参考資料3－1

（原文）

May 2006

Peer Review of the Estimation of Bovine Spongiform Encephalopathy (BSE) Prevalence in the United States

Final Report

Prepared for

U.S. Department of Agriculture
Animal and Plant Health inspection System
Riverdale, MD

Contracting Officer's Technical Representative

Chuanfa Guo
U.S. Department of Agriculture
Food Safety and Inspection Service
Washington, DC

Prepared by

Sumeet Patil
RTI International
Health, Social, and Economics Research
Research Triangle Park, NC 27709

RTI Project Number 0208893.023

RTI Project Number
0208893.023

Peer Review of the Estimation of Bovine Spongiform Encephalopathy (BSE) Prevalence in the United States

Final Report

May 2006

Prepared for

U.S. Department of Agriculture
Animal and Plant Health inspection System
Riverdale, MD

Contracting Officer's Technical representative

Chuanfa Guo
U.S. Department of Agriculture
Food Safety and Inspection Service
Washington, DC

Prepared by

Sumeet Patil
RTI International
Health, Social, and Economics Research
Research Triangle Park, NC 27709

---

*RTI International is a trade name of Research Triangle Institute.

Contents

Executive Summary

The Office and Management and Budget (OMB) requires a peer review for important scientific information to ensure the quality of scientific and technical research and guide improvements in the draft before federal agencies disseminate it (OMB, 2004). The Animal and Plant Health Inspection Service (APHIS) is interested in conducting a peer review of their BSE prevalence estimation methods based on Bovine Spongiform Encephalopathy (BSE) surveillance data. These data and analysis results can inform the surveillance plan and also provide inputs into the policy and decision making related to BSE risk mitigation efforts. APHIS requested RTI International's (RTI) support for conducting a peer review conforming to OMB's guidelines (OMB, 2002; 2004) under RTI's task order contract with the Food Safety and Inspection Service (FSIS).

Specifically, APHIS needs review of two prevalence estimation methods/models (BsurvE model and a Bayesian analysis) and a report describing the approach, data, assumptions, and conclusions drawn from the analysis. RTI identified three experts and conducted the peer review according to the statement of work. We present these three reviews in this report.

All reviewers agree with the estimated BSE prevalence of 1 in 1,000,000 live cattle in the United States. Although each reviewer made suggestions to improve the model, they all acknowledge that the prevalence estimate will likely be robust to the suggested changes given the detailed sensitivity analysis performed by APHIS. A couple of reviewers have suggested sensitivity analysis of additional parameters and assumptions to increase confidence in the results. Given that the prevalence estimate is robust, reviewers focused on other aspects of the model, such as using alternative input assumptions, acknowledging or further discussing the limitations in the surveillance samples, suggesting sensitivity analysis for additional parameters, using appropriate technical language and specification in the report, and clarifying other specific parts of the report.

In terms of suitability, transparency, and robustness of the models, all reviewers found the BsurvE and Bayesian Birth Cohort (BBC) models statistically and epidemiologically sound. For example, two reviewers clearly note that *the BsurvE model has received a favorable review by the European Food Safety Authority and is judged to be a suitable tool to estimate the BSE prevalence in a country*. All reviewers could satisfactorily reproduce the results by running the BsurvE model. The BBC model that uses information on the 1997 feed ban was also found suitable, transparent, and robust. However, a couple of reviewers use the fact that BBC results are derived from BsurvE model results to make two related inferences that the BBC results are equally important to the BsurvE model and that it is no surprise that the BBC results are correlated with the BsurvE results. All reviewers could run the BBC model using the WinBUGS program, but one of them has identified a need for more explanation of the code in the report.

In terms of completeness of the models, all reviewers agree that APHIS has appropriately considered the key factors, and the models appear to be appropriately parameterized. All reviewers are in general agreement with the assumptions especially because these assumptions are later evaluated in a sensitivity analysis and found to have little or no impact on the results. However, for a few input assumptions, the reviewers needed a more thorough treatment.

A couple of reviewers argued over the accuracy of using the United Kingdom data to obtain the relative rate of decline in incidence in the United States in the BBC model, but they did not make a strong recommendation against the current APHIS assumption. One reviewer argued that the rate of incidence decline in the U.S. may more closely approximate incidence decline represented at a later stage in the United Kingdom epidemic. Second reviewer suggested an extension to the BBC model where the yearly adjustment factor will be a model parameter with priors derived from the U.K. data.

Probably the most significant comment on the analysis is to conduct sensitivity analysis of additional parameters to evaluate their effect on the results. For example, two reviewers strongly suggested evaluating the effect of age distribution of the sample because sampling strategies were very different between early and later parts of the surveillance period. In addition, reviewers identified a need to evaluate the effect of the sample set's representative, the exit probabilities for infected animals, and the incubation period distribution. It is likely that these parameters may have negligible effect on the robustness of the results; nevertheless, such uncertainty analyses are recommended for the sake of completeness and further increasing the confidence on the results.

1 Background and Objective

RTI International (RTI) coordinated external peer review of the estimation of Bovine Spongiform Encephalopathy (BSE) prevalence in the United States as requested by the U.S. Department of Agriculture's Food Safety and Inspection Service (USDA, FSIS) and Animal and Plant Health Inspection Service (APHIS) under task order 0208893.023. In this report, we present the background information about the peer review, describe the review process, list key questions or the charge to the reviewers, and include the three peer review reports.

APHIS has conducted BSE surveillance in the United States cattle herd with increasing intensity since 1990. Beginning in June 2004, these efforts were enhanced significantly, with a goal of obtaining as many samples as possible from the targeted population in a 12- to 18-month period. Data from these surveillance efforts have been analyzed to estimate the prevalence of BSE in the United States. The conclusions of this analysis will serve several purposes. USDA will use data from the enhanced surveillance effort to design a maintenance surveillance plan. Understanding the prevalence of disease will contribute to science-based policy and regulatory decisions on BSE risk mitigation measures. The prevalence estimates will also contribute to trade discussions and efforts to regain export markets. Therefore, this analysis and the report are scientifically important and deserve an external peer review as per the guidelines by the Office and Management and Budget (OMB) (2004).

Peer review is an important process that can help ensure that the quality of scientific information meets the standards of the technical community, and it can help strengthen and clarify the analysis. APHIS requested RTI's support in conducting a formal and independent peer review of their BSE prevalence estimation methods conforming to OMB's guidelines for peer review and quality of information (OMB, 2002, 2004).

2 Methods Used to Estimate BSE Prevalence in the United States

The analysis used data obtained from BSE surveillance efforts over a 7-year period to estimate the prevalence of BSE in the United States. The prevalence provides an estimate of the amount of disease that is present in the cattle population at a single point in time. It represents the total number or proportion of infected cattle in the adult cattle population, including those that have undetectable levels of infection. Methods of estimation for BSE prevalence have been developed that account for population differences and demographics. Two methods were used in this analysis to estimate prevalence as described next. In addition, uncertainty analyses were conducted on several parameters with potential to have significant effects on the models used.

The BSurvE model was used as one method for prevalence analysis. The BSurvE model is a recently developed method that was designed to use cattle demographics, rate of exit from the populations, and known disease characteristics. This model directly estimates the likelihood of finding BSE and assigns a value to each surveillance sample compared with the information that might be gained from a random sample.

A Bayesian Birth Cohort (BBC) method was used to incorporate the information about the presence or efficacy of disease control measures, such as a feed ban, to estimates prevalence in the standing cattle population. Given the knowledge of an effective feed ban, one might arrive at informed conclusions about the prevalence at some future time. Bayesian methods of analysis allow mathematical inclusion of prior knowledge with surveillance information to give a final estimate reflecting the total of available information.

Uncertainty analyses were conducted on five factors that could have potential impact on the outcome of the estimates. The final conclusions are based on the results of both methods of estimating prevalence and the results of the uncertainty analyses.

3 Description of Review Process

RTI conducted the review process in accordance with the OMB guidelines (OMB, 2004). The review process consisted of selecting the reviewers, explaining the scope of the review, facilitating the review, and consolidating the reviews in a single report.

First, we selected three peer reviewers based on their expertise. We initially identified 14 potentially suitable reviewers after understanding the background and objectives of the peer review from FSIS and APHIS. Subsequently, we finalized three reviewers based on their availability and the desired overlap of expertise in the science of BSE, particularly related to animal health; mathematical modeling and Bayesian theory; and estimation procedures for determining disease prevalence. We also considered conflict of interest in the selection process.

Second, we explained the scope of the review in terms of the charge to the reviewers prepared by APHIS. Along with the charge, RTI provided the report, BsurvE Excel model, and WinBUGS program to run the BBC model. The charge consists of five set of questions as described in Section 4.

Third, RTI communicated and clarified any questions the reviewers had about the scope of the review or the analysis itself. We communicated the progress and status of the review to APHIS and FSIS regularly and ensured that the reviewers were meeting the objectives of the peer review. We also ensured that the reviewers describe possible ways to address their concerns instead of only describing the concerns.

Finally, we consolidated the three reviews in a single report. In this report, we provide brief background information on the analysis that was peer reviewed, the review process, and peer review reports. We attach the original peer reviews in Section 6.

To maintain the integrity of the reviews, we present the reviews as separate chapters in this report instead of consolidating the comments by the charge questions. Each reviewer focused on different aspects of the charge questions depending on his area of expertise, and their reporting formats and writing styles also differ. Therefore, reading each review separately can help readers better understand their comments. We have corrected minor typographical errors and slightly reformatted their reports to ensure a minimum level of uniformity of presentation in this report.

4 Charge to the Peer Reviewers

APHIS asked the reviewers to focus and structure their writings on the specific questions listed below.[1]

1. Given that the goal of the report is to determine the prevalence, examine the agency's conclusions of this analysis. Did the agency use the appropriate models and was the data valid? Focus on the suitability of the methods, the transparency of the approach, and the robustness of the results.
2. The goal of the analysis, as stated in the document, is as follows:

    The United States has conducted BSE surveillance with increasing intensity since 1990, including an enhanced effort following the identification of a Canadian cow that tested positive in 2003. The enhanced BSE surveillance program in the United States is summarized in a separate report: USDA, APHIS, CEAH, Summary of Enhanced BSE Surveillance, 4-6-06. This analysis uses surveillance data that have been collected over the seven-year period prior to March 17, 2006 to estimate the prevalence of BSE in the United States.

    This information will help guide and support any future requests for consideration of the overall BSE status of the United States in line with international guidelines adopted by the World Organization for Animal Health (OIE).

---

[1]The reviewers also provided general comments that can inform the six specific questions. However, they kept the general comments separate from the specific questions as per the request by RTI.

Within the context of the study goal, is the model complete and does it make sense? Does it capture the key factors that influence estimation of prevalence and the interaction of these factors in a real-world setting? To the extent that it does not, what parameters are missing or incorrectly specified? Are there alternative models that would be of value to investigate? Are the places in which the model departs from reality clearly defined?

3. Examine the assumptions about input parameters. Are they reasonable within the context of the goal of the analysis? Do the assumptions reflect reasonable most likely estimations of the parameters?

   Note: Monte Carlo modeling is part of the WIN BUGS software that was used, and the distributions chosen used a Poisson distribution that is explicit in the WINBUGS code (included in the document.)

4. Are the conclusions robust to the uncertainties in both the model and the input data? Are those uncertainties described completely? How robust are the conclusions? Are these implications described transparently?

5. Are the results of the analysis correct given the models used? Check the code and formulas by running the program to make sure the model does what it is intended to do. Given the models used, determine if the agency came out with the right answer.

5 References

The Office of Management and Budget (OMB). 2002. "Information Quality Guidelines." The Office of Information and Regulatory Affairs, the Office of Management and Budget, the Executive Office of the President, Washington, DC. October 1, 2002.

The Office of Management and Budget (OMB). 2004. "Final Information Quality Bulletin for Peer Review." A Memorandum for Heads of Departments and Agencies. M-05-03. The Office of Management and Budget, the Executive Office of the President, Washington, DC. December 16, 2004.

6 Peer Review Reports

Review of "An Estimate of the Prevalence of BSE in the United States"

by

Paul S. Morley, DVM, PhD, DACVIM
Associate Professor, Epidemiology and Biosecurity Director of Biosecurity, Veterinary Medical Center
Associate Director, Animal Population Health Institute Colorado State University

Dr. Morley is an Associate Professor in the Clinical Sciences Department and Associate Director of the Animal Population Health Institute in the College of Veterinary Medicine and Biomedical Sciences at Colorado State University. He is also the Director of Biosecurity for the James L. Voss Veterinary Teaching Hospital at CSU. In addition, while on sabbatical leave in 2005, Dr. Morley worked as a feedlot and research consultant at Feedlot Health Management Services in Okotoks, Alberta, Canada. Dr. Morley teaches epidemiology and infection control in the undergraduate, veterinary, and graduate programs at CSU, and maintains an active research program. Major focuses for his professional and research activities include analytical epidemiology related to infectious diseases of livestock, biosecurity issues important to agriculture and veterinary medicine, analysis of data related to beef production, investigating the ecology of antimicrobial resistance in animals, and identifying effective pre-harvest approaches for food-safety. Dr. Morley is the president elect of the Association for Veterinary Epidemiology and Preventive Medicine, and serves as the Associate Editor for the Journal of Veterinary Internal Medicine with emphasis on internal medicine in large animal species. He is also an active member of several professional organizations, including the American College of Veterinary Internal Medicine, the Association for Veterinary Epidemiology and Preventive Medicine, the Veterinary Infection Control Society, and the Academy of Veterinary Consultants. Dr. Morley served on the faculty at The Ohio State University for three years before joining the faculty at Colorado State University in 1998. He is a native of Nevada and received baccalaureate degrees and his veterinary medical degree from Washington State University. After graduating from WSU, he served as an intern and resident in the area of large animal internal medicine at the Western College of Veterinary Medicine in Saskatoon, Saskatchewan. He received his doctorate from that institution studying the epidemiology of infectious respiratory disease in horses. He is a Diplomate of the American College of Veterinary Internal Medicine. Dr. Morley is a recognized authority on epidemiology and infection control in animal populations and has consulted on infection control and biosecurity issues at several veterinary colleges in North America as well as for several intensive animal production facilities. He is an author on more than 75 peer reviewed scientific publications in addition to several book chapters.

Introduction: As requested in directions forwarded to me by RTI and in the charge to reviewers prepared by the USDA:APHIS, this document contains general and specific comments regarding the report titled, "An Estimate of the Prevalence of BSE in the United States," dated April 27, 2006 (hereafter referenced as "the report"), as well as for the two mathematical models used in preparation of the report. I have reviewed this primary document as well as supporting documents found on the USDA's web page (e.g., "Summary of Enhanced BSE Surveillance in the United States," dated April 27, 2006). I have also reviewed the spreadsheet containing the USDA BSE surveillance results and the BSurvE stochastic model (BsurvE_0603_USData.4.26.06.xls) in addition to the output from the WIN BUGS model that is obtained using the code provided in Appendix B of the report. I have also reviewed documentation provided with the BSurvE spreadsheet model, the EFSA Scientific report (2004) 17, 1-6 on the assessment of the BSE surveillance model (BSurvE), as well as information from the OIE Terrestrial Animal Health Code relative to BSE and surveillance (specifically article 2.3.13—Bovine Spongiform Encephalopathy, Appendix 3.8.4—Surveillance For Bovine Spongiform Encephalopathy), and Appendix 3.8.1—General guidelines for animal health surveillance).

General Comments: For the most part, I found the report and supporting documents and models clearly written and transparent relative to the assumptions and methods used to reach the USDA's conclusions. I believe that the USDA has considered and addressed most of the items on which readers would seek clarification in order to fully evaluate the merits of the scientific "argument" and the conclusions. However, there were some issues that I believe are worthy of consideration and discussion in this review. These issues include the following:

1. Were the surveillance samples truly representative of the reference population?
2. The prior question must interpreted in light of whether the OIE guidelines on this issue require a truly representative sample or whether the sampling recommendations are conservative enough to reach appropriate conclusions in the absence of representative, probability based sampling.
3. Whether extrapolations made from other sources to the USDA data were appropriate relative to:
   a. The age distributions that were assumed for animals when there were no recorded ages associated with a portion of samples collected in 1999-2003?

   b. The rate of decline in BSE incidence that might be expected as a result of the feed ban?

4. Whether age data recorded for specimens were truly accurate within 1 year for the animals tested as was represented in the model? Why was this assumption not evaluated in the uncertainties section of the report?
5. Are the sensitivity and specificity estimates for clinical signs appropriate as they are used in the BSurvE model for assigning animals to the clinical suspect surveillance stream?
6. For each of the previous questions, whether errors that might have been inadvertently created using the stated study methods would have biased the models such that an incorrect overall conclusion was reached?

Specific Comments:

USDA Charge 1) *"Given the goal of the report is to determine the prevalence, examine the agency's conclusions of this analysis. Did the agency use the appropriate models and was the data valid? Focus on the suitability of the models, transparency of the approach, and the robustness of the data."*

- **General approach:** Clearly there are many different ways that establishing the prevalence for BSE or any other disease. Therefore it is logical and responsible that the USDA used general approach outlined in the OIE's Terrestrial Animal Health Code (article 2.3.13—Bovine Spongiform Encephalopathy, and Appendix 3.8.4—Surveillance for Bovine Spongiform Encephalopathy). Further, one of the key principles behind the promotion and use of the World Trade Organization's (WTO) agreement on the application of sanitary and phytosanitary measures (SPS) is that disease control and surveillance measures be harmonized, equivalent, and transparent among trading partners. As such, use of the BSurvE model as the backbone of these analyses was also logical given its development was sponsored by the European Commission specifically for the purpose of estimating BSE infection prevalence in national herds and for national surveillance programs. Further, the European Food Safety Authority (EFSA) commissioned a Scientific Expert Working Group to conduct a review of the BSurvE model. The summary of this report states, *"…the model is an excellent development and potentially represents a major step forward in the development of appropriate national surveillance programmes for BSE. It is a very powerful tool and has been technically well designed. It is relatively user-friendly, very accessible and has a high level of transparency. It was further also concluded that the model structure is likely to be sound…"* In reviewing the BSurvE model and the documentation, I found the assumptions and methods to be highly transparent. Additionally, I also found the documentation that USDA provided regarding the application of this model to this specific analysis to be very clear.

  Comparison of results obtained from the BSurvE Model to those obtained with the BBC model in combination with the BSurvE model was also appropriate. Because there is other information is available regarding the prevalence of BSE infections in the U.S. that was not considered in the BSurvE model (i.e., that regardless of the initial prevalence, the incidence most likely declined after introducing the ruminant feed ban), use of a Bayesian approach to account for this information in an objective manner is logical.

- **Examination of the USDA's conclusions:** As stated in the executive summary of the report, I believe the USDA's conclusions represent and appropriate conclusion based upon the data that has been analyzed to date using the stated methods. Other items of the charge relate to the completeness of these investigations, and there are other considerations to be considered in this regard. However, as stated, I believe these conclusions are appropriate.

  There is a conclusion stated in a different part of the document that I do not believe is appropriate. In the "Results of Uncertainty Analysis" section of the report (page 24), it states, *"…the similarity of results* [obtained from the BSurvE and BBC models] *suggest that the methods, assumptions, and inputs used in the analysis are reasonable and accurately reflect the prevalence of BSE in the United States."* A contrarian might argue that it is equally valid to conclude that the similarity of results indicates that the models are equally biased, equally inaccurate, that assumptions are equally inappropriate, etc. I believe a more reasonable and defensible conclusion is that the point estimate and confidence interval for prevalence that was obtained from the BBC model was lower than those obtained from the baseline BSurvE model, but that they are not very dissimilar. While the feed ban is obviously a critical control and prevention measure, the similarity of results indicates that it is not necessary to make strong assumptions about the effectiveness of this measure in order to conclude that the BSE prevalence in 1996 is most likely less than 1 per 1,000,000 live cattle.

  Considering that the estimates obtained in the BSurvE model are used in the BBC analysis, and considering that the BSurvE results were less conservative than the BBC model, I believe that the USDA should not underemphasize the BSurvE results nor should they over emphasize the BBC results. As such, I believe it is absolutely critical that the validation of inputs and assumptions for the BSurvE model must be rigorous and complete, probably more rigorous than for the BBC model.

USDA Charge 2) *"The goal of the analysis, as stated in the document is as follows: The United States has conducted BSE surveillance with increasing intensity since 1990, including an enhanced effort following the identification of a Canadian cow that tested positive in 2003. The enhanced BSE surveillance program in the United States is summarized in a separate report: USDA, APHIS, CEAH, Summary of Enhanced BSE Surveillance, 4-6-06. This analysis uses surveillance data that have been collected over the seven-year period prior to March 17, 2006 to estimate the prevalence of BSE in the United States. This information will help guide and support any future requests for consideration of the overall BSE status of the United States in line with international guidelines adopted by the World Organization for Animal Health (OIE)."*

- As presented, this item is a brief statement of the background and objectives for the report. I could not identify a specific question or direction for reviewers that is not covered in subsequent items. Were you asking whether the goals were reasonable and appropriate, and whether the model specifically addressed this question? If so, then yes to both parts of my supposed question.

USDA Charge 3) *Within the context of the study goal, is the model complete and does it make sense? Does it capture the key factors that influence estimation of prevalence and the interaction of these factors in a real world setting? To the extent that it does not, what* *parameters are missing or incorrectly specified? Are there alternative models that would be of value to investigate? Are the places in which the model departs from reality clearly defined?"*

- In considering these questions I found it useful and pertinent to review Article 3.8.4.1 of the OIE Terrestrial Animal Health Code states:

  "The BSE status of a country, zone or compartment cannot be determined only on the basis of a surveillance programme but should be determined in accordance with all the factors listed in Article 2.3.13.2. The surveillance programme should take into account the diagnostic limitations associated with the above sectors and the relative distributions of infected cattle among them. The points targets and surveillance point values in this appendix were obtained by applying the following factors to a statistical model:

  – *A prevalence of one case per 100,000 of the adult cattle population;*
  – *A confidence level of 95%;*
  – *The pathogenesis, and pathological and clinical expression of BSE:*
    - *Sensitivity of diagnostic methods used;*
    - *Relative frequency of expression by age;*
    - *Relative frequency of expression within each subpopulation;*
    - *Interval between clinical pathological change and clinical expression;*
  – *Demographics of the cattle population, including age distribution;*
  – *Influence of BSE on culling or attrition of animals from the cattle population via the four subpopulations;*
  – *Percentage of infected animals in the cattle population which are not detected."*

- I believe that the USDA has appropriately considered these key factors elaborated by OIE in using the BSurvE and BBC models, and the models appear to be appropriately parameterized. That BSurvE model focuses on these same parameters and that an expert panel commissioned by the EU and the EFSA has reviewed the BSurvE model and found it to be appropriate and robust adds further support for appropriateness of the model. However, while the BSurvE template allows for a large amount of flexibility and specificity regarding specification of the empirical data, I believe that the USDA needs to carefully consider whether they have exceeded the limitations of the available surveillance data in some areas. This is not really an issue of whether the model is appropriately specified and parameterized, but rather it is an issue related to assumptions and application. As such I will discuss my questions and reservations further in the subsequent item.

- In considering whether the addition of the BBC model is a necessary or useful contribution to the analytical effort, I believe it is important to consider the first sentence of Article 3.8.4.1 quoted above. This sentence, in part, suggests that when estimating BSE prevalence, officials need go beyond surveillance results in isolation by considering historical and existing risk factors as well as whether appropriate control measures were in place for the relevant period of time. Also, consider that the world scientific community agrees that the three most important control and prevention factors for BSE are 1) appropriately controlling on importation of live cattle, 2) appropriately controlling importation of feedstuffs, and 3) rigorous implementation of a ban on feeding ruminant or mammalian source protein to cattle. Taken together, I believe these indicate the need for considering results from the BBC model when attempting to draw appropriate conclusions from this prevalence estimation effort. As a point of lesser importance, I believe this also supports the USDA's decision to use the Surveillance B worksheet in BSurvE.

- **Assumptions inherent in the analytical models (discussed on pages 10-11 of the report):** I do not disagree or have comments about the assumptions presented here, with the exception of two items.
  - Relative to the assumption about the effectiveness of the feed ban in the U.S. being at least as effective as the feed ban in the U.K., I find this assumption generally acceptable. However, one could argue that the relative rate of decline in incidence in the U.S. may not be exactly proportional to the relative rate of decline that was seen in the U.K. after introducing this measure. For example, considering that the prevalence of BSE was vastly greater and that the amount of contaminated feed was also likely vastly greater, one could argue that the U.K. feed ban had a larger proportional impact than did the U.S. feed ban given that there was already less contaminated feed to begin with. In other words, the rate of incidence decline in the U.S. may more closely approximate incidence decline represented later in U.K. epidemic. At the same time, as previously noted, the results of the BBC model are not tremendously different than the baseline BSurvE results, suggesting that the USDA's conclusion is robust relative to this assumption.
  - I believe the USDA needs to present a much more thorough consideration of how representative their sample set was, and whether this could have impacted the outcome or conclusions of the analyses. While the report states on page 11 that the BSurvE requires independence among samples but not that they are "randomly" selected, it is also emphasized that samples should be "representative" of the reference population. This is consistent with the general guidelines presented in the OIE Terrestrial Animal Health Code. What is the correct definition or interpretation of "representative" relative to this matter? I believe it implies that you would obtain approximately the same results if the sampling were conducted again. In this case because, there is an inherent assumption that the sampling was equally effective through all years of the surveillance program, it also implies that if the sampling were conducted again, moving sampling intensity from one part of the study to another, that you would also obtain essentially the same results. This would also include application of the different definitions that were applied in different parts of the sampling period. In this regard, I do not believe that the USDA has adequately discussed the whether their sampling was truly "representative" throughout the sampling period, or that these documented changes did not affect the conclusions of this study. For example, given that the sampling strategies changed dramatically during the 7 year sampling period, and that the surveillance points are strongly weighted into the latter sampling period represented by the Enhanced BSE Surveillance program, what would have happened if sampling efforts were redistributed more uniformly, or if the early period were emphasized? While I am not convinced that the USDA is wrong in their assumption on this matter, they have also not

presented information that strongly convinces me that they are correct. They present much stronger data in regard to other model assumptions. My apprehension is augmented by concerns about the validity of age assumptions about the sample sets (not the underlying cattle population), which is discussed relative to the next charge item.

This could be addressed by the USDA through including this parameter in those investigated as part of the uncertainty analysis, along with enhancing discussion of this issue.

USDA Charge 4) *Examine the assumptions about input parameters. Are they reasonable within the context of the goal of the analysis? Do the assumptions reflect reasonable most likely estimations of the parameters?"*

- **Demographics and age distributions:**
  - **Reference cattle population:** I believe the USDA has very good data for parameters relative to the general cattle demographics in the U.S. However, as they indicate, assumptions were needed in order to approximate an idealized cattle population for use in the model relative to exact age demographics and age related exit probabilities for birth cohorts. For the purposes of the analysis I was satisfied that assumptions were appropriate. Results of the uncertainty analysis were also reassuring on this matter.
  - **Surveillance population:** I was concerned about several issues relative to the age used for the surveillance population. First, as noted in the "Enhanced Surveillance Report," age data were collected categorically for animals sampled early in 1999-2003, and the USDA assumed that the true age distribution was represented by the age distribution for samples collected through the Enhanced Surveillance period. I believe further justification is warranted given that the sampling strategies were very different between early and later parts of the surveillance period.

    Second, I believe that it may be invalid to assume that ages of cattle in the surveillance population can be estimated or approximated with precision using 1-year age increments up to the age of 17. Table 1 from the "Enhanced Surveillance Report" shows that nearly half of the samples collected for this period were obtained from renderers, which most likely would be collecting cattle that would recorded as being dead with no other signs, and therefore would be allocated to the fallen stock surveillance stream.

    Examining the age-specific scheme for point allocation, it can be seen that 5 year old cattle receive the most points for all surveillance streams. While the exact age for cattle may have been requested during the Enhanced Surveillance period, I believe these data were biased as it is essentially impossible to accurately age cattle after they reach approximately 5 years of age using only externally visible physical parameters. The most common method used to estimate the age of cattle is by examining the dentition. Once all 8 permanent incisors have erupted and are in wear (usually assumed to occur at approximately 5 years of age), it becomes very difficult to accurately age cattle until they have begun to lose or wear out these teeth when it might be reasonably assumed that they have reached some advanced age. As such, I believe it may have been more

Table 1. Number of targeted samples tested by collection site type from June 1, 2004, through March 17, 2006

Collection Site	Targeted Samples	% of Total
Slaughter Plant*	32,560	5.03
Renderer	356,879	55.16
On-Farm	34,464	5.33
Public Health Lab	191	0.03
Diagnostic Lab	2,804	0.43
3D-4D	202,844	31.35
Other	17,303	2.67
TOTAL	**647,045**	**100.00**

* Does not include antemortem condemned animals transported to offsite facilities (3D/4D collection sites) for sampling.

reasonable to group surveillance data into age categories rather than to assume it was valid to report age in 1 year increments. I recommend that the USDA reconsider what the impact might have been regarding this potential misclassification. Ages for cattle could be aggregated for analysis in the models (perhaps using the scheme documented in the OIE Terrestrial Animal Health Code), and after reallocating points, results from these models could be compared to the baseline models to include this information in the uncertainty analysis.

- Regarding the likelihood ratio values used to assist in allocating animals in the sample set to the clinical suspect stream, I disagree with the argument provided in the USDA document, "Summary of Enhanced BSE Surveillance in the United States."[2] The USDA assumes that clinical signs are independent relative to the diagnosis of BSE, and yet they state in the same document that there is data which contradicts this assumption. Intuitively it seems nonsensical to assume, as the USDA has done, that affected cattle are not likely to have more than one sign if they show any sign of infection. However, this concern is addressed in the uncertainty analysis. While I believe this assumption is incorrect and it therefore affects sensitivity and specificity estimates along with the likelihood ratio estimates, results of the uncertainty analysis indicate that the model is robust to this assumption, and I am therefore satisfied that this did not affect results such that it created significant bias in the overall conclusions.
- I thought the assumptions relative to other input parameters were reasonable, especially in light of the results of the uncertainty analyses.

[2]RTI Note: The reviewer has commented on a supporting document that was referenced in the main report and not a part of this review. Although such review is not specifically requested, the result from the supporting document has implications for analysis presented in this report.

<u>USDA Charge 5)</u> *Are the conclusions robust to the uncertainties in both the model and the input data? Are those uncertainties described completely? How robust are the conclusions? Are these implications described transparently?"*

- I have previously elaborated on two areas that I believe need to be included in uncertainty analyses (representativeness of the surveillance samples, and age assumptions for the surveillance populations).
- I found the uncertainty analyses and the descriptions to be very reassuring regarding whether the models were robust to the assumptions of the model. I also found the methods to be very clearly elaborated and the methods transparent.
- As to the impact on the conclusions: given that the uncertainty analyses suggest that 3 additional cases could be included in the analyses without altering results such that the conclusions change (i.e., that the prevalence is less than 1 per 1,000,000 adult cattle), and that the model is highly robust relative to variation in the model inputs and assumptions, I believe this greatly strengthens the believability of the report and the conclusions. However, I would like still like to see the additional components described above.

<u>USDA Charge 6)</u> *Are the results of the analysis correct given the models used? Check the code and formulas by running the program to make sure the model does what it is intended to do. Given the models used, determine if the agency came out with the right answer."*

- I have run the models using the data and coding as provided to me, and I did not find any discrepancies between the output and what was presented in the report. However, as directed by the instructions received from RTI ("…please allocate more efforts to the methodological/technical aspects than to purely coding aspects of the model…"), I spent far more time reviewing the two mathematical models for the purpose of understanding how the basic model assumptions were generally being used in calculations; I did not review in detail all of the coding, formulas, or program-specific functions included in these models. Again, I believe that this is consistent intent of the instructions provided to me by RTI.
- I would like to see the USDA's response to my areas of concern before I give final judgment as to whether this is the "right answer." However, within the limits of what was provided to me, and assuming that this further analysis also showed the model was robust the assumptions of representative sampling and the age assumptions for the surveillance population, I believed that the conclusion was appropriate (i.e., that the prevalence is less than 1 per 1,000,000 adult cattle).

Review of An Estimate of the Prevalence of BSE in the United States

by

Larry G. Paisley, D.V.M., M.S., Ph.D.
Senior Researcher, Epidemiology and Risk Assessment Section Danish Veterinary Institute, Denmark

Dr. Paisley, a Montana native, is currently a Senior Researcher in the Danish Institute for Food and Veterinary Research, in Copenhagen, Denmark. Dr. Paisley has a DVM from Washington State University, a MS in Theriogenology from the University of Minnesota and a PhD in Epidemiology from the Royal Veterinary and Agricultural University, Copenhagen. Denmark.

Dr. Paisley spent his early career as a Therigenologist, teaching at Washington State University, Pullman, Washington and Ross University; Basseterre, St. Kitts, W.I.

Since receiving his Ph.D. in 1991, Dr. Paisley has worked in Puerto Rico for USDA:APHIS:VS as an Area Epidemiologist and at Albany, New York as the Northern Regional Epidemiologist. In 1996, Dr. Paisley began working at the National Veterinary Institute, Oslo, Norway as an Epidemiologist/Risk Analyst and in January, 2000 began working for the Danish Institute for Food and Veterinary Research as an Epidemiologist with focus on TSE surveillance, disease modelling and risk assessments. Dr. Paisley has conducted numerous TSE/BSE related risk assessments for the Danish Veterinary Authorities, the Plant Directorate and Foreign Ministry. He participates in several EU funded projects on TSE/BSE, FMD, CSF and aquatic animal risk assessments and has been an invited expert on European Food Safety Authority (EFSA) Working Groups.

INTRODUCTION[3]

I, Larry G. Paisley, have been commissioned by the Food and Agricultural Policy program (FAPR) at Research Triangle Institute (RTI), to conduct a peer reviewer for the Bovine Spongiform Encephalopathy (BSE) prevalence estimation models and associated report developed by USDA's Centers for Epidemiology and Animal Health (CEAH), National Surveillance Unit (NSU).

[3]RTI Note: This review is organized differently than the other two reviews. Dr. Paisley first focuses on the five specific charge questions in terms of summarizing APHIS's approach and providing any relevant comments in that context. Then towards the end, he consolidates and summarizes the comments in the section titled "Conclusions from the Review."

The charge for the review includes evaluation of:

- The suitability, transparency and robustness of the models used
- The validity of the data
- The model input assumptions
- The uncertainties in the model results
- The conclusions

Each of these points are discussed leading to an overall conclusion regarding prevalence estimate for BSE in the United States

The Models

The models used to estimate the BSE prevalence were the BSurvE model (Wilesmith et al. 2004), A Bayesian Birth Cohort (BBC) model and an extrapolation model (Cohen et al., 2001, 2003).

The BSurvE Model

In 2003 the European Commission requested that the Community Reference Laboratory to develop an epidemiologically valid integrated approach to evaluate the Bovine Spongiform Encephalopathy (BSE) status of individual countries. A statistical model (BsurvE) was developed that uses the demographic information and surveillance data in a country to estimate the prevalence of BSE in the standing population (Wilesmith et al., 2004). The BSurvE model received an overall favorable evaluation by a European Food Safety Authority (EFSA) Scientific Expert Working Group. The Working Group concluded that the model was well designed, relatively user-friendly, very accessible and transparent. The Working Group recommended that the BSurvE results be compared to the results of other back calculation methods using the same data and hypotheses to analyze the model's integrity. In order to assess the robustness of the model in the case of inaccurate data or assumptions, a series of sensitivity analyses would be required (EFSA, 2004).

The main assumptions are:

- The size and age structure of the national herd does not vary over years;
- The age at clinical onset follows a lognormal distribution with mean being five years of age;
- The distribution of exiting infected animals showing clinical signs apportioned between surveillance streams will not change over time;
- Only cattle less than one year of age are susceptible to the infection;
- The test is 100% sensitive during the period 0-3 months prior to the onset of clinical signs (In the updated version);
- There is no spontaneous case of BSE (EFSA, 2004).

Data required:

- Cattle population size
- Age distribution of the cattle population
- BSE testing data
  - By age
  - By surveillance stream
    - Healthy slaughter
    - Emergency slaughter
    - Fallen stock
    - Clinical suspect

The main outputs are:

- BSE Surveillance Assessment
- BSE Status Assessment
  - For endemic countries
    - The prevalence (with confidence limits) of BSE in the standing population
    - The prevalence (with confidence limits) of BSE in birth cohorts
  - For non-endemic countries
- Surveillance Resource Allocation

BAYESIAN BIRTH COHORT MODEL

The Bayesian Birth Cohort model utilizes outputs from the BSurvE model "The Bayesian Birth Cohort method provides a more precise estimate of U.S. prevalence by combining the epidemiologic theory underlying the BSurvE model with information about the effect of the feed ban on prevalence. As a starting point, this method assumes that prevalence could be any value between 0 and 100 percent, then uses the total number of surveillance sampling points for each birth cohort sampled in the United States to update the initial value" (CEAH, 2006). It is assumed that the feed ban in 1997 in the US was as least as effective as the feed ban instituted in the UK in 1988. Data based on the decreased incidence of BSE in the birth cohorts after 1988 in the UK were used to model the effect of the US feed ban.

The Extrapolation Model (CEAH, 2006)

The extrapolation model begins by estimating the BSE prevalence in the targeted population directly from the sampling data with a beta distribution. The prevalence in the non-targeted population is estimated by assuming it is proportional to the prevalence in the targeted population. The apparent prevalence is adjusted with an estimate of the surveillance sensitivity (~ 40%). Data from testing in the EU suggest the BSE is 28 times more likely in the high-risk population. The total population prevalence is estimated from the weighted average for the subpopulations using the assumption that the high-risk population constitutes 3% of the total population.

"Dividing the apparent prevalence distribution by the 40 percent sensitivity, then solving for the resultant beta distribution, gives a true estimated prevalence distribution in the higher-risk sub-population. Applying the 1:28 ratio to the estimated "true" BSE prevalence distribution in the higher-risk sub-population, then solving for the resultant beta distribution, estimates a "true" prevalence in the normal sub-population. Simulating the weighted average of the higher-risk and normal subpopulations' beta distributions estimates the mean, 5th, and 95th percentiles of the prevalence distribution" (CEAH, 2006).

COMMENTS ON THE DRAFT ESTIMATE OF THE BSE PREVALENCE IN THE UNITED STATES

Data Inputs for the Model

Samples

The report states that BSE surveillance samples in the analysis were collected in the 7-year period prior to March 17 2006. It states that a total of 735,213 samples were taken. There were 23,322 healthy slaughter samples, 571,888 fallen stock samples, 134,232 emergency slaughter samples and 5771 samples from clinical suspects. There is a small problem in the fact that in the Prevalence B spreadsheet, tables 17a and 17b there are 735,054 negative and 2 positive samples. This is only a minor irritant. The data in table 17 are derived from Test Results, Table 13. This data is for testing during 5 years. It is not clear how the data from 7 years of sampling are incorporated into the 5 years teat data.

Surveillance Streams

The samples from the enhanced surveillance scheme were allocated to surveillance streams according to the requirements of the BSurvE model. Samples collected prior to the enhanced surveillance scheme were allocated to surveillance streams by epidemiologists based on the clinical history on the laboratory submission forms. Many animals tested during enhanced surveillance were recorded with signs compatible with BSE but were not assigned to the clinical suspect category. These animals were re-allocated to the clinical suspect category if they showed signs 807 times more likely to be seen in a BSE case than the US targeted population. The justification for this procedure can be found in APHIS, 2006.

Model Parameters Affected by Population Age Distribution

The population age distribution affects the exit probabilities of the uninfected animals, the infected animals and the proportion of uninfected and infected animals that exit via each surveillance stream. The exact age distribution of the US cattle population is unknown but was estimated with a well-documented method (Appendix A in CEAH, 2006).

Assumptions in Model Methods

Proportion of Pre-Clinical Detection

It was assumed that 40% of BSE infected animals would become detectable in the year before they showed clinical signs. It is possible to change this value in the BSurvE model. Current information suggests that the rapid tests are able to detect BSE infection about 3 months prior to the onset of clinical signs such 25% might be a more reasonable value. However, the sensitivity analysis in the report suggests that the estimate has little effect on the prevalence estimate.

Infection in First Year of Life

This is a reasonable assumption.

Constant Versus Declining Prevalence

The BSurvE model assumes a constant incidence over time while the BBC model assumes a declining incidence following the feed ban. The different assumptions had little effect on the prevalence estimates.

The U.S. Feed Ban is at least as Effective as the 1988 UK Ban

This is a reasonable assumption.

Sensitivity of Testing Clinical Animals

The BSurvE model assumes the BSE tests are 100% sensitive if an animal is in the clinical stage of BSE. The sensitivity of the tests during the pre-clinical phase is 40%. As discussed previously, 25% may be a more appropriate estimate.

Disease Behavior

No comment.

Influence of Sample Distribution

No comment.

Prior Distribution for BBC

No comment.

ESTIMATION OF PREVALENCE WITH BSURVE PREVALENCE B METHOD

Although the BSurvE is designed to analyze several aspects of BSE surveillance programs only the Prevalence B estimation method is used in the report. The Prevalence B method is intended for use in countries with surveillance programs but no recorded BSE cases or *countries with very low numbers of BSE cases. The choice of the Prevalence B method is appropriate for the US. The method assumes there is a period of time across which the infection rate in a country remains relatively constant and estimates the prevalence across all surveillance data. The model description is clearly written and appears to be accordance with BSurvE documentation (Wilesmith et al., 2004).*

ESTIMATION OF PREVALENCE WITH BAYESIAN BIRTH-COHORT METHOD

The Bayesian Birth Cohort model is an extension to the BSurvE model that incorporates information on the 1997 feed ban to estimate the prevalence in the standing population. In this model the infection rate was assumed to be constant during the period prior to the feed ban then decreasing during the next 5 years. The decrease was assumed to occur at the same rate as was observed in the UK following the feed ban in 1988.

The model description is clear and the methodology is sound.

It should be pointed out that the results of the BBC model are highly dependent on the results obtained from the BSurvE model. It uses the surveillance points accumulated each year that were generated by the BSurvE model to estimate the prevalence.

Uncertainty Analysis

Uncertainty analysis provides a mechanism for evaluating the influence of model inputs and assumptions on the estimated BSE prevalence. Three potential sources of uncertainty were identified that could have the greatest impact on the prevalence estimate. They included:

- Sensitivity of the prevalence estimate to the BSurvE algorithm and its assumptions;
- Sensitivity of the prevalence estimate to inclusion of additional cases (for example, the Canadian origin case) with the same amount of negative surveillance;
- Sensitivity of the prevalence estimate to various alternatives for input parameters to the BSurvE model.

Two methods of uncertainty analysis were used:

- BSens, a software program designed to do sensitivity analysis on The BSurvE model.
- The extrapolation model (CEAH, 2006)

Results of the Prevalence Estimations

The results of the prevalence estimations by the three methods (Prevalence B, BBC and extrapolation) are presented. The results are consistent with the inputs in the models and the models run correctly. As expected, the prevalence estimate by the BBC model was lower than that of Prevalence B method because it incorporates the feed ban information. However, both methods suggest the prevalence of BSE in the US population is very low i.e., <1:1,000,000.

Results of Uncertainty Analysis

This is perhaps the most informative and useful part of the report.

Sensitivity of the Prevalence Estimate to BSurvE Algorithm and Assumptions

The sensitivity of prevalence estimate to BSurvE algorithm and assumptions was estimated by the extrapolation method. The extrapolation method produced a somewhat higher estimate than both the Prevalence B and BBC methods. However, the extrapolation results also suggest that the BSE prevalence is less than 1 per million.

Sensitivity of the Prevalence Estimate to 1, 2, or 3 Additional Positive Animals

The addition of 1, 2 or 3 additional cases had only minor effects of the prevalence estimates by the Prevalence B and BBC methods. This is not surprising considering the size of the US cattle population.

Sensitivity of the Prevalence Estimate to Alternatives for Assumptions and Input Parameters to the BSurvE Model

Likelihood Ratio

Considerable was expended to correctly allocate some of the BSE surveillance samples into the clinical suspect category. This was apparently done because the clinical suspects provide the most surveillance points from which the prevalence is estimated by the BSurvE and BBC models. Not surprisingly, this procedure resulted in uncertainty regarding the prevalence estimate. However, increasing or decreasing the cutoff value of 807 had little effect on the prevalence estimates.

A second source of uncertainty was the inclusion of all negative surveillance data in the denominator of the likelihood ratio. Exclusion of the samples classified as "dead-unknown cause" from the denominator resulted in fewer clinical suspect samples but had very little effect on the prevalence estimate.

Sensitivity of the Prevalence Estimate to Exit Parameters Dependent on Age Distribution (BSurvE variables $D_{j,t}$ and $C_{j,t}$), Pre-Clinical Detection, and Probabilities (Proportions) that Uninfected and Infected Cattle Will Exit Via each Surveillance Stream

Exit constants for uninfected and infected cattle ($D_{j,t}$ and $C_{j,t}$)

The age distribution of the cattle population has an effect on the exit probabilities of uninfected and infected cattle. The project group has expended considerable effort in describing the method for estimating the age distribution of the US cattle population. I agree with the statement "We have a high degree of confidence that this input is correct for the United States' cattle population and consider this potential source of uncertainty unlikely to have substantial impact on the results of the analysis."

Because of lack of data specific for the US the default values in the BSurvE model were used. The default estimates of the exit probabilities as done in the BSurvE model are somewhat problematic. First, there is no clear explanation or justification for the estimates in the documentation or instructions for the BSurvE Model (Wilesmith et al., 2004; Wilesmith et al., 2004) only that they were derived from UK or EU data. Theoretically, the vast majority of cattle over 24-30 months of age are tested in the EU countries. In 2004, the percentage of BSE tests done in the surveillance streams healthy slaughter, fallen stock, emergency slaughter and clinical suspects were about 88%, 10%, 0.7% and 0.03%, respectively (EC, 2005). The BSurvE suggests that 89%, 7%, 4% and 0.09% would enter the respective surveillance streams. The BSurvE estimates seem reasonable in this case.

The percentage of BSE cases in the EU in 2004 found in the surveillance streams healthy slaughter, fallen stock, emergency slaughter and clinical suspects were about 29%, 49%, 23% and 18%, respectively. However, the default values in the BSurvE model suggest that about 17%, 10%, 10% and 64% of the BSE cases would be found in the respective surveillance streams (see table 9 in the Parameters worksheet). It appears to me that something is wrong here. It seems to me that the surveillance stream where the BSE cases are actually found in a near total surveillance system should be a better estimator of the exit probabilities than those provided in the BSurvE model. It is somewhat surprising that differences of this magnitude would not have a significant effect on the prevalence estimates. Lack of time prevents me from determining what effect changes in the exit constants would have on the US prevalence estimate. I encourage the authors to explore this further. Thorough analysis of the EU BSE testing data may be useful.

Proportion of Exiting Preclinical Cattle that Are Detectable

I agree that changes in the proportion of exiting pre-clinical cattle that are detectable cause minor changes in the prevalence estimate.

Probabilities (Proportions) that Uninfected (BSurvE variable dj,t) and Infected (BSurvE variable cj,t) Cattle Will Exit via each Surveillance Stream (j) Given that They Exit at Time t (from BSurvE table 4 and 5)

There is a very thorough discussion of the effects of individual changes in the exit probabilities that suggest, in effect, that the prevalence estimates are relatively insensitive to changes in the exit probabilities. However, including the US surveillance data into the d_{jt} *proportions but maintaining the BSurvE proportions for the healthy slaughter stream had the effect of increasing the prevalence estimates. It would be of interest to know the effect of using exit probabilities based on the EU test data would have on the prevalence estimates. In addition, it would be informative if the effects of simultaneous changes in several inputs were known.*

Conclusions from the Uncertainty Analysis

As shown in Table 6 the Prevalence B and the BBC models are quite robust to changes in the inputs. Regardless of what changes were made in the inputs the prevalence estimates *remained quite low. Even when three addition cases were added to the total the prevalence estimates were < 1 per million.*

CONCLUSIONS FROM THE REVIEW

The charge for the review includes evaluation of:

- The suitability, transparency and robustness of the models used
- The validity of the data
- The model input assumptions
- The uncertainties in the model results
- The conclusions

The Suitability, Transparency, and Robustness of the Models Used

The BSurvE model is a statistically and epidemiologically sound model that has received a favorable review by the European Food Safety Authority and is judged to be a suitable tool to estimate the BSE prevalence in a country (EFSA, 2004). The BSurvE model is, for the most part, transparent and the transparency of the CEAH report is good. The BSurvE model is quite robust.

The Bayesian Birth Cohort (BBC) model that utilizes information on the 1997 feed ban is also suitable, transparent and robust. However, the BBC results are highly correlated with the BSurvE results because the prevalence is calculated from surveillance points generated by BSurvE model.

The extrapolation model, which was used in the uncertainty analysis, is a relatively simple model that requires only a few assumptions. The description in the report is transparent and the model seems to be suitable for what it was used, uncertainty analysis. The prevalence estimate from the extrapolation model is somewhat higher that those of the Prevalence B and BBC models. However, it also suggests that the BSE prevalence in the US cattle population is very low.

The Validity of the Data

The BSurvE model requires data on the population size, age structure and BSE test data allocated amongst four surveillance streams. This data may be difficult to obtain in some countries. The estimates of the US cattle population size and age structure seem adequate for the analysis. The quality of the surveillance data is not as good because relatively few tests were done in the healthy slaughter stream and the allocation of the tests between surveillance streams was not ideal. However, it seems that this data also is adequate for the prevalence estimates.

The Model Input Assumptions

For the most part the model assumptions are valid and justified. It appears that the exit probabilities for infected animals are somewhat questionable and should be analyzed further to determine if they are realistic. The exit probabilities determine to which surveillance stream infected and uninfected animals in the model will exit. The exit probabilities for infected animals have a large effect on the proportion of BSE cases found in the respective surveillance streams. However, it is unlikely that changes in the exit probabilities would have an effect large enough to change the overall conclusion that the BSE prevalence in the US is very low.

The Uncertainties in the Model Results

Many sources of uncertainty in the model results have been identified and analyzed. It appears that most changes to individual input parameters have no great effect on the prevalence estimates. It is unknown what effect simultaneous several changes in input parameters would have on the results. In addition, there are some uncertainties that have not been identified and analyzed including the age at infection distribution and the incubation period distribution. The use of a different age at infection assumption or a different incubation period distribution would have an effect on the prevalence estimates (Ferguson et al.,1997; Arnold and Wilesmith, 2004).

The Conclusions

Because of the uncertainties associated with the prevalence estimates there should not be much weight associated the actual numbers and confidence intervals in the report. However, the overall conclusion that the BSE prevalence in the US is < 1 per million in the adult population seems justified. Additional BSE testing, especially in fallen stock and older healthy slaughter cattle, would increase the confidence in the prevalence estimates.

REFERENCES

APHIS (2006). USDA:APHIS: VS, CEAH. Summary of Enhanced BSE Surveillance in the United States, April 28, 2006.

Arnold, M. & Wilesmith, J. (2004). Estimation of the age-dependent BSE infection of dairy animals in Great Britain. Preventive Veterinary Medicine, 66:35-47.

Centers for Epidemiology and Animal Health (CEAH) (2006) An Estimate of the Prevalence of BSE in the United States (DRAFT)

EC (European Commission).(2005) Report on the Monitoring and Testing of Ruminants for the Presence of Transmissible Spongiform Encephalopathy (TSE) in the EU in 2004. http://europa.eu.int/comm/food/food/biosafety/bse/annual_report_tse2004_en.pdf

EFSA (European Food Safety Authority). (2004) Scientific Report of the European Food Safety Authority on the BSE surveillance model (BSurvE) established by the Community Reference Laboratory for TSE. http://www.efsa.eu.int/science/tse_assessments/bse_tse/661/sr17_bsetse_bsurve_en1.pdf

Ferguson, N. M., Donnelly, C. A., Woolhouse, M.E.J. & Anderson, R. M. (1997). The epidemiology of BSE in cattle herds in Great Britain .2. Model construction and analysis of transmission dynamics. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 352:803-838.

Wilesmith J, Morris R, Stevenson M, Prattley D, Cannon R, Benard H (2005). BSurvE: A Model for Evaluation of National BSE Prevalence and Surveillance. User instructions for the BSurvE model, Version 05.04b, April 2005. Weybridge, England. Veterinary Laboratories Agency.

Wilesmith J, Morris R, Stevenson M, Cannon R, Prattley D, and Benard H (2004) Development of a Method for the Evaluation of National Surveillance Data and Optimization of National Surveillance Strategies for Bovine Spongiform Encephalopathy. Weybridge, England, European Union TSE Community Reference Laboratory, Veterinary Laboratories Agency.

Review of An Estimate of the Prevalence of BSE in the United States

by

Eric A. Suess, Ph.D.
Associate Professor Department of Statistics California State University

Prof. Eric A. Suess has worked on research projects in the field of biostatistics/epidemiology related to disease monitoring in livestock populations. His research in this area is related to applying Bayesian hierarchical modeling and MCMC methods to estimate infection rates at different levels within a country, herds, regions, and at the overall country level. Most of his statistical applications have been developed using R and WinBUGS. Prof. Eric A. Suess's other interests include applications of Regression and Time Series methods, Computational Statistics, Consulting, and Statistical Education.

1) Given the goal of the report is to determine the prevalence, examine the agency's conclusions of this analysis. Did the agency use the appropriate models and was the data valid? Focus on the suitability of the methods, the transparency of the approach, and the robustness of the results.

Models:

In the research area of disease freedom the models selected (BSurvE and BBC) are reflective of the literature. There are other models that could be selected that use Monte Carlo simulation or purely Bayesian models and MCMC. (If a reference list is needed I can gather some references.) The main advantage of the methods chosen is that they use calculation more than simulation and therefore compute answers more quickly.

The presentation of the BSurvE model on p.13 has numerous errors. The indices are important to include in the presentation of the model. The data are not labeled *x* in the model but are $x_{i,j}$. The *n* is not a parameter in the model, $n_{i,j}$ is the collection of parameters in the model. The *f* and the *a* are similarly not in the model and need the appropriate subscripts. More care needs to be take when presenting the model so model can be understood. Also, the subscripts should be defined separately for clarity.

The BSurvE model is not really an algorithm. It should be referred to as a model.

The BBC model should be referred to as a Bayesian Statistical model rather a mathematical model. Bayesian Statistics is a branch of Mathematics similar to Physics being a branch of Mathematics. The models use data so they should be referred to as statistical models.

The presentation of the statistics is done imprecisely and with lack of understanding of the meaning of their output. Below I go into the problems with the imprecision of the language used when discussing the statistics. In particular the confidence interval produces are incorrectly discussed in the paper. The same language should be used that is used in the documentation for the BSurvE model and software. With this said, all of the reported statistics seem appropriate (i.e., the numbers in the tables that were calculated by the methods seems correct), but the language to describe these reported values is not always correct.

The language used to describe the "point values" assigned to the categories in the BSurveE model is imprecise. A specific choice of words needs to be chosen and then used consistently. This suggestion should be considered with all of the technical language.

Some words that are used interchangeably are mathematical, statistical, analytical, the word statistical should be used. The words probability and proportion are used in all cases the word probability should be used. The word likelihood is used at least once to mean probability, this is bad use since the Likelihood function is used to determine part of the model.

One page 10 the first paragraph is unclear. The term "analytic models" should be "mathematical models" or "statistical models." I would suggest "statistical" since data is included in the models and the uncertainty in the parameters in measured and included in these models. The word "numbers" is used in this paragraph to refer to data. Note that a number and a data value from a survey are not really interchangeable. A number is a number on a number line and a data value is a measurement that may include error so a data value is sampled from a distribution. Using the word number here gives a false sense of certainty and is not what is assumed in these models (at least that is my interpretation). The data is collected and the sensitivity and specificity are discussed. I do not understand what was intended by the language "statistically large numbers." When discussing Statistical models they are technically based on "data" which is assumed to be randomly sampled. The data available for these models is relatively "large" but still a sample of the population.

Data:

The data are from appropriate sources and interpreted correctly to the best of my knowledge. Of course the validity of the data is dependent on the people who have collected the data. The policy to sample more likely to be infected animals does add some challenges to the data analysis. The "points" method chosen seems to deal with this issue well.

Methods:

The two main models discussed (BSurvE and BBC) are both suitable models that are transparent and easily understood. With the Excel spreadsheet the BSurvE model is easy to implement and the software has very nice documentation. The BBC model seems to need more documentation. The BBC also seems to be a partially Bayesian model. Much of the parameters are treated as fixed and the numbers are used from the BSurvE model. It seems that if this model was to be Bayesian the numbers taken from the BSurvE model would be used to determine priors on the parameters in the Bayesian model, for example, the values might be the mean or mode values of appropriate prior distribution on the parameters that would be included in the model. Without having complete knowledge of the model I am not certain if this would improve the model in any way, but the use of fixed values in this model reduces the extent that the model is actually Bayesian.

Results:

The results section of the paper seems not to be the main part of the paper. There seems to be much more emphasis on the Sensitivity Analysis of the models. The writing in the entire paper is quite loose when it comes to the use of technical terms and there are many errors in terms of the use of statistical language.

The use of imprecise statistical language is a problem. In many places the language is just wrong. Also, loose use of language in place of technical terms seems inappropriate. The main problem is in the interpretation of the Confidence Intervals in the Results section.

The other main issue is that this report is on what I would consider two statistical models. Referring to them as "mathematical" or "analytic" models is not appropriate.

The use of the word "algorithm" when referring to the BSurvE model is misused and inconsistent. The BSurvE "model" is implemented with a spreadsheet not really an algorithm. Also when conducting the sensitivity analysis the impression is give that the model is changed which is not the case. The input values to the model are changed or the data is changed, the model remains the same. The idea of sensitivity analysis is to see the effect changes in input values or the data have on the final estimates of the parameters, here the prevalence. In Table 6 the BSurvE model is referred to as both an algorithm and a model, it is the same thing and should be consistently referred to as a "model."

The use of the words "Uncertainty Analysis" seems to be imprecise and not the term commonly used to describe the idea of changing the inputs and the data to see the effect of such changes. The standard words in Statistics are "Sensitivity Analysis." The substitution of similar words for technical terms seems to confuse the writing.

Use of Confidence Intervals and Credible Intervals:

The most problematic part of the report is the lack of appropriate use of Statistical terminology. It seems that all of the methods are based on Statistical models, models that use data and model uncertainty as part of the model. Such technical terminology should be corrected. The most important distinction is the use of the Confidence Intervals and Credible Interval (a.k.a. Bayesian Confidence Intervals). A confidence interval is computed with a "confidence level" such as 95%. This is basically the central 95%. So the confidence level is not computed it is fixed and the confidence interval is calculated. The computed upper and lower limits of a confidence interval are not referred to as "confidence levels" as presented in the Results of the Prevalence Analysis section p.23 and in Table 4 and Table 6.

The other issue here is that if a 95% confidence interval is computed then the lower limit is computed for the 2.5 percentile and the upper limit is computed for the 97.5 percentile, so that the central 95% is the confidence level.

In Classical Statistics the confidence level is interpreted using the ideas of repeated sampling in Bayesian Statistics the Credible Interval or Bayesian Confidence Interval is a probability interval. So in the Classical setting a confidence interval can me interpreted as "95% of the time the intervals computed in repeated sampling will include the true parameter value." In the Bayesian setting, "the posterior credible interval has a 95% probability of containing the parameter value."

So the two methods are really different, the BSurvE method used the ideas of Classical Statistics and the BBC method used Bayesian Statistics.

This error is continued in the Results of Uncertainty Analysis section p.24. In this section the words "5th percentile" and "95th percentile" are used to refer to the limits of the 95% Confidence intervals. This change to a further incorrect set of terms needs to be fixed. Before "5 and 95 percent confidence levels" is used, which is also incorrect. In the figures (Fig. 2, Fig. 3) "5th" and "95th" are used, this is wrong. These are the upper and lower confidence limits for 95% confidence intervals. They might be considered 2.5 and 97.5 percentiles, but this is not the language that is used to describe confidence intervals. The standard language is "lower and upper 95% Confidence Interval limits" or "lower 95% CI limit" and "upper 95% CI limit."

The Figures 2 and 3 describe the plots incorrectly. In an xy-plot it is common to refer to the graph as plotting y versus x. In the description it is described as x versus y which is not the usual way such plots are described.

2) The goal of the analysis, as stated in the document is as follows:

The United States has conducted BSE surveillance with increasing intensity since 1990, including an enhanced effort following the identification of a Canadian cow that tested positive in 2003. The enhanced BSE surveillance program in the United States is summarized in a separate report: USDA, APHIS, CEAH, Summary of Enhanced BSE Surveillance, 4-6-06. This analysis uses surveillance data that have been collected over the seven-year period prior to March 17, 2006 to estimate the prevalence of BSE in the United States.

This information will help guide and support any future requests for consideration of the overall BSE status of the United States in line with international guidelines adopted by the World Organization for Animal Health (OIE).

The goals seem to be appropriately addressed by the models chosen and overall this report does describe each models strengths and weakness, and reports the overall findings from the application of these methods to the available data.

The need for the 3 model for comparison seems unnecessary. The use of the extrapolation method seems out of place. If it is a good comparison model for the other two models, BSurvE and BBC, why is it not just included as one of the models used.

3) Within the context of the study goal, is the model complete and does it make sense? Does it capture the key factors that influence estimation of prevalence and the interaction of these factors in a real world setting? To the extent that it does not, what parameters are missing or incorrectly specified? Are there alternative models that would be of value to investigate? Are the places in which the model departs from reality clearly defined?

Yes the BSurvE model seems complete and makes sense for the available data. It captures all of the important parameters that help to estimate the prevalence. The sensitivity seems to be includes in the model, the P(Test + | Infected) but it does not seem to address the idea of specificity P(Test—| not Infected). On the other hand, maybe assuming the specificity is 1 is ok.

The BBC model seems appropriate also. The weighting scheme using the point values derived from a likelihood seems like a reasonable way to come up with weights.

One alternative model is the one that I worked on see Suess, Gardner, Johnson, Prev. Vet. 2002. We proposed a full Bayesian model. There were some papers referenced in our paper that used MC simulation that might also be considered.

4) Examine the assumptions about input parameters. Are they reasonable within the context of the goal of the analysis? Do the assumptions reflect reasonable most likely estimations of the parameters?

Note: ***Monte Carlo modeling is part of the WIN BUGS software that was used and the distributions chosen used a Poisson distribution that is explicit in the WINBUGS code (included in the document.)***

BSurvE Model:

The assumptions in the BSurvE Model about the inputs make good use of the available data. Using the epidemiological model as defined seems to take into consideration all of the relevant aspects of the data collection. Having the information about the age groups seems to be very important and since it is included this seems appropriate.

The BSurvE spreadsheet works very simply and is well organized. It seems easy to use and should be easy to use year to year. The spreadsheet gives a nice presentation of the inputs to model and how the inputs related the calculations.

I did not check the coding of the model in Excel.

BBC Model:

The idea of the BBC Model makes conceptual sense. I can't judge exactly whether the point values used in the model are measured on the correct scale, someone with more knowledge of the epidemiology would be a better judge. The presentation in the paper seems clear what the points are but when it comes to the WinBugs program it is not clear where the values used to compute the expected infections come from. I understand how they are computed but I can't find exactly where they are computed. The concept of down weighting the yearly prevalence's makes sense. Using the values from England is a reasonable idea but in the model I suppose one extension would be to have the yearly factor be a model parameter with priors derived from the values from England since the US is most likely not going to have exactly the same values for the points. Technically these values are not known for the US so testing a model with the points as model parameters would be a next step in developing this model to better reflect reality. The use of the Poison model is reasonable.

In the report there is a reference to the beta(0,1) which should be the beta(1,1) which is equivalent to the Uniform(0,1).

Overall, assuming the values used are accurate the simulation from WinBugs should be fine.

The burn in as long as suggested is most likely not necessary since the simulation has few parameters and they are uncorrelated, the program will converge very quickly.

5) Are the conclusions robust to the uncertainties in both the model and the input data? Are those uncertainties described completely? How robust are the conclusions? Are these implications described transparently?

The extensive discussion of the Sensitivity Analysis does seem to show that the conclusions are robust to the model and the data. The use of the words Uncertainty Analysis for the more common Sensitivity Analysis is again loose use of technical language. The main discussion in the report is not about changing the model but the values of the inputs and changing the number of positive infected animals to see how the model estimates change. In both cases the presentation in the paper shows that the results are robust.

The results are not communicated clearly since the language used to describe the confidence intervals is in error, but with the language describing the results corrected the implications of the analysis will be clear.

6) Are the results of the analysis correct given the models used? Check the code and formulas by running the program to make sure the model does what it is intended to do. Given the models used, determine if the agency came out with the right answer.

The numerical results of both the BSurvE and the BBC model seem to produce comparable estimates of the prevalence. Assuming the model is epidemiologically correct and is not missing any other important variables or input, the model is well defined and produces sensible results. The BBC model for the overall prevalence, again assuming it includes the relevant variables and input, gives sensible answers.

The only remaining issue with the models is that the data that is being used is from test data. I do not know the details of how the animals are examined or determined to be infected before an animal starts to show clinical signs of BSE. But there is the possibility of the animal being determined false positive which relates to the specificity of the testing for symptoms occur. From the report there is a sensitivity of 0.4 for preclinical animals but what is the specificity? In thinking about this it may be possible that the specificity can be assumed to be 1, that is the probability that an animal tests negative given that the animal is free of disease is 1.

Suggested wording and other changes:

Page 8 In the paragraph about the Likelihood ratio, in the first sentence "many" is unclear. It should be clarified what data was collected for the animals tested. This makes it seem that not all of the data was collected for each animal tested.

Page 8 later in the paragraph the word "uncertainly" is used in place of "sensitivity."

Page 8 the BSurvE model is referred to as the downloaded model. This description may be true but no website is given. I would suggest just using the words BSurvE model.

Page 9 in the second paragraph the word "value" is used in place of "points" as introduced on page 7. A common language should be adopted for clarity maybe "point values" should be used everywhere to refer to likelihood values.

Page 9 last paragraph. When discussing Bayesian analysis it would be more clear to describe the model as incorporating the data and all prior knowledge about the parameters in the model. While the prior knowledge is a for of information it is not measured data from a test. On the other hand the prior used in the BBC model is uniform and therefore with this prior no prior knowledge is assumed.

Page 10 the title. The word "model" should be "modeling" since 2 primary models are discussed and a 3rd, the extrapolation model, is discussed for comparison.

Page 10 first paragraph. As described above the first paragraph is awkward.

Page 10 In the discussion of sensitivity, the sensitivity of the preclinical animals is not addressed. The assumption of sensitivity being one is assumed which should be clarified as to the validity of this assumption.

Page 11 In the paragraph on Influence of sample distribution. The title should be "sampling" I would guess. The paragraph is more about the sampling method than the distribution of the sample. There is discussion of targeted sampling but no discussion of the shape of the distribution. And the words "in both methods" needs to be clarified. I would guess this is the A and B methods in BSurvE.

Page 11 In the paragraph on the prior distribution for BBC. The first word "only" is awkward dropping this word would not change the meaning and would sound better. In the last sentence of the paragraph after the comma the word "it" is unclear. It seems that the starting value being overwhelmed is mistaken by the "it" which is the prior distribution. The starting value in the WinBugs program starts the simulation but it is the prior distribution on P that is overwhelmed by the data.

Page 12 Again the word "value" is used and should be clarified as "point values" of the likelihood.

Page 12 end of the 4^{th} paragraph is the first mention of the 95% Confidence Interval which is miss represented. The words "95^{th} percentile confidence level" is incorrect use of the words. A confidence interval is computed for a specific confidence level, if a 95% CI is computed the confidence level is fixed to be 95%. This percentage is the central 95% of the distribution in questions so it is the 2.5 and 97.5 percentiles of the distribution that are used to compute the CI. So there are what people refer to as the lower limit of the 95% CI and the upper limit of the 95% CI. The word percentile is missed used here and should be corrected throughout the paper for technical accuracy of the language used to describe the CIs.

Page 12 in the last paragraph the word "random" is used to describe the samples taken. From earlier in the paper there was discussion of targeted sampling. The sampling was not described at targeted random sampling. This should be clarified. I believe this is what is assumed in the model but it should be made clear earlier that the sampling was targeted random sampling if that was what the sampling was. I have not had time to read the paper on how the data sampling was done.

Page 13 More attention needs to be paid to the presentation of the BSurvE model as mentioned above. The loose presentation does not give an accurate description of the model. The manual for the BSurvE model should be used for presenting the model accurately with the correct notation.

Page 13 In the BBC model discussion. The words "Bayesian mathematical methods" might be more accurately described as "Bayesian statistical methods"

Page 13 this paragraph needs work:

Line 1 "knowledge of" in place of "knowledge about"

Line 4 "Empiric" should be "Empirical." The word empiric is a noun and an adj. is needed here.

Line 6 "prior to the feed ban the prevalence was increased and after the ban the prevalence was decreased"

Line 7 After the ref. there needs to be an "and."

Page 14 In the paragraph starting "The Bayesian Birth Cohort…" the word "epidemiologic" should be "epidemiological." In the presentation of the prior being uniform, the sentence uses the word "starting point" which is confused with the "point" used to refer to the "point value" of the likelihood. And in the next paragraph, the words "from this point forward" are used. The term point has a technical meaning in the report and for clarity it is suggest that either the word "point" is clarified everywhere as "point value" or the use of "point" in sentence transition be removed.

The prevalence is a probability in the BBC model and is a value between 0 and 1, it is not represented by a percentage value.

Page 15 The words "epidemic curve" are use to describe the plot in Figure 1. Is this a technical term in epidemiology? This figure is also referred to as a "cumulative" plot which is not true. It is a bargraph that has bars that represent percentages. It is a type of histogram not a cumulative histogram. The word cumulative is incorrect here.

Page 15 Is "infectivity" a common word to use in this discussion. It seems that the discussion is about the "rate of infection." Late the words "dose of infectivity" seems give the impression that there is some control over the "rate."

Page 15 last paragraph. Line 3 "post ban exposure would still be lower than the U.K. The use of U.S. above would lead to U.K. also for consistency.

Page 16 Again the word cumulative is used to describe the percentages in Figure 1.

Page 16 Table 1 title. The word "percent" is used when a "proportion" is given in the table. There is a footnote with a star that does not refer to anything in the table.

Page 16 last paragraph. The p is P in the model and it is assumed to have a uniform (0,1) density which is a beta(1,1).

Page 17 the software WinBugs needs to be referred to consistently in the report.

Page 17 in the first full paragraph the discussion about the use of the Poisson is presented. The reason "due to the large sample" should be "due to the large sample and the small prevalence." The Poisson is an approximation to the Binomial when the sample size is large and the probability is small.

At the end of this paragraph there is "one BSE point value is equivalent to one randomly sampled animal." If I understand the model correctly it seems the "one randomly sample animal that has BSE" should be included here. I am not 100% certain but this should be checked.

Page 17 The word "uncertainty" is usually "sensitivity." The word sensitivity is used to check the robustness of the model to changes in the data or input parameters. It is not clear what "uncertainty" means here. Maybe I have not seen this term used before. In Bayesian literature the word "sensitivity" is used and in statistics literature this is the common word used also.

Page 17 in the last full paragraph the work "likelihood" is used in place of "probability" or maybe "chance." The word likelihood refers to the likelihood function and this word is used when discussing the point values in the BSE model. For clarity it is suggested that probability is used here.

Page 18 when referring to the BSurvE model calling it an algorithm is not an accurate description. This a spreadsheet model.

In the first paragraph there is a new description of the "points" in the BBC model, "surveillance points" maybe these words should be used everywhere, but it seems changing the words is unclear.

The word "mechanics" is unclear. The sentence is unclear. With sensitivity analysis one would try to check the sensitivity of the BSE model prevalence estimate to changes in the data or prior knowledge. Using another model really is not conducting a sensitivity analysis this is more of model comparison. In fact the title of this section seems inappropriate. This is a discussion of a model comparison not a sensitivity analysis.

Page 18 at the end of the 3rd paragraph "likelihood" is again used in place of probability.

Page 19 in the first paragraph about fitting the beta the discussion of "the advantage of matching moments is that is provides an unbiased estimate of prevalence" is possibly wrong. First the fitting of a density give a parametric density estimate, so it might give an unbiased estimate of the distribution of prevalence, not an estimate of the prevalence itself. If the method of moments is used for the beta the estimate of the prevalence is most likely not unbiased because this method does not generally produce unbiased estimators. In the end the fitting of a density does not produce a point estimate as implied.

Page 20 the word "true" is used before estimated. The word true might be used to describe the "true" population prevalence. An estimate is an estimate of the true population prevalence. True is not commonly used to describe an estimator since an estimator has a distribution.

Page 21 in the last paragraph the word probabilities has (proportions) after it. This is unnecessary and just adds another possible lack of clarity to the presentation, it is suggested that probability be used throughout the report removing "proportion" and reserving "likelihood" the likelihood function that defines how the "point values" are calculated.

Page 22 again "download" needs to be removed.

Page 22 in the second full paragraph "large coefficient suggest inputs are …" In a sensitivity analysis it is the model that is being examined in terms of its inputs. So a model is sensitive the model input not the other way around. This sentence needs to be clarified.

Page 22. A model is fit to data. It is not usually describe that the data is fit to the model. So in the sentence with the gamma distribution, "BSens fits the cattle age distribution with a gamma model. And again on the top of the next page "to" should be replaces by "by."

Page 23 in the Results section, as mentioned above the discussion of confidence intervals is in accurate and confused. The word in the first sentence should be something like "… the expected prevalence and the 95% confidence intervals are shown in Table 4. And again at the end of the paragraph this needs to be corrected.

Page 24. Table 4 (5 and 95 percent confidence levels) need to be changed to (95% confidence intervals)

As a suggestion the work "level" should not be used anywhere in the report. When specifying that the confidence intervals will all be 95% the level has been set and there is no further discussion. The word level needs to be removed and replaced with "upper and lower limits."

Page 24 in the last paragraph the references to 5^{th} and 95^{th} are incorrect, just say the 95% confidence interval is.

Page 25 same errors with the confidence intervals. In the Figures the reference manual should be consulted to get the appropriate labels. The 5^{th} and 95^{th} are incorrect and unclear. These are the upper and lower 95% CI limits.

Page 25 and 26 Figures 2 and 3 use "versus" in the wrong order. The usual use is for y versus x.

Page 27 Figure 4 starts with the word distribution. This is a plot of likelihood ration values and it not a distribution. And the x scale is not drawn to scale so it is very misleading. Drawn on a correct scale the plot will look quite different. The spacing on the x axis need to be made equal or a log scale might be used.

At the end of the first paragraph the word "unremarkable" is not common language to use when discussing a sensitivity of an estimator from a model "was low" would be better.

In the last paragraph, the word "confidence" might better be described as "a high degree of certainty that." In statistics confidence refers to the method used to describe confidence intervals so reusing the word may lead to confusion.

Page 28 again the "unremarkable" should be "low" when referring to the sensitivity. And in the last paragraph the word "insensitive" should be changed to "low sensitivity" for further clarity and consistence of presentation.

Page 29 end of the first paragraph there are 2 periods when only one at the end is needed.

Page 29 the Table 5 is split. In the footnote the indexes j and t in the model should be in italics. In fact, very where in the report where the model is referred a consistent use of italics should be used.

Page 30 at the end of the paragraph with "94 percent" it is not clearly stated what the probability is for.

Page 32 Again the (5 and 95 percent confidence levels) should be (95% confidence intervals)

Page 33 third paragraph "work sheet" is "worksheet" before in the report.

In the fifth paragraph "estimations" should be "estimates."

And it is suggested that "uncertainty analysis" should be "sensitivity analysis."

On the other hand, here is a link to an EPA paper about guidelines for Uncertainty Analysis. But as a Bayesian Statistician this is more commonly call sensitivity analysis.

http://risk.lsd.ornl.gov/homepage/tm35r1.pdf

Page 34 Again the Confidence Interval discussion needs to be corrected.

Page 35 the use of "Exp" should be the function "exp." Try using the Equation editor to get the notation correct. And gain this should be in italics for consistence of the presentation of the models,

Again " the data were best fit by a Weibull."

In the sentence including " number of 1-year olds, minus the probability of culling for the first year heifers." Should this be "minus the expected number"?

Page 37 "BSurvE mechanics of point calculation" should be "BSurvE point estimation."