen for large ϵ—and can entirely be attributed to the increasing numbers of predicted maternal cases (figure 6*e*).

These results are of particular significance when evaluating the robustness of model predictions of future case numbers. By driving feed risk to low levels earlier than would occur in its absence, maternal transmission (even at low levels) dramatically

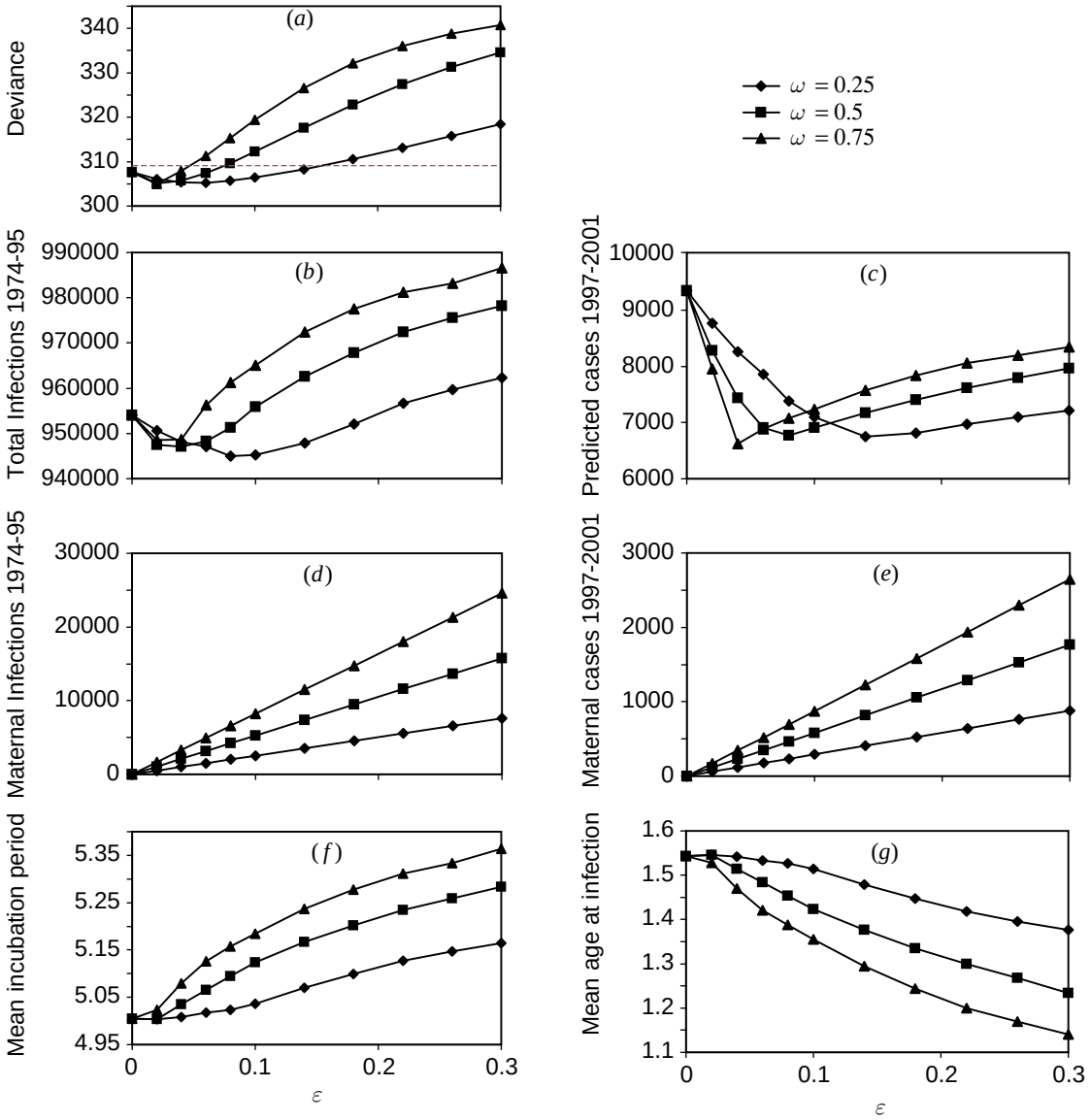


Figure 6. The sensitivity of the model results to the probability of maternal transmission of the aetiological agent of BSE, ϵ, and proportion of the last year of the incubation period in the dam during which maternal transmission can occur, ω_{M}, is displayed for: (*a*) the goodness-of-fit χ^2 value (the dashed line indicates the 95% confidence level for 1 d.f., relative to the minimum of the χ^2 profile); (*b*) the estimated total number of cattle infected between 1974 and 1995, (*c*) the predicted number of cases to arise 1997–2001; (*d*) the estimated number of infections due to maternal transmission between 1974 and 1995; (*e*) the number of cases due to maternal transmission predicted to arise 1997–2001; (*f*) the mean of the incubation-period distribution and (*g*) the mean age at infection.

improves the robustness of any prediction method based on extrapolation of a non-parameterically fitted feed-risk profile. Indeed, we can go as far as stating that, for $\epsilon \geqslant 0.05$ and $\omega_{\mathrm{M}} \geqslant 0.25$, 8000 represents a robust upper bound on the expected case numbers between 1997 and 2001. For $\epsilon = 0.1$, and $\omega_{\mathrm{M}} = 0.5$—the parameters used in Anderson *et al.* (1996)—our prediction of case numbers over 1997–2001 remains unchanged at approximately 6900—despite the improvements in model fit since publication of that paper.

(*e*) *Horizontal transmission*

The existence of direct (non-feed borne) horizontal transmission is a prerequisite for the endemic persistence of an infectious disease agent in its host population. In the case of BSE, such transmission might be postulated to occur through contact with contaminated pasture or placenta, or direct physical contact with an affected animal. However, it should be emphasized that to date no evidence exists supporting the hypothesis of direct horizontal transmission. Given the indirectly horizontal nature of feed-based transmission, analyses of the case data that attempt to determine whether horizontal transmission occurs are desirable but problematic. A realistic goal is to determine the maximum rate of direct horizontal transmission that is consistent with trends in the case data, and whether that rate is sufficient to sustain the epidemic ($\boldsymbol{R}_0^{(\mathrm{H})} > 1$).

Three factors determine the rate of horizontal

transmission: the transmission coefficient, β; the (age-dependent) susceptibility of an uninfected animal, $g_{\mathrm{H}}(a)$; and the (incubation-stage-dependent) infectiousness of an infected animal, determined by the function $\Omega_{\mathrm{H}}(v)$ (see § 2 *b*). In what follows, we make the simplifying assumption that age-dependent susceptibility to horizontal infection is identical to that for feed-borne infection, and that the incubation period distribution is identical in both cases; namely $g_{\mathrm{F}}(a) = g_{\mathrm{H}}(a) = g(a)$ and $f_{\mathrm{F}}(a) = f_{\mathrm{H}}(a) = f(a)$. While *a priori* this clearly cannot be justified in the absence of experimental data, it represents a conservative choice since the age-at-onset distribution arising from feed- and horizontally infected animals will then be identical. This is likely to improve model fits to moderate rates of transmission; by comparison, one of the reasons that a high level of maternal transmission is not well fitted by the model is the differing age-at-onset distributions for the maternal and feed-borne transmission routes. Finally, we assume $\Omega_{\mathrm{H}}(v)$ is step-like in form: namely Ω_{H} rises from zero to one at time ω_{H} before disease onset. We examine three values of ω_{H} here, 0.5, 1.0 and ∞ years (the latter representing constant infectivity throughout the incubation period). For the former two values, any epidemic of horizontal infections tracks the epidemic of feed-infected cases, while in the latter case both epidemics are more synchronized.

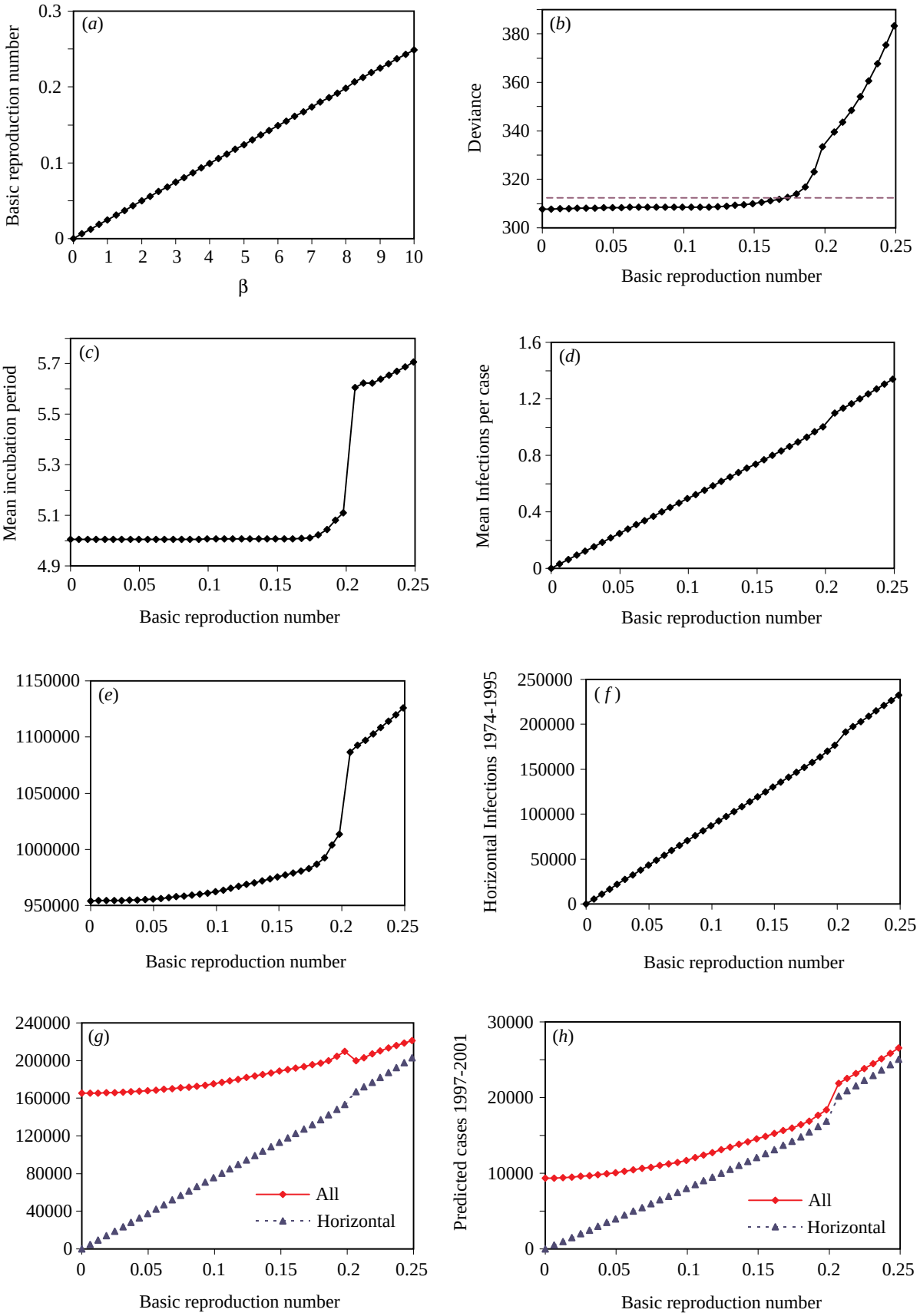
We use equation (20) to calculate $\boldsymbol{R}_0^{(\mathrm{H})}$, making the simplifying assumption that the birth rate, $B(t)$, is constant throughout the epidemic. Since the survivorship curve, $S(a)$, used in this model was time-independent, this results in a time-independent estimate of $\boldsymbol{R}_0^{(\mathrm{H})}$.

Figures 7 and 8 show the results of an extensive set of model runs for $\omega_{\mathrm{H}} = 0.5$ and $\omega_{\mathrm{H}} \to \infty$, respectively. The results for $\omega_{\mathrm{H}} = 1.0$ were nearly identical to those for $\omega_{\mathrm{H}} = 0.5$, except that the $\omega_{\mathrm{H}} = 0.5$ model requires twice the value of β to produce the same value of $\boldsymbol{R}_0^{(\mathrm{H})}$ than is needed in the $\omega_{\mathrm{H}} = 1.0$ model. For figure 8, the range $0 \leqslant \beta \leqslant 0.9$ was explored, while for figure 7 the range was widened to $0 \leqslant \beta \leqslant 10$. In the former case this resulted in exploration of the range $0 \leqslant \boldsymbol{R}_0^{(\mathrm{H})} < 0.25$, while in the latter the range was $0 \leqslant \boldsymbol{R}_0^{(\mathrm{H})} < 0.19$ (figures 7*a* and 8*a*). Figures 7*b* and 8*b* are critical, showing how the model goodness-of-fit varies as a function of $\boldsymbol{R}_0^{(\mathrm{H})}$. For $\omega_{\mathrm{H}} = 0.5$ and $\omega_{\mathrm{H}} \to \infty$ the upper 95% confidence bounds for $\boldsymbol{R}_0^{(\mathrm{H})}$ are approximately 0.15 and 0.09, respectively.

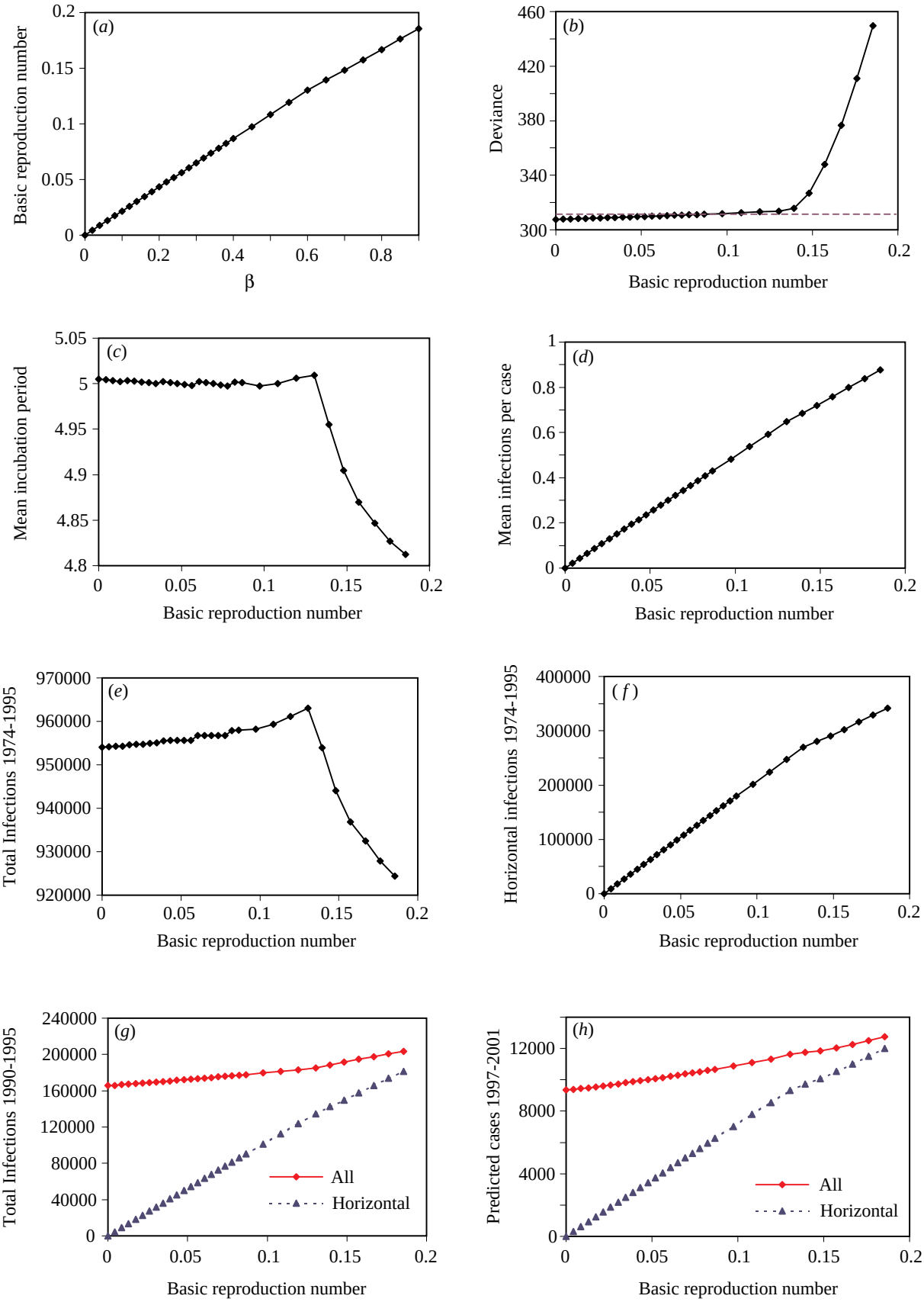
For $\omega_{\mathrm{H}} = 0.5$, when $\boldsymbol{R}_0^{(\mathrm{H})}$ rises above 0.15, the existing local minimum of the χ^2 hypersurface is no longer the global minimum, and the best-fit model 'flips' to a distinctly separate parameter region. This can be seen in figure 7*c*, where the mean incubation period, μ_{I}, suddenly jumps from 5 to around 5.65. This results in a corresponding jump in I_{T} (figure 7*e*). A similar effect is seen in the $\omega_{\mathrm{H}} \to \infty$ model, but in this case μ_{I} falls dramatically as the model fit worsens (figure 8*c*), resulting in a corresponding decline in I_{T} (figure 8*e*).

It is also instructive to examine how $\boldsymbol{I}^{(\mathrm{H})}$, the number of secondary *infections* produced by a single primary *case*, relates to $\boldsymbol{R}_0^{(\mathrm{H})}$ (figures 7*d* and 8*d*). For a disease such as BSE, the distinction between $\boldsymbol{R}_0^{(\mathrm{H})}$ and $\boldsymbol{I}^{(\mathrm{H})}$ is significant: due to the long incubation period of the disease and the survivorship distribution of the British cattle herd, most infected animals do not survive to become cases—resulting in $\boldsymbol{I}^{(\mathrm{H})}$ being nearly exactly five times larger than $\boldsymbol{R}_0^{(\mathrm{H})}$. The production of one secondary infection for each primary case is therefore inadequate to sustain the epidemic. Conversely, this also means that relatively low values of $\boldsymbol{R}_0^{(\mathrm{H})}$ can generate very significant numbers of infections (figures 7*f* and 8*f*)—of the order of 130–140 000 over the period 1974–1995 for $\boldsymbol{R}_0^{(\mathrm{H})} \simeq 0.15$ in the $\omega_{\mathrm{H}} = 0.5$ case. The potential role of horizontal transmission in the latter stages of the epidemic is even more striking, with up to 50–70% of infections in the period 1990–1995 (figures 7*g* and 8*g*) being attributable to horizontal transmission while having a minimal impact on the model goodness-of-fit. Since feed risk declined with time, the impact on the future of the epidemic is more dramatic still, with up to 85% of case numbers from 1997–2001 being attributable to horizontal transmission with $\boldsymbol{R}_0^{(\mathrm{H})} \simeq 0.15$ and $\omega_{\mathrm{H}} = 0.5$. Horizontal transmission also results in increased numbers of cases being predicted in that interval—up to approximately 15 000 for $\omega_{\mathrm{H}} = 0.5$ and 11 000 for $\omega_{\mathrm{H}} \to \infty$, compared with 9300 for the model with $\beta = 0$. Furthermore, the year-on-year excess (particularly for $\omega_{\mathrm{H}} = 0.5$) is greatest for the later years. This is to be expected, since we would expect horizontal transmission with $\boldsymbol{R}_0^{(\mathrm{H})} \simeq 0.15$ and a five year latent period to produce a sub-epidemic with a half-life (following removal of all feed-based risk) of approximately two years—prolonging the epidemic considerably.

Our results therefore produce no evidence to support the hypothesis that horizontal transmission is occurring at a rate sufficient to allow BSE to become endemic in the British herd. This is hardly surprising given the rapid decline in case numbers seen in the last few years, and bearing in mind that—assuming homogeneous mixing, exposure and susceptibility—infection incidence never exceeded 10% in any cohort. It may of course be argued that assuming homogeneous transmission in the entire GB cattle population is simplistic, since only one-third of British herds ever reported a BSE case, and that herds comprising some 20% of the GB national herd are responsible for some 80% of cases. That said, even under the extreme scenario that only 20% of the national herd was *ever* exposed/susceptible to BSE, then the results presented here would still put an upper 95% confidence bound on $\boldsymbol{R}_0^{(\mathrm{H})}$ of 0.75. However, excluding the possibility that $\boldsymbol{R}_0^{(\mathrm{H})} > 1$ for some smaller 'core' subset of animals is clearly impossible without extensive future modelling and analysis of the spatial and within-herd case clustering seen in the BSE database. Such work is in progress.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES



(*f*) *Variable susceptibility/exposure*

One proposed explanation for the results of the maternal cohort study (Wilesmith *et al.* 1997; Donnelly *et al.* 1997*c*) is that the design represented a selection bias (in the case of offspring of infected dams) for animals with higher, genetically determined susceptibility. Ferguson *et al.* (1997*a*) discuss this hypothesis in detail, showing that in the case of a simple single-locus two-allele system determining susceptibility, the minimum ratio of susceptibilities between the high and low susceptibility class is 20.

We therefore explore the effect of splitting the cattle population into two classes, one 20 times more susceptible than the other, and varying the proportion, f, in the high susceptibility class. Note that zero maternal and horizontal transmission was assumed here. Figure 9 shows the result for $0 \leqslant f \leqslant 0.3$. In essence, varying f makes very little difference to the quality of fit (figure 9*a*); for $f > 0.25$, χ^2 rises slowly to the $f = 1$ (or $f = 0$) value of 307.7. I_{T} does rise by around 10 000 in the $0.01 \leqslant f \leqslant 0.1$ region, due to an exactly correlated decrease in μ_a (figure 9*e*). This effect is due to the fact that it is only in this region that appreciable numbers of animals in both the high and low susceptibility classes become infected. The value of μ_a declines, and the $g(a)$ distribution becomes more narrowly peaked, in order to minimize the generation of significant differences in the age structure of infection between the two classes. In essence, for low f, the high susceptibility animals are subject to a very high force of infection, with nearly all animals in the most affected cohorts becoming infected early in life. This is in contrast to the low susceptibility class, where the model predicts that at most 8% in any one cohort become infected.

Once $f > 0.18$ approximately, the great majority of all cases arise from the high-susceptibility class, meaning that the low susceptibility class becomes increasingly irrelevant and the epidemiological behaviour of the model drifts back to the $f = 1$ case.

Clearly, this model does not really test the hypothesis of genetic susceptibility, as it contains no explicit genetics; a true genetic model needs to take account of the extreme variability in exposure between different herds, and would require detailed genetic/progeny information. To date simple versions of such models have failed to find any evidence for genetically variable susceptibility (Wijeratne & Curnow 1990; Curnow *et al.* 1994; Curnow & Hau 1996; Hau & Curnow 1996)—though this work has been hindered by the relative paucity of detailed data across a large enough sample of herds.

However, the results presented here do show that the model is robust against the hypothesis that only a minority of animals were actually exposed to significant feed risk. This hypothesis is supported by the observation that throughout the whole course of the epidemic, cases were only reported on one-third of holdings, with the majority of those only reporting a single case; i.e. cases are strongly clustered by herd. In this scenario, f represents the proportion of herds experiencing a significantly higher level of feed risk than the rest.

(*g*) *Under-reporting*

Quantifying the under-reporting rate is one of the most difficult aspects of any modelling of the BSE epidemic, as estimates of under-reporting, the incubation period and age-dependent susceptibility distribution exhibit a complex interdependency. We will therefore describe the effect of various under-reporting profiles on model fit, before discussing the confounding factors which may complicate any simple interpretation of the results.

In the absence of independent data on reporting rates, it is clearly impossible to fit a time-dependent probability of reporting, $\Lambda(t)$, across the whole epidemic. We therefore make the assumption that $\Lambda(t) \simeq 1$ following the introduction of compulsory notification of BSE cases in mid-1988. It might be argued that a more realistic cut-off date might be February 1990, when 100% compensation for BSE cases was introduced. However, fitting an additional constant under-reporting rate for the period July 1988 to February 1990 made a negligible difference to the model fit to the resampled data ($\chi^2_{217} = 306$), with the estimated under-reporting rate in that period being 3%. That said, fitting a July 1988 to February 1990 under-reporting rate to the raw case data made a significant difference ($\chi^2_{217} = 309$), with the best-fit rate being 13%. The difference may be explained as a result of the similar effects of resampling and under-reporting. Resampling tends to reassign animals to the previous cohort, thereby boosting the numbers of cases seen at relatively young ages at onset (4–6 years) across the 1981–1985 cohorts (since epidemic growth is rapid across those cohorts). This slightly ameliorates the rapidly changing age-at-onset distributions seen for those cohorts, thereby improving the model fit. As discussed below, this is basically the same effect that including an under-reporting profile has on the model fit.

As a starting point, we consider the simplest realistic under-reporting profile, namely that the under-reporting rate, $1/\Lambda(t)$, increased linearly with the time before July 1988 that disease onset occurred:

$$\Lambda(t) = \begin{cases} \dfrac{1}{1 + \delta_1(88.5 - t)}, & t < 88.5, \\ 1, & t \geqslant 88.5. \end{cases} \qquad (28)$$

This attempts to reproduce the fact that reporting became more likely as the disease became better known.

By varying δ_1, one generates the likelihood profile shown in figure 10*a*. We used the best fit model, C7, with no maternal transmission, but the trends seen can be reproduced with all models explored with $\chi^2 < 1500$. We see that if under-reporting is not taken into account at all ($\delta_1 = 0$), the model fits the data very poorly ($\chi^2_{220} = 1542$), being unable to reproduce the rapidly declining mean age at onset seen in the early birth cohorts. However, as δ_1 is increased

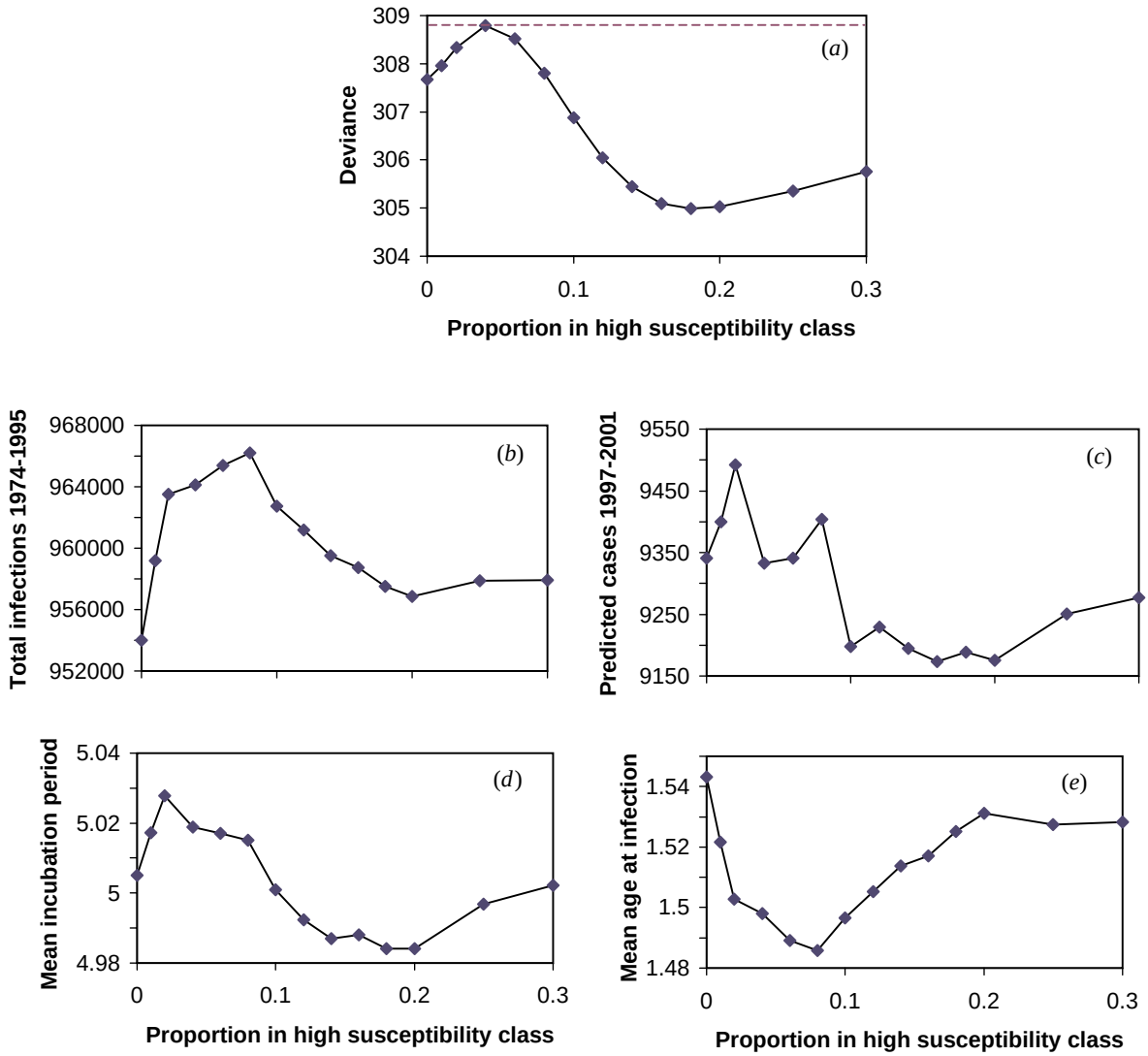


Figure 9. The sensitivity of the model results to the proportion of the population in the high susceptibility class is displayed for: (*a*) the goodness-of-fit χ^2 value (the dashed line indicates the 95% confidence level for 1 d.f., relative to the minimum of the χ^2 profile); (*b*) the estimated total number of cattle infected between 1974 and 1995; (*c*) the number of cases predicted to arise 1997–2001; (*d*) the mean of the incubation period distribution; and (*e*) the mean of the age-dependent susceptibility/exposure distribution, g(a).

above zero, the fit rapidly improves, with the best fit being at $\delta_1 \simeq 4$.

Striking variations are also seen in estimates of I_T and C_F: in the $\delta_1 = 0$ case the model is unable to fit the early (1981–1985 especially) birth cohorts well, and instead optimizes the fit to the later cohorts. However, as soon as even a low level of under-reporting is accounted for (though $\delta_1 = 0.5$ gives $\Lambda = 0.5$ for $t = 86.5$), it becomes optimal to balance fitting the typically older age profile of cases in early cohorts with maintaining a good fit to the later cohorts. This causes a significant rise in μ_a (figure 10*e*)—partly balanced by a corresponding decrease in μ_I (figure 10*d*)—which results in a significant decrease in I_T (figure 10*b*). As δ_1 is increased further, under-reporting increasingly fits the age structure of reported cases in early cohorts better, allowing the μ_a to gradually fall towards the optimal value for the later cohorts.

For the reasons discussed in earlier sections, predicted case numbers (figure 10*c*) are barely affected by changes in μ_a, so a rapid stabilization in C_F to around 9000 cases is seen as δ_1 increases. For small δ_1, C_F is larger due to a good fit to recent cohort data only being achieved by a large increase in the value of the last estimated point of the feed-risk profile—which confounds the feed-risk extrapolation procedure.

We can refine this simple model of under-reporting further, by allowing for a much higher under-reporting rate in the very early years of the epidemic, when BSE was largely unknown and previously un-

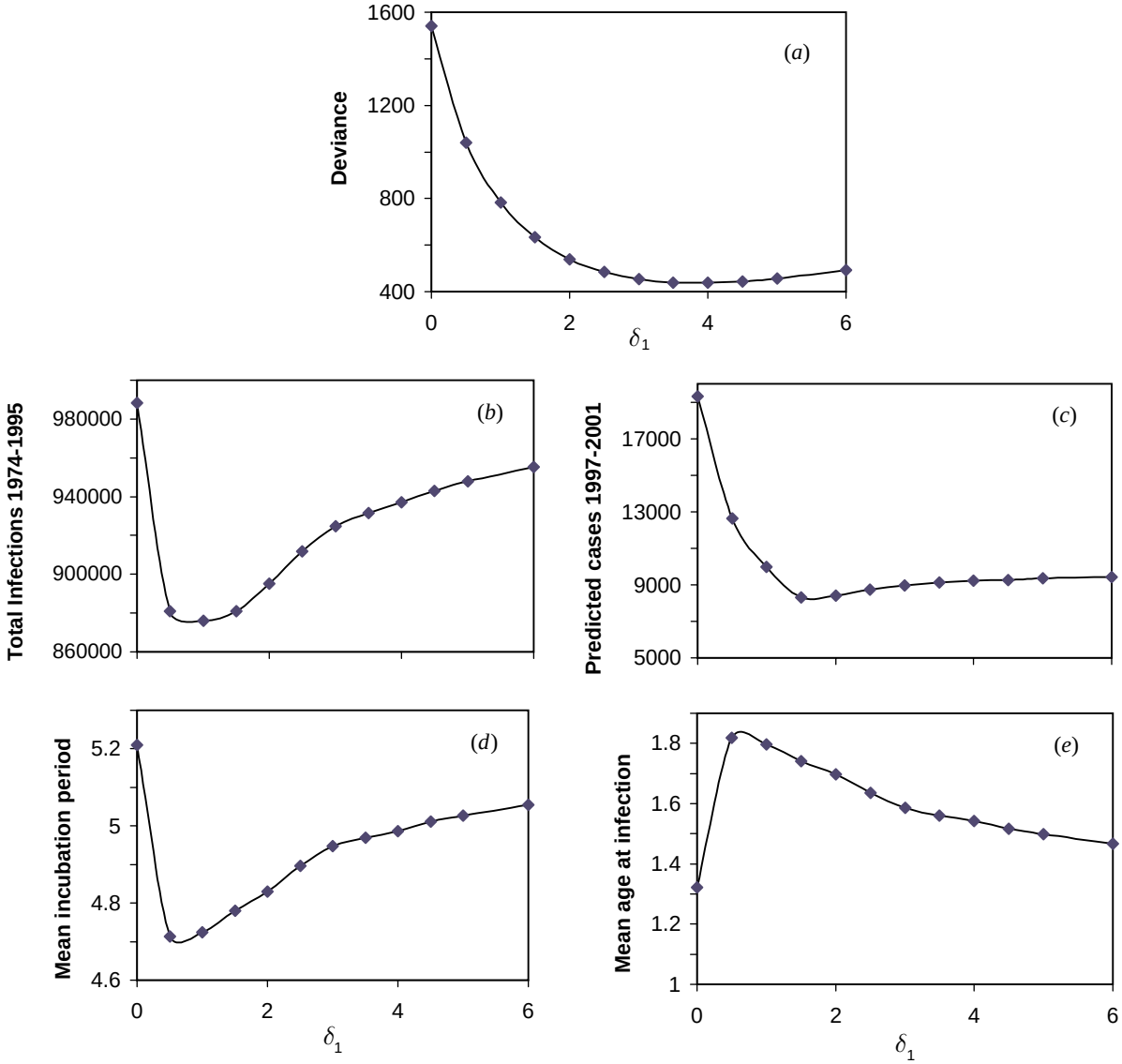


Figure 10. The sensitivity of the model results to the under-reporting parameter δ_1 is displayed for: (*a*) the goodness-of-fit χ^2 value; (*b*) the estimated total number of cattle infected between 1974 and 1995; (*c*) the predicted number of cases to arise 1997–2001; (*d*) the mean of the incubation period distribution; and (*e*) the mean of the age-dependent susceptibility/exposure distribution, g(a).

diagnosed by most veterinarians (i.e. before 1987):

$$\Lambda(t) = \begin{cases} \dfrac{1}{1 + 1.5\delta_1 + (\delta_2(87 - t))^{1.2}}, & t < 87, \\ \dfrac{1}{1 + \delta_1(88.5 - t)}, & 87 \leqslant t < 88.5, \\ 1, & t \geqslant 88.5. \end{cases} \tag{29}$$

Figure 11 shows the results of fitting δ_1 and varying δ_2. In all cases, the fitted value of δ_1 lies in the range (3.5,3.8). It can be seen that for $\delta_2 = 5$, χ^2 (figure 11*a*) starts at about the minimum value achieved using the simpler form of $\Lambda(t)$ (figure 10*a*). However, as δ_2 increases to above 100, a further dramatic fall in χ^2 is seen. For $\delta_2 > 100$, χ^2, I_{T} (figure 11*b*), C_{F}, (figure 11*c*), μ_{I} (figure 11*d*) and μ_a (figure 11*e*) rapidly converge to their asymptotic ($\delta_2 \to \infty$) values. It was this form of $\Lambda(t)$ (with both δ_1 and δ_2 being fitted) that was used in all model fits outside this section.

Part of the reason that very large values of δ_2 give the best fit is that the model has to take account of the fact that no cases were reported before 1986. Since the model fits feed risk non-parameterically, with no assumptions about the start date of infections, some small number (less than one) of cases are predicted for years before 1986—meaning infinite under-reporting best fits the observed absence of cases. This can lead to somewhat artefactual results—namely that the feed-risk profile rises to high levels around 1979–1980, and then temporarily declines before the main epidemic growth. It is for this reason that the power of 1.2 was used in (29), as beyond this value such artefacts occur frequently in model fits.

The effect of the estimated high levels of under-reporting in the early years of the epidemic is illustrated in figure 12, showing estimated reported

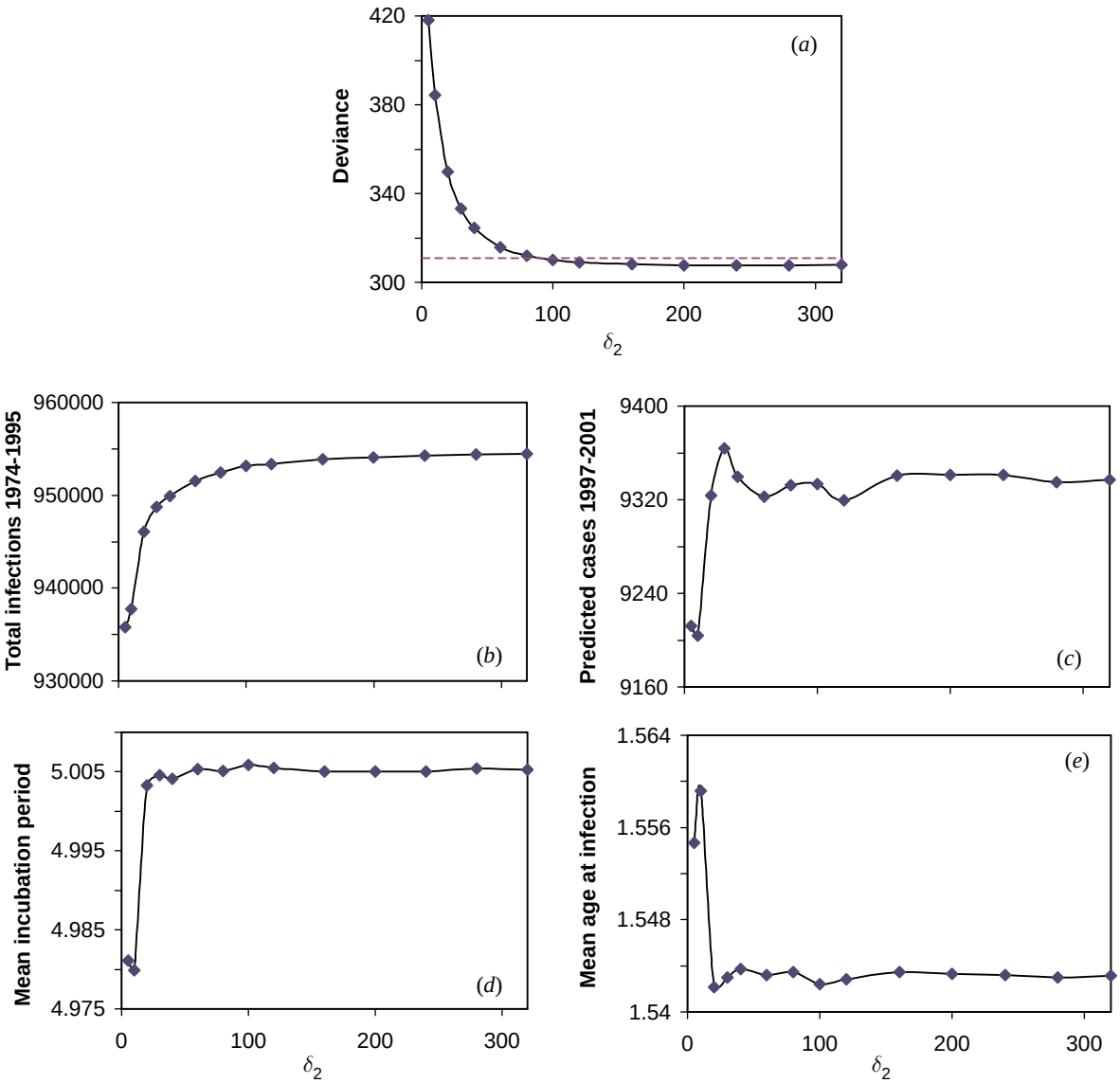


Figure 11. The sensitivity of the model results to the under-reporting profile slope δ_2 is displayed for: (*a*) the goodness-of-fit χ^2 value (the dashed line indicates the 95% confidence level for 1 d.f., relative to the minimum of the χ^2 profile); (*b*) the estimated total number of cattle infected between 1974 and 1995; (*c*) the predicted number of cases to arise 1997–2001; (*d*) the mean of the incubation-period distribution; and (*e*) the mean of the age-dependent susceptibility/exposure distribution, g(a).

and total case numbers for the 1981–1984 birth cohorts (these were taken from the best-fit model C7, and the estimated reported case numbers nearly exactly match those observed). It can be seen that the overall shape of the distribution of disease onset ages changes little between the cohorts, but is shifted—in steps of one year—to earlier ages as one steps through the cohorts. An under-reporting profile allows this pattern to be fitted by attributing the changes in age structure nearly entirely to under-reporting—the age structure of 'total cases' in the four cohorts remaining relatively constant.

The very high rates of under-reporting predicted by the model beg the question of whether they truly reflect the initial stages of the epidemic or are partly artefactual—implying some deficiency in model design. An obvious hypothesis is that the changing age structure is a result of a rapidly increasing force of infection during the early years of the epidemic. However, for this to work would require a much flatter age-dependent susceptibility distribution than can fit the data, given the tight age-at-onset distribution and relatively low incidence observed in even the most affected herds.

A more feasible hypothesis is that, while most animals were always infected early in life, the incubation period dramatically shortened in the initial stages of the epidemic. It is indeed possible to fit models with a variable mean incubation period to the data, but the best fit of such models still only accounts for a small proportion of the changing age structure; they still produce very high estimates of early under-reporting. It is also difficult to speculate on a biological mechanism by which such a dramatic change in incubation period could have occurred over such a short time. Serial passaging of a TSE agent in rodents can

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

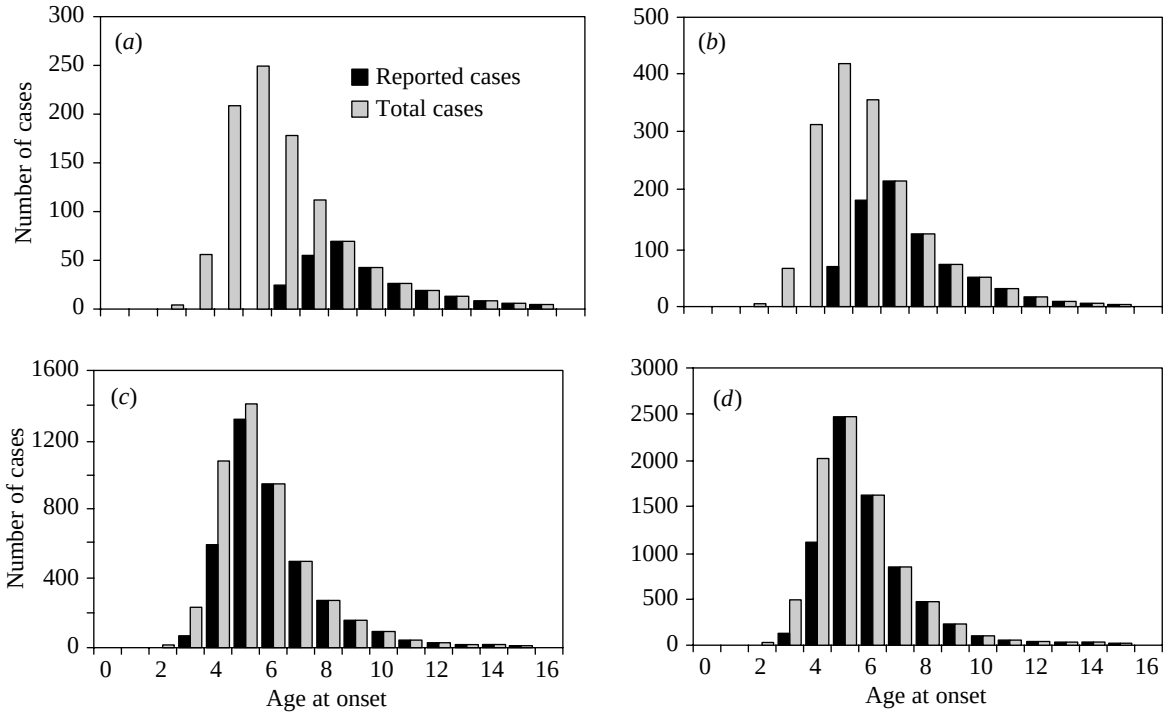


Figure 12. The reported and total (reported and unreported) cases from the best fitting model for birth cohorts: (*a*) 1981; (*b*) 1982; (*c*) 1983; and (*d*) 1984.

certainly produce dramatic changes in incubation period (Weissmann 1991; Kimberlin 1993), but the time between passages is of the order of the mean incubation period. In cows, therefore, the typical serial passage time would be five years or more—making a four year change in incubation period over four sequential birth cohorts difficult to explain. Another possibility—accepting the mechanistic model of the incubation period described in § 4 *a*—is that changes in incubation period were caused by a rapidly increasing mean infective dose of the BSE aetiological agent. This is certainly more feasible, but would require a detailed model of the feed recycling and manufacturing process (to determine the key factors affecting the aggregation distribution of infectious material in feed) to justify. It is also difficult to explain, under this hypothesis (namely, that the mean infective dose scales with overall feed risk), why no increase in the mean incubation period is seen in animals born after MBM restrictions were put in place in mid-1988.

Finally, in any discussion of epidemiological trends in the early stages of the epidemic, it is also necessary to consider the origins of BSE, and the exact nature of the epidemic amplification process. Backcalculation models of the type described here are unable to give insight in this case; it will be necessary to develop models that explicitly describe the recycling process to better bound the range of epidemiological and biological mechanisms that can explain the observed pattern of cases by age and time. The problem with simple (non-spatial) time-delayed SEIR version of such models is that it becomes necessary to fit another time-varying profile—the efficiency with which infectious material is 'transmitted' through the rendering process—which can give rise to many of the same interpretational problems discussed above. It will therefore be necessary to develop herd-based models—with more explicit descriptions of the infection process (i.e. whether one-off or cumulative)—before deeper insight can be gained.

However, while more work needs to be done before a thorough understanding is gained of the early stages of the epidemic, the stabilization in the age-at-onset distribution for animals from the 1985 and later cohorts (from which over 95% of cases arose) means that we can still have confidence in model estimates of the number and pattern of infections, and in predictions of future cases. Further confidence can be gained from figure 6, which shows that even models with low levels ($\delta_1 \simeq 1.5$–2) of under-reporting produce estimates of I_T and C_F within 5% of the best-fit values.

(*h*) *Feed-risk profile*

While the incubation-period distribution, age-dependent susceptibility distribution and under-reporting profile largely determine the age-at-onset distribution, it is the feed-risk profile, $r_F(t)$, which primarily determines the temporal pattern of the epidemic. As stated previously, we chose to fit the profile non-parameterically to avoid making prior assumptions about its form. The approach adopted was spline based, with geometric interpolation between adjacent knots. Knot locations were not fitted, but were chosen so as to increase resolution where $r_F(t)$ was most rapidly changing.

Table 5 lists a few of the knot placement options explored, together with the best fit χ^2 values ob-

tained using model C7 (without maternal or horizontal transmission). A plot of the relationship between numbers of knots fitted, n, and χ^2 is given in figure 13*a*, showing that goodness-of-fit increases rapidly as n is increased to around 20, but that beyond that point rapidly diminishing returns are achieved.

The implementation of the backcalculation model used discrete time to evaluate the convolution integrals, effectively evaluating infection risk for a discrete set of birth times (though within each step, all probabilities from continuous distributions were evaluated exactly, or approximated using robust numerical methods). For all the results published here, a step size, Δt, of 0.25 years was used. For any particular choice of n, this was found to produce results within 1% of those obtained using 12 steps per year. However, since the computational requirements of the model scale as $(1/\Delta t^2)$, it proved impossible to check convergence for smaller values of Δt. It is possible that the magnitude of n for which further increases produce diminishing returns in χ^2, together with the minimum value of χ^2 reached, do vary with Δt, though probably not significantly. Again, with the computational resources available, it proved impractical to explore such asymptotic behaviour thoroughly.

Figures 13*b–g* show the exact form of the feed-risk profile for the options listed in table 5. The most noticeable behaviour is seen in figure 13*b*, for $n = 31$, where the model fit apparently shows a seasonal effect in feed risk. Typically, feed risk is seen to be at a minimum in the second quarter of each year—in line with what is known about cattle feeding practices. The exact quarter in which feed risk is at a peak does vary somewhat, however. Given the small χ^2 difference between the fits shown in figures 13*c*, *d*, and the fact that the model was fitting to data stratified by year of birth and age (rather than finer stratifications), it is obviously difficult to exclude the possibility that the effect observed is somewhat artefactual. It is, however, independent of the data on seasonality in birth rate, $B(t)$, that was used in the model—the same effect is seen when $B(t)$ is assumed to be constant throughout each year. It should be noted that the exact form of the feed-risk profile before 1983 is intimately tied up with the fitting of the under-reporting profile. Similarly, feed-risk estimates for times beyond mid-1992 are less well controlled due to the relatively few cases seen in animals born after that time.

Another noticeable feature of the profiles in figures 13*b–g* is the very rapid rate of change in feed risk just before and after its peak—and the fact that, without exception, the best fit model always estimates the time of maximum risk to be mid-1988—at exactly the time the feed ban was implemented. Indeed, this offers some support for the form of $g(a)$ used—in that if the age of maximum susceptibility is lowered, the location of peak feed risk (and, indeed, the entire feed-risk profile) is shifted, exactly proportionately, forward in time. This is to be expected, as the model always tends to fit to the same age-at-onset distribution, regardless of the age of peak susceptibility (see §§ 4 *a* and 4 *c*).

Lastly, it is informative to compare the goodness-of-fit achieved using a simple parametric model of the feed-risk profile. The model adopted used back-to-back exponentials, with a variable peak location and no requirement for continuity at that point:

$$r_{\mathrm{F}}(t) = \begin{cases} 0, & t < \eta_1 - \eta_6, \\ \eta_2 \mathrm{e}^{-\eta_3(\eta_1 - t)}, & \eta_1 - \eta_6 \leqslant t < \eta_1, \\ \eta_5 \mathrm{e}^{-\eta_4(t - \eta_1)}, & t \geqslant \eta_1. \end{cases} \qquad (30)$$

The maximum likelihood fit achieved with this six parameter feed-risk profile model was $\chi^2_{232} = 1964$. The resulting profile is shown in figure 13*h*—being a crude representation of the non-parameterically fitted versions.

(*i*) *Demographic factors*

The demographic data and estimates utilized by the backcalculation model are described in the companion paper to this work (Donnelly *et al.* 1997*b*). Here, therefore, we briefly comment on the sensitivity of the model to the three key elements of demographic information used.

Optimally, one would want to sample from the set of survivorship distributions consistent (with 95% confidence) with the cross-sectional herd age-structure data from which survivorship is estimated. By then fitting the model for each such survivorship distribution, it would be possible to generate a distribution of χ^2 values. This would allow a detailed analysis of the sensitivity of model results to survivorship variability to be performed. Unfortunately, however, such an analysis is beyond the means of currently available computational resources. That the results of the model are critically dependent on the overall form of the survivorship curve is without question; the fact that most animals are slaughtered at two years of age, before disease onset, is one of the key epidemiological features of the observed BSE epidemic. With regard to the more detailed form of the survivorship curve, only ad hoc explorations of sensitivity have been performed; e.g. by non-fitted adjustments of the survivorship for 8- and 11-year-olds of less than 10% of their initial values, the goodness-of-fit of model C7 is dramatically increased ($\chi^2_{218} = 255$, $p > 0.05$)—though without more systematic investigation, such arbitrary modifications clearly cannot be justified.

The GB national herd decreased in size by some 30% over the period 1974–1995. However, while this trend is important for the accurate estimation of the numbers of BSE infections over time, it is not critical to the model goodness-of-fit, qualitative results on the age-structure of infection, or predicted number of future cases. Fitting model C7 (in the absence of maternal or horizontal transmission) with a constant annual birth rate set to the recorded 1985 level (3 195 000) resulted in $\chi^2_{218} = 313$, with $I_{\mathrm{T}} = 950\,000$, and $C_{\mathrm{F}} = 9202$.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

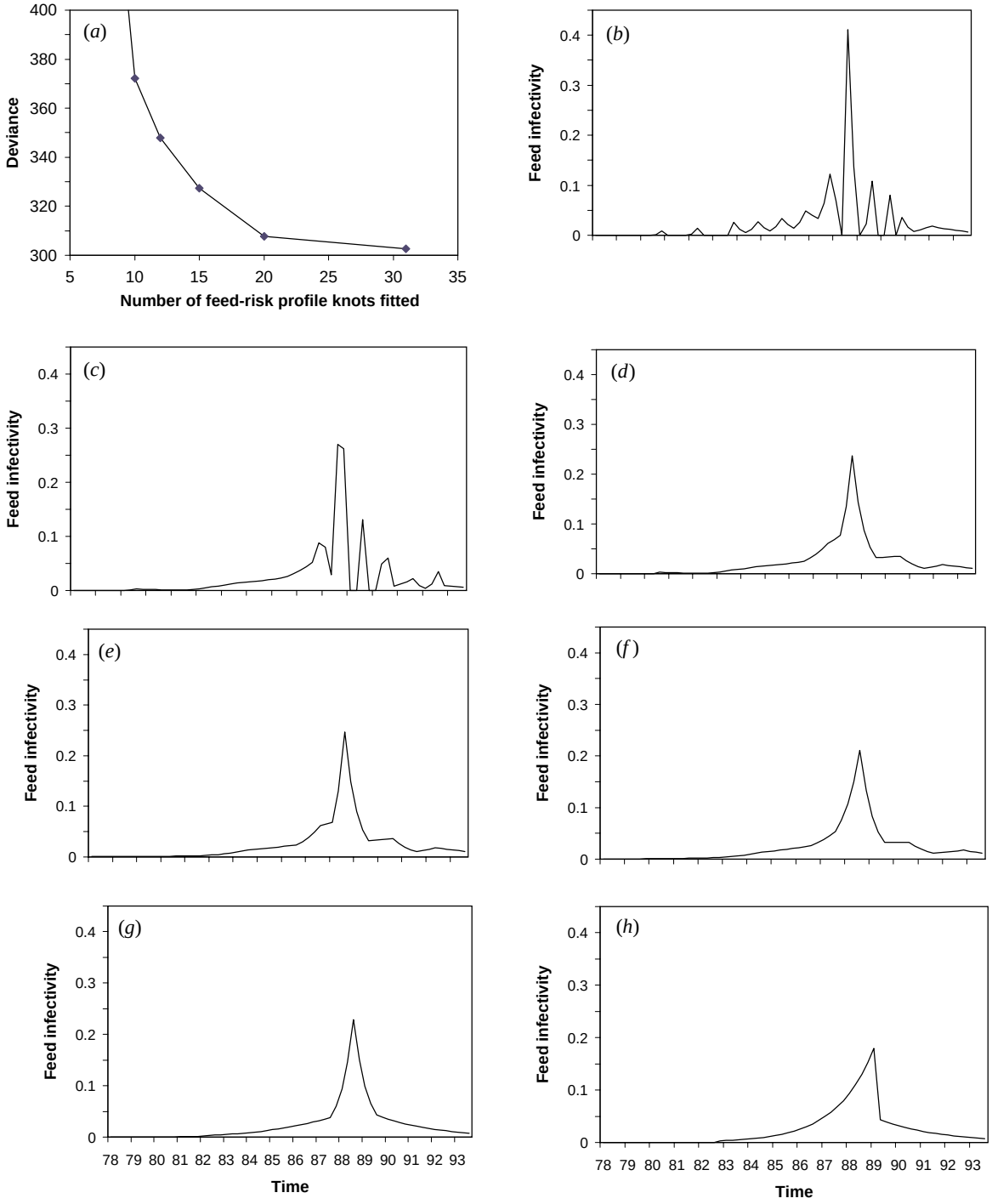


Figure 13. (*a*) Goodness-of-fit χ^2 as a function of the number of knot locations used in the non-parametric fit to the feed-risk profile, $r_F(t)$. The feed-risk profiles resulting from fitting different numbers of knots: (*b*) 31; (*c*) 20; (*d*) 15; (*e*) 12; (*f*) 10; (*g*) 7; and (*h*) a parametric fit with six parameters.

The birth seasonality estimates described in the companion paper (Donnelly *et al.* 1997*b*), derived from the GB National Milk Records (NMR) dairy cattle database, were incorporated into the time varying birth rate, $B(t)$ used in the model. However, while assuming no seasonality in births results in a significantly worse fit to the case data ($\chi^2_{218} = 355$), model estimates and predictions are relatively unaffected ($I_T = 945\,000$, $C_F = 9562$).

(*j*) *Cross-validation of disease parameters*

An independently collected database on BSE cases was available from the Department of Agriculture, Northern Ireland. By 1 January 1997 there had been 1746 BSE cases in Northern Ireland (NI). The course of the NI epidemic to date has been described by Denny *et al.* (1992) and Denny & Heuston (1997). The herd of NI is much smaller than that of GB but has shown the same trend of declining size over the last 20 years. The observed incidence rate has also been much lower in NI than in GB, possibly due

Table 5. *Feed risk knot location options explored, showing best fit achieved with model C7*
(n is the number of knots fitted. Note that the $n = 20$ option was used for all other model fits in this paper.)

n	χ^2	knot locations used
31	302.6	73.5, 80.5, 81, 81.5, 82, 82.5, 83, 83.5, 84, 84.5, 85, 85.5, 86, 86.5, 87, 87.5, 87.75, 88, 88.25, 88.5, 88.75, 89, 89.25, 89.5, 89.75, 90, 90.25, 90.5, 91, 91.5, 92
20	307.7	73.5, 80.5, 81.5, 82.5, 83.5, 84.5, 85.5, 86.5, 87.5, 87.9, 88.3, 88.7, 89.1, 89.5, 89.9, 90.3, 90.7, 91.5, 92, 92.6
15	327.4	73.5, 80.5, 81.5, 82.5, 83.5, 84.5, 85.5, 86.5, 87.5, 88, 88.5, 89.5, 90.5, 91.5, 92.5
12	347.8	73.5, 80.5, 82.5, 84.5, 86.5, 87.5, 88, 88.5, 89.5, 90.5, 91.5, 92.5
10	372.1	78.5, 82.5, 84.5, 86.5, 87.5, 88.5, 89.5, 90.5, 91.5, 93
7	699.6	78.5, 82.5, 85.5, 87.5, 88.5, 89.5, 92.5

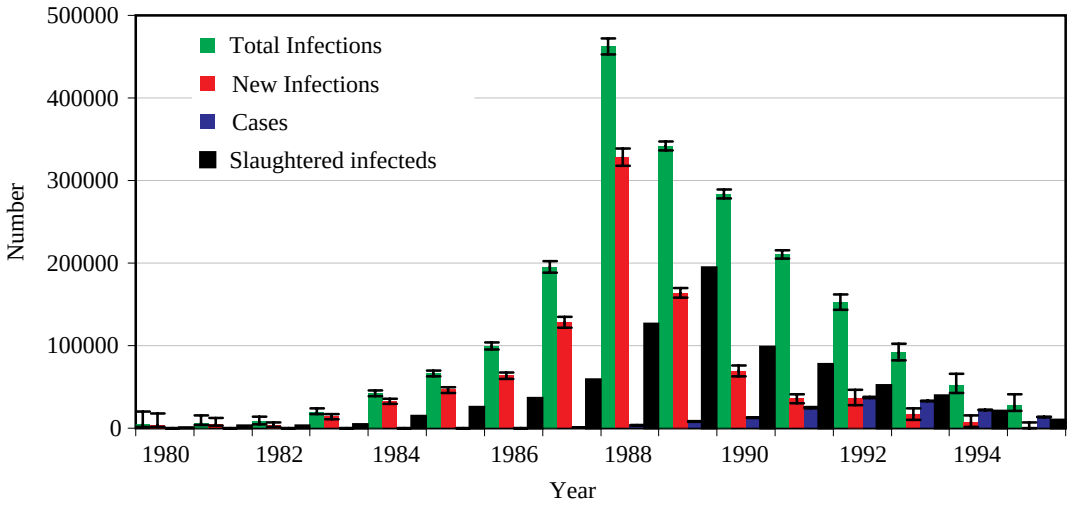


Figure 14. Estimated trends in the annual incidence of infection and disease in addition to the total number of cattle infected with the aetiological agent of BSE alive at the year end and the number of infected animals slaughtered by year for Great Britain assuming no maternal and no horizontal transmission.

to differences in rendering practices, or better implementation of controls, as well as relative isolation from the rest of the UK.

The NI case data were fitted using the disease (incubation-period and age-dependent-susceptibility distribution) parameters estimated from the GB case data (using model C7, in the absence of maternal or horizontal transmission), but estimating the feed-risk and under-reporting parameters for NI independently. It was found that the GB parameter estimates fitted the NI data well ($\chi^2_{188} = 147.9$, $p = 0.986$). Cases with onset dates before 1997 were analysed, since confirmation delays in NI are sufficiently short (approximately two weeks) for 1996 case reports to be virtually complete.

While not conclusive, this good fit to NI data provides additional confidence in the parameter estimates and model formulation used here.

5. MODEL RESULTS AND IMPLICATIONS

In this section we review the results from a subset of the models discussed above in more detail, concentrating on those aspects most relevant to animal- and public-health considerations, and on a discussion of the possible future course of the epidemic in GB and NI. In the case of GB, we consider four basic model scenarios: (I) no maternal or horizontal transmission ($\chi^2_{218} = 308$); (II) 10% maternal transmission over the last six months of the maternal incubation period ($\epsilon = 0.1$, $\omega_M = 0.5$, $\chi^2_{218} = 312$); (III) 10% maternal transmission over the last 12 months of the maternal incubation period ($\epsilon = 0.1$, $\omega_M = 1.0$, $\chi^2_{218} = 326$); and (IV) horizontal transmission over the last six months of the incubation period sufficient to give $\boldsymbol{R}_0^{(H)} \simeq 0.15$ ($\beta = 6.0$, $\omega_H = 0.5$, $\chi^2_{218} = 310$). For the NI case data we restrict our discussion to the model with no maternal or horizontal transmission. More detailed analyses of the NI BSE epidemic are presented in Ferguson *et al.* (1997*b*).

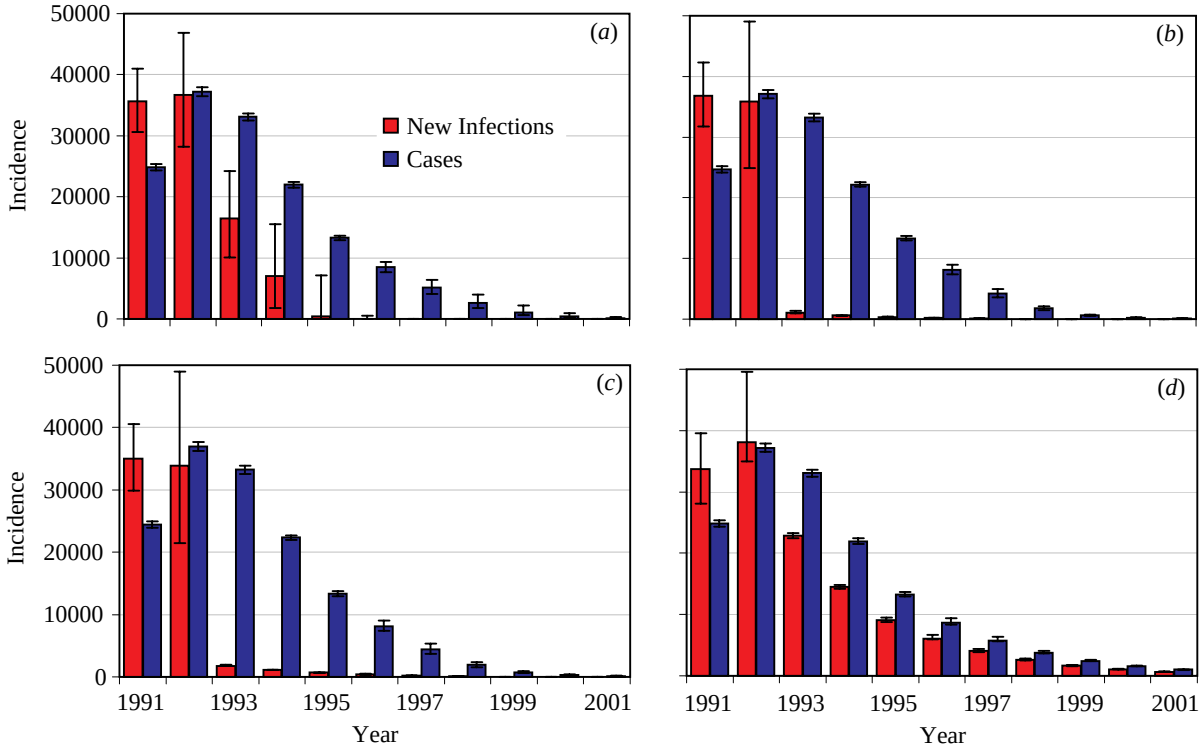


Figure 15. Predictions for the future trends in the annual incidence of cases of BSE and infections: (*a*) for the model with no maternal or horizontal transmission; (*b*) with 10% maternal transmission for the last six months of the maternal incubation period; (*c*) with 10% maternal transmission for the last 12 months of the maternal incubation period; and (*d*) with horizontal transmission for the last six months of the incubation period.

(*a*) *Temporal trends*

The estimated trends in past infection and case incidence (figure 14), assuming no maternal and no horizontal transmission, illustrate the time-lag and the difference in scale—caused, respectively, by the long incubation period and the survivorship of cattle—between the epidemic of infections and the resulting epidemic of cases of BSE. The 95% confidence intervals shown were calculated from the hypercube obtained from the bounds of the univariate 95% likelihood ratio confidence intervals, under the assumption that the risk of feed-borne infection drops to zero by July 1996. Note the numbers of infected animals that are estimated to have been slaughtered before the peak of the case epidemic and before the imposition of the specified bovine offal ban in November 1989—467 000 (460 000–482 000) in total. However, of these only 8000 are estimated to have been in the last year of incubation, when it is hypothesized that affected tissue is at its most infectious. In the period 1990–1995 we estimate 299 000 (285 000–317 000) infected animals were slaughtered for meat, with 43 500 being in the last year of the incubation period (all numbers quoted are from model C7 with no maternal or horizontal transmission).

Figure 15 gives a more detailed view of the recent estimated and predicted trends in infection and case incidence. Models II and III give fewest new infections from 1993 onwards—nearly all of which are maternally transmitted—due to the effect of small levels of maternal transmission being able to drive feed risk to low levels earlier than would otherwise occur. Horizontal transmission (figure 15*d*) also drives feed risk to very low levels, but, for model IV, is occurring at a rate that generates many more infections than do the models with maternal transmission (0.74 infections per case compared with 0.05 for model II and 0.1 for model III). The long tail to the infection epidemic for model IV is reflected in the resulting slower decline of the case epidemic in this case, relative to the other scenarios.

Table 6 gives case and infection predictions (with 95% prediction intervals) over the years 1997–2001 for the four models. In each case the model is constrained so that the risk of feed-borne infection drops to zero by July 1996. Note that by comparison with the equivalent table in Anderson *et al.* (1996), predicted case numbers are larger for the model with 10% maternal transmission over 12 months than for that with maternal transmission over six months. Overall infection numbers are also higher than in Anderson *et al.* (1996) in the case of model III. This is because the models now also fit the mean incubation period, as well as the variance, and this results in the fitted mean increasing with the rate of maternal transmission (see figure 6 and § 4 *d*). Due to the effects of survivorship, a longer incubation period requires larger estimates of infection numbers to reproduce the same number of resulting cases. If the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 6. *The predicted number of new infections and cases for the years 1996–2001 for the model with no maternal and no horizontal transmission (I), 10% maternal transmission for the last six months of the maternal incubation period and no horizontal transmission (II), 10% maternal transmission for the last 12 months of the maternal incubation period and no horizontal transmission (III) and horizontal transmission for the last six months of the incubation period and no maternal transmission (IV)*

		new infections		cases	
	year	predicted value	95% prediction interval	predicted value	95% prediction interval
(I)	1996	0	(0,533)	8452	(7673,9355)
	1997	0	(0,0)	5125	(4136,6381)
	1998	0	(0,0)	2628	(1819,4011)
	1999	0	(0,0)	1090	(659,2169)
	2000	0	(0,0)	380	(211,943)
	2001	0	(0,0)	118	(63,337)
(II)	1996	212	(188,241)	8075	(7356,8944)
	1997	100	(84,119)	4197	(3583,4944)
	1998	39	(33,47)	1741	(1450,2098)
	1999	14	(12,17)	641	(534,772)
	2000	5.3	(4.5,6.3)	235	(198,280)
	2001	2.0	(1.7,2.4)	89	(76,105)
(III)	1996	398	(343,466)	8126	(7363,9054)
	1997	188	(154,229)	4404	(3677,5289)
	1998	77	(63,94)	1929	(1568,2370)
	1999	31	(25,37)	764	(628,930)
	2000	13	(11,15)	309	(259,370)
	2001	5.5	(4.6,6.5)	131	(111,156)
(IV)	1996	6088	(5918,6721)	8654	(8383,9440)
	1997	4019	(3912,4415)	5792	(5624,6413)
	1998	2581	(2514,2770)	3765	(3683,4069)
	1999	1644	(1599,1756)	2409	(2356,2569)
	2000	1050	(1019,1133)	1555	(1517,1669)
	2001	661	(640, 718)	1012	(988,1094)

mean incubation period is held constant, the model predictions remain very close to those published in Anderson *et al.* (1996). Indeed, the fit to longer incubation periods may be affected by the inclusion of 1996 case reports in the model fitting, since, correcting (approximately) for confirmation delays, current estimates of case numbers with 1996 onset are of the order of 7500, compared with the the 8000 or more predicted in table 6. Furthermore, inclusion of 1996 case reports may well cause the goodness-of-fit of models with horizontal transmission at the levels assumed in scenario IV to deteriorate markedly. That said, it is also possible that the agricultural crisis precipitated in 1996 by the hypothesized link between BSE and new variant CJD has caused a slight increase in under-reporting. In either case, the predictions published in table 6 can be viewed as being conservative.

The proportions of infections and cases arising from maternal or horizontal transmission over the period 1989–1999 are presented in figure 16 for model scenarios II–IV. In each case, feed risk is estimated to have dropped to negligible levels by 1993–1994, but it is only in the case of horizontal transmission that a significant fraction of cases before 1999 are due to non-feed-borne transmission. Indeed, the pattern seen in figure 16*c* is quite dramatic, with horizontal transmission being able to explain a large fraction of the latter part of the epidemic (see § 4 *e*).

(*b*) *Age structure of the epidemic and human exposure*

Examination of the age structure of infection prevalence through time also gives insight into the transmission dynamics of BSE and the potential effect of maternal or horizontal transmission routes (figure 17). In the early stages of the epidemic the age structure reflects the exponential growth in infection numbers, with the majority of all affected animals being young and recently infected. However, following the introduction of the MBM ban in mid-1988, infection incidence dropped sharply, resulting in a steady (and near linear) growth of the mean age of an infected animal from that time on (figure 17*a*). In the latter stages of the epidemic, maternal transmission somewhat complicates this picture (figures 17*b*, *c*), producing a secondary much-smaller epidemic of maternally transmitted infec-

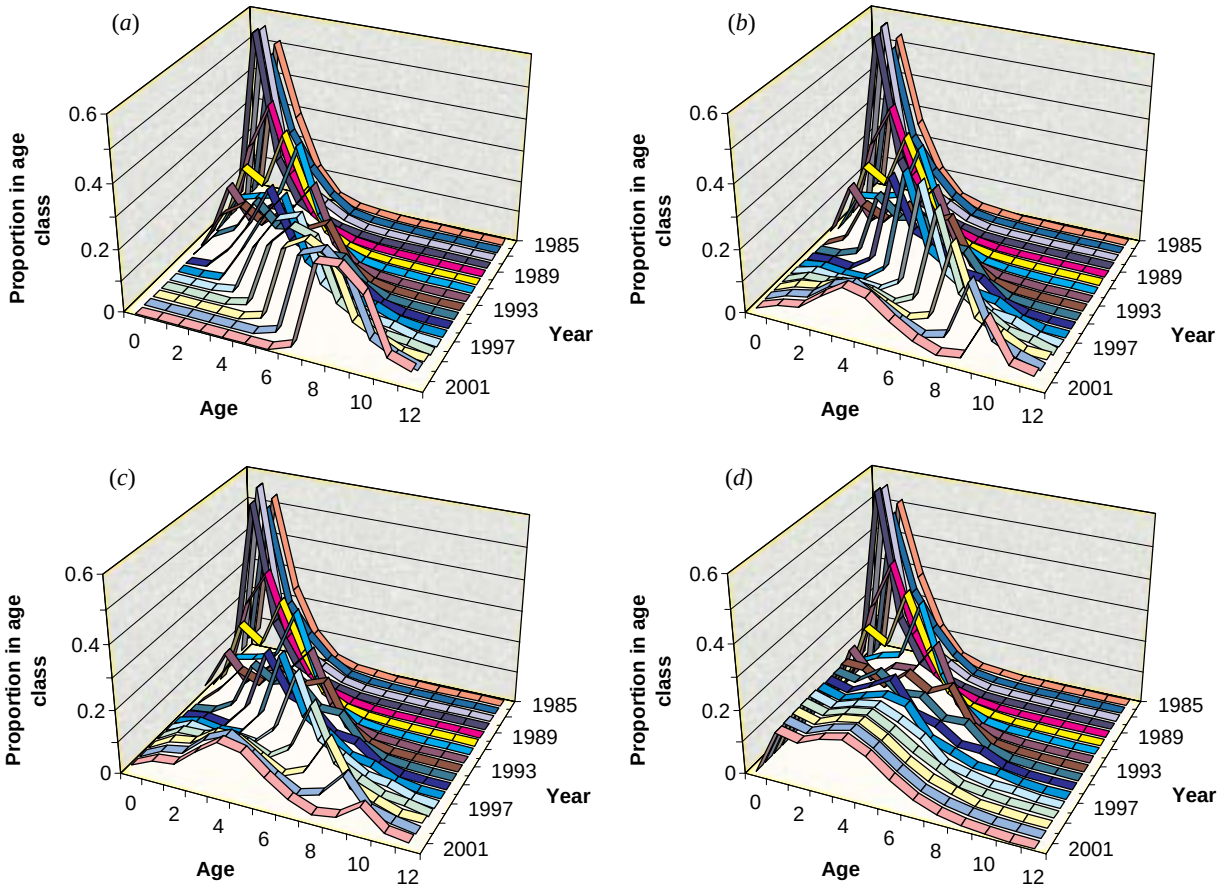


Figure 17. The predicted age distribution (proportion in each age class) of infected animals by year for the model with: (*a*) no maternal and no horizontal transmission; (*b*) 10% maternal transmission for the last six months of the maternal incubation period; (*c*) 10% maternal transmission for the last 12 months of the maternal incubation period; and (*d*) horizontal transmission for the last six months of the incubation period.

tions that becomes clearly visible only once feed risk has dropped to negligible levels. Horizontal transmission (figure 17*d*) produces a constant age structure in the latter stages of the epidemic. This may appear to be a non-intuitive result at first glance, but is caused by the extremely narrow age-dependent susceptibility distribution, $g(a)$, which means that—unlike many human diseases—the mean age at infection changes little as the force of infection declines. Hence with the incidence of new infections declining proportionally with the overall prevalence of infection in the population, a constant age structure is obtained. The last point to note from examination of figures 15 and 17 and table 6 is that, in the absence of horizontal transmission, negligible numbers of infected animals are predicted to be under 30 months of age from 1997 onwards—a result of obvious relevance to public concerns about the safety of beef in GB.

Insight into the past pattern of human exposure to BSE-infected material can be gained from figure 18, which shows the numbers of infected animals slaughtered through time stratified by their incubation stage. The relevance of this stratification is that it is thought that tissue from affected animals is at its most infectious around the time of disease onset. The estimated pattern reflects the numbers quoted above: while very large numbers of infected animals were slaughtered, most were in the early stages of incubation, especially before November 1989 when the SBO ban was introduced. As more information on disease pathogenesis in cattle becomes available, information of the type shown in figure 18 will enable refinement in the estimates of past human (relative) exposure to be made. Given longer time-series of the incidence of new variant CJD in humans, this may facilitate more reliable prediction of the overall scale of any future human epidemic.

(*c*) *Northern Ireland*

Finally, figure 19 shows the estimated pattern of the BSE epidemic in NI. Figure 19*a* gives the results from model C7 (with no horizontal or maternal transmission) using the biological parameters obtained by fitting to the GB case data ($\chi^2_{188} = 147.9$, $p = 0.986$), while figure 19*b* shows the results obtained by fitting the biological parameters to the NI data separately ($\chi^2_{188} = 94.9$, $p > 0.999$). The only significant difference between the two fits is that fitting the incubation-period distribution to the NI data gives a mean incubation period somewhat lower (4.65 years)

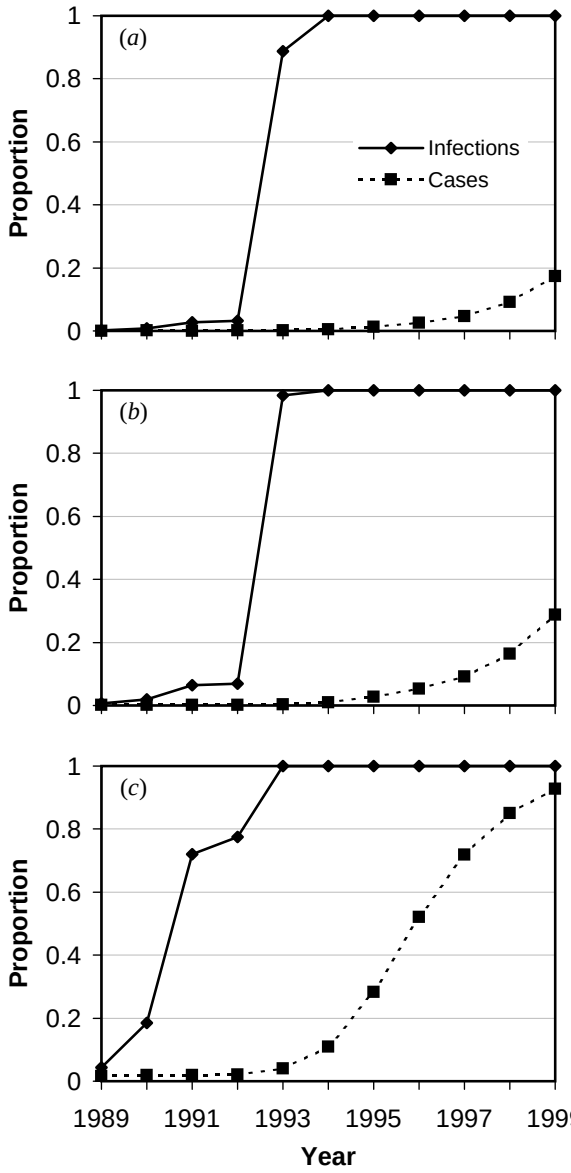


Figure 16. Proportions of infections and cases that are predicted to arise from maternal transmission at a rate of 10% for: (*a*) the last six months of the maternal incubation period; and (*b*) the last 12 months of the maternal incubation period. (*c*) Proportions of infections and cases arising from horizontal transmission in the last six months of the incubation period.

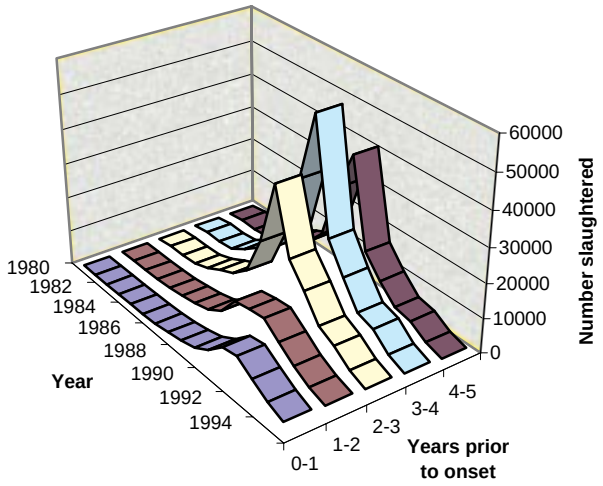


Figure 18. The estimated number (from the model with no maternal or horizontal transmission) of BSE infected animals that were slaughtered, by year of slaughter and the time to disease onset at slaughter.

6. CONCLUSIONS

The results of our earlier work (Anderson *et al.* 1996) have been subject to a very high degree of scrutiny due to their significance to policy formulation within the European Union and the relevance to any calculation of risk in terms of human consumption of contaminated meat and meat products. In particular, criticism (much unscientific) has been made that the published predictions of future case numbers were over-optimistic. In that paper, we estimated that, in the absence of maternal transmission, 12 100 cases would occur over 1997–2001, with 7000 over the same time interval under the scenario of 10% maternal transmission over the last six months of the maternal incubation period.

In this work we have shown, using a comprehensive set of sensitivity analyses, that both of these estimates remain fully justified, and are robust to changes in nearly all model parameters and distributions. Specifically, examining all model variants producing $\chi^2 \leqslant 400$ gives an upper bound on case numbers over 1997–2001 of approximately 11 000 in the absence of maternal transmission, with the great majority of models producing predictions in the range 9000–9500. Bearing in mind that the best-fit model has $\chi^2 = 308$, we believe this demonstrates remarkable robustness, especially considering the intrinsic difficulties involved in extrapolating a non-parameterically estimated feed-risk profile. Moreover, we have demonstrated that even low levels of maternal transmission allow even more robust predictions of future case numbers to be made—producing results in line with those published in Anderson *et al.* (1996). The increased robustness is due to the tendency of maternal transmission to explain an increasing proportion of cases as one moves closer to the present, thereby making the best-fit estimates of feed-risk decline more rapidly than is otherwise seen.

than that obtained from the GB data (5 years). This explains the lower estimated infection numbers in figure 19*b* compared with 19*a*. It is important to note that while the overall pattern of the epidemic in NI is comparable to that seen in the rest of the UK, the observed per capita incidence is actually an order of magnitude smaller. From 1997 to 2001, 103 (37–342) cases are predicted to occur using the GB fitted parameters, though it should be noted that this may be a slight underestimate due to a possible increase in under-reporting following March 1996.

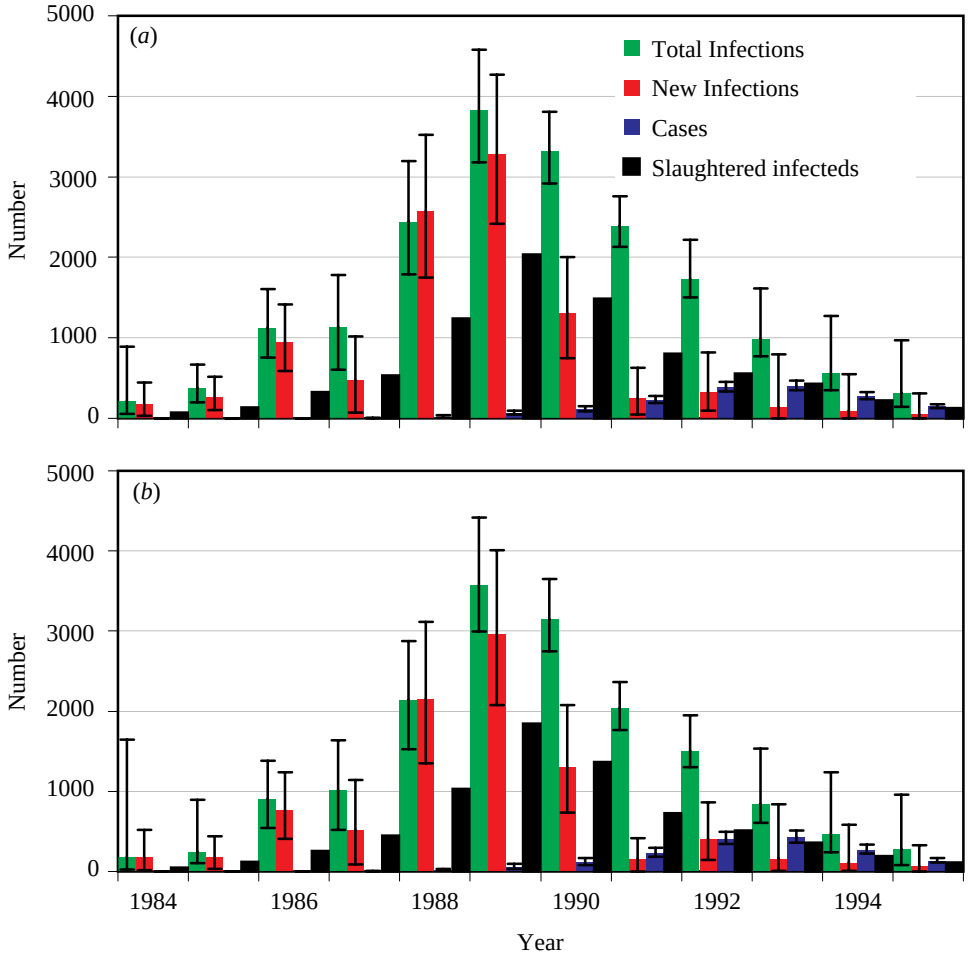


Figure 19. Estimated trends in the annual incidence of infection and disease in addition to the total number of cattle infected with the aetiological agent of BSE alive at the year end, and the number of infected animals slaughtered by year for NI assuming no maternal and no horizontal transmission for the models which: (*a*) use the best fit values of biological parameters obtained from the GB case data, but fit the feed-risk and under-reporting parameters; and (*b*) fit the feed-risk, under-reporting and biological parameters.

Estimates of the past number and pattern of BSE infections prove even less sensitive to variation in model parameters. All the models presented in this work produce estimates of total infections in the range 920 000–1 050 000.

The sensitivity analyses presented here give greater insight than previously into the key epidemiological determinants of the BSE epidemic. Key among these is the distinctive, and consistent (for the later cohorts) age-at-onset distribution seen in the case data. The form of this distribution, coupled with knowledge of the relatively low peak incidences (even at the herd level, less than 30%) seen, allow much tighter bounds to be (simultaneously) put on both the forms and means of the incubation-period and age-dependent-susceptibility distributions. Indeed, we explored an exhaustive range of forms of both distributions (much wider than the subset presented here), and all produced best-fit estimates of the mean of the incubation-period distribution in the range 4.5–5.5 years, agreeing convincingly with the limited experimental data available (Anderson *et al.* 1996; Donnelly *et al.* 1997*b*). This result was robust even when the population was split into two groups with a 20-fold difference in susceptibility/exposure—a first approximation to a model reflecting herd-based case aggregation or genetically variable susceptibility—excluding the possibility that the observed age-at-onset distribution reflects very high infection prevalence early in life in some small group of susceptible animals.

The detailed analysis of maternal transmission presented here also provides interesting insights into the epidemiology of BSE. Maternal transmission, when limited to the end of the maternal incubation period, produces a sub-epidemic of *infections* temporally correlated with the *case* epidemic. This has the inevitable effect of lowering estimates of feed-risk in the recent past, and therefore reducing predictions of future case numbers. It is also worth reiterating the epidemiologically obvious point: maternal transmission alone can never sustain an epidemic. Finally, while the result is not statistically significant, it is interesting to note that the overall best-fit model has 4–10% maternal transmission with a 3–6 month maternal infectious period—exactly in line with the best

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

estimates of maternal transmission obtained from ongoing analyses of the maternal cohort study data (Wilesmith *et al.* 1997; Donnelly *et al.* 1997*c*). The models presented here clearly reject the hypothesis of high rates (greater than 20%) of maternal transmission.

Much work remains to be done before a comprehensive understanding of the detailed epidemiology of the BSE epidemic is gained, however. While clearly outside the scope of the population level models discussed here, it will be important to develop models explaining the high degree of clustering of BSE cases in a small subset of GB herds. Of equal importance is research to explore the exact nature of the feed recycling and infection processes that drove the rapid growth of the epidemic. Indeed, as discussed previously, it is only through such research that insight will be gained into the origins of the epidemic. The analysis of under-reporting in the early years of the epidemic—and its possibly artefactual role in explaining the dramatic changes seen in the age-at-onset distribution in early cohorts—beg many questions concerning the exact (and possibly changing) nature of the infectious agent. The case data might therefore provide some insight into the role of passaging effects in determining the early epidemiological pattern. This has obvious relevance to discussion of the hypothesis that BSE originated from cross-species transmission of sheep scrapie. Alternatively, the early case data may provide evidence for changing temporal trends in the mean infective (or infectious) dose of the BSE aetiological agent. While such analysis will require a detailed model of the infection process, which incorporates the potential effect of cumulative exposure, it is interesting that the incubation-period distributions that best fit the case data are those that are based on a mechanistic model of disease pathogenesis (Medley & Short 1996). This suggests that understanding dose-dependent effects should be a key priority of future research.

The question of whether or not horizontal transmission of the aetiological agent of BSE, independent from exposure to contaminated feed, has and will play a role in the overall pattern of the epidemic, has been of much concern amongst the scientific and policy-making communities. We presented an analysis of a model that incorporates this transmission route and conclude that it appears unlikely to contribute significantly to past trends. We cannot exclude the possibility of its occurrence but on the basis of the analyses presented here and separate other study of its possible contribution to the overall basic reproductive number, R_0, of the agent throughout the course of the epidemic, we believe that, even if it does occur, its magnitude is insufficient to maintain BSE endemically in the GB cattle herd. In an ideal world, carefully designed and controlled experiments are needed to assess the likelihood of horizontal transmission but these may not be possible at this late stage of the epidemic. Furthermore, even if initiated in the near future, results would not be available for analysis for at least five to six years given the long incubation period of the disease and the absence of an ante-mortem test for the presence of the aetiological agent. Finally, it should be noted that the results on maternal transmission reveal a significantly enhanced risk of disease developing in calves born to dams who subsequently develop BSE (with a further enhancement in those born after the onset of BSE in the dam). As noted in the papers detailing the evidence for a maternally enhanced risk (Wilesmith *et al.* 1997; Donnelly *et al.* 1997*a*, *c*; Gore *et al.* 1997; Curnow *et al.* 1997), it is not possible at present to distinguish the relative contribution of maternal transmission of the aetiological agent from that of genetic predisposition to feed-borne infection and subsequent disease. Additional research is required to examine the genotypes (particularly the sequences of the *PrP* gene and flanking regions in diseased and non-diseased animals). Published work hints at a genetic association (Neibergs *et al.* 1994) as a contributory factor but more research is required. This is urgent given that the rapid decay in the epidemic in the British cattle herd will make it increasingly difficult to collect the required numbers of biological samples.

C.A.D., N.M.F. and R.M.A. thank the Wellcome Trust and MAFF for research grant support. M.E.J.W. thanks the Royal Society, MAFF and the BBSRC. We are grateful to John Wilesmith and other members of the BSE research teams of the Epidemiology Department, Central Veterinary Laboratory, and Owen Denny at the Veterinary Service, Department of Agriculture for Northern Ireland, for the provision of data and technical advice. We are also grateful to Robert May, Azra Ghani, Brian Ripley and Garrett Fitzmaurice for their assistance.

REFERENCES

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control.* Oxford University Press.

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.

Bacchetti, P., Segal, M. R. & Jewell, N. P. 1993 Back calculation of HIV-infection rates. *Statist. Sci.* **8**, 82–101.

Brookmeyer, R. & Gail, M. H. 1986 Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* **2**, 1320–1322.

Brookmeyer, R. & Gail, M. H. 1988 A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J. Am. Statist. Ass.* **8**, 301–308.

Brookmeyer, R. & Liao, J. 1990 Statistical modelling of the AIDS epidemic for forecasting health care needs. *Biometrics* **45**, 325–335.

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.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Curnow, R. N., Wijeratne, W. V. S. & Hau, C. M. 1994 The inheritance of susceptibility to BSE. In *Proc. European Commission Consultation on Transmissible Spongiform Encephalopathies* (ed. R. Bradley & B. Marchant), pp. 109–124. Brussels.

Curnow, R. N., Hodge, A. & Wilesmith, J. W. 1997 Analysis of the BSE maternal cohort study: the discordant case-control pairs. *Appl. Statist.* (In the press.)

Denny, G. O. & Hueston, W. D. 1997 Epidemiology of Bovine Spongiform Encephalopathy in Northern Ireland 1988 to 1995. *Vet. Rec.* **140**, 302–306.

Denny, G. O., Wilesmith, J. W., Clements, R. A. & Hueston, W. D. 1992 Bovine spongiform encephalopathy in Northern Ireland: epidemiological observations 1988–1990. *Vet. Rec.* **130**, 113–116.

Donnelly, C. A., Ferguson, N. M., Ghani, A. C., Wilesmith, J. W. & Anderson, R. M. 1997*a* Evidence for direct maternal transmission—the analysis of dam–calf BSE pairs. (Submitted.)

Donnelly, C. A., Ferguson, N. M., Ghani, A. C., Woolhouse, M. E. J., Watt, C. J. &Anderson, R. M. 1997*b* The epidemiology of BSE in cattle herds in Great Britain. I. Epidemiological processes, demography of cattle and approaches to control by culling. *Phil. Trans. R. Soc. Lond.* B **352**, 781–801. (Preceding paper.)

Donnelly, C. A., Ghani, A. C., Ferguson, N. M., Wilesmith, J. W. & Anderson, R. M. 1997*c* Analysis of the BSE maternal cohort study: evidence for direct maternal transmission. *Appl. Statist.* (In the press.)

Ferguson, N. M., Donnelly, C. A., Woolhouse, M. E. J. & Anderson, R. M. 1997*a* A genetic interpretation of heightened risk of BSE in offspring of affected dams. *Proc. R. Soc. Lond.* B **264**. (In the press.)

Ferguson, N. M., Donnelly, C. A., Ghani, A. C., Denny, G. O. & Anderson, R. M. 1997*b* Transmission dynamics and epidemiology of BSE in Northern Ireland. (In preparation.)

Gail, M. H. & Rosenberg, P. S. 1992 Perspective on using backcalculation to estimate HIV prevalence and project AIDS incidence. In *AIDS epidemiology: methodological issues* (ed. N. P. Jewell, K. Dietz & V. T. Farewell), pp. 1–38. Boston, MA: Birkhauser.

Gore, S. M., Gilks, W. R. & Wilesmith, J. W. 1997 Analysis of the BSE maternal cohort study: exploratory analysis. *Appl. Statist.* (In the press.)

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.

Haydon, D. T., Woolhouse, M. E. J. & Kitching, R. P. 1997 An analysis of foot-and-mouth-disease epidemics in the UK. *IMA J. Math. Appl. Med. Biol.* **14**, 1–9.

Jacobs, D. A. H. 1977 *State of the art in numerical analysis.* London: Academic.

Kimberlin, R. H. 1993 Bovine spongiform encephalopathies: an appraisal of the current epidemic in the United Kingdom. *Intervirology* **35**, 208–218.

Kirkpatrick, S. 1984 Optimization by simulated annealing—quantitative studies. *J. Statist. Phys.* **34**, 975–986.

Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. 1983 Optimization by simulated annealing. *Science* **220**, 671–680.

McCullagh, P. & Nelder, J. A. 1989 *Generalised linear models*, 2nd edn. New York: Chapman & Hall.

McKay, M. D., Beckman, R. J. & Conover, W. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.

Medley, G. F. H. & Short, N. R. M. 1996 A model for the incubation period distribution of transmissible spongiform encephalopathies and predictions of the BSE epidemic in the United Kingdom. Preprint.

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.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 *Numerical recipes in C.* Cambridge University Press.

Statutory Instrument 1988 *The bovine spongiform encephalopathy order 1988*, no. 1039. London: HMSO.

Vanderbilt, D. & Louie, S. G. 1984 A Monte Carlo simulated annealing approach to optimization over continuous variables. *J. Comp. Phys.* **56**, 259–271.

Weissmann, C. 1991 The prion's progress. *Nature* **349**, 569–571.

Wells, G. A. H., Scott, A. C., Johnson, C. T., Gunning, R. F., Hancock, R. D., Jeffrey, M., Dawson, M. & Bradley, R. 1987 A novel progressive spongiform encephalopathy in cattle. *Vet. Rec.* **121**, 419–420.

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

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. 1992 Bovine spongiform encephalopathy: case-control studies of calf feeding practices and meat and bonemeal inclusion in proprietary concentrates. *Res. Vet. Sci.* **52**, 325–331.

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 27 March 1997; accepted 2 May 1997

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES