参考資料7－1

APHIS Response to Peer Review of the Sample Size Estimate for BSE Ongoing Surveillance in the United States

July 20, 2006

(原文)

APHIS response

The document *Sample Size Estimate for BSE Ongoing Surveillance* prepared by the National Surveillance Unit of USDA:APHIS:VS, was completed May 31, 2006. The principal purposes of the ongoing BSE surveillance plan stated in the document are:

1. To continue to monitor the BSE status of U.S. cattle.
2. To provide mechanisms for detection of rising BSE prevalence among U.S. cattle.

Because this document contains important scientific information, it was subjected to a peer review in line with Office of Management and Budget (OMB) guidelines to help ensure the quality of information and to guide improvements prior to final dissemination by federal agencies. At USDA request, RTI International (RTI) performed the peer review. RTI contracted with three independent reviewers, and RTI summarized the results in a report. OMB guidelines also address transparency, and therefore recommend that a response to a peer review be completed explaining the agency's agreement or disagreement, the actions the agency will undertake, and the reasons the agency believes these actions satisfy any key concerns or recommendations in the report.

The following excerpt from the RTI report summarizes the three peer reviewer's comments:

> Two of the three reviewers agreed that the sample size estimate of 40,000 samples per annum is sufficient and exceeds the requirement for "type A" surveillance by OIE. They agreed with most of the assumptions APHIS made in estimating the sample plan, except for a few suggestions to improve their plan.
>
> The third reviewer, Dr. Morris, disputed the reliability of the sample size, based mainly on his reservations about the accuracy of BSE prevalence estimation in the United States. Overall, he expects the plan to meet OIE "type B" surveillance requirements and possibly "type A" requirements. His main concerns are related to the accuracy of determining the age of animals and BSE regulatory practices in the United States.

APHIS' response to each reviewer, actions that will be taken, and reasons for those actions are presented below.

Reviewer Stuart C. MacDiarmid, BVSc, PhD
Principal International Adviser (Risk Analysis), the International Coordination Group of Biosecurity, New Zealand, the New Zealand Ministry of Agriculture and Forestry.

Dr. MacDiarmid comments:

> *"… Given the perfectly reasonable assumption that the proportion of samples drawn from each of the three surveillance streams (clinical suspects, casualty slaughter and fallen stock) in the next 7 year period will be essentially the same*

as in the last 7 year period, it is valid to assume that the estimated sample size of 40,000 animals per year will provide a very high degree of confidence that BSE will be detected, if it is present at a rate equal to or greater than 1 per million."

"The planned sample size will thus provide assurances that BSE is either not present or is present at a prevalence too low to sustain an epidemic. The planned surveillance will also provide an effective mechanism for detecting an increasing prevalence of BSE in US cattle, in the highly unlikely event that current control measures are inadequate."

"Again, on the basis of sampling in the past 7 years, it is evident that the planned sample size will provide surveillance assurances which far exceed the requirements of OIE's 'Type A' surveillance. The document correctly estimates that OIE's 'Type A' surveillance requirements would be met by an annual sample of around 10,500 animals, selected from the three surveillance streams in proportions similar to that sampled in the past 7 years."

MacDiarmid continues:

"I believe this is a reasonable estimate. However, I suggest that USDA could either increase the sensitivity of its ongoing surveillance program, or reduce the number of samples, by working to change the proportion of samples harvested from each of the three surveillance streams."

APHIS agrees with the assessment that the sample size is conservative and that the sensitivity may be increased or sample size reduced with increased focus on the higher point value animals as described in the OIE Terrestrial Animal Code (Article 3.8.4.2). The design of the ongoing surveillance plan assumes that the same proportion of animals in each surveillance stream will continue to be collected for each of the upcoming 7 years. However, since a large portion of the samples collected during the Enhanced Surveillance came from lower point value sources (i.e., dead animals from renderers or food inspection condemnations and others not having specific clinical signs compatible with BSE), the sampling strategy for ongoing surveillance is designed to limit the number of animals from these sources to approximately 15,000. APHIS believes that limiting these sources while maintaining efforts[1] to collect all animals from higher value sources will achieve the suggestions made to increase the sensitivity. Nonetheless, APHIS believes that sampling animals from all three surveillance streams is necessary to provide a representative sample and therefore will continue to collect a portion of the total samples from animals in the lower value streams.

Reviewer: Ian Gardner, BVSc, MPVM, Ph.D.
Professor of Epidemiology, School of Veterinary Medicine at the University of California, Davis

[1] Ongoing surveillance will target all clinical suspects from any source, rabies negative animals from public health and diagnostic laboratories, and animals displaying clinical signs compatible with BSE.

Dr. Gardner comments:

> *"The estimate of 40,000 samples per annum is conservative given the fact that prior information about prevalence and the effects of the feed ban have not been incorporated into the sample size estimation procedure. First, this reviewer recommends consideration of a Bayesian approach based on modification of previously-developed methods (references 1 and 2). Such an approach is scientifically justifiable and will yield smaller sample sizes, if prior knowledge about prevalence is modeled. Second, the sensitivity of the entire BSE surveillance system should be estimated and incorporated into the calculations, regardless of whether a frequentist or Bayesian approach is used. Third, sample size calculations for the second purpose for ongoing BSE surveillance should be added."*

APHIS agrees that the 40,000 sample size is a conservative estimate. However, the sampling strategy is designed to provide cumulative information over a total of 7 years and may be adjusted accordingly if necessary. Further, if the results of surveillance are reviewed during or at the end of each year and indicate that a higher or lower number of samples are necessary to maintain confidence in the low prevalence of BSE in the United States, APHIS prefers to err conservatively.

APHIS acknowledges that a Bayesian model for sample size could reduce the estimated sample size and will consider this approach for future analyses.

APHIS believes that surveillance sensitivity is already incorporated into its calculation of sample size. Surveillance sensitivity is defined as the likelihood of detecting infection with a particular surveillance system given that infection exists at a given design prevalence (e.g., 1 infected cattle per million). Surveillance sensitivity for BSE is a complex parameter to estimate. Because of this complexity, the BSurvE model was developed to assist analysts in determining the value of samples collected. BSurvE incorporates the likelihood that infected cattle will be available for sampling at different ages within different surveillance streams, that such cattle will be within one year of developing (or already exhibiting) clinical signs of BSE, and the likelihood that such cattle will be detectable given that they are sampled (i.e., test sensitivity). Within BSurvE, there are three test sensitivities considered; 0% sensitivity if an infected animal is sampled more than one year before it would exhibit clinical signs of BSE infection, 40% sensitivity if an infected animal is sampled within one year of exhibiting clinical signs and 100% if an infected animal is sampled after it begins exhibiting clinical signs. Given the dependency of infection status on the surveillance stream sampled (e.g., infected animals are more likely in the clinical suspect stream), the differential probabilities that cattle exit the standing population at different ages (e.g., infected cattle may exit because of the effects of BSE or may be culled for other reasons before the end of their incubation period), the age dependency of incubation periods for cattle infected in the first year of life (e.g., a 5-year old infected animal is nearly 6 times more likely to exhibit clinical signs than a 3-year old infected animal), and the dependency of detection on stage of infection, we determined that sample size calculations based on a model

previously reviewed and approved by the international animal health community would gain greater acceptance than developing our own alternative model. Nevertheless, we will consider estimating surveillance sensitivity via such an alternative model for future analyses.

APHIS understands the concern regarding our stated second purpose for on-going BSE surveillance. It is not the intent of this surveillance to statistically demonstrate an increase (or decrease) in the estimated prevalence. Nevertheless, it is understandable how one might interpret such a purpose from our statement. Instead, we intended our second purpose to be the detection of infection above the previously-stated design prevalence of one per million. Therefore, we will re-word the second purpose to read "To provide mechanisms for detection of BSE prevalence if it were to increase above 1 infected animal per million adults." Given the calculations already provided in our report, we believe the proposed on-going surveillance plan will meet this objective.

Reviewer: Roger S Morris MVSc, PhD, F Amer CE, FACVSc, FRSNZ, CNZM
Co-Director, Massey University EpiCentre, New Zealand

Dr. Morris makes the following points regarding the numerical calculations:

> *"It is stated that the number of points required over 7 years is 2,900,000. In BSurvE, the number of points required is 2,995,730, but OIE rounds the numbers up, so the figure would be 3 million to give 95% confidence that the prevalence is less than one in a million. The number of points is wrongly stated in the document under review to be 2,973,804 and is wrongly attributed to the OIE Code Chapter, instead of to BSurvE."*
>
> *"The document also states (on page 2) that "during the 7 consecutive years prior to March 17, 2006 the US collected 735,213 BSE samples ….and accumulated 2,973,804 points (APHIS 2006b)" but on page 3 states "The prevalence analysis …..reports 6,745,010 points resulting from 735,213 samples (APHIS 2006a)". Such inconsistencies need to be addressed to help the credibility of the document."*

In deriving the number of points required over 7 years, APHIS cites a Cannon and Roe formula (Cannon and Roe, 1982). The figure of 2,900,000 samples from the general population results from that formula. Nevertheless, APHIS acknowledges that OIE or BSurvE requirements are relatively conservative and would be more appropriate for use in this document. Therefore, we will use 3 million points in our revised calculations. This adjustment results in a slight increase in the required number of points collected per year from 43,747 to 45,113.

APHIS disagrees with the second point. The different analytic point totals that are noted in the document simply reflect different calculation methods. There are two methods for calculating the total analytic points represented by the U.S. surveillance; one method uses

the OIE point values and the other uses point values generated by the BSurvE model. It should be understood that the two methods generate different values and that one "point" from the OIE table is not the same value as one "point" from the analytic model. Likewise, the interpretation of the point total must be made within the context of the system that generates the point value. APHIS has previously reported its calculations using both of these methods. These calculations are appropriately referenced in the document reviewed.

Dr. Morris suggests the following with respect to terminology:

> *"Use of the term random sampling in relation to BSE is best avoided, because it does not help understanding."*
>
> *"All reference to "detectable" prevalence should be removed, since it is incorrect in the context of the document."*

APHIS agrees with these recommendations. We will change the term "random" to "general population" to more correctly reflect the nature of the sampling and we will remove the term "detectable" from our document.

Dr. Morris suggests the following with respect to sampling strategy:

> *"Further emphasis should be given in the plan to focusing special effort on high-risk sub-populations of animals, selected both geographically and on the basis of potential exposure to contaminated feed material as calves."*

APHIS acknowledges this point, but the document under review only intended to discuss sample size issues. Other documents provide greater detail regarding the allocation of sampling with respect to geography and higher-risk subpopulations.

Dr. Morris makes the following point with respect to surveillance objectives:

> *"I expect that the plan as currently proposed would meet OIE Type B surveillance requirements and possibly Type A requirements, but I cannot directly check this with the information I have available."*

APHIS scientists have no doubt that the proposed plan exceeds OIE Type A and B surveillance requirements. OIE's Type A surveillance requires collection of 300,000 analytic points across 7 years to be 95% confident of detecting a prevalence of at least 1 per 100,000. OIE Type B surveillance requires sampling at one-half the intensity of Type A surveillance (i.e., 150,000 surveillance points across 7 years). If we collected truly random samples – which would equal one analytic point per sample if sensitivity were perfect - we would meet the OIE Type A surveillance by collecting 43,000 random samples per year. However, we are collecting targeted samples, that on average will be greater than one analytic point per sample and therefore will exceed OIE Type A surveillance.

Dr. Morris also makes the following point regarding surveillance objectives:

> *"The sample size estimate of 40,000 to 43,747 BSE tests per year "to achieve a high degree of confidence that the prevalence is less than 1 per million" is not soundly based on epidemiological and statistical principles and is invalid for reasons described above. It is simply an arithmetic calculation which extrapolates from past testing to future requirements."*

APHIS disagrees with this comment. First, the sample size estimate of 40,000 to 43,737 derived by APHIS is not intended to be APHIS' surveillance objective. The objective of detecting a prevalence of 1 per million or more is only achieved after combining the sample size with the point values generated by targeted sampling. The calculation of sample size is, in fact, a simple "arithmetic calculation" once the total requisite points (e.g., 3 million), time period (e.g., 7 years) and points per sample (e.g., 9.5) are determined, and therefore perhaps the comment was meant to identify a discrepancy in the average point value per targeted sample.

The estimated analytic points per sample are based on APHIS' experience of sampling the U.S. cattle population across the past several years. From the BSurvE model, we calculated the total number of analytic points represented by the total number of samples we collected; the analytic points per sample are simply the quotient of these two numbers. Comments questioning the accuracy of the estimated ages of cattle sampled were provided by this reviewer, although they were outside the charge. If inaccurate ages inflate the analytic points per sample, this could question the robustness of the total number of analytic points calculated.

Reviewers of the document; "An estimate of the prevalence of BSE in the United States" have questioned the age distribution of samples collected in the past in the United States. In response to those questions, we reiterate discussion from the final prevalence analysis.

> ***Age determination***
> *At the time of sample collection, documentation of exact cattle ages was often not available, so cattle were primarily aged via dentition. With the exception of a five-month period in 2004, ages were recorded in a continuous fashion in years or months. The BSurvE model uses 17 different ages ranging from 1 to 17 in one-year increments (with age "1" representing cattle less than 2 years old, and "17" representing cattle aged 17 or older). Thus, our data were placed into one of these 17 age categories for use in the BSurvE model, as shown in Table 1. Ages for cattle of unknown age and for those collected during the 5-month period in which ages were recorded categorically were imputed as described in the document, Summary of Enhanced BSE Surveillance in the United States (APHIS 2006)*
>
> *Dentition is a relatively reliable indicator of age up to age 5, but it is difficult to accurately age cattle within one year of its true age for cattle aged 5 or older.*

Thus, it is likely that some ages were misclassified for cattle 5 or older. Because points in the BSurvE model are affected by cattle age, the sensitivity of the prevalence estimate to age is assessed in the sensitivity analysis.

Additionally, to test the effect of possible age misclassification, ages recorded for cattle 5 years or older were adjusted based on the probability of cattle of each age existing in the U.S. general population (i.e., using data for cattle 5 and above in Table A2, the number for each age group was reallocated proportionate to the number of animals in the general population). This adjustment was performed prior to imputation of unknown cattle ages so the imputation could be based on the new age distribution.

In the sensitivity analysis section of the prevalence estimate, results are reported demonstrating that the U.S. prevalence estimate from the BSurvE model was only slightly increased by making the changes described above. A small increase in the estimated prevalence implies that the total number of analytic points is somewhat reduced after reallocating samples into ages more consistent with the BSurvE exit probabilities. The calculated total number of points generated by the 735,213 samples previously collected is ~5 million (instead of 6.7 million). Nevertheless, the relevance of this sensitivity analysis to the sample size calculation is doubtful.

APHIS believes the best estimate of points per sample is directly derived from past sampling evidence in the United States. Those data support 9.5 points per sample. If ages of cattle sampled in the past were inaccurately determined, then reducing the points per sample in the sample size calculations would be appropriate. Alternatively, APHIS expects that refocusing of surveillance toward higher point value animals during the ongoing surveillance will elevate the points per sample substantially. Although one might argue that the points per sample should be higher or lower than the data suggest, APHIS is unaware of evidence documenting inaccuracies in the collection of cattle ages and believes that the best estimate should be derived from the existing data. Furthermore, we are unaware of valid techniques (e.g., tooth calcium deposition rings) for aging cattle other than via dentition or use of animal identification and production records. However, APHIS agrees that a national animal identification system could assist in documenting cattle ages in the future.

These comments seem to imply a need for extreme conservatism in determining a sample size. APHIS believes its calculations are sufficiently conservative for this stage of surveillance, but we will continue to monitor our surveillance activities and make adjustments as needed.