Predictability of the UK Variant Creutzfeldt-Jakob Disease Epidemic

Jerome N. Huillard d'Aignaux,* Simon N. Cousens, Peter G. Smith

Back-calculation analysis of the variant Creutzfeldt-Jakob disease epidemic in the United Kingdom is used to estimate the number of infected individuals and future disease incidence. The model assumes a hazard of infection proportional to the incidence of bovine spongiform encephalopathy in the United Kingdom and accounts for precautionary control measures and very wide ranges of incubation periods. The model indicates that current case data are compatible with numbers of infections ranging from a few hundred to several millions. In the latter case, the model suggests that the mean incubation period must be well beyond the human life-span, resulting in disease epidemics of at most several thousand cases.

Variant Creutzfeldt-Jakob disease (vCJD) is caused by an agent that is currently indistinguishable from that responsible for bovine spongiform encephalopathy (BSE) in cattle. However, 5 years after the identification of vCJD, great uncertainty remains over how many individuals have been infected with the agent and how many of these individuals will go on to develop clinical disease (*1*–*5*).

In the absence of a test for infection, one approach to estimating the number of infected individuals is provided by back-calculation, a statistical technique developed in the context of the HIV/AIDS (human immunodeficiency virus/acquired immunodeficiency syndrome) epidemic (*6*, *7*). This approach makes use of the fact that the number and timing of cases of disease that occur depend on three factors: (i) how many people were infected, (ii) when they were infected, and (iii) how long it takes from infection for disease to become apparent—the incubation period. To use this approach to estimate the number of individuals infected with the vCJD agent, it is necessary to make assumptions about when people were exposed to infection and how long it takes them to develop disease. Previous work has shown that the estimated number of infections/cases produced by this approach is very sensitive to the assumptions made about the incubation period distribution (*2*, *8*, *9*).

We have developed a family of back-calculation models (*10*) to explore the prevalence of infection with the vCJD agent and the incidence of clinical vCJD in the UK. The main features of these models are as follows: (i) The hazard of infection was assumed to have been proportional to the incidence of BSE (*11*). We did not consider onward, human-to-human transmission of the infectious agent. (ii) The incubation period of the disease was assumed to follow an offset generalized F distribution, which has five parameters. We also investigated the lognormal, Weibull, and gamma distributions as special cases of this distribution (*12*). (iii) We assumed that the incubation period was independent of age at infection (*13*). (iv) The model was restricted to the 40% (approximately) of the UK population assumed to be methionine homozygous at codon 129 of the prion protein (PrP) gene. (All cases of vCJD identified to date have been of this genotype.) (v) To minimize the impact of reporting delays, we fitted the models to the data on the 82 cases with onsets before 2000 that had been identified by 31 December 2000.

The back-calculation model had seven parameters in total [five for the incubation period distribution, one for the hazard of infection, and one for the effect of the specified bovine offals (SBO) ban in 1989]. The model was fitted by the maximum likelihood method, assuming a Poisson likelihood. Because of a very severe parameter identifiability problem, we estimated the incubation period distribution parameters for fixed values of the hazard of infection (corresponding to total numbers of infections ranging from 100 to 12 million) and for fixed effects of the SBO ban ranging from 0 to 90%. Allowing a very flexible incubation period distribution (offset generalized F), we found that the cases observed to date were almost equally compatible with any number of infections up to several millions. However, when a very large number of infections was considered, the model indicated that the average incubation period was likely to be extremely long and, in most instances, well beyond the normal human life-span. As a result, the corresponding epidemic sizes (clinical cases) lay within a much narrower range, from a few hundred to a few thousand cases (Table 1).

When making stronger assumptions about

London School of Hygiene and Tropical Medicine, Infectious Disease Epidemiology Unit, Keppel Street, London WC1E 7HT, UK.

*To whom correspondence should be addressed. E-mail: jerome.huillard@lshtm.ac.uk

Table 1. Estimates of numbers of infections, numbers of clinical cases, incubation period parameters, and prediction intervals by assumed incubation period (IP) distribution. Values shown are calculated assuming a SBO ban efficiency of 80%.

Incubation period distribution (number of parameters)	−Log likelihood (+constant term)	Estimated number of infections (expected number of clinical cases)	Median IP (90% IP) (years)	Upper limit on the number of infections for three levels of confidence (corresponding expected number of clinical cases)			Expected annual numbers of cases (upper 95 and 99% prediction limits)*		
				95%	99%	99.9%	2005	2010	2020
Offset Generalized F (5)	−134.50	Unrestricted (250–3,000‡)	9.8–∞ (16.5–∞)†	Unrestricted (3,000‡)	Unrestricted (3,000‡)	Unrestricted (3,000‡)	16 (50,65)	12 (70,130)	8 (70,165)
Generalized F (no offset) (4)	−133.60	250 (200)	11.6 (16.8)	1,000 (900)	7,000 (2,500)	Unrestricted (30,000‡)	8 (40,65)	2 (70,115)	1 (65,130)
Offset Lognormal (3)	−134.45	400 (305)	16.8 (63.1)	25,000 (2,000)	350,000 (5,300)	Unrestricted (13,000‡)	14 (50,70)	9 (70,110)	5 (105,200)
Offset Weibull (3)	−134.46	440 (330)	20.0 (67.9)	25,000 (4,000)	260,000 (12,000)	Unrestricted (35,000‡)	16 (55,75)	10 (75,150)	6 (120,380)
Offset Gamma (3)	−134.43	380 (346)	15.7 (38.5)	6,000 (2,500)	80,000 (8,000)	Unrestricted (40,000‡)	20 (40,60)	8 (60,105)	3 (40,180)

*For the offset generalized F distribution, the prediction limits have been calculated assuming the most likely number of infections to have been 190,000, resulting in an expected total of 500 clinical cases. †Infinity symbol (∞): Well beyond human life-span. ‡Point estimate of the number of clinical cases for 12 million infections.

the form of the incubation period distribution (i.e., reducing the number of parameters from five), the range for the number of infections with which the observed cases are compatible becomes narrower. For example, using a simple generalized F distribution led to a point estimate of the total number of infections of a few hundred, with an upper 95% confidence limit of about 1000 (and upper 99% confidence limit of 7000). Using an offset lognormal distribution again led to a point estimate of the total number of infections of a few hundred. The predicted course of the vCJD epidemic was calculated under different assumptions about the incubation period distribution (*14*). No matter which incubation period distribution is used, the point estimates obtained from the model suggest that the epidemic of cases of vCJD is very close to its peak. However, the expected numbers of cases corresponding to the upper limits of infection (*14*) indicate that the data are also compatible with an epidemic whose peak, many years hence, is determined by mortality among infected individuals from competing causes. Table 1 also presents approximate prediction intervals (*15*) for annual numbers of cases at different times in the future. These indicate that the annual incidence of vCJD is unlikely ever to be much more than 100 cases (*14*).

None of our models suggest that the number of primary cases of vCJD in methionine homozygotes is likely to be more than a few thousand, even though the number of primary infections could be anything from a few hundred to many thousands or even millions. In interpreting these results, and extrapolating them to other codon 129 genotypes, we must bear in mind our model assumptions. Our key finding that, regardless of the number of infections that have occurred, the number of clinical cases is unlikely to exceed a few thousand (in any one genotype) is sensitive to a number of assumptions.

First, we have assumed that in codon 129 methionine homozygotes, the incubation period for vCJD has a unimodal distribution. This is a key assumption that is open to question (*16*). In mice, there are genetic factors lying outside the coding region of the PrP gene that have an important influence on the incubation period of transmissible spongiform encephalopathies (*17–20*). It is possible, therefore, that among human codon 129 methionine homozygotes there are other, presently unknown genetic factors that influence the vCJD incubation period. We have used the generalized F distribution, which can take a wide range of unimodal forms. If, across the methionine-homozygous population, the mixture of other genetic factors affecting incubation period results in an overall incubation period distribution that is close to unimodal, we would be confident that, broadly, our findings with respect to the numbers of clinical cases hold. If, however, the overall incubation period distribution is strongly multimodal, there might be many more clinical cases of vCJD than our models predict. If the latter is the case, then the development of reliable back-calculation models will be possible only when the relevant genetic factors have been identified and measured in the population. Strong multimodality is most likely to apply if only a small number of other genetic factors are involved and there was little variation between infected individuals in the infecting dose to which they were exposed.

Second, we have assumed that the incubation period distribution does not vary with age at infection. Experimental evidence in mice indicates that, for a fixed dose, incubation period does vary with age at inoculation (*21*). However, this variation is small, with young mice having incubation periods 7 days longer than older mice, compared with mean incubation periods of several hundred days. If vCJD infections occurred through diet, as we have implicitly assumed, infected individuals may have been exposed to a very wide range of infectious doses whose impact on incubation period is likely to dwarf any small age effects.

Third, to extrapolate from codon 129 methionine homozygotes to other genotypes, we need to assume that across codon 129 genotypes the relation between the mean and the variance of the incubation period distribution does not vary greatly. If other genotypes have longer mean incubation periods but with lower variance, then we might observe larger numbers of cases in these genotypes. It is, however, unusual for the variance of a distribution to decrease as the mean increases. If this is not the case, then to extend our results to include all genotypes one could, as a worst-case scenario, multiply our predictions by about 2.5 to obtain a figure for the whole population.

A further assumption of the model is that infection was essentially through diet and that the amount of infectivity consumed in food during any given period was proportional to the number of BSE cases occurring up until 1996. In the absence of ongoing human-to-human transmission of the vCJD agent, our findings are likely to be much less sensitive to this assumption than they are to the assumptions about incubation period.

The upper limits of our estimates differ from those presented by Ghani *et al.* (*1*). These authors used simulation to identify a range of scenarios compatible with the actual observed incidence. One advantage of this approach is that it allows the incorporation into the model of several parameters that could not be estimated. However, it does not enable any probability statement to be made about the coverage of the range of epidemics that it produces, and running more simulations can only increase the range of scenarios that are plausible. We believe the most likely explanation for the different ranges of cases coming out of our work and that of Ghani *et al.* is that the coverage probabilities of those intervals are different.

Our models suggest that the number of primary cases of vCJD in methionine homozygotes is unlikely to exceed a few thousand, but that considerably greater uncertainty surrounds the number of primary vCJD infections that have occurred. Whether a few hundred or many more people have been infected has important consequences for public health and, in particular, for the risk of secondary transmission (*22*). If secondary transmission does occur, the mean incubation period in secondary cases may be much shorter than in primary cases (*23*). In the absence of a reliable test for asymptomatic infection, considerable uncertainty about the number of infected individuals may remain for a number of years.

References and Notes

1. A. C. Ghani, N. M. Ferguson, C. A. Donnelly, R. M. Anderson, *Nature* **406**, 583 (2000).
2. S. N. Cousens, E. Vynnycky, M. Zeilder, R. G. Will, P. G. Smith, *Nature* **385**, 197 (1997).
3. M. Balter, *Science* **289**, 1663 (2000).
4. D. Bonn, *Lancet* **356**, 570 (2000).
5. J. Jones, *Br. Med. J.* **321**, 469B (2000).
6. R. Brookmeyer, M. H. Gail, *Lancet* ii, 1320 (1986).
7. R. Brookmeyer, *Science* **253**, 37 (1991).
8. A. C. Ghani, N. M. Ferguson, C. A. Donnelly, T. J. Hagenaars, R. M. Anderson, *Proc. R. Soc. Lond. Ser. B* **265**, 2443 (1998).
9. J. D. Cooper, S. M. Bird, D. De Angelis, *J. Epidemiol. Biostat.* **5**, 209 (2000).
10. The incidence of vCJD in the UK was modeled by an age-structured back-calculation equation:

$$n(t,a\,|\,\rho,\theta) = \int_{t_i}^{t_f}\int_{u=0}^{a} i\,(s,u\,|\,\rho) \times f\,(t-s\,|\,\theta)S(u,a)\,du\,ds \qquad (1)$$

where $n(t,a)$ represents the number of cases of age a occurring at time t, $i(s,u|\rho)$ represents the number of individuals aged u newly infected at time s, $f(t|\theta)$ represents the probability density function of the incubation period, and $S(u,a)$ represents the proportion of individuals in the population aged u who survive to age a. The function $S(u,a)$ can be derived from census data and is not estimated within the back-calculation model itself. Age-specific mortality rates for causes of death other than vCJD were obtained from national census data. ρ and θ are the unknown parameters of the hazard of infection and the incubation period distribution, respectively; t_i and t_f define the start and end of exposure of the population. In the calculation, t_i was set to the estimated year of the first case of BSE (1982) and t_f was set to 1996.
11. A plausible assumption regarding the hazard of vCJD infection over time is that it was approximately proportional to the incidence of clinical cases of BSE in cattle, with the proportionality factor varying over time as precautionary control measures were introduced to prevent highly infectious material entering the human food supply. We assumed that the proportionality factor is a step function [denoted $h(s|\rho)$], which is equal to ρ up to 1989, with a partial step down at the end of 1989, when the specified SBO ban was introduced, and a further step down to zero in

1996, after the identification of variant CJD. Additionally, we assumed some degree of BSE underreporting by using maximum estimates of underreporting based on a model of the BSE epidemic published previously (*24*).

12. A positive random variable T is said to have a generalized F distribution with μ and σ as location and scale parameters, respectively, and s_1, s_2, as shape parameters, if $W = [\ln(T) - \mu]/\sigma$ has an F distribution with $2s_1$ and $2s_2$ degrees of freedom. The density of W is given by
$$f(w,s_1,s_2) = \left(\frac{s_1 e^w}{s_2}\right)^{s_1}\left(1 + \frac{s_1 e^w}{s_2}\right)^{-(s_1+s_2)} B(s_1,s_2)^{-1},$$
where $B(x,y)$ is the beta function [i.e., $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$]. The generalized F distribution includes many commonly used distributions as special cases, such as the lognormal, Weibull, gamma, and log-logistic distributions.

13. The quantity $i(s,u|\rho)$ in Eq. 1 can be expressed as
$$i(s,u|\rho) = r(s,u|\rho) \times \exp[-R(s,u|\rho)] \times N(s,u) \quad (2)$$
where $N(s,u)$ is the number of individuals aged u at time s, $r(s,u|\rho)$ is the hazard of infection for these individuals (rate of infection at a given time conditional on having survived uninfected up to that time), and $R(s,u|\rho)$ is the cumulative risk of infection in these individuals up to time s [i.e., $R(s,u) = \int_0^s r(t,u|\rho)dt$]. To take into account variation in exposure/susceptibility to infection with age, the hazard of infection, $r(s,u)$, is allowed to vary with age (u). To estimate the parameters of the model, we must make some assumptions about the way in which $r(s,u)$ varies with age (u). We first made the simplifying assumption that $r(s,u|\rho)$ can be rewritten as $h(s|\rho) \times \phi(u)$; i.e., we assume that the age dependency of the hazard function does not vary over time (likely to be an oversimplification). This would yield the following expression for the distribution of infection over time: $i(s,u|\rho) = h(s|\rho) \times \phi(u) \times \exp[-\phi(u) \times H(s)] \times N(s,u)$, where $H(s)$ stands for the cumulative value of $h(s)$. We further simplified our equation, by approximating the above expression by $i(s,u|r) = h(s) \times \exp[-H(s)] \times k(u) \times N(s,u)$, with $k(u)$ chosen in such a way that the age distribution of cases in the model matches the observed age distribution of cases. In this expression, $h(s)$ was assumed to take the form of a step function [see (*2*)].

14. Supplementary Web material is available on *Science* Online at www.sciencemag.org/cgi/content/full/1064748/DC1.
15. To obtain approximate 95% prediction intervals for the number of cases in a given year, we fitted models in which the annual incidence for that year was fixed, and we compared the likelihood of this model with the maximum likelihood until a log likelihood ratio of 1.92 was obtained.
16. S. E. Lloyd *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6279 (2001).
17. R. I. Carp, R. C. Moretz, M. Natelli, A. G. Dickinson, *J. Gen. Virol.* **68**, 401 (1987).
18. A. Bossers, B. E. Schreuder, I. H. Muileman, P. B. Belt, M. A. Smits, *J. Gen. Virol.* **77**, 2669 (1996).
19. L. Cervenakova *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13239 (1998).
20. K. Manolakou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7402 (2001).
21. A. R. McLean, C. J. Bostock, *Philos. Trans. R. Soc. London Ser. B* **355**, 1043 (2000).
22. A. Frosh, R. Joyce, A. Johnson, *Br. Med. J.* **322**, 1558 (2001).
23. C. I. Lasmezas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4142 (2001).
24. R. M. Anderson *et al.*, *Nature* **382**, 943 (1996).
25. This work was supported by the UK Department of Health. The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health. We thank the Creutzfeldt-Jakob Disease Surveillance Unit and Pr. R. Will for providing the data analyzed in this study.

25 July 2001; accepted 11 October 2001
Published online 26 October 2001;
10.1126/science.1064748
Include this information when citing this paper.

REPORTS

The Influence of the Proinflammatory Cytokine, Osteopontin, on Autoimmune Demyelinating Disease

Dorothée Chabas,[1]* Sergio E. Baranzini,[2]* Dennis Mitchell,[1] Claude C. A. Bernard,[3] Susan R. Rittling,[4] David T. Denhardt,[4] Raymond A. Sobel,[5] Christopher Lock,[1] Marcela Karpuj,[1,2] Rosetta Pedotti,[1] Renu Heller,[6]† Jorge R. Oksenberg,[2]† Lawrence Steinman[1]†‡

Multiple sclerosis is a demyelinating disease, characterized by inflammation in the brain and spinal cord, possibly due to autoimmunity. Large-scale sequencing of cDNA libraries, derived from plaques dissected from brains of patients with multiple sclerosis (MS), indicated an abundance of transcripts for osteopontin (OPN). Microarray analysis of spinal cords from rats paralyzed by experimental autoimmune encephalomyelitis (EAE), a model of MS, also revealed increased OPN transcripts. Osteopontin-deficient mice were resistant to progressive EAE and had frequent remissions, and myelin-reactive T cells in $OPN^{-/-}$ mice produced more interleukin 10 and less interferon-γ than in $OPN^{+/+}$ mice. Osteopontin thus appears to regulate T helper cell–1 (T_H1)–mediated demyelinating disease, and it may offer a potential target in blocking development of progressive MS.

Multiple sclerosis (MS) is often characterized by relapsing episodes of neurologic impairment followed by remissions. In about one-third of MS patients, this disease evolves into a progressive course, termed secondary progressive MS (*1*). In a minority of patients, progressive neurologic deterioration without remission occurs from the onset of disease, and this is called primary progressive MS. The pathophysiologic and genetic causes underlying primary versus secondary progressive MS remain unclear (*2–4*).

Osteopontin, also called early T cell activation gene-1 (*5*, *6*), has pleiotropic functions (*7–9*), including roles in inflammation and in immunity to infectious diseases (*8*). OPN costimulates T cell proliferation (*8*) and is classified as a T helper cell–1 (T_H1) cytokine, because of its ability to enhance interferon-γ (IFN-γ) and interleukin 12 (IL-12) production, and to diminish IL-10 (*10*). We investigated a role for OPN in MS and an experimental model for MS, experimental autoimmune encephalomyelitis (EAE) in mice.

Initially, we set out to identify gene transcripts involved in the inflammatory response that might be increased in the central nervous system (CNS) during active EAE and that returned to normal when EAE was successfully treated after the onset of paralysis. Customized oligonucleotide microarrays were produced to monitor transcription of genes involved in inflammatory responses (*11–14*). These initial microarray experiments showed that osteopontin transcripts were elevated in the brains of rats with EAE but not in brains of rats protected from EAE. Details of these experiments are available at *Science* Online (*14*).

In parallel, we performed high-throughput sequencing of expressed sequence tags (ESTs), using nonnormalized cDNA brain libraries (*15–17*), generated from MS brain lesions and control brain (*18*). Using this protocol, the mRNA populations present in the brain specimens are accurately represented, enabling the quantitative estimation of transcripts and comparisons between specimens (*18*) [Table 1 ,and Web table 1 (*14*)]. Molecular mining of two sequenced libraries and their comparison with a normal brain library, matched for size and tissue type and constructed with an identical protocol, revealed that OPN transcripts were frequently detected and were exclusive to the MS mRNA population, but not found in control brain mRNA (Table 1).

We sequenced more than 11,000 clones

[1]Department of Neurology and Neurological Sciences, Beckman Center for Molecular Medicine, B002, Stanford, CA 94305, USA. [2]Department of Neurology, University of California at San Francisco School of Medicine, San Francisco, CA 94143, USA. [3]Neuroimmunology Laboratory, Department of Biochemistry, La Trobe University, Bundoora, Victoria, 3083 Australia. [4]Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ 08854, USA. [5]Department of Pathology (Neuropathology), Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA.

*These authors contributed equally to this work.
†These senior authors contributed equally to this work.
‡To whom editorial correspondence should be addressed. E-mail: steinman@stanford.edu