INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

An Overview of the Statistical Analyses of the Bovine Spongiform Encephalopathy Epidemic

by
R. Sturm-Beiss

Submitted to the Department of Mathematics and Statistics on August 31st, 2000, in partial fulfillment of the requirements for the degree of Master of Science

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-70556-0

Canada

Abstract

In this thesis the statistical analyses that were used to study the by now well known bovine spongiform encephalopathy (BSE) epidemic are reviewed. Central to the analysis is a backcalculation survival model whose development is discussed in detail. Various techniques applied to examining the likelihood of a maternal infection route (in addition to the main feed infection route) are discussed. It is found that maternal transmission is likely to occur at low rates. Measures taken to eliminate meat and bone meal feed supplements, the main infection source, have essentially eliminated BSE. However, the magnitude of the latent effect of tainted meat on humans in the form of the linked new variant Creutzfeldt-Jacob disease is yet to be assessed.

Résumé

L'objectif de cette thèse est d'examiner les méthodes statistiques utilisées dans traitement et l'analyse des données provenant de l'épidémie de l'encéphalopathie spongiforme bovine (ESB). Ces analyses sont formées à partir d'un modèle de survie employant une technique de retrocalcule, dont le développement est scruté. Nous examinons aussi plusieurs estimés du taux de transmission mère-enfant de l'ESB. Les résultats démontrent qu'en général, ce taux est faible. Les precautions visant à éliminer les suppléments nutritifs dans l'alimentation pour bétail (la source principale d'infection), ont pratiquement éliminé le problème de l'ESB. Toutefois, la gravité de l'effet latent des viandes contaminées chez les humains, qui pourrait éventuellement se manifester par la maladie de Creutzfeldt-Jacob, reste à être évaluée.

Acknowledgement

I am deeply grateful to my advisor, Professor David Wolfson, for suggesting the thesis topic, for his constant encouragement, and for his continued help throughout. I would also like to thank all my professors in the Department of Mathematics and Statistics for the knowledge they imparted.

Chapter 1

Introduction

1.1 Bovine Spongiform Encephalopathy

The disease affecting cattle termed bovine spongiform encephalopathy (BSE) was first diagnosed in Britain in November 1986. As of July 2000 there have been 176,922 occurrences in Great Britain. Afflicted cattle show symptoms of increased nervousness, lack of coordination, and weakness leading to death approximately 6-12 months after the onset of symptoms.

The infectious agents are believed to be "prions", proteins that are chemically close to the healthy variety produced in nerve tissue. It is widely accepted that BSE prions introduced (orally or through injection) into a cow's body alter healthy proteins producing more prions, thus causing a chain reaction. The subsequent build-up of prions in the nerve tissue brings on disease symptoms after an average lag of 4-5 years (Anderson et al., 1996). Disease diagnosis is based on observation of symptoms and autopsy brain tissue. Prion accumulation in brain tissue gives the brain a sponge-like appearance, hence the term "spongiform". These protein based infectious agents differ from viruses and bacteria in that they are not "alive". That is they do not contain genetic material; high temperatures and radiation do not affect their potency. Other well-known diseases related to prions are scrapie in sheep, TSE in rodents, CJD (Creutzfeldt-Jakob Disease) and Kuru in humans. Stanley Prusiner's controversial work (Prusiner, 1995) led to the discovery of the infectious proteins which he coined "prions". In January 1998 Prusiner won the Nobel prize for his discovery.

The main source of BSE is meat and bone meal feed tainted with BSE infected cow tissue

and scrapie infected sheep tissue. The cow epidemic received much media attention in 1996 after two patients in England were diagnosed with a new variant of CJD and the possibility was raised that their disease was caused by the consumption of BSE tainted meat. Medical studies have confirmed that the diseases are causally linked (Bruce et al., 1997).

1.2 The BSE epidemic in Great Britain

The quasi-carnivorous practice of feeding cattle meat and bone meal (MBM) is an old one. In Europe and North America (before various countries imposed bans) MBM formed a substantial part of the diet of cows. Rendering plants convert animal remains into feed and oils, recycling carcasses whose disposal would otherwise be problematic. In Britain over 14% of MBM is recycled sheep. It is believed that scrapie infected nerve tissue broke the species barrier sparking the BSE epidemic. Large quantities of MBM were fed to British cattle in the early 80's. Around this time certain flammable fat separating chemicals were eliminated from the rendering process because of explosions at some plants. It is believed that these flammable substances may have deactivated the infectious proteins and that their removal resulted in prions entering the food chain. Although the disease was not recognized until 1986, infection may have occurred much earlier since BSE is known to have a long incubation period (4-5 years on the average (Anderson et al., 1996)), and isolated preliminary cases may have gone undiagnosed. In November of 1986 the clinical and pathological investigation of a diseased cow led to the identification of BSE. Epidemiological studies in 1988 and 1989 revealed that MBM from infected cattle and sheep was the most likely cause of the persistent and frequent outbreak of the disease in Great Britain's cattle. In June of 1988 the disease was made notifiable to the Central Veterinary Laboratory, and a database of confirmed cases was maintained by them. In July 1988 an MBM ban was imposed. However, the ban was not fully effective until several years later. On August 1st of 1996 a stricter more effective ban was initiated. However, at least one cow born after the August 1, 1966 lived to experience the disease.

Deaths from the human form of BSE, new variant CJD, appear to be increasing. So far this year 14 Britons have died and 5 others are known to be dying from the disease. Since it first appeared in humans in 1996, a total of 74 people in Britain, 2 in France and1 in Ireland

have contracted the disease.

1.3 Statistical Issues

The suspected and recently confirmed link between BSE and CJD has sparked intense study of the dynamics of the BSE epidemic and its implications to public health. Of great concern is the perpetuation of the cow disease, and the possibility of a latent hazard to humans due to past tainted meat consumption. Addressing these concerns scientists have focused on the study of BSE transmission dynamics and estimates of infection rates in past and future cow cohorts. Infection levels and information concerning the infectiousness of a cow at various incubation stages are needed to estimate the magnitude of risk to cows and to humans. A cow's infection hazard has been observed to depend on both its birth cohort (due to changing feed practices) and age. Feed supplement consumption and possibly protein absorption vary with age, creating an age dependency. Data used to estimate rates of infection are in terms of rates of disease onset. Infection rates are related to onset rates via an incubation period density and the probability of cow survival till onset. Hence, the incubation period distribution and the cow survival distribution have an important role in the estimation of infection levels. Disease perpetuation depends on the existence of alternate routes of infection. If routes other than feed exist, then the MBM ban will not eradicate the disease. In particular the maternal route, (dam infects fetus), has been investigated and the horizontal route, (cow to cow), has not yet been ruled out.

The main tool used to address these concerns was a survival model developed by Anderson et al., (1996) and enhanced by Ferguson et al., (1997b) and that takes into account cohort membership, age at infection and age at disease onset. The model contains an age and time dependent infection hazard. The age at infection and age at onset are related via a parameterized incubation period distribution. The feed and maternal infection routes are presented as competing risks. A likelihood for the onset data is formulated in terms of the survival model and the cow survival distribution. Maximum likelihood techniques are used to estimate model parameters. The model, through backcalculation is applied to estimate past and future infection, and numbers of future cases. It is also applied to the analysis of data arising from

a paired "maternal cohort" study designed to investigate the maternal transmission assumption (Donnelly et al., Maternal Cohort Study, 1997b), (Donnelly et al. Analysis of Dam/Calf pairs, 1997c). In the maternal cohort study animals were matched according to feed related covariates; a pair consisted of an "exposed" and a "control" animal. The exposed cow's dam developed BSE near the time of the cow's birth. Since the study animals were exposed to some MBM feed (the paired nature of the study partially controlled for this confounding factor), the question arises as to whether the observed heightened risk of the exposed group was due to inherited genetic susceptibility and not to maternal transmission. Various genetic models have been formulated but lack of data has made it impossible to confirm the existence of varying susceptibility classes. So far all humans who have contracted new variant CJD have possessed a particular genetic trait that predisposed them to the disease. It is estimated that at least 40 percent of the British population shares that trait which involves a variation of the prion protien.

1.4 General Outline

This thesis presents the work of a group of British scientists who analyzed BSE related data and presented their results in a series of papers.

Chapter 2 explains the main technique used in the development of Ferguson et al.'s (1997b) BSE survival model: the backcalculation technique. Discrete and continuous examples of backcalculation as well as variations in the technique used to model the BSE onset density are presented with increasing complexity, with illumination of the ideas being the goal.

Chapter 3 develops the backcalculation BSE onset survival model. This chapter explains and elaborates the methods in Ferguson et al.,(1997b).

Chapter 4 presents methods used by Donnelly et al. (1997b) to analyze observations from the maternal matched-pair cohort study. The aim of the study was to determine the likelihood of a BSE vertical (dam to calf) infection route. Several likelihood methods are developed. The main likelihood relies on the survival model discussed in Chapter 3.

Chapter 5 is based on a paper by Donnelly et al. (1997c) that continues the discussion of the existence of a vertical BSE infection route. Data from the main data base of BSE cases is

analyzed to confirm results of the maternal matched-pair cohort study. The outcome variable is not BSE status given dam's status (as in the maternal cohort study), rather it is BSE status of the dam of a diseased calf. A likelihood form containing parameters related to maternal transmission is optimized and resulting estimates confirm those of the maternal-cohort study.

The purpose of this thesis is to take the reader through the "statistics story" of the BSE epidemic.

Chapter 2

Backcalculation

The technique of backcalculaton was developed by Brookmeyer (1988) to obtain a lower bound on the number of future AIDS cases by estimating present levels of infection. Knowledge of the AIDS virus incubation distribution is used to "backcalculate" past infection rates based on present disease levels. A variation of this technique (Ferguson et al.1997b) is applied to the problem of estimating BSE infection levels. Like AIDS, BSE has a long incubation period, but, lacking a test for BSE infectivity and having recognized the disease fairly recently, the incubation period distribution is not well known. Early statistical analyses of the AIDS epidemic also encountered the latter problem. Unlike AIDS most infected cows do not survive till disease onset as they are, in most cases, slaughtered before. The backcalculation adaptation to BSE involves assuming a parametric form for the incubation period distribution whose parameters are estimated along with past infection rates and incorporates the cow survival distribution into the model, to explain disease onset data. In this chapter we explain the backcalculation technique applied to any disease with a long incubation period and present several examples that build in complexity.

2.1 The Basics of Backcalculation

A disease with a long incubation period has two time dependent random variables of interest: time of infection and time of disease onset. Let T = time of infection and S = incubation period (then, $T + S$ = time of disease onset). Assume that T and S are independent. That is,

assume that the length of the incubation period is independent of time of infection. This is a reasonable assumption if environmental conditions affecting the progress of the disease remain constant. Suppose the density of the incubation period, $f(s)$, is known and the density of the time of infection, $I(t)$, is unknown. Furthermore, assuming that $0 \leq t \leq t_*$, where t_* is the present time,

$$k(t, s) = I(t)f(s),\ 0 \leq t \leq t_*,\ 0 \leq s < \infty$$

is the joint density of time of infection and incubation period.

If U = time of onset, then, $U = T + S$. The joint density of T (time of infection) and U (time of disease onset) is then given by:

$$h(t, u) = I(t)f(u - t),\ 0 \leq t \leq t_*,\ t \leq u < \infty.$$

If the time of infection density, $I(t)$, is assumed to belong to a parametric family of distributions indexed by the parameter vector θ, then the (marginal) density of U is given by the following backcalculation formula or convolution integral:

$$g(u) = \int_0^{\min(u, t_*)} I(t;\theta)f(u - t)dt \tag{2.1}$$

or

$$g(u) = \sum_{t=0}^{\min(u, t_*)} I(t;\theta)f(u - t) \tag{2.2}$$

if T and S are discrete.

Now, often as is the case with both AIDS and BSE, given the quantity of data, it is not practical to include every observed onset date in the likelihood as it would be highly computationally intensive to fit. A simplified approach is to categorize onset times within contiguous time intervals and work instead with count data.

Let $0 < t_1 < \ldots < t_k = t_*$ be a partition of interval $[0, t_*]$. Define count variables X_i,

$i = 1, \ldots, k+1$.

X_i = the number of observed onset times in time interval $(t_{i-1}, t_i]$, and

X_{k+1} = the number of future cases resulting from an infection time in interval $[0, t_*]$.

Let N = the total number of future cases resulting from an infection in time interval $[0, t_*]$.

$$N = \sum_{i=1}^{k+1} X_i.$$

If N is unknown then one of the aims of backcalculation is to estimate N. (Backcalculation was originally applied to estimate the number of AIDs infections obtained within a set time interval. We will see in Chapter 3, section 3.6 that the parameter N, in the BSE likelihood has a different meaning, representing the number of cattle in a given cohort and is a known quantity.) The first challenge is to build a likelihood based on count observations, the unknown parameters in the onset density $g(u)$ and parameter N (the number of infections occuring in fixed interval $[0, t_*]$). Brookmeyer's backcalculation technique achieves this. We divide this "challenge" into two steps: first, in the section that follows, we discuss likelihoods of interval count observations based on parametric densities, and then in section 2.3 count observation likelihoods based on parametric densities arising from (2.1) and (2.2), the backcalculation densities and probability functions respectively.

2.2 Density estimation if the presence of complete and incomplete interval count data

Let U represent any random variable with density $g(u;\theta)$, $u_0 \leq u \leq u_*$, where θ is an unknown parameter. Our object is to estimate parameter θ from observed data. Let $u_0 < u_1 < \cdots < u_k = u_*$ be a partition of $[u_0, u_1]$. Suppose rather than values of the random variable U being observed our observations are counts of the form: $x_1, \ldots, x_k$ where x_i = the number of sample U values occurring in $[u_{i-1}, u_i]$, $i = 1, \ldots, k$.. To obtain a maximum likelihood estimate of θ one must build a likelihood for the count observations. The likelihood is based on $f(u;\theta)$ and

will therefore incorporate θ. Count random variables $X_1, \ldots, X_k$ have a multinomial distribution with parameters: $p_1, \ldots, p_k$, where p_i depends on θ :

$$p_i(\theta) = \int_{u_{i-1}}^{u_i} g(u;\theta)du, \quad i = 1, \ldots, k.$$

A likelihood for our data is of the form:

$$L(\theta) = \frac{N!}{x_1! \ldots x_k!} p_1(\theta)^{x_1} \cdots p_k(\theta)^{x_k}. \tag{2.3}$$

(2.4)

$$where,\ p_k(\theta) = 1 - \sum_{i=1}^{k-1} p_k(\theta)\ and\ N = x + \ldots + x_k.$$

Let,

$$l(\theta) = x_1 \log p_1(\theta) + \cdots + x_k \log p_k(\theta).$$

The function $l(\theta)$ differs from the log likelihood $\log L(\theta)$ by a constant. As is standard practice function $l(\theta)$ is maximized with respect to θ to obtain the maximum likelihood estimator, $\hat{\theta}$, leading to the estimated density of U :

$$\widehat{g} = g(u;\hat{\theta}).$$

If however, our data are "incomplete", that is if the observation x_i is unknown, then $N = x_1 + \ldots + x_k$, may be regarded as an unknown parameter and needs to be estimated together with θ. Introducing the new parameter N, define,

$$l(\theta, N) = \log N! - \log(N - \sum_{i=1}^{k-1} x_i)! + x_1 \log p_1(\theta) + \cdots + x_k \log p_k(\theta). \tag{2.5}$$

The function $l(\theta, N)$ differs from $\log L(\theta, N)$ by a known constant. If N and $N - \sum_{i=1}^{k-1} x_i$ are assumed to be large the log likelihood is treated as a continuous function of N and $d(\log N!)$ may be approximated by $\log N$ (Sanathanan, 1972). The likelihood equations $\frac{\partial l(\theta, N)}{\partial \theta_r} = 0$, and $\frac{\partial l(\theta, N)}{\partial N} = 0$ may then be solved simultaneously to yield the MLE's $\hat{N}$ and $\hat{\theta}$.

In the example below a discrete density is estimated given a complete set of count observations and then again given incomplete count data.

Example 1 *Density estimation is the presence of complete and incomplete count data. Let.*

$$g(u;\theta) = \begin{cases} \theta. & 0 \le u \le 1 \\ 1-\theta\ . & 1 \le u \le 2 \\ 0. & elsewhere \end{cases} \quad . \text{ where } \theta \text{ is unknown.}$$

Suppose we have two count observations x_1 and x_2. where x_1 = the number of sample observations of the random variable U in $[0.1)$. x_2 = the number of sample observations in $[1.\frac{3}{2}]$ and x_3 = the number of sample observations in $(\frac{3}{2}.2]$. Suppose one observes: $x_1 = 10$. $x_2 = 8$. and $x_3 = 5$ Observations x_1, x_2.and x_3 have a multinomial likelihood with parameters p_1, p_2. and N where $N = x_1 + x_2 + x_3 = 23$. Parameters p_1 and p_2 can be expressed in terms of θ through the density $f(u;\theta)$:

$$p_1(\theta) = \int_0^1 g(u;\theta)du = \theta$$
$$p_2(\theta) = \int_1^{\frac{3}{2}} g(u;\theta)du = \frac{1}{2}(1-\theta).$$

Thus θ is the only unknown parameter. The log likelihood is (up to a constant) equal to:

$$l(\theta) = 10\log\theta + (23-10)\log\frac{1}{2}(1-\theta)$$

Differentiating l with respect to θ and setting the derivative equal to zero we obtain the maximum likelihood estimates:

$$\frac{dl}{d\theta} = \frac{10}{\theta} - \frac{13}{1-\theta} = 0.$$
$$\hat{\theta} = \frac{10}{23}.$$

If x_3 is unknown. the log likelihood contains the additional unknown parameter N:

$$l(\theta, N) = \log N! - \log(n-18)! + \log\theta + 8\log\frac{1}{2}(1-\theta) + (N-18)\log\frac{1}{2}(1-\theta)$$
$$= \log N! - \log(N-18)! + 10\log\theta + (N-10)\log\frac{1}{2}(1-\theta).$$

Differentiating with respect to N and θ (using the approximation $d(\log N!) \cong \log N$) one obtains

$$\frac{\partial l}{\partial \theta} = \frac{10}{\theta} + (N - 10)(\frac{-1}{1-\theta}) \text{ and}$$
$$\frac{\partial l}{\partial N} = \log \frac{1}{2}(1-\theta) + \log N - \log(N - 18).$$

Setting the two equations equal to zero and solving simultaneously yields the maximum likelihood estimates:

$$\hat{\theta} = \frac{5}{13} \text{ and } \hat{N} = 26.$$

□

2.3 Backcalculation examples with a known incubation period density.

The technique of backcalculation has two basic ingredients: the backcalculation (convolution) formula (2.1) and the interval data methods of section 2.2. In the examples that follow the two ingredients are combined yielding estimates of disease onset densities and future disease levels starting with the simplest possible case: a disease contracted at one point in time.

Example 2 *Suppose a disease is contracted at one point in time, $t = 0$, by some of a group of exposed individuals, and that the disease has a known incubation period $s \geq 0$, with density $f(s)$, for $0 \leq s < \infty$. Then the infection density, $I(t)$, for T = time of infection, is the trivial one:*

$$I(t) = \begin{cases} 1. & t = 0 \\ 0. & \text{elsewhere.} \end{cases}$$

Using backcalculation formula (2.2) the onset time density, $g(u)$, of U = time of disease onset , is

$$g(u) = \sum_{t \leq u} I(t) f(u-t) = I(0) f(u-0) = f(u).$$

In this case the time of onset density is the same as the incubation period density, and is therefore known. We are interested in estimating N, the number of cases resulting from infection at time $t = 0$. If we have at least one count observation: $x_1 =$ the number of disease occurrences in $[0, T_1]$, then we can estimate N, the number of disease occurrences in $[0, \infty)$ by forming a binomial likelihood for observation x_1 involving the unknown parameter N. The likelihood is:

$$L(N) = \frac{N!}{x_1!(N - x_1)!} p^{x_1} (1 - p)^{N - x_1}.$$

$$\text{where } p = \int_0^{T_1} f(u)du \text{ is known.}$$

The log likelihood, up to a constant, is given by:

$$l(N) = \log N! - \log(N - x_1)! + x_1 \log p + (N - x_1) \log(1 - p).$$

The derivative with respect to N (obtained using the continuity approximation of $\log N$) is:

$$\frac{\partial l}{\partial N} \approx \log N - \log(N - x_1) + \log(1 - p).$$

Setting the above to zero, and solving for N we obtain the MLE: $\hat{N} = \frac{x_1}{p}$.

□

We continue with an example of a disease contracted at two points in time. The discrete infection density has one unknown parameter representing the weight at one of the two points in its domain. The onset density is a mixture of the infection density and the (known) incubation density. Backcalculation yields estimates of the infection density's unknown parameter and of N, the number of individuals infected at the two points.

Example 3 *Suppose a disease is contracted at two points in time by a group of exposed individuals, and the disease has a known incubation period density $f(s)$. The infection density $I(t)$*

is given by :

$$I(t) = \begin{cases} \theta, & t = 0 \\ 1 - \theta, & t = 1 \\ 0, & elsewhere \end{cases}.$$

The time of onset density $g(u)$, calculated using backcalculation formula (2.2), where U = time of disease onset, is

$$g(u) = \begin{cases} \theta f(u), & u \leq 1 \\ \theta f(u) + (1-\theta) f(u-1), & u > 1 \end{cases}$$

Suppose we have two observations x_1 and x_2, where

x_1 = *the number of disease occurrences in* $[0, 1) = 20$.

x_2 = *the number of disease occurrences in* $[1, 2) = 25$.

(x_3 = *the number of disease occurrences in* $[2, \infty)$ *is unknown*).

With the goal of estimating θ and N, where N = the number of individuals infected at $t = 0$ or $t = 1$ (which is the same as the number of disease occurrences in $[0, \infty)$ resulting from infection at $t = 0$ or $t = 1$) we form the multinomial likelihood:

$$L(\theta, N) = \frac{N!}{x_1! x_2! (N - x_\cdot)!} p_1(\theta)^{x_1} p_2(\theta)^{x_2} p_3(\theta)^{N - x_\cdot}$$

where $x_\cdot = x_1 + x_2$, and $p_3(\theta) = 1 - p_1(\theta) - p_2(\theta)$.

Expressions for $p_i(\theta)$, $i = 1, 2$ are obtained from the backcalculation formula for $g(u)$:

$$p_1(\theta) = \int_0^1 g(u)du = \theta \int_0^1 f(u)du = \theta a. \text{ where } a \text{ is known.}$$

$$p_2(\theta) = \int_1^2 g(u)du = \theta \int_1^2 f(u)du + (1-\theta) \int_1^2 f(u-1)du =$$

$$\theta \int_1^2 f(u)du + (1-\theta) \int_0^1 f(u)du$$

$$= \theta b + (1-\theta)a. \text{ where } b \text{ is also known.}$$

Let us assume that $a = \frac{1}{2}$ and $b = \frac{1}{4}$. The multinomial likelihood of θ and N is then given by:

$$L(\theta, N) = \frac{N!}{20!25!(N-45)!}(\frac{1}{2}\theta)^{20}(-\frac{1}{4}\theta + \frac{1}{2})^{25}(\frac{1}{2} - \frac{1}{4}\theta)^{N-45}.$$

Up to a known constant the loglikelihood is given by:

$$l(\theta, N) = \log N! - \log(N-45)! + 20\log(\frac{1}{2}\theta) + 25\log(-\frac{1}{4}\theta + \frac{1}{2}) + (N-45)\log(\frac{1}{2} - \frac{1}{4}\theta).$$

Setting the two partials of l with respect to θ and N equal to zero and solving simultaneously yields the estimates: $\hat{N} = 70$ *and* $\hat{\theta} = \frac{4}{7}$.

□

The next two examples assume that the disease was contracted over an interval of time. The first example below assumes that infection is equally likely throughout the interval. Thus the incubation density is uniform and has no unknown parameters. Backcalculation yields an estimate of N, the number of cases resulting from infection during the interval. The second example below assumes that the infection density is a step function with two steps whose heights are represented in terms of unknown parameter θ. Backcalculation yields estimates of θ and N.

Example 4 *Suppose that the density of the instant of contraction of a disease in uniform on the interval* $(0, 1)$. *Furthermore we will assume that the incubation density is exponential with*

mean 1. The infection density $I(t)$ and the incubation density $f(s)$ are given by:

$$I(t) = \begin{cases} 1. \ 0 \le t \le 1 \\ 0. \ elsewhere \end{cases} . \ and$$

$$f(s) = \begin{cases} e^{-s}. \ 0 < s < \infty \\ 0. \ elsewhere \end{cases} .$$

The onset density $g(u)$ is formed using backcalculation formula (2.2):

$$g(u) = \begin{cases} \int_0^u I(t)f(u-t)dt = \int_0^u e^{-(u-t)}dt = 1 - e^{-u}. \ u < 1 \\ \int_0^1 I(t)f(u-t)dt = \int_0^1 e^{-(u-t)}dt = (e-1)e^{-u}. \ u \ge 1 \end{cases}$$

Suppose one observes 20 cases between times 0 and 1, so that

$$x_1 = \text{the number of cases in } [0,1] = 20$$
$$(x_2 = \text{the number of cases in } (1,\infty) \text{ is unknown}).$$

The probability function of the binomial random variable X_1 has parameter p = the probability of disease onset occurring between time 0 and time 1:

$$p = \int_0^1 g(u)du = \int_0^1 (1 - e^{-u})du = e^{-1} \cong .37.$$

If N = the number of cases resulting from infection in $[0,1]$, then N has likelihood:

$$L(N) = \frac{N!}{20!(N-20)!}(.37)^{20}(1-.63)^{N-20}$$

The loglikelihood, up to a known constant, is given by:

$$l(N) = \log N! - \log(N-20)! - N\log(.63)$$

and, hence the maximum likelihood estimate is given by

$$N = 54.$$

□

If one has more than one count observation, then one is able to discern more about the shape of the infection density. It is then reasonable to assume that the infection density is a step function where the height of each step is an unknown parameter.

Example 5 *Suppose a disease is contracted over a known interval of time and the infection density $I(t)$, is assumed to be a step function with two steps. The incubation density $f(s)$, is assumed to be exponential with mean equal to 1.*

$$I(t) = \begin{cases} \theta, & 0 \le t < 1 \\ 1-\theta, & 1 \le t \le 2 \end{cases}$$

$$f(s) = e^{-s}, \quad s > 0.$$

Onset density $g(u)$ is calculated with backcalculation formula (2.2):

If $u \le 2$,

$$g(u) = \int_0^u I(t)e^{-(u-t)}dt = \begin{cases} \int_0^u \theta e^{-(u-t)}dt = \theta(1-e^{-u}), & 0 \le u \le 1 \\ \theta(1-e^{-1}) + \int_1^u (1-\theta)e^{-(u-t)}dt = (1-\theta e^{-1}) - (e-e\theta)e^{-u}, & 1 \le u < 2 \end{cases}$$

if $u > 2$,

$$g(u) = \int_0^2 I(t)e^{-(u-t)}dt = \int_0^1 \theta e^{-(u-t)}dt + \int_1^2 (1-\theta)e^{-(u-t)}dt = \{\theta[e^1 - 1] + (1-\theta)[e - e^2]\}e^{-u}.$$

Suppose we have two observations

$x_1 =$ *the number of cases diagnosed in interval* $[0, 1) = 10$

$x_2 =$ *the number of cases diagnosed in interval* $[1, 2) = 20$

$(x_3 =$ *the number of cases diagnosed in interval* $[2, \infty)$ *is unknown).*

The multinomial likelihood $L(\theta, N)$ has parameters $p_i(\theta)$, $i = 1, 2, 3$ representing the probabilities that a case is diagnosed in one of the three intervals.

$$p_1(\theta) = \int_0^1 \theta(1 - e^{-u})du \cong (.37)\theta$$
$$p_2(\theta) = \int_1^2 \{(1 - \theta e^{-1}) + (e - e\theta)(e^{-2} - e^{-1})\}du \cong .37 + (.26)\theta$$
$$p_3(\theta) \cong .63 - (.63)\theta.$$

The loglikelihood up to a known constant is given by

$$l(N, \theta) = \log N! - \log(N - 30)! + 10\log(\theta) + 20\log(.37 + .26\theta) + (N - 30)\log(.63 - .63\theta).$$

We form the two partials of $l(N, \theta)$.

$$\frac{\partial l}{\partial \theta} = \frac{10}{\theta} + \frac{5.2}{.37 + .26\theta} + \frac{N - 30}{\theta - 1}$$
$$\frac{\partial l}{\partial N} = \log N - \log(N - 30) + \log(1 - \theta) + \log(.63).$$

Leading to the maximum likelihood estimates

$$\hat{\theta} = .215 \text{ and } \hat{N} = 59.$$

□

2.4 A look at backcalculation when the incubation density is unknown.

The backcalculation technique developed by Brookmeyer (Brookmeyer, 1988) assumes knowledge of the incubation period distribution. This knowledge together with disease onset observations yield an estimate of the infection onset densities. In the case of Bovine Spongiform Encephalopathy the incubation density is not known. It is difficult to indiependently estimate since the incubation period seems to be related to the dose of the aetiological agent and the

means of infection (oral or injection to the brain). If one tries to parameterize the incubation density and estimate its parameters along with those of the incubation density then the model becomes overparameterized yielding more than one set of values that maximize the likelihood. In Ferguson et. al.'s (1997b) article, backcalculation is applied with a parameterized form for the incubation density yielding more than one set of parameter estimates or more than one BSE disease model. In this section we examine backcalculation without complete knowledge of the incubation density through a series of examples.

A disease contracted at one point in time has a known infection density (the trivial one). In this case lack of knowledge of the incubation density leads to a model where the onset density, $g(u)$, is equal to the incubation density $f(u;\theta)$ (as in example 2 of this chapter). Thus, $f(u;\theta)$ (and $g(u)$) are estimated using the method of density estimation in the presence of incomplete interval count data as seen in section 2.2.

Example 6 *Consider a disease contracted at two points in time and a parameterized discrete incubation probability function which can take on two values. In the example that follows we see that backcalculation yields two sets of solutions for the model parameters. Suppose a disease was contracted at times, $t = 0$ and $t = 1$, by some of a group of exposed individuals.*

Let the following discrete probability function $f(s;\theta)$ be the assumed form for the incubation period probability function:

$$f(s;\theta) = \begin{cases} \theta, & s = 0 \\ 1-\theta, & s = 1 \\ 0, & elsewhere. \end{cases}$$

Let $I(t;\alpha)$ be the assumed infection probability function:

$$I(i;\alpha) = \begin{cases} \alpha, & t = 0 \\ 1-\alpha, & t = 1 \\ 0, & elsewhere \end{cases}$$

Then, U = time of onset can equal 0, 1 or 2, and the backcalculation formula for $g(u)$ yields:

$$g(0) = \alpha\theta$$
$$g(1) = \alpha(1-\theta) + (1-\alpha)\theta = \alpha - \theta - 2\alpha\theta$$
$$g(2) = (1-\alpha)(1-\theta) = 1 - \alpha - \theta - \alpha\theta.$$

After time $t = 2$ we observe the values of x_0, x_1, x_2.

$x_0 =$ *the number with onset at* $t = 0$

$x_1 =$ *the number with onset at* $t = 1$

$x_2 =$ *the number with onset at* $t = 2$.

(note: $N = x_0 + x_1 + x_2$ = the total number of infected individuals. In this example our count observations are "complete".)

Observations x_0, x_1, and, x_2 have a multinomial likelihood with parameters p_0, p_1, and, $p_2 = 1 - p_0 - p_1$. The parameters are dependent upon α and θ:

$$p_0 = g(0) = \alpha\theta$$
$$p_1 = g(1) = \alpha - \theta - 2\alpha\theta$$
$$p_2 = g(2) = 1 - \alpha - \theta + \alpha\theta.$$

The loglikelihood has the form:

$$L(\alpha, \theta) = \frac{N!}{x_0! x_1! x_2!} p_0(\theta)^{x_0} p_1(\theta)^{x_1} p_2(\theta)^{x_2}$$

and, the system of equations:

$$\frac{\partial L}{\partial \theta} = \left(\frac{\partial L}{\partial p_0}, \frac{\partial L}{\partial p_1}, \frac{\partial L}{\partial p_2}\right)\left(\frac{\partial p_0}{\partial \theta}, \frac{\partial p_1}{\partial \theta}, \frac{\partial p_2}{\partial \theta}\right)^T = 0$$
$$\frac{\partial L}{\partial \alpha} = \left(\frac{\partial L}{\partial p_0}, \frac{\partial L}{\partial p_1}, \frac{\partial L}{\partial p_2}\right)\left(\frac{\partial p_0}{\partial \alpha}, \frac{\partial p_1}{\partial \alpha}, \frac{\partial p_2}{\partial \alpha}\right)^T = 0.$$

is reduced to the system:

$$\frac{\partial L}{\partial p_0} = \frac{\partial L}{\partial p_1} = \frac{\partial L}{\partial p_2} = 0.$$

which yield multinomial parameter MLE's:

$$\hat{p}_0 = \frac{x_0}{N}, \quad \hat{p}_1 = \frac{x_1}{N}, \quad \hat{p}_2 = \frac{x_2}{N}.$$

Estimators of α and θ are found by solving:

$$\begin{aligned} \hat{p}_0 &= \alpha\theta \\ \hat{p}_1 &= \alpha + \theta - 2\alpha\theta \\ \hat{p}_2 &= 1 - \alpha - \theta + \alpha\theta. \end{aligned} \tag{2.6}$$

which lead to the quadratic:

$$\theta^2 - (2\hat{p}_1 + \hat{p}_2)\theta + \hat{p}_1 = 0.$$

having solutions

$$\theta = \frac{(2\hat{p}_1 + \hat{p}_2) \pm \sqrt{(2\hat{p}_1 + p_2)^2 - 4\hat{p}_1}}{2}.$$

Notice that the system of equations in lines 2.6 are symmetric in α and θ, which implies that if (θ_0, α_0) is a pair of solutions of the system, then $(\theta_1, \alpha_1) = (\alpha_0, \theta_0)$ is also a solution. In order to obtain unique maximum likelihood estimators additional constraints must be introduced. This is desirable if one wants to estimate $I(t)$ or $f(s)$.

□

The next example generalizes the one above and introduces a method for generating sets of solutions that maximize the likelihood and estimate the parameters of discrete infection and incubation densities.

Example 7 *Suppose a disease could be contracted at $n+1$ points in time: $t = 0, 1, \ldots, n$, and*

that the incubation period s can be any of the $m+1$ values $s = 0, 1, \ldots, m$. That is,

$$f(s;\theta) = \begin{cases} \theta_0, & s = 0 \\ \theta_1, & s = 1 \\ \vdots & \\ \theta_m, & s = m \\ 0, & \text{elsewhere.} \end{cases}$$

and,

$$I(t;\alpha) = \begin{cases} \alpha_0, & t = 0 \\ \alpha_1, & t = 1 \\ \vdots & \\ \alpha_n, & t = n \\ 0, & \text{elsewhere.} \end{cases}$$

After time $n+m$ we observe a (complete set of) counts $x_0, x_1, \ldots, x_{n+m}$ where x_i = the number of cases with onset at $t = i$. Let $N = \sum_{i=0}^{n+m} x_i$. The multinomial likelihood of the observed counts is of the form:

$$L(\theta, \alpha) = \frac{N!}{x_1! \ldots x_{n+m}!} p_1^{x_1} \cdots p_{n+m}(\theta)^{n+m}.$$

The MLE's are $p_i = \frac{x_i}{N}$, $i = 0, \ldots, n+m$. The relationship between p_r (p_r = the probability of disease onset at time $t = r$) and parameters α and θ is given by the backcalculation formula:

$$p_r = g(r) = \sum_{i+j=r} \alpha_i \theta_j, \quad r = 0, \ldots, m+n. \tag{2.7}$$

Maximum likelihood estimators for α and θ are found by solving the $m+n$ system of equations in (2.7). Let $\alpha(x)$ and $\theta(x)$ represent polynomials:

$$\alpha(x) = \alpha_0 + \ldots + \alpha_n x^n, \text{ and}$$

$$\theta(x) = \theta_0 + \ldots + \theta_m x^m.$$

Finding a real solution for the system of equations in (2.7) is equivalent to finding a real factorization of the polynomial:

$$p(x) = p_0 + \ldots + p_{n+m}x^{n+m}$$

having positive coefficients. The equations in (2.7) are recovered by equating the coefficients of the polynomials on both sides of the equality: $p(x) = \alpha(x)\theta(x)$.

Since polynomials over the real numbers factor into irreducible factors of degrees 2 and 1, one would expect many such factorizations, $\alpha(x)\theta(x)$ *to exist by rearranging or permuting the irreducible factors. A unique solution would require the addition of many added assumptions or of constraints on the functions* $I(t)$ *and* $f(s)$.

□

In the next example the technique introduced above is applied to generate all the sets of maximum likelihood estimates for a disease that was observed to have four onset times and is assumed to have two distinct incubation periods.

Example 8 *The following example uses the notation introduced in the previous example. Suppose that a disease is observed to have onset at 4 distinct points in time with the following estimated probabilities:*

$$p_0 = \frac{6}{24},\ p_1 = \frac{11}{24},\ p_2 = \frac{6}{24},\ p_3 = \frac{1}{24}.$$

Then, the polynomial, $p(x)$, *introduced in the previous example, is given by:*

$$p(x) = \frac{1}{24}\{x^3 + 6x^2 + 11x + 6\} = \frac{1}{24}(x+1)(x+2)(x+3) = \alpha(x)\theta(x).$$

If it is known that infection occurred at two distinct points in time, and the incubation period can be one of three distinct periods, then the number of factorizations of $p(x)$ *into polynomials:* $\alpha(x)$, *of degree two, and* $\theta(x)$ *of degree one, will correspond to the number of sets of parameter estimates :* $\{\alpha_0, \alpha_1\}$, *and* $\{\theta_0, \theta_1, \theta_2\}$. *There are* $\begin{pmatrix} 3 \\ 1 \end{pmatrix} = 3$, *such factorizations yielding three*

sets of equally likely estimates:

$$p(x) = \{\frac{1}{6}(x^2+3x+2)\}\{\frac{1}{4}(x+3)\} = \alpha(x)\theta(x).$$
$$\alpha_2 = \frac{1}{6}, \alpha_1 = \frac{1}{2}, \alpha_0 = \frac{1}{3} \quad and \quad \theta_1 = \frac{1}{4}, \ \theta_0 = \frac{3}{4}.$$

$$p(x) = \{\frac{1}{8}(x^2+4x+3)\}\{\frac{1}{3}(x+2)\} = \alpha(x)\theta(x).$$
$$\alpha_2 = \frac{1}{8}, \alpha_1 = \frac{1}{2}, \alpha_0 = \frac{3}{8} \quad and \quad \theta_1 = \frac{1}{3}, \ \theta_0 = \frac{2}{3}.$$

$$p(x) = \{\frac{1}{12}(x^2+5x+6)\}\{\frac{1}{2}(x+1)\} = \alpha(x)\theta(x).$$
$$\alpha_2 = \frac{1}{24}, \alpha_1 = \frac{5}{24}, \alpha_0 = \frac{1}{2} \quad and \quad \theta_1 = \frac{1}{2}, \ \theta_0 = \frac{1}{2}.$$

□

Finally we consider incubation and infection random variables having continuous densities.

Example 9 *The previous examples were all discrete. The following is a simple example of a backcalculation model based on an infection time and incubation period which are both assumed to be continuous exponential random variables.*

Let.

$$I(t) = \begin{cases} \theta e^{-\theta t}, \ t > 0 \\ 0, \ elsewhere \end{cases}, \ and$$
$$f(s) = \begin{cases} \alpha e^{-\alpha s}, \ s > 0 \\ 0, \ elsewhere \end{cases}.$$

Then.

$$\begin{aligned} g(u) &= \int_0^u I(t;\theta) f(u-t;\alpha) dt \\ &= \int_0^u \theta e^{-\theta t} \alpha e^{-\alpha(u-t)} dt \\ &= \frac{\theta\alpha}{(\alpha-\theta)}[e^{-\theta u} - e^{-\alpha u}]. \end{aligned}$$

Notice that $g(u)$ is symmetric in θ and α. The symmetry makes it impossible to have unique MLEs, since if (θ, α) maximizes the likelihood: $L(\theta, \alpha | u_1, \ldots, u_n)$, then (α, θ) will be a maximum as well.

□

Although parameters may not be uniquely estimated when the incubation density is unknown, one can choose a set of likely parameters based on prior knowledge, or one can examine the resulting range of densities and the range of parameter estimates that are equally likely.

2.5 Backcalculation in the presence of low survivability

In the examples of Chapter 2, section 2.4 we assume that after a final time (equal to the maximum time of infection plus the maximum incubation period) all exposed individuals will have experienced onset.

In the case of BSE, due to the low survivability of cattle (i.e. slaughtering patterns), most infected animals do not live to experience disease onset. Thus, the observations of disease onset counts are lower than they would be if cattle were allowed to survive.

Anderson et al., (1996), and Ferguson et al., (1997b), have modified the backcalculation formula to incorporate an (independently estimated) cow survival distribution which explains the discrepancy between the number of infections and cases.

In this section we shall assume that we observe a birth cohort from time of birth until death due to disease or otherwise. Observations are disease counts in subintervals of the cohort's life span. Furthermore we assume that members of the cohort have low survivability (due to slaughtering practices) and their survival distribution is known. We will see how the survival distribution is included in the likelihood.

Let $[0, t_*]$ be the maximum life span of cohort members. Let $0 = t_1 < \ldots < t_k = t_*$ be a partition of $[0, t_*]$ and $x_1, \ldots, x_k$ be disease count observations. x_i = the number of disease onsets observed in $[t_{i-1}, t_i)$, $i = 1, \ldots, k$. The onset density of U = time of onset, $g(u)$, $0 < u < \infty$, in

the absence of mortality (i.e. slaughtering practices), is given by:

$$g(u) = \int_0^{\min(u,t_*)} I(t)f(u-t)dt, \text{ and}$$

$$p_i = \int_{t_{i-1}}^{t_i} g(u)du = \Pr\{\text{onset in } [t_{i-1}, t_i), \text{ in the absence of mortality}\}.$$

In the presence of mortality or low survivability,

$$p_i \cong \Pr\{\text{survivorship until } t_i\} \cdot \Pr\{\text{onset in } [t_{i-1}, t_i) \mid \text{survivorship until } t_i\}$$
$$= S(t_i) \cdot \int_{t_{i-1}}^{t_i} g(u)du,$$

where the survival distribution is $S(a)$ = the probability that a cow survives until age a (in non-epidemic times). More precisely,

$$\Pr\{\text{onset in } [u, u+du)\}$$
$$= \Pr\{\text{survival until time } u\} \cdot \Pr\{\text{onset in } [u, u+du) \mid \text{survival until } u\}$$

Hence,

$$p_i = \int_{t_{i-1}}^{t_i} S(u)g(u)du. \tag{2.8}$$

Note: If $p = p_1 + \ldots + p_k$, then p = the probability that a member of the birth cohort is infected and lives until disease onset. If $p_{k+1} = 1 - p$, then p_{k+1} = the probability that a cohort member does not experience disease onset.

Define multinomial variables $X_1, \ldots, X_{k+1}$:

X_i = the number of cohort members having experienced disease onset in $[t_{i-1}, t_i)$, $i = 1, \ldots, k$, and X_{k+1} = the number of cohort members not having experienced disease onset.

The multinomial variables $X_1, \ldots, X_{k+1}$, have a distribution depending on the parameters N, $p_1, \ldots, p_k$, where N = the size of the cohort and p_i, $i = 1, \ldots, k$ are defined by equation (2.8) as

functions of the unknown parameters of $g(u)$.

In Chapter 3, section 3.6 we see that the likelihood of BSE observations incorporates a survival distribution to explain the discrepancy between low onset levels and high infection rates.

In the next chapter the techniques seen in this chapter are applied towards the construction of the BSE survival model. Background information, examples and details are added to Ferguson et al.'s (1997b) presentation, beginning with an overview of the development of the BSE model.

Chapter 3

The BSE Model

Now that we reviewed backcalculation and have (in Chapter 2) seen several hypothetical examples, we are ready for the development of the BSE backcalculation survival model. The model originally appeared in Anderson et al. (1996) and was later enhanced in Ferguson et al. (1997b). This presentation is based on Ferguson et al.'s enhanced model.

3.1 An Overview of the Development of the BSE Model

The BSE epidemic started in Britain in the late 1980's, peaked in the early 90's and has been rapidly decreasing since then. The pattern of disease is a reflection of meat and bone meal (MBM) feed practices. Large quantities of recycled meat and bone meal were fed to cows in the early 80's. In 1988 the MBM feed ban, imposed to curtail BSE , caused the number of new infections to rapidly decrease. Thus *the infection hazard function depends on a time variable.* The epidemic has not ended and today (January 2000) France continues to ban the import of British beef.

An additional dependency of infection risk on age of cow has been observed. This may be linked to age dependent feeding patterns and perhaps to age dependent absorbancy rates of the infectious prion agent (Anderson et al., 1996, Ferguson et al., 1997b).

Although tainted feed is believed to be the main infection route, it has been demonstrated that the existence of a maternal route (in the womb or during the birthing process) is likely (Donnelly et al.,1997b), and the possibility of a horizontal infection route (cow to cow) has

not been yet been eliminated. *Multiple infection routes are included in Ferguson et al.'s BSE model as "competing risks"*: the feed and horizontal additive components are continuous risks and the maternal component is a discrete risk, assumed be positive only at age zero. *The magnitude of maternal risk is estimated iteratively.* The first generation of exposed cattle were feed infected. Each successive generation had a maternal risk proportional to the number of maternally infectious dams in the previous generation. Thus, maternal risk is estimated iteratively based on the feed risks of the previous generations.

Infection ages of cohort members are unknown. Observations available to estimate the age at infection density parameters are counts x_i expressed more precisely in the form (t_0, x_i), where t_0 = the cow's birth year , and x_i = the number of cohort t_0 BSE cases diagnosed in the i^{th} year of the cohort's life span. The difference between the age at onset and age at infection is the incubation period. The backcalculation formula (2.1) (which requires some knowledge of the incubation period density) expresses the age at onset density in terms of the age at infection density and the incubation period density.

Since most cows are slaughtered (for meat) before age 3, it must be assumed that most infected cows do not survive until disease onset. Therefore, the available onset counts are small in comparison to the number of infections. *Knowledge of the survival distribution of British cows (in the absence of an epidemic), together with backcalculation formula (2.1) enable the formation of a likelihood for the available disease onset counts.* Maximization of the likelihood results in the estimation of the joint age at infection/age at onset density parameters.

We begin the presentation of Ferguson et al.'s model with the definition of competing risks.

3.2 Multiple routes of infection presented as competing risks

The backcalculation formula in the BSE model (N.M. Ferguson, 1997b), considers the time of infection to be a failure time variable, where failure in this case means infection. If failure can be attributed to more than one cause then the various causes are termed competing risks. The multiple causes or multiple infection routes are represented in the failure time density through

the hazard function. Let $\lambda(t;\theta)$ denote the hazard function at time t. That is

$$\lambda(t;\theta)dt \cong \Pr\{\text{ failure occurs in } [t, t+dt) | \text{ failure has not yet occurred}\}.$$

Then, every failure time density, $f(t;\theta)$, and corresponding survival distribution $S(t;\theta)$ have hazard form representations

$$f(t;\theta) = \lambda(t;\theta)\exp(-\int_0^t \lambda(t';\theta)dt'), \text{ and}$$

$$S(t;\theta) = \exp(-\int_0^t \lambda(t';\theta)dt').$$

If there are competing risks of failure then the hazard function can be written as a sum of competing hazard functions. These functions (defined below) estimate the failure rates for a specific cause given the "removal" of all other causes.

Definition 10 *(Kalbfleisch and Prentice, p.167) : Let T be a continuous failure time variable. Suppose that when failure occurs it may be attributed to exactly one of m distinct causes. Let J denote the cause for failure, where J is an integer in the set $\{1,2,....,n\}$. Then $\lambda_j(t;\theta)$, the j^{th} competing risk component, is defined by the equality:*

$$\lambda_j(t;\theta)dt = prob\{T \in [t, t-dt), J = j | T \geq t\}.$$

Assuming J must be a unique element from set $\{1,...,m\}$, we have the competing risks hazard decomposition:

$$\lambda(t;\theta) = \sum_{j=1}^{m} \lambda_j(t;\theta), \text{ and}$$

$$f(t;\theta) = (\lambda_1(t;\theta) + \cdots \lambda_m(t;\theta))S(t;\theta).$$

The density $I(t;\theta)$ can be decomposed into a sum of subdensities as follows:

$$I(t;\theta) = \sum_{j=1}^{m} I_j(t;\theta) \quad \textit{where.}$$

$$I_j(t;\theta) = \lambda_j(t;\theta)\exp(-\int_0^t \lambda(t';\theta)dt').$$

Given observations of the competing risks failure time variable T, and each observation's cause of failure, each $\lambda_j(t;\theta)$ can be estimated (Kalbfleisch and Prentice, p.168).

□

Now, the main BSE infection route (i.e. the primary "risk" or cause of failure) is believed to be tainted feed. A second likely infection route is maternal transmission: infection occurring in the womb or during the birthing process. If the cow population at risk is a given birth cohort, then the infection time variable, T, is an age at infection where $t = 0$ is age 0. Furthermore, a cow has a positive *maternal* risk only when $t = 0$. Thus the maternal competing risk component is discrete and T is a mixed continuous/discrete failure time variable. In the next section we will look at continuous/discrete failure time variables in more detail. If $\lambda(t) = \lambda_F(t) + \lambda_M(t)$, where $\lambda_F(t)$ is the feed risk component and $\lambda_M(t)$ is the maternal risk component, then,

$$\lambda_M(t) = \begin{cases} R, & t = 0 \\ 0, & elsewhere \end{cases}.$$

and R = the probability that a calf in the given cohort will be maternally infected.

In section 2.3 we give an example of a mixed density arising from a competing risks model where one risk function is discrete.

3.3 Mixed Continuous/Discrete Failure Time Densities.

Let $I(t)$, $t \geq 0$ be a mixed failure time density. Suppose that $\Pr\{T = a_i\} > 0$ for a finite set of points $\{a_1, \ldots, a_n\}$. Let $S(t)$, $t \geq 0$, be the survivor function of the random variable T, where

$$S(t) = \Pr\{T \geq t\}. \text{ and}$$
$$S(t) = \int_t^\infty I(t)dt + \sum_{a_i \geq t} I(a_i).$$

If $\lambda(t)$ is the hazard of T, then

$$S(t) = \exp[-\int_0^t \lambda(t')dt'] \Pi_{a_i < t}(1 - \lambda(a_i)). \quad \text{and}$$
$$I(t) = \lambda(t)S(t) = \lambda(t)\exp[-\int_0^t \lambda(t')dt'] \Pi_{a_i < t}(1 - \lambda(a_i)) \quad \text{Kalbfleisch \& Prentice, p.8).}$$

Example 11 *A competing risks model where one risk function is discrete*

Suppose $\lambda(t) = \lambda_1(t) + \lambda_2(t)$, where λ_1 and λ_2 are competing risk functions. Furthermore, suppose,

$$\lambda_2(t) = \begin{cases} R, & t = 0 \\ 0, & elsewhere. \end{cases}$$

The random variable T is a mixed continuous/discrete failure time variable with a positive density at T=0. The infection density $I(t)$ is expressed in hazard form:

$$I(t) = \lambda_1(t)\exp[-\int_0^t \lambda_1(t')dt'](1 - R), \ t > 0, \text{ and}$$
$$I(0) = \lambda_1(0) = R.$$

A combined form for $I(t)$ is:

$$I(t) = \lambda_1(t)\exp[-\int_0^t \lambda_1(t')dt'](1-R) + R\delta(t). \quad \textit{where}$$

$$\delta(t) = \begin{cases} 1. & t = 0 \\ 0. & \textit{elsewhere} \end{cases}.$$

□

3.4 The BSE Joint Infection/Onset Competing Risks Density

Let A, $A \geq 0$, be a random variable representing the age at BSE infection for a cow of birth cohort, t_0, where t_0 = the birth year of the cohort. The age at infection density $I(a|t_0)$, for cohort t_0 can be written in hazard form:

$$I(a|t_0) = \lambda(a|t_0)\exp[-\int_0^a \lambda(a'|t_0)da'].$$

Suppose that farmer-determined age dependent feeding patterns of cows do not change over time, and that *age* dependent protein absorption rates are, similarly *time* invariant. Furthermore, it is known that the amount of MBM given to cows varied over time. Thus one can justify (Anderson et al.1996) factoring the hazard $\lambda(a|t)$, into time dependent and age dependent factors

$$\lambda(a|t) = r(t)g(a). \tag{3.1}$$

The function $g(a)$ reflects age dependent protein absorption and feeding practices. The function $r(t)$ reflects time dependent MBM feeding patterns.

The hazard function for cohort t_0 is defined to be:

$$\lambda(a|t_0) = r(t_0 + a)g(a). \text{ where}$$

$$\lambda(a|t_0)da \cong \Pr\{A \in [a, a+da) | A \geq a, T = t_0 + a\}.$$

At the time that this model was proposed it was accepted that there existed two and possibly

three routes of infection: oral feed route, vertical (dam to calf) infection route and horizontal (cow to cow) infection route. The three infection routes are competing risks of infection and the hazard is written as a sum of the competing risk components (see section 3.2):

$$\begin{aligned}\lambda(a|t) &= \lambda_F(a|t) + \lambda_M(a|t) + \lambda_H(a|t) \\ &= r_F(t)g_F(a) + r_M(t)g_M(a) + r_H(t)g_H(a).\end{aligned}$$

where the maternal risk is non-zero only at birth yielding the discrete function,

$$g_M(a) = \begin{cases} 1, & a = 0 \\ 0, & a > 0 \end{cases}.$$

Consequently,

$$\lambda_M(a|t_0) = \begin{cases} r_M(t_0), & a = 0 \\ 0, & a > 0 \end{cases}, \text{ where}$$

$$\begin{aligned}\lambda_M(0|t_0) &= \Pr\{\text{a dam born at time } t_0 \text{ is maternally infected}\} \\ &= \Pr\{A = 0|t = t_0\}.\end{aligned}$$

The components $r_F(t)g_F(a)$ and $r_H(t)g_H(a)$ are not formulated as easily. The authors (Ferguson et al. 1997b) express the time dependent feed factor, $r_F(t)$ as a piecewise quadratic curve and estimate its unknown parameters when the likelihood is optimized. Furthermore, the function $g_F(a)$ is assumed to have a parametric form whose unknown parameters are also estimated when the likelihood is maximized. At the time of the study there was no evidence of horizontal transmission. Any parameters related to horizontal transmission are fixed and not estimated. For most of the discussion the authors assume that horizontal transmission does not exist and we will at a later point continue with this assumption. The presence of the discrete competing risk component $\lambda_M(a|t_0)$ makes the age at infection variable, A, a mixed continuous/discrete random variable, with a positive probability at $A = 0$. A form for the age at infection density corresponding to the above hazard is derived in a similar way to Example

11 of this chapter and is given by:

$$I(a|t_0) = [\lambda_F(a|t_0) + \lambda_H(a|t_0)] \exp[-\int_0^a (\lambda_F(a'|t_0) + \lambda_H(a'|t_0))da'](1 - r_M(t_0))(1 - \delta(a))$$
$$+ r_M(t_0)\delta(a).$$

where, $\delta(a) = \begin{cases} 1. & a = 0 \\ 0. & elsewhere \end{cases}$.

Assuming that the incubation period, S, is independent of birth cohort and of the age of a cow, it follows that the joint density of A and U is given by:

$$J(a, u|t_0) = I(a|t_0)f(u - a) \tag{3.2}$$
$$= \{[\lambda_F(a|t_0) + \lambda_H(a|t_0)] \exp[-\int_0^a (\lambda_F(a'|t_0) + \lambda_H(a'|t_0))da'](1 - r_M(t_0))(1 - \delta(a))$$
$$+ r_M(t_0)\delta(a)\}f(u - a) \tag{3.3}$$

where, $a \geq 0$, and $u \geqslant a$.

The density $J(a, u, t_0)$, is the major component of the backcalculation formula 2.1). It is used to form a likelihood of count observations reflecting BSE onset rates. Maximization of the likelihood leads to parameter estimates and an estimate of the time of infection density, $I(a|t_0)$. The density $J(a, u|t_0)$ has many unknown parameters. If some parameters can be expressed in terms of others then the likelihood equation will not yield unique MLEs. The authors go through great lengths to express the maternal risk component, $r_M(t)$, in terms of feed hazards of previous generations. In the next section we see the relationship between a maternal hazard and past feed hazards, furthermore, we see how this relationship yields the desired expression.

3.5 Estimation of the maternal risk component

The many components of the joint infection/onset density make it impossible to estimate from onset data alone. One does not know the infection route of a case and therefore cannot separate the three risk components. The authors Ferguson et al., 1997b), simplify the model by assuming that there is no horizontal transmission (the existence of horizontal transmission

has never been demonstrated leaving only feed and maternal components. The authors point out that the maternal risk component can be traced back to a feed risk. The first generation of exposed cattle were only feed infected. Cattle of the second generation were exposed to tainted feed and a maternal risk from dams that were feed infected. Subsequent generations had a maternal risk proportional to the number of infectious dams in the previous generation. Every infectious dam can be traced to a feed infection in a previous generation. In this section we see how the maternal risk is expressed in terms of feed risks of previous generations.

Recall that the maternal risk component is a discrete component, and is nonzero only at $a = 0$.

$$\lambda_M(a|t) = \begin{cases} r_M(t), & a = 0 \\ 0, & a > 0 \end{cases} \quad \text{and,}$$

$$r_M(t) = \Pr\{\text{a calf born at time } t \text{ is maternally infected}\} \tag{3.4}$$

$$= \Pr\{\text{a dam is infectious and preonset at time } t \text{ and infects her calf}\}. \tag{3.5}$$

The probability of a calf being maternally infected is proportional to the proportion of infectious, preonset, dams. (A dam that is post onset can give birth; however, its calf is likely to be culled.) The proportionality constant is the probability that an infected preonset dam transfers the infection to its calf. Therefore,

$$r_M(t) = \epsilon\{\text{ the proportion of infectious preonset dam s at time } t\},$$

where, $\epsilon = \Pr\{\text{ a dam infects its calf } | \text{the dam is infectious and preonset at time } t\}$.

The infection/onset density (3.2) is used to obtain an expression for the proportion of infectious preonset dams at time t. The following observations play a role in forming an expression for the proportion of infectious preonset dams at time t :

- It is clear that a dam must be infected to be infectious.
- The probability of a dam being infected depends on its age at time t and is a function of infection/onset density (3.2).
- The level of infectiousness of a dam will be expressed as a function of the time left until

disease onset in the dam, where level of infectiousness is assumed to increase as onset time approaches.

Suppose an infected preonset dam is of age r at time t, and has v years left until onset. Then $u = v + r$ is the age at which disease onset will occur. The probability that an age r dam is infected is

$$\int_r^\infty \int_0^r J(a, u|t-r)dadu$$

where $J(a, u|t-r)$ is BSE infection/onset density (3.2). To express the probability of infectiousness we define function

$$\Omega_M(v) = \Pr\{\text{an infected cow is maternally infectious } | v \text{ years left until disease onset}\}.$$

The probability that an age r dam is maternally infectious and preonset at time t is

$$y_M(t, r) = \int_r^\infty \Omega_M(u-r) \int_0^r J(a, u|t-r)dadu. \tag{3.6}$$

We present two special cases of $y_M(t, r)$ that elucidate the general case.

Example 12 *Assume that an infected dam is always maternally infectious, then*

$$\begin{aligned}
\Omega_M(v) &\equiv 1, \text{ and} \\
y_M(t, r) &= \Pr\{\text{age } r \text{ dam is maternally infectious and preonset at time } t\} \\
&= \Pr\{\text{age } r \text{ dam is infected and preonset at time } t\} \\
&= \Pr\{A < r \text{ and } L > r\} \\
&= \int_r^\infty \int_0^r J(a, u|t-r)dadu \ . \ (t - r = \text{ year of birth of the dam}) \\
&= \int_r^\infty \Omega_M(u-r) \int_0^r J(a, u|t-r)dadu.
\end{aligned}$$

□

Example 13 *Suppose an infected cow is maternally infectious only during the last half year of the incubation period. That is:*

$$\Omega_M(v) = \begin{cases} 1. & 0 \le v \le 1/2 \\ 0. & elsewhere \end{cases}.$$

Then the probability that an age r dam at time t is infectious and preonset is:

$$y_M(t,r) = \Pr\{0 < A \le r,\ r \le U \le r + 1/2\}$$

$$\Pr\{0 < A \le r,\ 0 \le U - r \le 1/2\} = \int_r^{r+1/2} \int_0^r J(a,u\,t-r)dadu$$

$$= \int_r^{\infty} \Omega_M(u-r) \int_0^r J(a,u|t-r)dadu.$$

□

Now that we have a formula for the probability that a dam is infectious, we are one step closer to the aim of obtaining a formula for the maternal risk component $r_M(t)$. Recall that

$$r_M(t) = \epsilon\{\text{the proportion of maternally infectious, preonset, dams at time } t\}.$$

Which leads to a heuristic justification for expression (3.7) below:

$$r_M(t) \cong \epsilon \sum_{r \ge 2} \Pr\{\text{a dam is infectious } | r \le \text{age of dam } < r + dr\} \cdot$$

$$\Pr\{r \le \text{age of dam } < r + dr\}$$

$$\cong \epsilon \sum_r \sigma_M(t,r) y_M(t,r) dr.$$

or, more precisely,

$$r_M(t) = \epsilon \int_2^{\infty} \sigma_M(t,r) y_M(t,r) dr. \tag{3.7}$$

where $\sigma_M(t,r)$, the age distribution, reflects the fact that the probability that a dam is of age r depends on t as well as r. The following points are worth noting about equation 3.7:

- Since a dam is a cow of age ≥ 2, the lower limit of integration in (3.7) is 2.
- All cattle of age ≥ 2 are assumed to be dams since almost all bulls are slaughtered before age 1.
- Formula (3.7) is, in a sense, recursive since formula (3.6) for $y_M(t,r)$ contains the density (3.2) which contains $r_M(t-r)$.

The authors use an iterative approach to "unravel" formula (3.7) and obtain the nonrecursive equivalent. The following example derives a nonrecursive formula for $r_M(t)$ under simplifying assumptions for illustrative purposes.

Example 14 *Suppose that in a certain cow population cows give birth to one calf at age 2, and all births occur in the same season. In that population a dam is defined to be a cow of age 2. Furthermore assume that every infected dam is infectious and that there is no horizontal transmission of infection. Then.*

$\Omega_M(v) \equiv 1,\ v \geq 0$, *since every infected dam is infectious.*

$$\sigma_M(t,r) = \begin{cases} 1, & r = 2 \\ 0, & elsewhere \end{cases}, \quad \textit{since a dam is a cow of age 2.}$$

$y_M(t,r) = \int_r^\infty \int_0^r J(a,u|t-r)dadu$, *(the probability of being infected and preonset by age r).*

$$J(a,u|t-r) = \lambda_F(a|t-r)\exp[-\int_0^a \lambda_F(a'|t-r)da']f(u-a)(1-r_M(t-r))(1-\delta(a))$$
$$+ r_M(t-r)\delta(a)f(u-a).$$

and, $r_M(t) = \epsilon\, y_M(t,2)$ *($\sigma_M(t,r)$ is nonzero only at r=2).*

Hence,

$$r_M(t) = \epsilon\, y_M(t,2) = \epsilon \int_2^\infty \int_0^2 J(a,u|t-2)dadu$$
$$= \epsilon \int_2^\infty \int_0^2 \{\lambda_F(a|t-2)\exp[-\int_0^a \lambda_F(a'|t-2)da']f(u-a)(1-r_M(t-2))(1-\delta(a))$$
$$+ r_M(t-2)\delta(a)f(u-a)\}dadu.$$

Leading to the recursive formula

$$r_M(t) = \epsilon \int_2^\infty \int_0^2 \{\lambda_F(a|t-2) \cdot \tag{3.8}$$

$$\exp[-\int_0^a [\lambda_F(a'|t-2)da']f(u-a)\}dadu(1-r_M(t-2)) \tag{3.9}$$

$$+ r_M(t-2)\epsilon \int_2^\infty f(u)du.$$

Observe that the above is the sum of the probabilities

$$r_M(t) = \epsilon \cdot \Pr\{\textit{a dam is preonset feed infected and not maternally infected}|t\}$$

$$+ \epsilon \cdot \Pr\{\textit{a dam is maternally infected and preonset}|t\}.$$

Let

$F(t) = \Pr\{$*a dam is feed infected and preonset* $|t$*. no maternal infection*$\}$*. and*

$G(t) = \Pr\{$*a dam is preonset* $|$*maternally infected*$\}$.

Then substituting $F(t)$ *and* $G(t)$ *into expression (3.8)*

$$r_M(t) = \epsilon F(t)(1 - r_M(t-2)) + \epsilon G(t) r_M(t-2).$$

In summary we have shown that

$$r_M(t) = \epsilon F(t)(1 - r_M(t-2)) + \epsilon G(t) r_M(t-2).\ t = 2, 4, 6, 8\ldots \tag{3.10}$$

The above recursive formula expresses the maternal transmission probability of one generation in terms of the probabilities of feed infection and maternal infection in the previous generation. Rearranging the terms we have.

$$r_M(t) = \epsilon F(t) - \epsilon r_M(t-2)[F(t) - G(t)].$$

Thus .

$$r_M(2) = \epsilon F(2).$$
$$r_M(4) = \epsilon F(4) - \epsilon^2[F(4) - G(4)].$$

Leading to the nonrecursive formula

$$r_M(t) = \epsilon F(t) - \epsilon^2 F(t-2) - \epsilon^3 F(t-4) - \cdots - \epsilon^{t/2} F(2)[F(4) - G(4)].$$

Hence, $r_M(t)$ is expressed only in terms of ϵ and feed infections of previous generations reducing the task of estimating a maternal risk function to that of estimating one parameter, ϵ.

$\square$

The authors Ferguson et al. (1997b), have a parallel formula for $r_M(t)$ in the general case. We now show how the general formula is obtained the details of which were omitted by the authors.

In Example 14 above, the cow population was partitioned into hypothetical generations spaced 2 years apart and recursive formula (3.10) for $r_M(t)$ expressed the maternal risk, $r_M(t)$, in terms of $r_M(t-2)$: the maternal risk of the previous generation. In the general case the authors partition the cow population into actual generations:

$\text{Gen}_1 =$ the first generation exposed to tainted feed

$\text{Gen}_2 =$ the offspring of Gen_1, etc.

Let

$$C_n = Gen_n \cup Gen_{n-1} \cup \ldots \cup Gen_1.$$

That is

$C_n =$ the cow population whose most recent generation is Gen_n.

The maternal risk depends of the current cow population. Population C_1 had only a feed risk, population C_2 had a feed risk and a maternal risk from dams of C_1 that were feed infected. To emphasize the dependence of $r_M(t)$ on the cow population and the dependence of the cow population on the current year t, we will use the notation $r_M(C_{n(t)})$ where $n(t)$ is an integer and $C_{n(t)}$ is the cow population at time t. Using this notation and generalizing (3.10) above, we can express the maternal risk of population $C_{n(t)}$ in terms of the maternal risk of $C_{n(t)-1}$.

$$\begin{aligned} r_M(C_{n(t)}) &= \epsilon\{\text{the proportion of infectious, preonset dams in population } C_{n(t)}\} \\ &= \epsilon\{\text{the proportion of feed infected infectious preonset dams in } C_{n(t)}\} \\ &+ \epsilon\{\text{the proportion of maternally infected preonset dams in } C_{n(t)}\} \\ &= \epsilon\{\text{the proportion of feed infected infectious preonset dams in } C_{n(t)}\} \\ &- \epsilon^2\{\text{the proportion of feed infected infectious preonset dams in } C_{n(t)-1}\} \\ &+ \epsilon^2\{\text{the proportion of maternally infected infectious preonset dams in } C_{n(t)-1}\}. \end{aligned}$$

The last term can be expressed in terms of feed and maternal risks of population $C_{n(t)-2}$, etc. In this more general setting we can derive an expression for $r_M(t)$ analogous to expression (3.10). Let,

$$I_F(a|t) = \lambda_F(a|t)\exp[-\int_0^a \lambda_F(a'|t)da'], \quad \text{the age at feed infection density,}$$

$$J(a,u|t) = I_F(a|t)f(u-a)(1-r_M(t))(1-\delta(a)) + f(u)\delta(a)r_M(t),$$

the joint (feed or maternal) infection onset density, $a \geq 0$, $u > a$.

Starting with formula (3.7) we have

$$\begin{aligned} r_M(t) &= \epsilon\int_2^\infty \sigma_M(t,r)y_M(t,r)dr = \epsilon\int_2^\infty \sigma_M(t,r)\int_r^\infty \Omega_M(u-r)\int_0^r J(a,u|t-r)dadudr \\ &= \epsilon\int_2^\infty \sigma_M(t,r)\int_r^\infty \Omega_M(u-r)\int_0^r \{I_F(a|t-r)f(u-a)(1-r_M(t-r)) + f(u)\delta(a)r_M(t-r)\}dad \end{aligned}$$

Rewriting as a sum of integrals

$$r_M(t) = \epsilon \int_2^\infty \sigma_M(t,r) \int_r^\infty \Omega_M(u-r) \cdot \tag{3.11}$$

$$\int_0^r I_F(a|t-r) f(u-a)(1-r_M(t-r)) da du dr \tag{3.12}$$

$$+ \epsilon \int_2^\infty \sigma_M(t,r) \int_r^\infty \Omega_M(u-r) f(u) r_M(t-r) du dr$$

$= \epsilon[\Psi \cdot (1 - r_M)](t) + \epsilon[\Phi \cdot (r_M)](t)$, where Ψ and Φ are intergral operators representing the factors that multiply ϵ in (3.11).

On the left side we have $r_M(t)$ (or equivalently $r_M(C_{n(t)})$). On the right side we have $r_M(t-r)$ or equivalently $r_M(C_{n(t-r)})$ where $r \geq 2$. Due to slaughtering practices most cows do not live beyond age 3. Using this information together with the fact that most cows have one calf a year we make an approximating substitution: $r_M(t-r) = r_M(C_{n(t)-1})$ and introduce the notation: $r_M^{(n)} = r_M(C_n)$. Hence,

$$r_M^{(n)} = \epsilon[(\Phi - \Psi) \cdot r_M^{(n-1)}](t) + \epsilon[\Psi \cdot (1)](t).$$

Repeated substitutions of the above recursive equation into itself leads to a non recursive formula for $r_M^{(n)}$:

$$r_M^{(n)} = \sum_{i=1}^{n} \epsilon^{i-1} (\Phi - \Psi)^{i-1} \cdot \epsilon \Psi \cdot (1). \tag{3.13}$$

Thus, the current maternal risk is expressed in terms of ϵ and feed risks of previous generations. The only unknown feed risk parameter left in density (3.2) is ϵ, the probability of maternal transmission of infection. This simplifies the backcalculation likelihood based on density (3.2) by reducing the number of unknown parameters and eliminating the need to estimate the maternal hazard component when maximizing the likelihood.

3.6 Maximum Likelihood methods

After establishing the form of the parametric joint infection/onset density one would like to estimate the unknown parameters from available observations.

Data used in the model was taken from the Central Veterinary Laboratory's BSE case reports, stratified by birth cohort and age at onset. The failure time variable was age at disease onset with covariate birth cohort. Thus, observations were grouped into yearly categories determined by birth cohort and within each birth cohort by yearly age at onset intervals. The great number of observations make the inclusion of a likelihood term for every observation too computationally intensive. A count variable evaluated within each cell summarized the data yielding multinomial observations used in the model's likelihood. Assuming that the life span of a cow is 18 years, let

X_{i,t_0} = the number of diseased cattle of cohort t_0 with age at onset in the i^{th} year, $i = 1, \ldots 18$.

X_{19,t_0} = the number of cohort t_0 cattle that did not experience the disease.

When backcalculation was applied to estimating future AIDS infection levels the multinomial parameter "N", representing all infections acquired during a set time interval, was unknown and was estimated when the likelihood was maximized. In this case the parameter, "N", representing the size of a birth cohort is known and is equal to N_{t_0}. Let

N_{t_0} = the size of cohort t_0, so that

$$\sum_{i=1}^{19} X_{i,t_0} = N_{t_0}.$$

Let the multinomial parameters

$$p_{i,t_0} = \Pr\{\text{onset in year } i \mid \text{cohort } t_0\}$$
$$= \frac{E(X_{i,t_0})}{N_{t_0}}, \quad i = 1, \ldots 18$$

and

$$p_{19,t_0} = \Pr\{\text{a cow does not live to experience disease onset (cohort } t_0\}.$$

Let

$$x_{i,t_0},\ i = 1, \ldots, 19$$

be the observed values of the count variables defined in (3.17) where.

$$x_{19,t_0} = N_{t_0} - \sum_{i=1}^{18} x_{i,t_0}.$$

If U = age of disease onset. then a probability related to p_{i,t_0} that is easier to express. is defined as an intermediate step in the formulation of an expression for p_{i,t_0}. Let

$$p_i(t) = \Pr\{i - 1 < U < i | \text{ calf born at time } t\}\ i = 1, \ldots, 18$$

and

$$p(t,u) = \Pr\{\text{a cow becomes a case by age } u \text{ |born at time } t\}.$$

$$p(t) = p(t,i) - p(t,i-1).$$

Since most cattle are slaughtered before age three. and since in the derivation of the expression for density $J(a,u;t)$ it is assumed that there is an "absence of slaughtering practices" the separately estimated survival distribution. $S(t)$. is included in the expression for $p(t,u)$ to account for the discrepancy between numbers of infections and cases. Let

$$S(a) = \Pr\{\text{survivorship until age } a\}$$

then.

$$p(t,i) = \int_0^t S(u) \int_0^u J(a,u|t) da du. \tag{3.14}$$

It is assumed that the parameter p_{i,t_0}, is close to $p_i(t)$ for t close to t_0. If T = the random variable representing time of birth, then:

$$p_{i,t_0} = E(p_i(T)|t_0 < T < t_0 + 1)$$

(each cohort includes calves born during a one year span).

Let R represent the restriction: $t_0 < T < t_0 + 1$. Then.

$$p_{i,t_0} = E_T(p_i(T)|R)) = E_{T|R}(p_i(T)). \tag{3.15}$$

Let $B(t)$ be the (time of birth) density of random variable T. The conditional density of $T|R$ has density $B_{T|R}$:

$$B_{T|R}(t|R) = \frac{B(t)}{\int_{t_0}^{t_0+\Delta} B(t)dt}.$$

Using (3.15) the multinomial parameters are given by.

$$\begin{aligned} p_{i,t_0} &= E_{T|R}(p_i(T)) \qquad (3.16)\\ &= \int_{t_0}^{t_0+\Delta} p_i(t) B_{T|R}(t) dt \\ &= \int_{t_0}^{t_0+\Delta} \frac{p_i(t)B(t)}{\int_{t_0}^{t_0+\Delta} B(t)dt} dt \\ &= \int_{t_0}^{t_0+\Delta} \frac{p(t,i)B(t)}{\int_{t_0}^{t_0+\Delta} B(t)dt} dt - \int_{t_0}^{t_0+\Delta} \frac{p(t,i-1)B(t)}{\int_{t_0}^{t_0+\Delta} B(t)dt} dt, \ i = 1, \ldots, p(t_0) - 1. \end{aligned}$$

and, $p_{19,t_0} = 1 - \sum_{i=1}^{18} p_{i,t_0}$.

It is assumed that the union of all birth cohort observations:

$$\cup_{t_0}\{x_{1,t_0},\cdots x_{n(t_0),t_0}\}$$

have a multinomial distribution. The loglikelihood written up to additive constants) is equal to:

$$l \propto \sum_{t_0}\{x_{18,t_0}\ln(p_{18,t_0}) + \textstyle\sum_{i=1}^{18} x_{i,t_0}\ln(p_{i,t_0})\}. \tag{3.17}$$

From (3.14) and (3.16)The quantities p_{i,t_0} are expressed in terms of the unknown parameters of $J(a,u|t)$. Not all parameters of $J(a,u|t)$ can be estimated from onset observations alone.. The likelihood may be overparameterized. If both the infection density, $I(a,t)$, and the incubation density, $f(s)$, are parameterized with unknown parameters then the model is likely to be overparameterized or unidentifiable. If this is the case then some quantities (such as the mean incubation period) must be guessed at or estimated from independent data. The maternal transmission rate, ϵ, can be estimated from data accumulated during the maternal transmission study (see Chapter 4). In particular, parameters in the maternal infectiousness distribution $\Omega(v)$, were assigned a few hypothesized values determining at which point in the incubation period infectiousness is assumed to begin (e.g.: within 6 months of onset) and separate sets of MLEs' were calculated for each assigned value. Some information on the incubation period distribution is provided by the maternal cohort study and the oral dosing study (Anderson et. al 1996). Maximization of the loglikelihood after some parameters are fixed, yields parameter estimates of the remaining unknown parameters of $J(a,u|t)$. The resulting estimates will be summarized in the last section of this chapter. The parametric forms assumed for some of the functions that appear in $J(a,u|t)$ are reviewed in the next section.

3.7 Incubation period and age-dependent susceptibility distributions

Recall by equation (3.1) that the feed hazard, $\lambda_F(a|t)$, is assumed to factor into a time dependent factor and an age dependent factor:

$$\lambda_F(a|t) = r_F(t)g(a).$$

The function $g(a)$ reflects the age dependent absorption rates of the aetiological agent in addition to exposure to the agent due to age dependent feeding practices of meat and bone meal. The authors (Ferguson et al., 1997b) tried fitting the likelihood (3.17) with various functional forms for $g(a)$, the age dependent susceptibility density, and $f(s)$, the incubation period density. It was found that the basic model results were robust to changes in the functional forms.

The following three functional forms were tried for the incubation period density: A Gamma density with a delay

$$f(u) = \begin{cases} 0, & u \leq (1-\alpha_1)\alpha_2 \\ (u-(1-\alpha_1)\alpha_2)^{\alpha_1^2\alpha_2^2/\alpha_3 - 1} & \\ \times \exp\left[-\frac{(u-(1-\alpha_1)\alpha_2)\alpha_1\alpha_2}{\alpha_3}\right], & u > (1-\alpha_1)\alpha_2 \end{cases}.$$

a Weibull density with a delay

$$f(u) = \begin{cases} 0, & u \leq (1-\alpha_1)\alpha_2 \\ (u-(1-\alpha_1)\alpha_2)^{\alpha_3 - 1} & \\ \cdot \exp\left[-\left(\frac{(u-(1-\alpha_1)\alpha_2\Gamma(1+1/\alpha_3)}{\alpha_1\alpha_2}\right)^{\alpha_3}\right], & u > (1-\alpha_1)\alpha_2 \end{cases}.$$

and one referred to as a "mechanistic incubation period density" which we derive below) to model the incubation period of BSE (Medley & Short, 1996). The delays in both the Gamma and Weibull densities reflect the fact that there are almost no known cases of BSE in cattle below the age of 2. Therefore the incubation period is believed to be greater than or equal to 2. The mechanistic density was found to give the best model fit and its derivation is based on the incubation dynamics of the BSE prion aetiological agent. It is believed that the incubation

period is related to the initial infecting prion dose and that onset is triggered when the prion level reaches a specific level. We now discuss the "mechanistic density" in detail.

Let, d_0 represent the proportion of the triggering prion dose that entered the body upon infection. The authors assume that d_0 has density $h(d_0)$. Let d represent the prion level at time t. It is assumed that d increases exponentially. Let

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

describe the deterministic growth of the prion substance over time. Recall that $d(t)$ is the *proportion* of the triggering prion dose. Therefore, at time t for which $d(t) = d_0 e^{\gamma_1 t} = 1$, disease symptoms appear and the incubation stage has ended making time t equal to s, the incubation period. Thus the incubation period, s, is a function of the initial dose d_0 :

$$d_0 = e^{-\gamma_1 s} \text{ and } s = \frac{-\ln(d_0)}{\gamma_1}.$$

A change of variables expresses density $f(s)$ in terms of h, the density of the initial dose

$$\begin{aligned} f(s) &= -h(e^{-\gamma_1 s}) \cdot \frac{d(e^{-\gamma_1 s})}{ds} \\ &= h(e^{-\gamma_1 s})\gamma_1 e^{-\gamma_1 s}. \end{aligned}$$

If h is assumed to be a Gamma density with parameters α and β, then,

$$\begin{aligned} f(s) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}(e^{-\gamma_1 s})^{\alpha-1} e^{-(e^{-\gamma_1 s})/\beta} \gamma_1 e^{-\gamma_1 s} \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\alpha_2 e^{-s/\alpha_1}}{\alpha_3}\right)^{\alpha_2^2/\alpha_3} \exp\left(-\frac{\alpha_2 e^{-s/\alpha_1}}{\alpha_3}\right) \end{aligned}$$

yielding the functional form of the mechanistic incubation period density.

Eight (un-normalized) functional forms are explored for the age-dependent susceptibil-

ity/exposure distribution. $g(a)$:

1. $g(a) = e^{-a/\gamma_1}$
2. $g(a) = e^{-a/\gamma_1} + \gamma_2$
3. $g(a) = a^{\gamma_2 - 1} e^{-a/\gamma_1}$
4. $g(a) = e^{-a/\gamma_1} + \gamma_2,\ a \leq 2$
   $= 2(e^{-a/\gamma_1} + \gamma_2),\ a > 2$
5. $g(a) = 1. a \leq \gamma_2$
   $= e^{-a/\gamma_1} + \gamma_2,\ a > \gamma_2$
6. $g(a) = 1$
7. $CDF = (1 - \exp[-(\gamma_1 a)^{\gamma_2}])(1 - \exp[-(\gamma_3 a)^{\gamma_2 + \gamma_4}])$
8. $CDF =$ as above but with a step at $a = 2$. doubling after 2 years of age.

The first form assumes exponentially decaying susceptibility and constant exposure. The second form also assumes exponentially decaying susceptibility with constant feed exposure. The third form assumes that the exposure is constant and the susceptibility is Gamma distributed. The fourth form is similar to the second but the level of susceptibility/exposure doubles at age 2. The fifth form assumes constant susceptibility until age γ_2, and exponential decaying with constant exposure after age γ_2. The sixth form assumes constant exposure and susceptibility. The seventh form is a cumulative density function, empirically derived (Anderson et al. 1996) and assumes constant exposure. This parametric form is very flexible in that it can take on many shapes.

The likelihood was found to be robust to changes in the functional forms of the incubation period density and the susceptibility/exposure distribution. Only the most extreme form 6 of constant exposure and susceptibility had an unacceptable goodness of fit. The combination of the "mechanistic incubation period density", and the empirically derived age/susceptibility density, form 7, yielded the best goodness of fit (Ferguson et al., 1997b). We conclude this

section with the infection/onset density and the related "best fit" forms:

$$J(a,u|t_0) = I(a|t_0)f(u-a) \tag{3.18}$$
$$= \{\lambda_F(a|t_0)\exp[-\int_0^a \lambda_F(a'|t_0)da'](1-r_M(t_0))(1-\delta(a))$$
$$+ r_M(t_0)\delta(a)\}f(u-a)$$

where, $a \geq 0$, and $u \geqslant a$.

$$\lambda_F(a|t_0) = r_F(t)g(a).$$

$$\int_0^a g(a')da' = (1-\exp[-(\gamma_1 a)^{\gamma_2}])(1-\exp[-(\gamma_3 a)^{\gamma_2+\gamma_4}]),$$

$$f(s) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\alpha_2 e^{-s/\alpha_1}}{\alpha_3}\right)^{\alpha_2^2/\alpha_3}\exp\left(-\frac{\alpha_2 e^{-s/\alpha_1}}{\alpha_3}\right).$$

$$r_M(t) = \sum_{i=1}^{n}[\epsilon^{i-1}(\Phi-\Psi)^{i-1}]\cdot\epsilon\Psi\cdot(1), \quad \text{see (3.13)},$$

and $r_F(t)$ is a piece-wise quadratic function.

3.8 Basic Results

The likelihood created with the mechanistic distribution and form 7 above for the incubation distribution and the age/susceptibility density, respectively leads to an estimate of 954.000 infections between the years 1974-1995 and to a prediction that there would be 9340 new cases for the years 1997-2001. (The actual number of case between January 1997 and July 2000 is 9786.) Estimates of the mean incubation period are consistently between 4.7-5.3 years for all models with relatively satisfactory goodness of fit. Estimates of the total number of animals infected lie in the range 900.000-1.130.000. Predictions of the number of cases between 1997 and 2001 show much variation. Future case predictions are highly effected by changes in the tail end of the feed risk profile. If the future feed risk is assumed to be nonexistent then all predicted cases would be maternal infections. Maternal cases are easier to estimate since the risk of infection is easier to predict. Under the assumption of horizontal transmission the number of predicted cases for the year 2001 was over 100.000. However, having no experimental evidence to back up assumed parameters related to horizontal transmission, the prediction is

only speculative.

Chapter 4

Maternal Transmission - The Paired Study

The maternal cohort study, initiated in July 1989 which, owing to the long incubation period, concluded in 1997, examined the possibility of a maternal transmission BSE infection route. Maternal transmission refers to the transmission of infection from an infected (usually preonset) dam to the calf in the womb or during the birthing process. Three groups of statisticians analysed the study's observations and presented their conclusions (Gore et al., Maternal Cohort Study, 1997, Donnelly et al., Maternal Cohort Study, 1997b, Curnow et al. Maternal Cohort Study, 1997). A vertical infection route (dam to calf) would prolong the BSE epidemic but would not sustain it indefinitely for several reasons: dams have an average of only one calf a year, slaughtering practices insure that the cattle population is more or less constant, and the probability of maternal transmission from an infectious dam to calf is likely to be much less than one. However, maternal transmission may imply the presence of disease in a wider range of body tissues and perhaps the existence of vertical transmission of other prion diseases such as CJD in humans.

The study design was 301 matched pairs of maternally exposed and control animals. The exposed cows were born of dams who had BSE at the time of calving or developed the disease within 13 months. The control animals were born of dams that were free of BSE up to age

6. Almost all infected cows that live to experience disease, have disease onset before age 6. This is partly because susceptibility to infection is age dependent.. Between 1986 and 1988 calves were recruited and placed on one of three study farms. They were followed for 7 years to observe BSE pathology and then slaughtered and examined to determine their disease status. The maternally exposed group were observed to have a significantly enhanced risk of disease. Had the study calves not been exposed to tainted feed, then observed enhanced risk in the maternally exposed group could be attributed to maternal transmission. However, study calves were given MBM both before recruitment and, on at least one study farm, after recruitment. Thus, MBM is a confounding factor which makes it difficult to distinguish between an enhanced risk due to maternal transmission and one due to varying genetic susceptibility to BSE. Matching was based on natal herd, calving period and time of recruitment into the study. The matching variables are related to the feed risk of a study animal. Hence, paired animals are assumed to have similar exposure to tainted feed. Donnelly et. al. (1997b) found that risk of disease increases with calves born to dams in a later incubation stage. This finding supports the maternal transmission theory but does not rule out the theory of varying genetic susceptibility or some combination of both.

We discuss Donnelly et al.'s (1997b) methods, which build on the infection/onset survival model described in Chapter 3. The authors explore three models utilizing covariates related to the study's confounding effect of feed exposure and to the relative contribution of maternal exposure. The first is a logistic regression model having BSE prognosis as a dependent variable. The covariates are variables related to feed exposure, maternal exposure, and maternal incubation stage. However, genetic susceptibility is not represented in this model. In order to estimate the relative effects of maternal transmission and genetic susceptibility one needs a model that incorporates parameters reflecting risks of different modes of disease transmission and of genetic susceptibility. The second model categorizes the study observations by BSE status and maternal exposure status and assigns likelihood terms according to category. The likelihood terms contain parameters related to maternal transmission and to genetic susceptibility. The third model utilizes information concerning modes of disease transmission, genetic susceptibility, the incubation density and , when appropriate, the dam's incubation stage at time of birth. Parameters related to these effects are represented in the likelihood whose terms

are formed from infection/onset density (3.2). The authors start with an exploratory data analysis looking for evidence of genetic susceptibility, and for a relationship between the maternal transmission rate and the dam's incubation stage, supporting the theory of maternal transmission.

4.1 Exploratory Data Analysis

Of the 301 study pairs, 18 were prematurely censored. Among the 602 animals, 42 maternally exposed and 13 control animals developed the disease. The proportion of BSE-affected exposed calves was .139, and the proportion of BSE-affected controls was .043. The authors found the difference to be significantly different from zero, p=value <.0001. The difference, .139-.043 = .096, can be used to estimate the maternal transmission parameter ϵ (see Chapter 3, section 3.5), where ϵ = the probability that an animal born to a BSE infected dam experiences maternal transmission.

It is important to take into account the paired nature of the study since doing so gives insight to the possibility of the enhanced risk being partially due to genetic susceptibility. Suppose feed exposure alone was the cause of the infected study animals' BSE status. Then the exposed animals' enhanced risks may be attributed to an enhanced genetic susceptibility to feed infection, since animals are paired based on variables related to feed exposure. If the BSE status of the exposed animal is independent of the BSE status of the matched control, then one can conclude that the enhanced risk is due to maternal transmission alone. However, if pairing is related to disease risk then one cannot rule out the possibility that enhanced risk may be partly due to enhanced genetic susceptibility of the exposed calf. The following is a contingency table of observed and expected pair outcomes indicating that pairing is related to BSE status. For example the table entry value of 36 indicates that there were 36 pairs whose maternally exposed calf had + BSE status at the study's conclusion and whose control calf had - BSE status at the study's conclusion.

Observed(Expected) maternally exposed

		+	-
Observed(Expected) control	+	6 (1.8)	7 (11.2)
	-	36 (40.2)	252 (247.8)

There is a significant difference between observed and expected BSE status counts under the null hypothesis that BSE statuses of paired animals are independent.. Fisher's exact test has a two sided p-value of .004. Thus, feed related variables may be contributing to the exposed animals' increased disease risk. The connection between increased risk and exposed status may be explained by the presence of a heightened genetic susceptibility in the exposed group.

4.1.1 Maternal incubation stage

If one established a dependence of BSE-infectivity in exposed calves on the dam's incubation stage, then one could claim that the observed enhanced risk amongst exposed study calves must be at least partly due to maternal transmission. A simple calculation of relative risks illustrates that such a dependence is plausible:

Relative risk of cows in a category $= \frac{\text{(porportion of disease affected exposed calves the category)}}{\text{(porportion of disease affected controls the category)}}$.

Relative risk of exposed calves $= \frac{.139}{.043} = 3.23$

Relative risk of exposed calves born after onset. $= 5.00$

Relative risk of exposed calves born during the incubation period. $= 2.91$

The increase in relative risk amongst exposed calves born after the dam experiences disease onset suggests that the incubation stage of a calf's dam is related to transmission of infection. The authors performed two tests to determine if BSE-risk varies significantly with incubation stage. Both tests use the paired nature of the data to control for the exposed calves' feed risk:

Test - 1

Pairs are stratified by year of birth (1987, 1988 or 1989) of the exposed animal and by incubation stage of the BSE infected dam (onset less than 50 days after birth, 51-100 days,

101- 150 days. more than 150 days). For each classification "excess risk of exposed animals" is estimated. Excess risk for a cell is defined as the mean of the random variable x_c, evaluated at each pair in cell c, where:

$$x_c = \begin{cases} 1, \text{ if exposed only is infected} \\ -1, \text{ if control only is infected} \\ 0, \text{ otherwise.} \end{cases}$$

The sample mean estimates together with 95% confidence intervals, are presented in the next table:

year of birth	onset ≤ 50	onset 51 - 100	onset 101 - 150	onset >150
1987	0.50±0.50	0.50±0.50	0.67±0.33	0.33±0.33
1988	0.12±0.04	0.14±0.05	0.13±0.15	-0.08±0.06
1989	0.02±0.03	0.04±0.24	0.00	0.00

No conclusions can be drawn from 1987 data since the sample sizes are too small. In 1988 one does see a drop of excess risk in animals born more that 150 days before onset in the dam. The drop of excess risk supports the existence of maternal transmission. On the other hand if the excess risk is due to maternal transmission then one would expect the 1989 estimates to be similar to those of 1988. However, the 1989, less than 50 days estimate is low and contradicts the increased risk at late incubation stage theory.

Test - 2

The second method splits the 301 pairs into two groups of pairs, where division is based on the dam's incubation stage at the birth of the exposed animal. The pair BSE-status (+,-), (+,+), (-,+), or (-,-) frequency distributions for the two groups are compared using Fisher's exact test. Eight different divisions were tested. The frequency distributions of the two groups divided at the "150 day before onset" point are given below.

$\leq$ 150 days, maternally exposed

		+	-	Total
Control	+	6	3	9
	-	34	214	248
	Total	40	217	257

> 150 days. maternally exposed

		+	-	Total
Control	+	0	4	4
	-	2	38	40
	Total	2	42	44

Pairing helps lessen the possible effect of genetic susceptibility since one is comparing groups of exposed calves whose matches have the same BSE status. If. for example. we are comparing the two (- control. +exposed) groups. then one can assume that genetic susceptibility does not have a significant role.

Fisher's exact test yielded a p-value of .011 suggesting that the frequencies were different. and that there is an enhanced risk of disease in animals born during a later incubation stage.

Results of the exploratory data analysis suggested evidence of a relationship between enhanced risk and maternal incubation stage. Hence. genetic susceptibility to feed infection may have contributed to the increase in risk amongst exposed calves. In order to understand the various factors contributing to the exposed group's enhanced risk the authors build three likelihoods for study observation which incorporate parameters related to disease transmission. The first is a logistic regression model with BSE status as the outcome variable. The second is a multinomial likelihood for observed disease counts within categories determined by birth period and the third is a full survival likelihood whose terms depend on infection/onset density (3.2).

4.2 Logistic Regression Model

A logistic regression model having BSE status as a dependent variable and covariates reflecting feed exposure and maternal exposure is a natural model that one would try given the study observations. The authors found that feed risk varied a great deal between herds. Animals

came from a wide range of herds with few animals per herd, making it not feasible to include a fixed herd effect in the model. Hence, a random effects term was introduced where by it was assumed that the effect due to the herd is drawn randomly from some population. Normal in this case. In order to represent feed exposure as accurately as possible the authors estimated the Pr{positive BSE pathology |herd}. The following model was explored:

$$\begin{aligned}
\text{logit}\{\Pr(Y_{ij}|U_i)\} &= \alpha + U_i + x'_{ij}\beta \\
Y_{ij} &= \text{BSE status for the j}^{\text{th}} \text{ animal in the i}^{\text{th}} \text{ pair. j} = 1,2 \\
U_i &= \text{the assumed (herd dependent) random intercept effect} \qquad (4.1) \\
&\text{for observations } Y_{i1} \text{ and } Y_{i2},\ U_i \text{ is assumed to be } N(0,\sigma_u^2). \qquad (4.2) \\
x'_{ij} &= \text{vector of covariates for j}^{th} \text{ pair member in i}^{th} \text{ observation} \\
\beta &= \text{vector of regression coefficients} \\
\alpha &= \text{the intercept}
\end{aligned}$$

The covariates of interest are age at purchase (used to reflect feed exposure) and the maternal incubation stage of dams of exposed calves.

4.3 Basic Mechanistic Likelihood Model

To test the hypotheses of increased risk amongst exposed calves being due to maternal transmission or genetic susceptibility, one needs to estimate parameters that are measures of each risk. Any model containing a genetic susceptibility parameter must also contain information about feed risk since short of identifying a susceptibility gene, a calf's genetic predisposition can only be ascertained if one has information related to the amount of tainted feed consumed by the calf and the BSE status of its dam. Period of birth is the covariate related to feed consumption: calves weaned in winter months can be assumed to have been exposed to more food supplements such as MBM and calves born after the feed ban are assumed to have less exposure to MBM.

The mechanistic model is a likelihood of count data derived from study observations.

Matched pairs are stratified by birth period. Within each birth period pairs are split into an exposed group and a control group. It is assumed that within a birth period the feed exposure is more or less constant so that pairing is no longer necessary. Within each group the count of BSE positive cows is a binomial variable whose probability is expressed in terms of the following three parameters:

π_b = feed risk within birth period b

ϵ = the probability that an exposed calf is maternally infected

s = the relative risk for all birth periods of a genetically susceptable cow ($s \geq 1$).

Let,

p_{bE} = the probability of an exposed birth period b calf developing the disease.and

p_{bC} = the probability of a control birth period b calf developing the disease.

Then,

$$p_{bE} = \epsilon + (1-\epsilon)\{1 - \exp(-s\pi_b)\}$$
$$= \Pr\{\text{disease is maternally transmitted}\} +$$
$$\Pr\{\text{disease is not maternally transmitted} \cap \text{ calf is feed infected}\}.$$

and,

$$p_{bC} = 1 - \exp(-\pi_b)$$
$$= \Pr\{\text{calf is feed infected}\}.$$

Note: Since study cows are observed for seven years it is assumed that all infections lead to disease onset.

The joint likelihood for all counts over all birth periods is the product of the binomial probability functions of the BSE count variables in each category. The log likelihood is given

below:

$$\sum_b x_{bE}\log[\epsilon+(1-\epsilon)\{1-\exp(-s\pi_b)\}]+(n_{bE}-x_{bE})\log[1-\epsilon-(1-\epsilon)\{1-\exp(-s\pi_b)\}]$$

$$+x_{bC}\log\{1-\exp(-\pi_b)\}+(n_{bC}-x_{bC})\log\{\exp(-\pi_b)\}.$$

Note: The effect of maternal incubation stage cannot be examined by this model.

4.4 Full Survival Mechanistic Likelihood Model

The full survival mechanistic likelihood model assigns likelihood terms to each member of the observered pairs based on infection/onset density (3.2), natal herd, time of birth and exposure status. The inclusion of the density introduces more parameters related to the infection process. The authors have three likelihood forms for exposed animals and three for controls. The forms depend on the animals' BSE statuses at the end of the study. The infection/onset density has a maternal transmission parameter and a genetic susceptibility parameter. However it does not have a parameter related to incubation stage in the dam at the time of calving. Thus, the significance of both parameters can be tested.

Let $f(s)$ represent the incubation period density. Let $g(a)$ be the likelihood of feed infection for a cow of age a, given a constant level of feed infectivity. Let $K_h(t)$ denote the risk of feed infection at time t in herd h and let $\Omega(v)$ represent the probability that a dams' infection could be transferred to its calf if the dam is v time units away from disease onset. Let ϵ be the rate of maternal transmission and s a genetic susceptibility parameter, where $s = 1$ means no genetic susceptibility and $s > 1$ indicates the presence of genetic susceptibility.

Each observed pair member is assigned one of three forms for the likelihood term based on the three possible end of study BSE statuses: onset before the end of the study, no clinical signs of onset but positive clinical pathology, and no signs of onset. We call the three forms for the likelihood terms A, B, and C. The three forms depend on function τ. Function τ depends on the age at onset in a calf (u) and in the case of exposed animals it also depends on the dam's incubation stage (v). It is formulated under the assumption that a study calf could not be feed

infected after recruitment into the study (after age a_k). For exposed animals.

$\tau(u, v)du \cong$ Pr {exposed calf is infected at birth by a dam with v years till onset and has onset in $(u, u + du]$} +

Pr {exposed calf is not infected at birth $\cap$ calf is feed infected before age a_k and has onset in $(u, u - du]$}.

More precisely,

$$\tau(u, v) = \epsilon\Omega(v)f(u) + \{1 - \epsilon\Omega(v)\}\int_0^{a_k} sK_h(t_0 + a)g(a) \cdot \exp\{-\int_0^a sK_h(t_0 + a')g(a')da'\}f(u - a)da, \text{ where}$$

a_k = age at recruitment.

For control animals,

$\tau(u)du \cong$ Pr{cow is feed infected before age a_k and has onset in $(u, u + du)$}.

More precisely,

$$\tau(u) = \int_0^{a_R} K_h(t_0 + a)g(a)\exp\{-\int_0^{a_R} K_h(t_0 + a')g(a')da'\}f(u - a)da.$$

Now, we can define the likelihood forms A, B, and C under the assumption that feed infection did not occur after recruitment into the study.

$Adu \cong$ Pr{exposed calf is infected at birth and has onset in $(u, u - du]$} +

Pr{exposed calf is not infected at birth $\cap$ calf is feed infected before age a_k and has onset in $(u, u + du]$}

$Bdu \cong$ Pr{cow has onset at age $>a_r$ |onset age is $>a_k$}

$Cdu \cong$ Pr{cow does not have onset between ages a_k and a_r | onset age is $>a_k$}

where a_k = the age of recruitment into the study, and a_c = the age at censoring.

More precisely,

$A = \frac{\tau(u)}{1 - \int_0^{a_k} \tau(u)du}$, for an animal with onset at age u where a_k = age at recruitment.

$B = \frac{\int_{a_c}^{18} \tau(u)du}{1 - \int_0^{a_k} \tau(u)du}$, for an animal without onset of clinical signs by age of censoring,

but with positive clinical pathology where a_c is age at censoring and 18 is a cow's life span.

$C = \frac{1 - \int_k^{a_c} \tau(u)du}{1 - \int_0^{a_k} \tau(u)du}$, for an animal without signs of clinical onset by age of censoring.

As in Chapter 3, function $f(s)$ is the incubation density and function $g(a)$ is the age at infection density. Both functions were not fitted to Maternal cohort study data, rather existing estimated parametric forms similar to those estimated in Anderson et al.(1996) were used. The denominators in terms A, B and C represent the probability that a cow has onset after age of recruitment. Since almost all cases of BSE are in cows above the age of two, and recruitment age was for the most part less than two, the denominator is very close to one and was assumed to be equal to one by the authors. The likelihood is constructed as a product of tems of type A, B, C.

The full likelihood model was compared via likelihood ratio tests to simpler models in order to test model assumptions. The following simpler models were explored:

- The incubation period distribution was eliminated resulting in a the likelihood term equal to the probability that BSE pathology is observed by age $\bar{7}$.
- Feed risk function $K_h(t)$ was assumed to be constant.
- The age susceptibility and absorption function g(a) was eliminated, thus in this model feed infection was assumed not to depend on age

The results obtained from the optimization of the logistic regression, mechanistic and full survival mechanistic models are summarized in the next section.

4.5 Study Results

4.5.1 Logistic Regression

The following fixed covariates were taken into consideration in the logistic regression model:

- maternal exposure to BSE
- age at recruitment
- birth cohort
- study farm
- sex.

A covariate allowing for incubation stage effect was added. The covariate was set equal to 1 if the dam's disease onset occurred X days before calving. The model was fitted for $X = 100, 110, \ldots, 170$ days and for $X = \infty$ (maternal effect throughout the incubation period). The maternal effect was most significant within 130 days of birth. The fact that incubation stage is significant suggests the presence of maternal transmission.

4.5.2 Basic Mechanistic Likelihood Model

Model results were sensitive to the number of birth classes used to stratify pairs and to the birth cut offs determining classes. The case of 2 birth periods: before and after July 18, 1988, the date of the MBM feed ban, yielded a model favouring $H_A : \epsilon = 0$ and rejecting $H_B : s = 1$. This suggests no maternal transmission and the existence of genetic susceptibility. However, when the cut off date was August 18, 1988 both H_A and H_B were accepted. When the data were divided into 10 birth classes the composite null hypothesis $H_A : \epsilon = 0$ and $H_B : s = 1$ was rejected.

The authors conclude that the grouping of pairs caused a loss of statistical power and that this simple model was insufficient.

4.5.3 Full Survival Mechanistic Likelihood Model

Models were fitted under the following assumptions:

I. Maternal transmission and no genetic susceptibility ($s = 1$).

II. Full duration maternal transmission and genetic susceptibility.

III. Genetic susceptibility only

V. Full duration maternal only.

All five model were fitted using two forms I and II for $g(a)$ and $f(s)$. Form I consists of the combination: mechanistically derived incubation period distribution and an empirically derived feed/susceptibility distribution (forms C and 7 of chapter 3. section 7) Form C consists of the combination: mechanistically derived incubation period distribution and a gamma distributed feed/susceptibility (forms C and 3). The models containing form I had a maternal only model that was significantly better than the other two However, the models containing form II had the combination maternal and agenetic model as the best fit. Parameter ϵ was estimated at .099 by the first and at .062 by the second. The second model estimated s at 2.39.

The results suggest that both a full maternal model and a combination of maternal and genetic are most likely.

Chapter 5

Maternal Transmission - Analysis of Dam/Calf pairs of BSE Cases

5.1 Introduction

Results from the maternal cohort paired study (discussed in Chapter 4) revealed the likely presence of a maternal transmission risk whose magnitude depends on the incubation stage of the dam at the time of calving. To confirm the paired study results Donnelly et al. (1997c). analyzed dam/calf data available in the large database of confirmed BSE cases initiated in 1986 by Britain's Central Veterinary Laboratory. The authors concentrated on cases born after the July 1988 Meat and Bone Meal (MBM) feed ban. Information related to the BABs' (born after ban BSE cases) dams, dams' incubation stage during calving, dams' herd and holding were recorded in most cases. There were approximately 30,000 BABs with disease onset before June 1996, the time of data collection. A dam of a BAB is labeled: $+$, if the dam had onset within 48 months of calving. The authors used information from the CVL database to determine if a dam's incubation stage influences disease transmission to its calf, and if feed risk may be enhanced by genetic susceptibility. Like in the paired cohort study feed infection is still a confounding factor as the use of meat and bone meal decreased but did not disappear after the feed ban.

The authors employed two strategies to achieve their results. Firstly, quantities estimating

the expected number of BABs born to dams in various disease incubation stages were computed under the null hypothesis that disease status in a dam is not related to disease status in its calf and the expected numbers were compared to the observed cases. Second a likelihood model similar to the mechanistic model in the paired cohort study (Chapter 4. section 4.3) was developed. However. the dependent variable was not disease outcome of the calf given incubation stage of the dam. rather it was the incubation stage of the dam given future disease onset in the calf. Thus Bayes Rule was used to calculate the reversed conditional probabilities. The model is an improvement on the mechanistic model as it categorizes BABs by cohort and holding. An observed clustering of cases within holdings (Ferguson et al.. 1997b) suggests the need to estimate feed risk within holding as well as within cohort. The mechanistic model of Chapter 4 was ineffective since feed risk was calculated within cohort. The model reviewed in this chapter estimates feed risk within holding and cohort thus taking into account the clustering of cases. This chapter's model incorporates the covariate: maternal incubation period of the dam at the time of calving, whose significance with respect to the calf's future disease status would imply the existence of maternal transmission. The authors started with a simple calculation of the expected number of BABs born at various incubation stages in the dam under the null hypothesis that incubation stage does not influence disease transmission.

5.2 Calculation of expected number of BABs born during various incubation stages in the dam

The authors divide the maternal incubation period into seven stages. k = 1.2.....6.-1. where $k = -1$ indicates lack of disease in a calf's dam. Expected numbers of BABs corresponding to the seven stages are estimated within birth cohort and holding. For the remainder of this section let us imagine. for the sake of clarity. that there is only one holding and one cohort.

Let,

D = the number of dams giving birth to calves in "the cohort" (Each birth cohort is one year in the length, and dams give birth to approximately one calf a year, so D is also the size of the cohort.)

D_k^+ = the number of dams in incubation stage k at time of calving.

O_k = the number of observed BABs with +k dams (i.e. dams who were in incubation stage k, at time of calving; $O_k < D_k^+$.)

B = the number of BABs in the given holding/birth cohort.

Under the null hypothesis of independence the occurrence / non occurrence of BSE occurring in either dams or calves, the expected value of O_k depends on the proportion of D_k^+ dams in the dam population:

$Prob(+\text{k dam} \cap + \text{calf}) = Prob(+\text{k dam}) \cdot Prob(+ \text{ calf})$, which is estimated by

$$\frac{D_k^+}{D}\frac{B}{D}$$

Multiplying both sides by D we have:

$$\hat{E}(O_k) = D_k^+ \frac{B}{D}.$$

The estimated expected value of O_k under independence is compared with the observed value ratio

$$\hat{R}_k = \frac{O_k}{\hat{E}(O_k)}.$$

A ratio significantly greater than one indicates a greater risk of disease amongst calves born to dams in incubation stage k than if there were independence. When several holdings and cohorts are considered the expected values of O_k are computed within holding/cohort categories and summed. Bootstrap confidence intervals for $\hat{R}_k$ did show significantly enhanced risks amongst

calves born to dams in late incubation stages. Only a likelihood approach would yield estimates of parameters related to disease transmission. The next section explains how likelihood terms for BABs were derived.

5.3 A likelihood model used to assess the relative contributions of maternal transmission and genetic susceptibility.

The great number of observations extracted from the CVL data base make a full likelihood approach, such as the one employed in the maternal paired cohort study, computationally intensive. Categorizing observations is more practical given the size of the data set. A model similar to the mechanistic model of the paired cohort study is developed and improved upon by estimating the feed hazard within holding as well as birth cohort and by incorporating a parameter that represents the dam's incubation stage at the time of calving. If feed hazard varies within holding and cohort, then a study of the corresponding case levels amongst exposed and control calves would give insight to the presence of heightened genetic feed susceptibility amongst exposed calves. As in the mechanistic model a cow survival distribution is not employed to explain the relatively few numbers of cases compared with the numbers of infections since estimation of separate distributions within each holding would be necessary. Instead a factor, α, representing the reciprocal of the probability that an infected animal will be observed to become a case, is introduced.

Feed hazard is calculated based on the total number of cases within a holding/cohort category. Since almost all cases occur amongst BABs whose dams are free of disease, BABs that may have been maternally infected were not removed. Let (failure time) random variable T represent the time of infection. We assume that in the absence of slaughtering practices the life span of a cow is 18 years so that $0 \leq T < 18$. Let $\lambda_{ij}(t)$ = the infection hazard in the i^{th} holding within the j^{th} cohort.. Then the cumulative feed hazard is given by

$$\pi_{ij} = \int_0^{18} \lambda_{ij}(t)dt$$

and the failure time survival function in the i^{th} holding, j^{th} cohort is

$$S_{ij}(t) = e^{-\int_0^t \lambda_{ij}(s)ds}$$

which is equal to the probability of not being feed infected by age t for a cow that has not been maternally exposed. The probability of being feed infected within the cow's life time can be estimated from the number of cases and the probability that an infected cow lives to experience onset. If y_{ij} is the probability of feed infection in holding i, cohort j, then,

$$\mathcal{Z}y_{ij} = \Pr\{\text{disease onset} \mid \text{holding } i, \text{ cohort } j\} \cong \frac{B_{H_i \frown C_j}}{D_{H_i \frown C_j}}$$

where, $B_{H_i \frown C_j}$ = number of BABs in holding i, within cohort j

$D_{H_i \frown C_j}$ = size of i^{th} holding within the j^{th} cohort.

$\mathcal{Z}$ = the probability that an infected cow lives to experience onset.

The parameter $\mathcal{Z}$ is estimated separately using the backcalculation model in Ferguson et al. (1997b). Let $\widehat{\mathcal{Z}}$ denote the estimate. Now

$$\widehat{y}_{ij} = \frac{1}{\widehat{\mathcal{Z}}} \frac{B_{H_i \frown C_j}}{D_{H_i \frown C_j}} \cong \Pr\{\text{disease infection} \mid \text{holding } i, \text{ cohort } j\}.$$

We use $\widehat{y}_{ij}$ to estimate the cumulative feed hazard in the absence of maternal exposure, by noting that

$$\mathcal{Z}y_{ij} \cong 1 - e^{-\int_0^{18} \lambda_{ij}(s)ds}.$$

This leads to the estimate of the cumulative feed hazard

$$\pi_{ij} = \int_0^{18} \lambda_{ij}(s)ds \cong -\log(1 - \widehat{\mathcal{Z}}\widehat{y}_{ij}) = \widehat{\pi}_{ij}.$$

Maternal transmission probabilities, estimated within each of six incubation stages, are

represented by:

$$\epsilon_k, \quad k = 1, \ldots, 6 \text{ and } -1,$$

where $k = -1$ means that the dam did not experience disease onset.

As in the mechanistic model the probabilities of infection within the k^{th} incubation stage for herd i and cohort j are given by:

$$\begin{aligned} p_{ij}^{-} &= \Pr\{\text{infection in herd } i, \text{ cohort } j \mid K = -1\} \\ &= 1 - S(18) = 1 - e^{-\pi_{ij}}. \\ p_{kij}^{+} &= \Pr\{\text{infection in herd } i, \text{ cohort } j \mid K = k\} \\ &= \Pr\{\text{maternal infection}\} + \Pr\{\text{feed infection} \cap \text{not maternally infected}\} \\ &= \epsilon_k + (1 - \epsilon_k)(1 - e^{-s\pi_{ij}}). \end{aligned}$$

The parameter s is the susceptibility factor where, $s > 1$ implies greater susceptibility to disease. Unlike the mechanistic model of Chapter 4 the observed random variable is not disease status of calf. Rather it is the dam's incubation stage given the calf's future disease onset. Bayes' Rule is used to invert the above probabilities to conditional probabilities corresponding to the observed data. In addition one needs survival probabilities for infected cows born of infected dams and born of non-infected dams. For the sake of simplicity we will drop the ij subscripts and assume that all calculations of probabilities refer to the i^{th} holding and the j^{th} cohort.

Let,

s^{+} = the probability that an infected cow born of an infected dam survives until disease onset.

s^{-} = the probability that an infected cow born of a non-infected dam survives until disease onset.

D = the size of the holding/cohort.

D_k^{+} = the number of $+k$ dams of calves in the holding/cohort who were in incubation sage k at the time of calving.

D^{-} = the number of dams of the holding/cohort who were not infected at the time of calving.

Using Bayes' Rule

$$\Pr\{K = k \mid \text{future onset in calf}\} =$$

$$\frac{s^+ \Pr\{\text{infection} \mid K = k\} \cdot \Pr\{K = k\}}{s- \Pr\{\text{infection} \cap (K = -1)\} + s^+ \Pr\{\text{infection} \cap (K = 1)\} + \ldots + s^+ \Pr\{\text{infection} \cap (K = 6)\}}$$

$$\cong \frac{s^+ p_k^+ \frac{D_k^-}{D}}{s^- p^- \frac{D^-}{D} + s^+ p_1^+ \frac{D_1^-}{D} + \ldots + s^+ p_6^+ \frac{D_6^-}{D}}$$

$$= \frac{s^+ p_k^+ D_k^+}{s^- p^- D^- + s^+ \sum_{k=1}^{6} p_k^+ D_k^+}, \quad K \neq -1.$$

$$\Pr\{K = -1 \mid \text{future onset in calf}\} = 1 - \sum_{k=1}^{6} \frac{s^+ p_k^+ D_k^+}{s^- p^- D^- + s^+ \sum_{k=1}^{6} p_k^+ D_k^+}.$$

The joint likelihood of all observations within a cohort/holding depends on the ordering of the observations. There are a finite number of dams in each incubation stage and for the r^{th} cow $\Pr\{K = k\}$ is dependent on the dams' incubation stages of calves $1, \ldots, r-1$.

Let K_r be the random variable that assigns the dam's incubation stage at calving to the r^{th} cow. The likelihood of all observations in the i^{th} holding and j^{th} cohort is a product of conditional probabilities summed over all possible orderings:

$$L_{ij} = \sum_{\text{all orderings}} \prod_{r=1}^{\substack{\text{size of} \\ \text{cohort/holding}}} \zeta(K_r = k_r | k_1, \ldots, k_{r-1})$$

Maximum likelihood estimates for parameters s and ϵ_k, $k = -1, 1, \ldots, 6$, are obtained by maximizing

$$L = \sum_{i,j} L_{ij}.$$

The results of the study confirm the presence of an enhanced risk of disease amongst calves born of dams that had the disease or were in a late incubation stage. Estimates of ϵ_k increased as k decreased (closer to onset), suggesting direct maternal transmission. The hypothesis $H_0 : \epsilon_k = \epsilon \ \forall k$ was rejected and the hypothesis $H_0 : \epsilon_k = 0$ $k = 1, 2, \ldots, 6$ and -1 was also rejected. Both suggest that BSE status of a calf is influenced by the incubation stage of its

dam at the time of calving. The maximum likelihood estimate of parameter s was <1. which is not biologically plausible. This may have occurred since calves born to dams in very early incubation stages had a low observed cases over expected cases ratio. One cannot conclude that $s \not> 1$ without information about the genotypes of dams and their calves.

The analysis of dam/calf pairs from the CVL database served as a confirmation of the enhanced risk of exposed calves and of the dependence of risk on the incubation stage of the dam observed during the maternal cohort study. The model does not demonstrate evidence of enhanced genetic susceptibility. The authors conclude that without genotype data one cannot exclude the possibility of a genetic component.

Conclusion 15 *The papers reviewed are intricate examples of the modeling of disease transmission dynamics. Knowledge gained through observation of the BSE epidemic and experimentation with spongiform diseases together with survival analysis and backcalculation techniques form the basis of the BSE infection/onset density (3.2) and provide a beautiful example of natural sciences and mathematical techniques working together. The infection/onset density is applied to estimate infection levels in past and present populations and to predict future case numbers. The density is used once more in the analysis of data from the maternal matched-pair cohort study addressing the question of the existence of maternal transmission of infection from a dam to its calf.*

The BSE epidemic continues to be a major concern of European health officials, the British meat industry and many others. An apparent cluster of four or possibly five cases in the Village of Queniborough in Leicestershire, England was recently discovered and is now being investigated. In the past year (1999) there was a sharp rise in the number of cases of new variant CJD (25 cases) bringing the total number of cases up to 75. In addition a cow born on August 25th, 1996 , twenty five days after further strict controls related to the use of meat and bone meal was enforced, lived to subsequently develop BSE. This suggests a maternal transmission infection route which was shown to be probable in the papers reviewed in this thesis. There is now growing concern that the disease may appear in the United States. In the state of Wisconsin the deer and elk populations are experiencing a serious epidemic where an estimated fifteen percent have chronic wasting disease (CWD), the deer form of BSE. The disease is believed to have been spread in gaming farms which use MBM feed supplements.

There is still a great deal left to learn about prion diseases. The existence of horizontal BSE transmission has not been confirmed. Maternal transmission was statistically shown to be probable but the mechanics of transmission is not understood. The magnitude of the future threat to humans of CJD infections caused by past tainted meat consumption is of course the major concern of all involved. One hopes that continued research and careful tracking of cases will lead to a better understanding of this risk.

References

Anderson, R. M., Donnelly, C.A, Ferguson, N.M., Woolouse. M.E.J., Watt, C.J., Udy, H.J., MaWhinney, S., Dunstan, S.Pl, Southweed, 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 dyanmics and epidemiology of BSE in British cattle. *Nature* **382**, 779-788

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

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

Bruce, M.E., Will, R.G., Ironside, J.W., McConnell, I., Drummond, D., Suttie, A., McCardle, L., Chree, A., Hope, J., Birkett, C., Cousens, S., Fraser, H., Bostock, C.J. 1997 Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent *Nature* **390**, no 6650

Curnow, R.N., Hodge, A. & Wilesmith, J.W. 1997 Analysis of the bovine spongiform encephalopathy maternal cohort study: the discordant case-control pairs. *Appl. Statist.* **46**, 345-349

Donnelly, C.A., Ferguson, N.M., Ghani, A.C., Woolhouse, M.E.J., Watt, C.J. &Anderson, R.M. 1997a The epidemiology of BSE in cattle herds in Great Britain I. Epidemiological processes. Demography of cattle and approachel to control by culling. *Phil. Trans R. Soc. Lond.* **B352**, 781-801

Donnelly, C.A., Ghani A.C., Ferguson, N.M. 1997b Analysis of the bovine spongiform encephalopathy maternal cohort study: evidence for direct maternal transmission *Appl. Statist.* 1997b **46**, 321-344

Donnelly, C.A., Ferguson, N. M. ., Ghani A. C., Wilesmith, J.W., Anderson, R.M. 1997c Analysis of dam-calf pairs of BSE cases: confirmation of maternal risk enhancement *Proc. R. Soc. Lond.* B **264**, 1647-1656

Ferguson, N.M., Donnelly, C.A., Woolhouse, M.E., Anderson, R.M. 1997a A genetic interpretation of heightenend risk of BSE in offspring of affected dams *Proc. R. Soc. Lond.* B **264**, 1445-1455

Ferguson, N.M., Donnelly, C.A., Woolhouse, M.E.J., Anderson, R.M. 1997b The epidemiology of BSE in cattle herds in Great Britain. II. Model construction and analysis of transmission dynamics *Phil. Trans. R. Soc. Lond. B* **352**, 803-838

Gore, S.M., Gilks, W.R. & Wilesmith, J.W. 1997 Analysis of the bovine spongiform encephalopathy maternal cohort study: exploratory analysis. *Appl. Statist.* **46**, 305-320

Kalbfleisch, J.D., Prentice, R.L., The Statistical Analysis of Failure Time Data *John Wiley & Sons*, 1980

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

Prusiner, S.B., 1995, The Prion Diseases *Scientific American* **272** 48-57

Sanathanan, L. (1972), "Estimating the Size of a Multinomial Population", *Annals of Mathematical Statistics*, **43**, 142-152