REPLY TO PEER REVIEW COMMENTS FOR FSIS RISK ASSESSMENT FOR GUIDING PUBLIC HEALTH RISK-BASED POULTRY SLAUGHTER INSPECTION

In June and July 2012, the 2011 *FSIS Risk Assessment for Guiding Public Health Risk-Based Poultry Slaughter Inspection* was independently peer reviewed under a contract with the Research Triangle Institute in accordance with the Office of Management and Budget peer review guidelines.[1] A list of peer reviewers is found in Appendix I; and the charge to the reviewers is found in Appendix II. Based on this peer review, the November 2011 risk assessment has been revised.

Below are itemized replies for each of the four peer review comment documents received for the *FSIS Risk Assessment for Guiding Public Health Risk-Based Poultry Slaughter Inspection.* Though slight editing was done to the peer review comments for corrections in spelling and grammar, reviewer comments are otherwise reproduced in this document verbatim.

Itemized FSIS Replies to Reviewer #1

Reviewer #1's comments:

The risk assessment uses generally appropriate data, with the exception of the attributable fraction values. It uses appropriate probability models (logistic regression and prevalence:expected incidence proportionality). The regression analysis has been done thoroughly. The Report is well written, focused on informing the decision questions, and sufficiently thorough for the intended audience.

FSIS Response: please see below for specific response to the issue of attributable fraction values used.

In my view, the Model needs to be written in a different environment. It is currently a Monte Carlo simulation model written in Excel/@RISK. This modeling environment is only capable of producing forward forecasts, and is not capable of normalizing the model to the observed data. This is important because one is forecasting two parallel models and finding the difference in their results. The first model is forecasting a version of the current state in which all samples are post chiller. However, much of the original data is post-chiller and therefore known (no uncertainty). A simple Monte Carlo model is not capable of anchoring a forecast to known values.

My recommendation is that the model is rewritten as a Markov Chain Monte Carlo model that does not have these limitations. The regression, forecast, and anchoring to known data can all then be performed together. Aside from being more logically correct, this should result in a narrower uncertainty in results and therefore provide a clearer guidance for decision-makers. I think that the regression model should have fewer parameters, in particular quarterly indices (spring, summer, fall, winter) instead of the many monthly indices that have no seasonal structure.

FSIS Response: although we understand the reviewer's concern, practically it is extremely difficult, if not infeasible, to move this complex problem into an MCMC framework. Reviewer 1 suggests that the MCMC modeling approach would be more logically correct and result in narrower uncertainty. However, the current FSIS model is simulating the baseline and alternative scenarios in parallel, so it is not clear that the FSIS modeling approach is logically incorrect. The recommended anchoring approach can be useful, but a decision not to use an anchoring approach is not fundamentally incorrect. (Hanley JA **, *1982)*Terms**

'Model' refers to an Excel spreadsheet model provided to me titled 'PSRA RA 2012 Review - new models-41_changeAnalysis (7wtd)'

'Report' refers to pdf document titled 'FSIS Risk Assessment for Guiding Public Health-Based Poultry Slaughter Inspection' prepared by Risk Assessment Division, Office of Public Health Science, Food Safety and Inspection Service, U.S. Department of Agriculture and updated November 2011.

'SAS Code' refers to Word document titled 'PSR RA 2012 Review SAS Code and Output - Four Logistic Models'.

1. Evaluate if the overall approach for modeling the public health benefits potentially realized from the change in inspection system examined is fundamentally sound.

   a. Is the overall approach used in the analysis to evaluate the linkage between inspection activities and potential reductions in annual human illnesses fundamentally sound? The regression model used to estimate changes in establishment prevalence should be addressed separately from the model used to estimate reductions in annual human illness.

   b. If not fundamentally sound, in each case, what problems exist and how should they be addressed?

Response

The technical evaluation of the model is somewhat involved, so I have added it in an Appendix. The answers provided here are a brief summary.

a. The general approach is sound, but the implementation of the logistic regression component is incorrect because it does not anchor the comparison between current and alternative states to the observed current state. This creates too much uncertainty in the results. I also question the use of some of the explanatory variables.
FSIS Response: we recognize this problem with the November, 2011 version of the model, and have corrected it in the November, 2012 version of the model by incorporating the entire range of input data for the explanatory variables into our forecasted estimates for current and alternative states. As a result, our uncertainty estimates for predicted changes in attributable human illnesses due to increased off-line inspection activities has tightened.

b. The anchoring needed cannot be achieved with the current Monte Carlo simulation approach. However, without any alteration of the assumptions or the data, I am confident that it can be implemented using a Markov Chain Monte Carlo (MCMC) approach using the free software OpenBUGS (http://www.openbugs.info/w/). SAS also has an MCMC capability, but I am unfamiliar with this program so cannot comment on whether it would be able to perform the same analysis.
FSIS Response: We were able to correct the error of propagating the mean through a nonlinear model. as described by the reviewer, without changing modeling platforms. The recommended anchoring approach can be useful, but a decision not to use an anchoring approach is not fundamentally incorrect

MCMC is a Bayesian approach, which means that one needs to define uninformed priors for each parameter of the model that needs to be estimated. The potential influence of the choice of prior distributions on the final results should be minimal as there are substantial amounts of data.

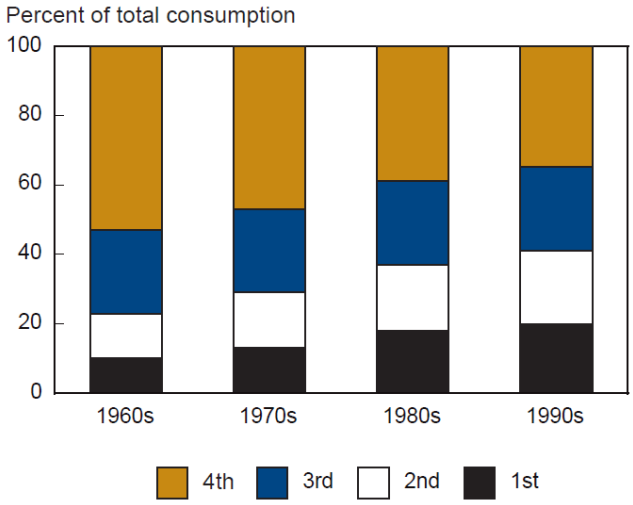
I have discussed in the Appendix why I think some of the variables should be removed, in particular the often large array of month variables. In my view, these may be replaced by up to four season variables if the prevalence of contamination shows some seasonal pattern (this may not be the case). Incidence rates of salmonellosis and campylobacteriosis have a strong seasonal pattern with a peak in the summer months, probably due to changes in the way people prepare their meals during the summer (barbecues, etc. where then is less control of hygiene) and, for *Salmonella*, the greater potential for growth as food is left in an unrefrigerated

environment. Unlike *Salmonella*, *Campylobacter* do not survive freezing, so only fresh meat is of relevance. Although turkey has become more popular year-round, it is still widely consumed in the United States at Christmas and Thanksgiving, as shown in the following graph from the USDA:

FSIS Response: It is not readily apparent that the model suffers from over-parameterization. Note also that Reviewer 2 argues that the model is under parameterized. The reasons for not replacing the monthly parameterization used originally in the regression model with the shortened seasonal parameterization suggested have been detailed in the risk assessment. This is a new analysis that was not included before. The analysis uses a statistical argument that indicates that the original monthly parameterization is superior to the abbreviated seasonal parameterization suggested. It is realized that the seasonal parameterization seems more logical because the data has a seasonal component. However, because the data used for the model has a more complex time structure than four seasons repeated over the years of the study this parameterization was not used. The data structure is complicated by using baseline data that includes post-chill and rehang and sampling verification data that includes only post-chill data. This means that prevalence is parameterized for rehang, post-chill, and the average over time. Due to this asymmetry the time variable was evaluated as a cyclic quarterly (seasonal), cyclic 12-month, and monthly time series. In order to evaluate the model for too many parameters the balance between increased variance explained by each parameterization and a decreasing validation statistic but only after the test for a logistic distribution of the estimated prevalence was passed. Using the newly added AIC, BIC, R-squared, and validation statistics now described in the risk assessment appendix the monthly parameterization originally used is shown to be the best parameterization that describes the data. In addition, the reasonableness for using the monthly categorization results from the observation of the variability in the months within each year of the study. There is an inconsistent but observable pattern to the monthly averages for prevalence over years not obvious from the individual parameter estimates. A time series analysis reveals that there is a weak repetitive pattern in the monthly data characterized by weakly consistent peaks and valleys. The reviewer points out the parameter inconsistency between the chicken and turkey Salmonella data at months 27, 29, and 33 mainly due to uncharacteristically low turkey Salmonella parameter estimates. However, recall that the monthly parameter estimates have been derived from a maximum likelihood estimation algorithm that simultaneously estimates all of the model coefficients. The low monthly parameter estimates have been adjusted by other variables in the model and do not represent to actual mean values in the data. These parameters are relative estimates to month 39 which corresponds to

September 2010. The extreme variation in the turkey monthly parameters at months 27, 29, and 33 has been verified to correspond to the economic downturn that had a severe effect on the turkey market in the 2009 holiday period. Similar matching of the monthly chicken parameters to economic data reveals similar but less obvious negative values related to the recession in 2008 and 2009. Therefore, it was rationalized that the data are unique to the period of analysis and should be modeled as such. Also, using a rank correlation estimate between the chicken and turkey monthly parameters is misleading because there was no expectation that chicken and turkey salmonellosis prevalence should exactly coincide. And there is the additional possibility of and ecological fallacy due to the fact that the actual sample size is much greater than 38 if the individual data points rather than the mean estimates are taken into account. Therefore parameter rank correlation may not be the best way to evaluate parameter significance in this case.

Figure 17
Per capita turkey consumption by quarter



Source: http://www.ers.usda.gov/media/490175/aer807g_1_.pdf

The nature of the implied Poisson process used for calculating the expected change in illness rates

2. Evaluate the complexity of the model in areas where the reviewer identifies limitations, weaknesses, or inadequacies; the reviewer must provide alternative data, data analysis, and/or modeling approaches.

   a. Is the model too complex, or not complex enough, to adequately address the risk management questions?

   b. Is the model over- or under-parameterized?

   c. Does the model adequately characterize the uncertainty present?

   d. Is variability sufficiently addressed?

Response

A more detailed analysis of the Model is provided in the Appendix. The answers provided here are a brief summary.

a. The Model is not too complex. The level of complexity is, in fact, quite small compared to other farm-to-fork models that one might have created. The relative simplicity of the model, making few assumptions and focusing on the specific problem, is a key positive attribute of the Model.

b. In my view, the regression model is over-parameterized, as explained in the Appendix.

FSIS Response: As explained above the temporal parameterization used originally in terms of different monthly indices was shown to provide the best statistical fit to the data. Because of this the amount of variability explained by the model is not exaggerated. Further, it is not readily apparent that the model suffers from over-parameterization. Note also that Reviewer 2 argues that the model is under parameterized.

.

c. From a statistical viewpoint, the way that the Model estimates the change in illness rates with new policies exaggerates the amount of uncertainty present given the

regression results. A revised regression analysis with fewer explanatory variables may produce more statistical uncertainty, but I think the correct implementation of the predictive part of the model that anchors to the observed data may well produce less uncertainty than currently presented in the predicted effect of the analyzed policy changes.

FSIS Response: the predictive part of the model now produces less uncertainty from the policy changes. The regression, analysis, however, was not altered from the Nov 2012 version.

d. In terms of the temporal effects, the use of many different month indices exaggerates the variability. Otherwise, variability has been sufficiently addressed.

3. Evaluate whether the model source code and mathematics are correct. If not, the reviewer must provide alternative modeling techniques.

   a. Are the modeling techniques (model mathematics and equations) appropriate?

   b. Are the methodologies used in the risk assessment for estimating parameters from the data appropriate (i.e., follow scientifically accepted methodologies)?

   c. Are the data analyses and source code accurate?

Response

A more detailed analysis of the Model is provided in the Appendix. The answers provided here are a brief summary.

a. Logistic regression and the ratio approach relating prevalence and illness rates are appropriate. The Monte Carlo simulation approach that interprets the statistical uncertainty in the logistical regression and predicts the effect of policy change is not appropriate, as described in the Appendix.
FSIS Response: we have modified the use of regression parameter estimates and the explanatory data in the MC simulation to properly interpret the statistical uncertainty in the logistic regression as suggested by the reviewer.

b. Logistical regression is the appropriate method for estimating prevalence. I question the choice of explanatory variables, specifically the month indices, since there is no causal argument underpinning this.

FSIS Response: see our earlier response.

The estimate of the current illness rates is based on a previous FSIS analysis which, in turn, is partially based on outbreak data. This previous analysis very significantly underestimates the amount of illness attributable to chicken and turkey. I have explained this in more detail in the Appendix.

FSIS Response: Our choice of attribution fractions was based on consistency and transparency. The fractions cited here are consistent with attribution fractions used in previous analyses and the development of those fractions is transparently explained in the referenced material. Nevertheless, we recognize there is substantial uncertainty about the true attribution fractions for chicken and turkey. As the reviewer explains, there are other attribution fractions developed from other countries or approaches that do not necessarily match those reported here.(References:

EFSA (European Food Safety Authority). 2008 A quantitative microbiological risk assessment on Salmonella in meat: Source attribution for human salmonellosis from meat.

EFSA (European Food Safety Authority). 2011. Analysis of the baseline survey on the prevalence of Campylobacter in broiler batches and of Campylobacter and Salmonella on broiler carcasses in the EU, 2008.

According to Reviewer 1, the effect of increasing the attributable fraction of illness due to poultry would be to increase the magnitude of the predicted reduction in illnesses. However, the prevalence ratio (prev(policy)/prev(baseline)) is independent of the baseline number of illnesses. Because the probability of increased illness = prob(prev(policy) > prev(baseline), it remains the same regardless of the magnitude of the estimated baseline number of illnesses. In this instance, the probability of increased illness is arguably a more informative summary statistic than the mean number of illnesses avoided.

, It should be noted that, in cases where illnesses avoided is negative (i.e., when illnesses increase following implementation of the rule), the number of increased illnesses will also be larger if attribution fractions are increased. Because the

probability of increased illnesses is generally small for most scenarios, the implication of such effects is also small.

c. I do not have access to the original data used in the regression analysis, nor would I be able to verify if those data were accurate if I had them. The binomial data reflect whether the pathogens in question were *detected* in each sample. Pathogens may be present but undetected. This is a general problem faced in food safety risk assessment and not a criticism of this analysis. *Salmonella* and *Campylobacter* reside in fecal matter on the carcass, i.e. they are clustered not homogeneously distributed over the carcass, so one cannot know with any certainty whether the carcass is contaminated when it tests negative. One cannot even say with certainty that the level of contamination is low in test-negative carcasses (i.e. that there are few enough bacteria to make the expected number of illnesses the carcass could produce very small relative to carcasses that test positive). The Model therefore implicitly assumes that the expected (mean) number of illnesses is proportional to the *observed* prevalence of contamination – a common assumption since one has little in the way of alternatives. This assumption has relatively little impact

FSIS Response: we agree with the reviewer – this model is a prevalence-based model.

4. Evaluate whether adequate sensitivity analysis has been provided. If not, the reviewer must provide an alternative approach or application for sensitivity analysis and/or identify those parameters that should have been included.

   a. Have the most important variables in the model been identified?

   b. Has an important variable been left out?

   c. Has the impact of including or excluding scientific studies or other data been adequately explored?

Response

Sensitivity analysis has been performed on all variables.

a. In terms of modeling, all uncertainties I can think of have been included. A Tornado plot may have been a helpful addition to allow the identification of the

most influential uncertainties. The author(s) have performed several experiments, including random splitting of data, to investigate the robustness of their analysis.

FSIS Response: the Nov 2012 version now includes a Sensitivity Analysis section in the results, including a tornado plot of most significant influences.

b. No

c. I cannot comment in terms of the regression data, but imagine that there were no alternatives. In terms of the illness rate data, I think they made a poor choice because the attributable-fraction estimates are too low, as explained in the Appendix.

FSIS Response: The attributable fraction values that are used in this risk assessment should be similar to the estimated values provided by CDC, when they are eventually published. They are also similar to the values estimated from Canadian data (Ravel et al. 2009). We understand that attributable fraction estimates derived from expert elicitation studies in the U.S. have consistently been higher than the values used here. Nevertheless, additional analyses comparing patterns in human illnesses to patterns observed in all FSIS regulated products (meat and poultry) suggest that the lower attributable fraction estimates used here are probably more accurate than higher attributable fractions values given in other studies. For these reasons we are more comfortable using the lower attributable fraction values.

Ravel, A., Greig, J., Tinga, C., Todd, E., Campbell, G., Cassidy, M., Marshall, B., & Pollar, F. 2009. Exploring historical Canadian foodborne outbreak data sets for human illness attribution. Journal of Food Protection, 72, 963–1976.

5. Evaluate the available data and the underlying assumptions used in this risk assessment. Are they complete and correctly analyzed and interpreted? If not, the reviewer must provide additional data sources and citations (where appropriate) or provide alternative interpretations, analysis, or suggested use of the data.

   a. Have all key studies and data been identified?

   b. Have the data been correctly interpreted, analyzed, and used in the risk assessment?

Response

a. As explained above and in the Appendix, better estimates of attributable fraction are available. The figures currently used are significant underestimate. All other data look to be appropriate.

FSIS Response: please see our previous responses above.

b. I believe that the regression data have been correctly interpreted, though have not seen the original data or how they were collected. I don't believe the regression data have been correctly analyzed as explained in the Appendix. I don't believe the resultant use of the regression coefficient estimates has been correctly used in the risk assessment, again as described in the Appendix.

FSIS Response: we have modified the use of regression parameter estimates and the explanatory data in the simulation to predict changes in attributable human illness to properly interpret the statistical uncertainty in the logistic regression as suggested by the reviewer.

6. Evaluate the regression analysis used to estimate baseline and scenario aggregate establishment prevalence.

   a. Is the technique accurately described, utilized, and appropriate for its intended use?

   b. If not, reviewer must provide rationale for why not and detail better alternatives.

   c. Are the conclusions drawn from the regression analysis appropriate?

   d. If not, reviewer must provide alternative interpretation of the results derived from this analysis.

Response

a. The logistic regression is accurately described and utilized. Logistic regression is appropriate for the problem, but the SAS method used does not fit with the needs of the model, as described in the Appendix.

FSIS Response: The logistic model parameters were not altered. But, in order to provide more accurate prevalence estimates the prevalence estimating equations were reformulated to model the parameter interdependence. Therefore, in order to make the logistic regression more appropriate to the problem the parameters were modeled to be dependent rather than originally independently modeled. Modeling the dependence structure involved using the SAS estimated variance-covariance matrix in the estimation procedure. The dependent logistic regression model used a multivariate normal distribution for the covariance structure of the model. In addition to considering the dependency in the model the average estimates originally used in the prediction equations were replaced by using the original data in making the prevalence estimates. This required iterating through each entire dataset for each random multivariate normal parameter set. The result of each iteration through the dataset provided a prevalence estimate that was weighted with each establishment's daily production volume. The final prevalence estimate over 100,000 iterations provided the final prevalence estimate from the regression model. This sufficiently addressed the model deficiencies addressed in the reviewer's appendix.

b. See Appendix.

c. No, I don't think so, because a number of explanatory variables used do not make sense to me (the month indices). A seasonal index would be more useful.

FSIS Response: please see our previous responses to questioned use of too many structural parameters in the regression model.

d. See Appendix.

7. Evaluate the scenario approach taken to quantify changes in establishment prevalence due to additional off-line inspection activities.

   a. Is this scenario approach reasonable, given the limited amount of data available?

   b. If not, what flaws do you perceive in the rationale and what information is lacking to make the case as proposed?

   c. What alternatives exist and how could they be incorporated?

Response

a. The scenario approach is reasonable both because it has the potential to answer the set of posed questions, and because it is practical given the limited amount of available

data. It makes relatively few assumptions. I cannot think of a better way to provide the analysis given the scope and constraints.

b. N/A.

c. I don't see any alternatives given the limited knowledge available about the steps between the slaughter plant and consumer illness (e.g. preparation, storage, cooking, dose-response by sub-population, etc.). Other risk assessments in both Europe and the USA have shown a linear relationship between prevalence and expected incidence, so the method used here skips these steps without any loss of accuracy.

FSIS Response: we agree.

8. Evaluate whether the documentation, discussion, and interpretation of results is appropriate. If not, the reviewer must provide an alternative outline and/or approach for adequately and clearly documenting this risk assessment.

a. Is the report clearly written?

b. Is it complete?

c. Does it follow a logical structure and layout?

d. Is it useful?

e. Does the risk assessment support the conclusions reached?

Response

a. Yes

b. Yes, though I would have liked to see some graphical representation (spider or tornado plot) of the contribution of each uncertain component to the output uncertainty.

FSIS Response: we have now included this graphical representation in a sensitivity analysis section of results.

c. Yes.

d. Yes, because one can easily follow what was done.

e. The risk assessment makes no conclusions as to the decision that should be made, which is appropriate. It does give a good description of the results and how they inform each question that was asked. It also provides a good and objective description of the vulnerability of the analysis.

Appendix

Technical review of the human health incidence component of the model

The formula for estimating the reduction in human illness is:

The basis for this formula is:

where is interpreted as the mean of a Poisson distribution. This is an approximation that will be very precise when:

a. There is no cross-contamination between food items after the measurement point. For post-chiller data this is more appropriate than for pre-chiller data where the chiller is a water bath, since bacteria can be spread between the carcasses at that point;
b. The prevalence estimate is accurate. This comes down to the sensitivity of the methods used to detect the presence of bacteria;
c. The proposed changes to inspection do not significantly alter the load distribution on the carcasses that pass inspection. For example, if heavily contaminated carcasses are more likely to be removed with a proposed change in inspection, the above formula will underestimate the human health benefit;
d. The illnesses that occur are sporadic, rather than in outbreaks. This is appropriate for *Campylobacter* more than for *Salmonella*, since *Campylobacter* are thermophilic and tend not to grow outside the host animal, whereas *Salmonella* can grow in the environment;
e. The bacteria do not create reservoirs. Again this is more appropriate for *Campylobacter*;
f. There is no seasonal effect that makes the illness rate change for the same prevalence; and
g. The attributable fraction refers to the illnesses that can be attributed to domestically reared chickens and turkeys, since the risk from imported poultry meat would not be affected by the proposed control changes.

If we relax the interpretation of as the mean of a Poisson distribution, but instead just describe it as the mean number of illnesses that may occur, then conditions d. and e. no longer apply. Condition f. can be accounted for by applying four separate equations for each season, and adding the results together:

In fact, if a regression analysis with quarterly indices shows no significant variation in prevalence by quarter attributable to the season, the first simple version of the model can be used:

because of the additive property of a Poisson process, i.e. that Poisson(a) + Poisson(b) = Poisson(a+b).

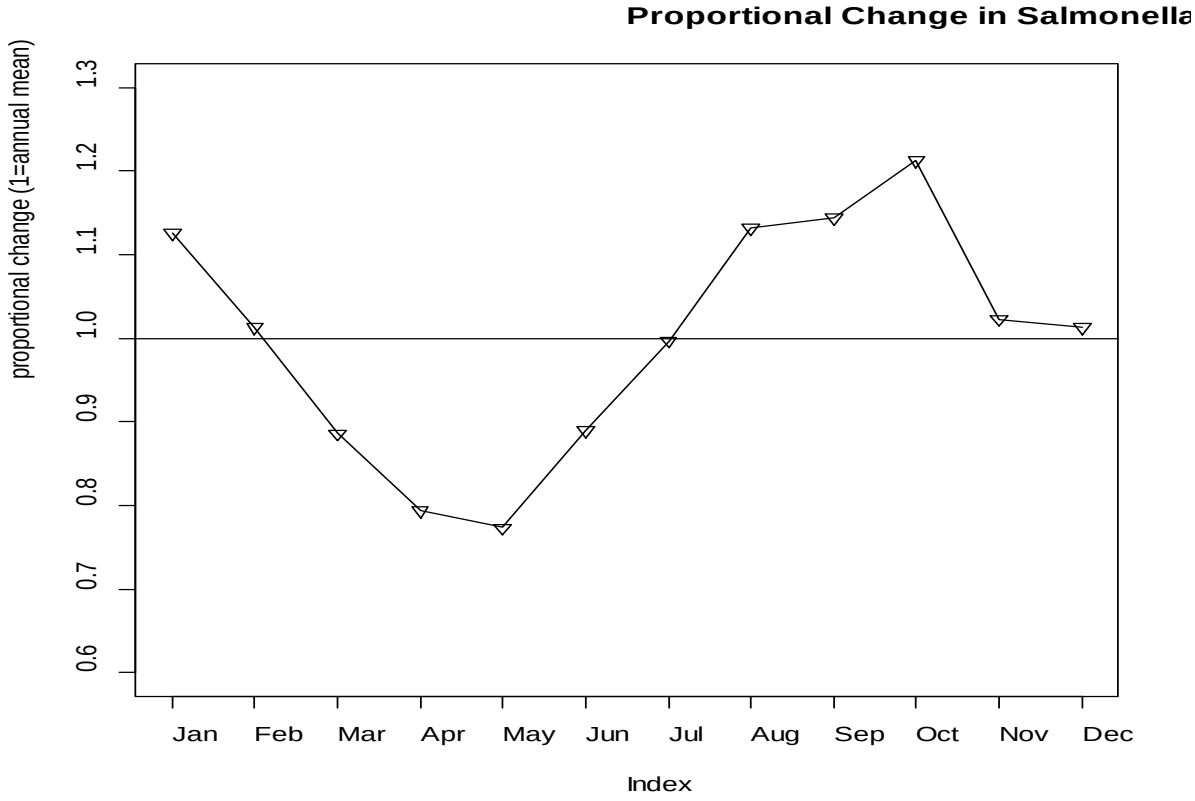
Recent simulation model studies for *Campylobacter* and *Salmonella* for EFSA have shown that despite the theoretical non-linearity between and because of non-linear dose-response relationships and cross-contamination, the linear approximation holds well if one accounts for seasonal effects. In practical terms, there isn't any alternative to this formula anyway given the scope of the assessment without building a far more complex, assumption-laden and data deficient farm-to-fork model. In general, the approximation used by this formula would be very good *provided any seasonal effect is accounted for*.

It is important to note that, in this model, is the fraction of all carcasses produced in the United States that are contaminated with the pathogen in question. The distinction between this and the binomial probability estimated by the logistic regression is described later.

Seasonal effects are important for both *Salmonella* and *Campylobacter*. Studies in the USA and the EU show that prevalence of contamination does not vary greatly by season of the year. However, there is a marked increase in campylobacteriosis and salmonellosis incidence rates during the summer months. These are currently not accounted for in the model. In fact, they would not have to be accounted for if the logistic regression did not include months as factors.

FSIS Response: This collection of comments raises some questions regarding the underlying assumptions of the basic modeling approach. The reviewer outlines the necessary conditions for the assumptions of the model to be valid (e.g., the reviewer states "If we relax the interpretation (of the model) as the mean of a Poisson distribution, but instead just describe it as the mean number of illnesses that may occur, then conditions d. and e. no longer apply. Condition f. can be accounted for by applying four separate equations for each season, and adding the results together…).

In general the reviewer comments are supportive of the modeling structure with the exception of the seasonality issue. The reviewer points out the seasonal change in *Salmonella* illnesses (a rise during the summer), but states that a seasonal fluctuation is not seen in pathogen contamination on poultry. This, however, is the not a correct assertion for US-produced poultry. A separate time series analysis of FSIS HACCP testing data shows a strong seasonal component to *Salmonella* contamination in poultry. To illustrate, consider the figure below that demonstrates the seasonal fluctuation in the proportion of test-positive young chicken carcasses.



This seasonal pattern matches the seasonal pattern for human illnesses. While the risk assessment does not fully explore the effect of the seasonal pattern on the underlying model, a

similar analysis of the seasonal pattern observed in *E. coli* O157:H7 and ground beef found that the *P*(*ill*|*exp*) term in the dose-dependent model (Williams et al. 2010) is essentially constant across all months so, for example, *P*(*ill*|*exp, annual*)= *P*(*ill*|*exp, July*). This lends further support to the use of the prevalence-based model because the fixed nature of *P*(*ill*|*exp*) and the uniformly low levels of *Campylobacter* and *Salmonella* found on finished carcasses support the concept that the presence of the pathogen at the end of production is the primary driver of food-borne illnesses (i.e., seasonal changes in growth and attenuation during distribution and consumer behavior do not dramatically change the probability of illness given exposure throughout the year).

Williams, M.S., Withee, J.S., Ebel, E.D., Bauer, N.E. Jr., Schlosser, W.D., Disney, W.T., Smith, D.R., Moxley, R.A. 2010. Seasonal occurrence of Escherichia coli O157:H7 in live cattle, ground beef and humans. Foodborne Pathogens and Disease. 7:1247-1254

Technical review of the logistic regression component of the model

The logistical regression is an appropriate and widely used method for statistically evaluating the factors that affect prevalence. The regression analysis appears to have been executed correctly from a statistical analysis viewpoint in terms of the estimation of the coefficients of the assumed regression equation, though I am not familiar with the SAS program to be able to identify any omissions from the script provided to me.

In my view, there are a number of important issues related to how the results of the regression analysis have been implemented. In order to explain these issues and help implement a corrected version, I begin with some explanation of logistic regression and its relationship to prevalence.

FSIS Response: The reviewer has provided many insightful and useful comments in this technical review. Two general insights prompted changes in the modeling approach. First, to address the non-linear conversion of the logit to prevalence, we now integrate the probability of a positive sample across the entire data set to generate a population prevalence. Second, to address excess variability in the predicted logit, we now model vectors of beta coefficients using the variance-covariance structure estimated from the regression. To implement this approach, we used the Cholesky decomposition method to model a multivariate Normal distribution for the beta vector. These changes are explained in the Nov 2012 revision of the risk assessment report.

FSIS Response: The consequences of these changes are the elimination of the "bath tub" shaped prevalence distributions shown in the reviewer's comments and a general reduction in the variance of estimated baseline and post-policy prevalence.

FSIS Response: This technical review includes a discussion of the logistic coefficients of the decision variables (i.e., SP, U, SNP, NC). The reviewer may imply that the spread of these coefficient – that in some cases overlaps zero – is not ideal, but this phenomenon is responsible for the occasional prediction that prevalence might increase following changes in these variable and illnesses will correspondingly increase. Therefore, we have not made changes to reduce the influence of these.

FSIS Response: A more complete explanation of the continuous and categorical structural variable selection method is now given in the risk assessment text and appendix. This provides more convincing statistical evidence for the rationale of including the large number of structural parameters in each model. There is a special focus on the determination of the number of categorical monthly parameters which is additionally explored in additional comments in the reviewer's appendix below. The arguments presented allowed that no changes were made to the original model parameter estimates such that all four original models have retained the same number of structural parameters.

The logistic function

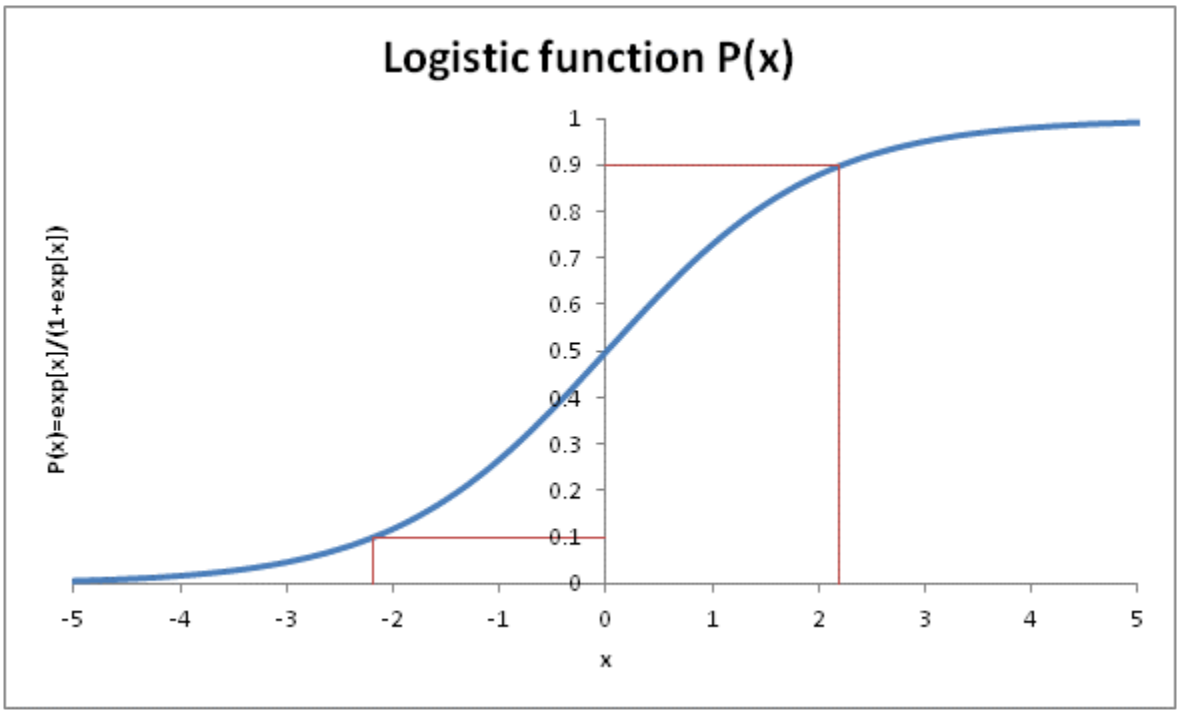
The logistic function takes the form:

$$\frac{\quad}{\quad}$$

The inverse equation is:

$$\frac{\quad}{\quad}$$

The following figure plots out this function:



It is symmetric so, for example, P(-2.19722) = 0.1 and P(2.19722) = 0.9 (i.e. 1 - 0.1) as shown by the red lines. When x is below -4, P(x) is close to 0 and when x is above 4 P(x) is close to 1.

How logistic regression should be used

Logistic regression attempts to evaluate the factors that affect whether an individual will have some condition or not, in this case whether or not a carcass will be contaminated.

The logistic regression is composed of the following logic:

---

where the observed values for one individual carcass *i* are (= 0, not contaminated or 1, contaminated) and , the values of the factors associated with the *i*th carcass.

The observed prevalence is then:

673

674

675 where *m* is the number of carcasses in the data set. If *m* is sufficiently large, as is the case
676 here, it is a good approximation to the prevalence that would be observed if all carcasses
677 in a year were tested. An extension to this model is possible where each observation is
678 given a weighting that relates to its relative importance, e.g. the fraction of a population
679 that the sample represents.
680

681 **How logistic regression is used in the Model**

682 From the SAS Code the logistic regression appears to have been carried out correctly,
683 though I am not familiar with this program. It uses slaughter volumes in each plant to
684 weight observations (Report, page 34). The Model then estimates the prevalence as
685 follows:
686

687

688 where are estimates from the regression model. The Model simulates these values
689 as independent, normally distributed with a mean equal to the maximum likelihood
690 estimate and a standard deviation equal to the standard error, both from the regression
691 results;
692 and where the values are fixed values. The Model does not specify where these fixed
693 values come from and provides no description (some description would have helped with
694 a review, in general the spreadsheet Model is poorly annotated), but it would most
695 logically be the weighted mean of all observations for that pathogen:poultry type pair.
696

697 This formula is incorrect, since it is mixing two different concepts – the logistic
698 regression seeks to explain the factors that effect the probability that an *individual* carcass
699 will be contaminated. In contrast, the Model uses the estimated coefficients to estimate a
700 *prevalence* for the carcass *population* by using mean values for each factor. In effect, it is
701 estimating the probability that a carcass of that bird:pathogen under the weighted average
702 circumstance will be contaminated. However, since the logistic function is non-linear it
703 cannot be used in this way.

704 **Non-linear effect of variance in the regression equation**

705 Consider the following simplified representation of the logistic model:
706

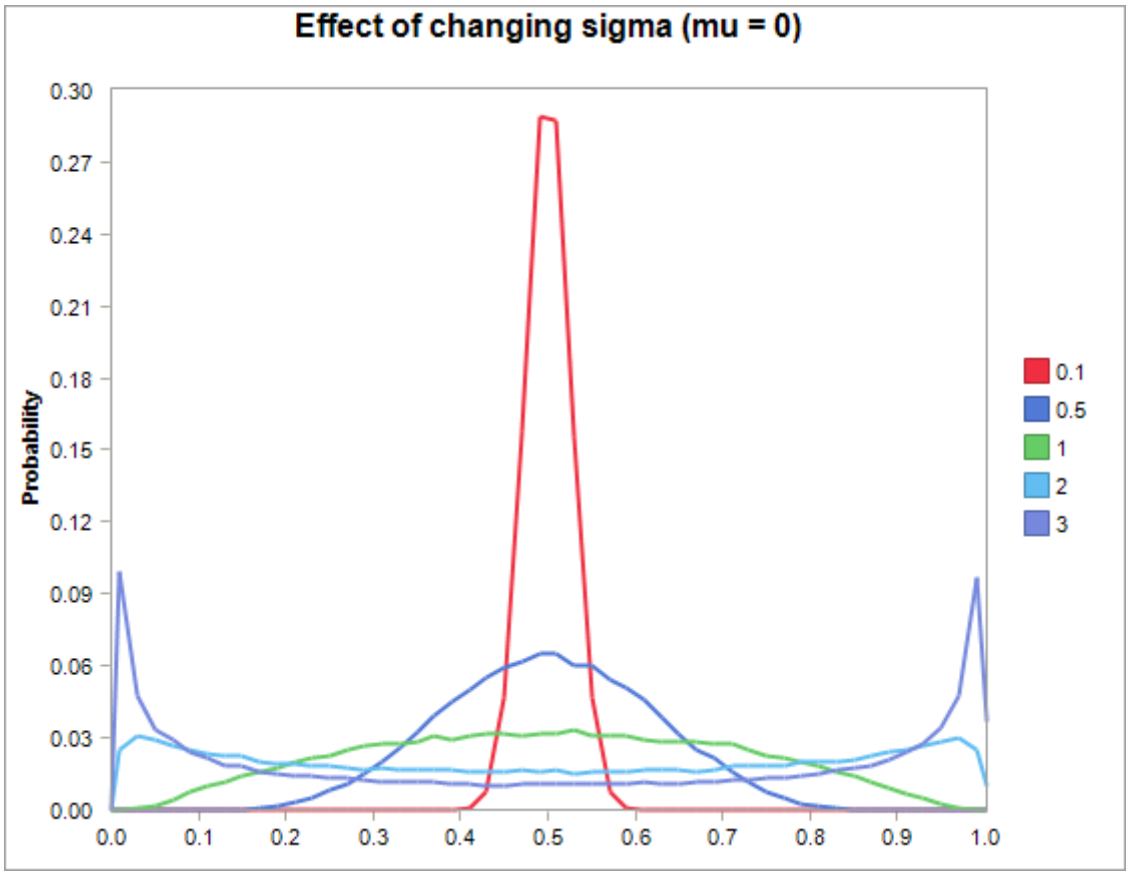
where is some prevalence, is the normal distribution mean (equivalent to in the notation of the Report) and is the aggregated variance of these terms.

This equation is the inverse of the logistic function, and the logistic function thus returns the estimate of .

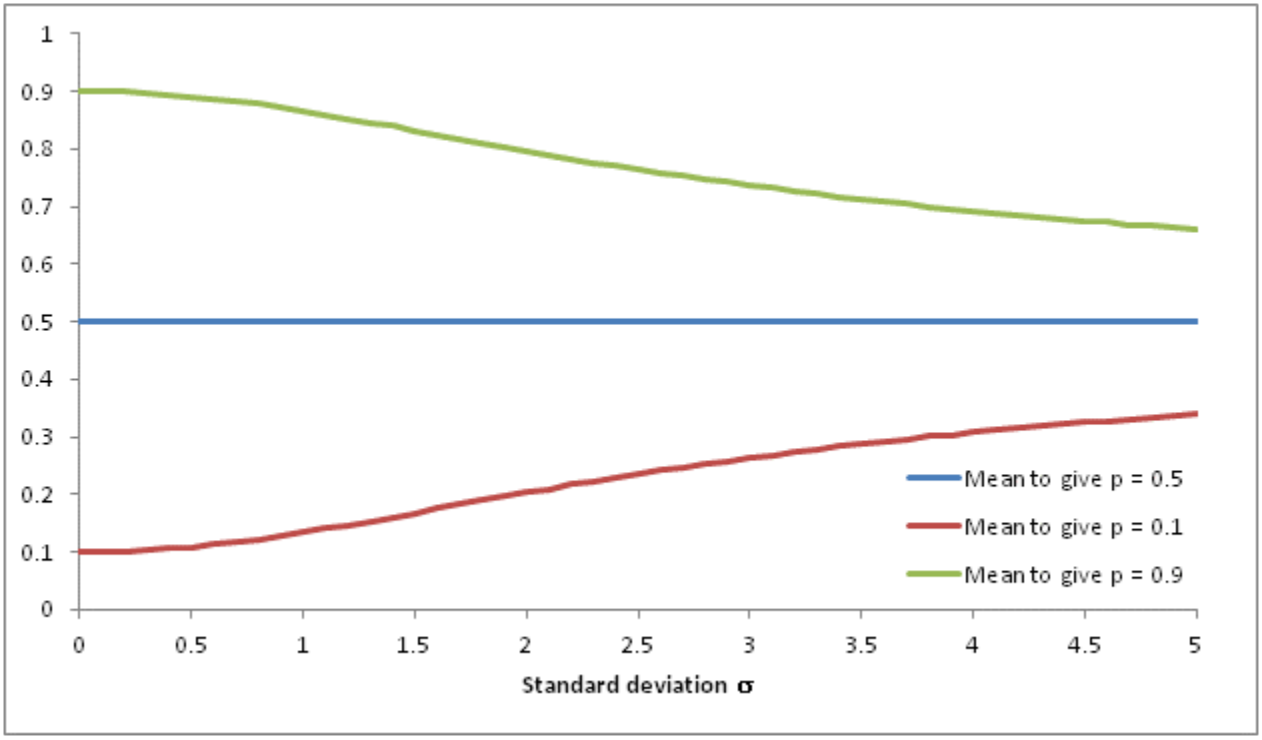
As increases the Normal distribution will have longer tails and thus greater probability of producing values large negative and positive values. From the form of the logistic function graphed above, one can see that this in turn will produce estimates of that have increasing probabilities near 0 and 1.

FSIS Response: The reviewer is making a relevant point regarding the error distribution that does apply to the logistic regression model in general. However, the normal distribution is not used as the error distribution because we modeled binary responses as individual Bernoulli trials as input data. The error distribution is not modeled as continuous but rather as a discrete binomial distribution. Asymptotically, the error distribution approaches the normal distribution so the reviewer's comments still have relevance. The distinction between the type of error distribution modeled is important when considering statistical inferences on the model parameters and the model significance. This is the reason that the logistic regression model used is characterized as based on quasi-likelihood estimation. This means correction has been made to the parameter error estimates which are based on deviance estimators that are different from the customary regression error estimates based on the normal distribution. The size of correction is based on the amount of disparity between the expected model error based on the logistic distribution and the observed model error termed overdispersion error. Additionally, perhaps unfamiliar statistics have been used for model evaluation. These are the Akaike Information Criterion (AIC), the Bayesian Information Criterion-Schwarz (BIC), the Hosmer-Lemeshow statistic, and the Nagelkerke R-squared statistic. These statistics account for the use of deviance estimators rather than the accustomed regression estimators used for normal distribution error based models. The basic reference for the logistic regression method used can be found at the SAS website:
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#logistic_toc.htm

The following plot illustrates the effect of increasing where has been set to 0. The red line shows that when is small (here = 0.1), the estimate for is roughly normally distributed with a small range. However, as gets larger the tails start to be constrained by the [0,1] extremities to the point where the estimate peaks at 0 and 1 and is concave in the middle.



Similarly, the *mean* (expected) value of is dependent on as shown in the following chart. In this plot, the three lines show the effect of for different values of . The three values of used are -2.19723, 0 and 2.19723 which give values of equal to 10%, 50% and 90% respectively when = 0. The reason for the *mean* estimate of value to move towards 0.5 with increasing is that the distribution is bounded at 0 and 1. So, for example, if = -2.19723 the estimate for is close to 0.1 for very small but as gets larger the left tail of the estimate of 'butts up' against the minimum of 0, while the right tail can spread way out before it 'butts up' against the maximum of 1. The median however remains constant at 0.1. That produces a right-skewed distribution for , and a commonly known probability identity says that a right-skewed distribution has a mean greater than its median, greater than its mode. Thus, as the skewness increases so the mean estimate of diverges further from the median estimate.



Thus, if one intends to pick a single value from the distribution of it is generally preferable to use the median rather than the mean.

Estimates of coefficients of decision variables

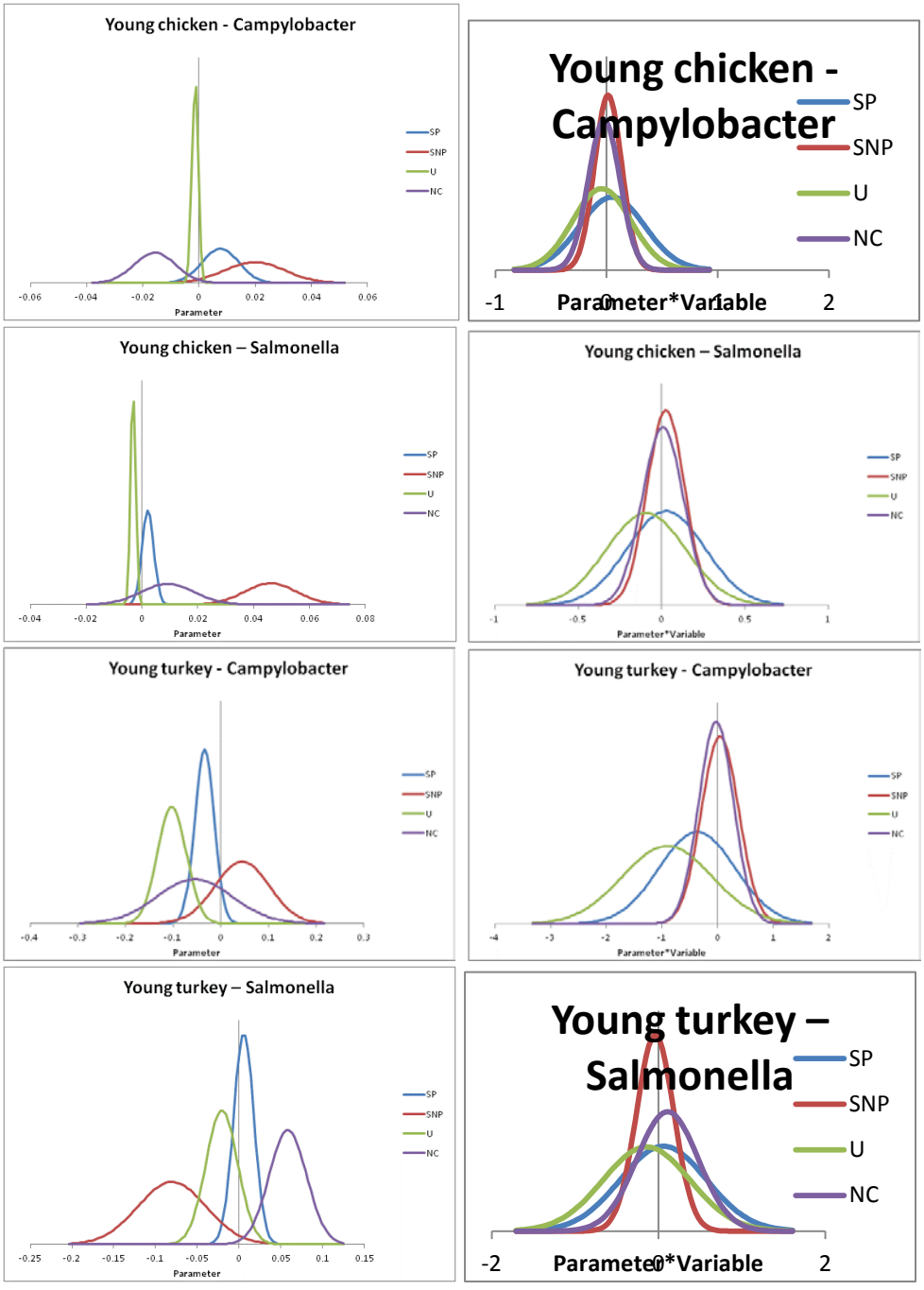
The four Model decision variables under consideration are labeled SP, SNP. U and NC. The robustness of their influence in the logistic regression is a key requirement for the Model to produce results that have value for the decision maker.

The following plots investigate the consistency and robustness of each regression coefficient estimate, using the same color key. The left hand graphs plot out the assumed Normal distributions of uncertainty for the regression coefficients. The right hand plots the Normal distribution of the [Coefficient*Variable].

In both types of plot, we should hope to see that the distributions for the same coefficient (or coefficient*variable) lie almost entirely to the left or right of zero (otherwise there is no sense to its influence) and lies on the same side of zero for all pathogen:meat type combinations.

The left plots show that the coefficient for U is consistently negative, and no other coefficient estimates are as consistent. Putting the spread of these coefficient estimates

into context, we multiply them by the variables they relate to for the base case. One can see in the right plots that this produces a fairly wide range of values, bearing in mind that a 0 +/- 1 for a logistic function give a prevalence range of 25% to 75%.

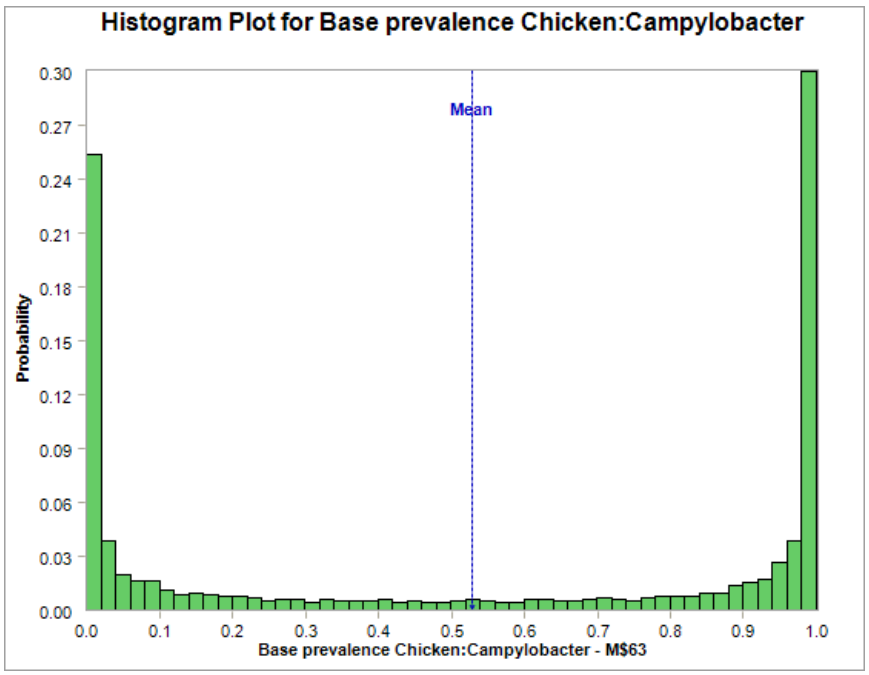


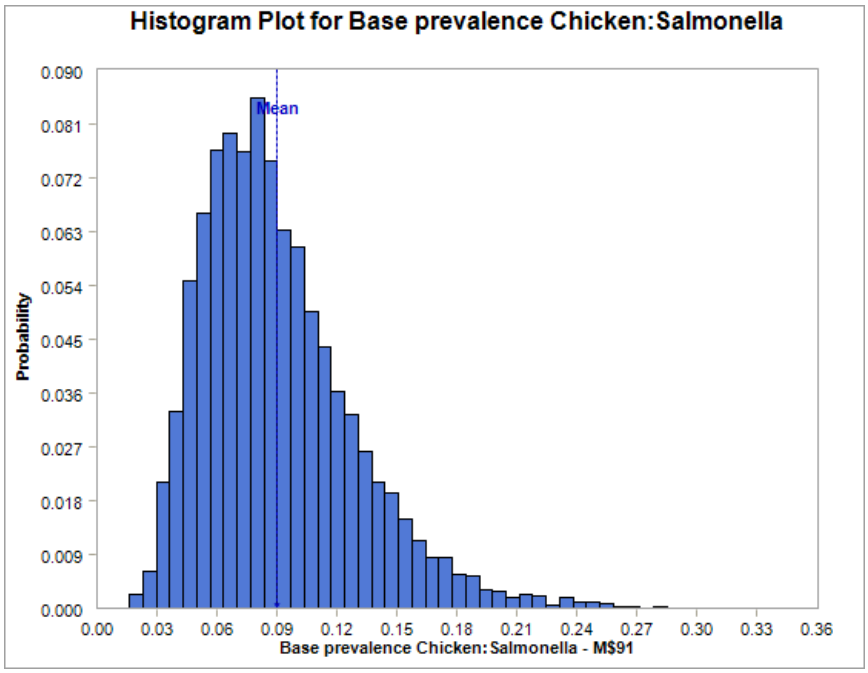
In the above parameter*variable plots, I have used the known identities:

Baseline estimates of prevalence

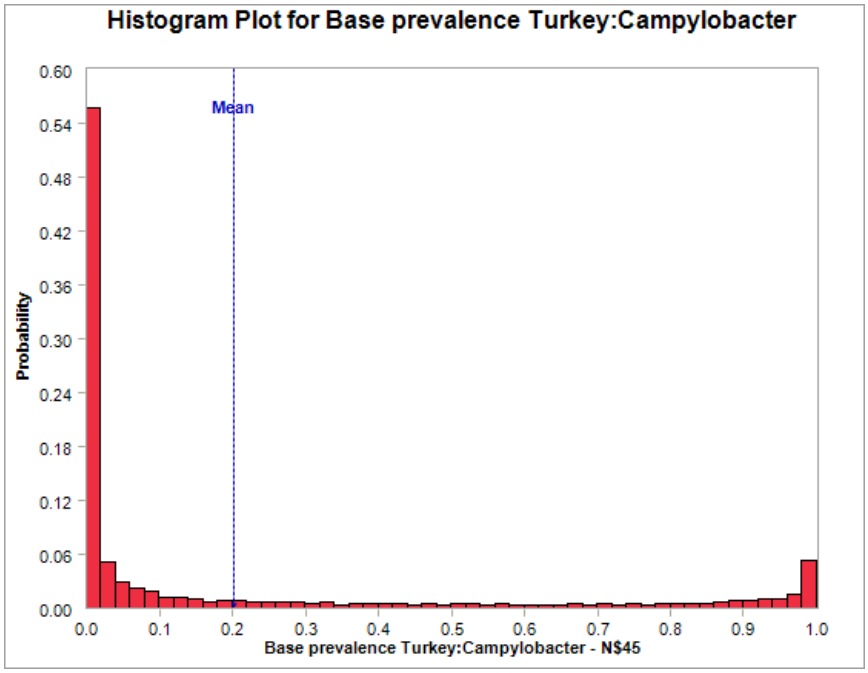
Bearing in mind the above description of the behavior of the logistic function when the input value follows a distribution, we can now look at the results generated by the Model when running the base case.

The Model describes an option to simulate the current state (base case) by replacing with a value of 1 all of the Pert distributions simulating the variation of the variables modeling the effect of policy change. Doing this, one gets the following results:

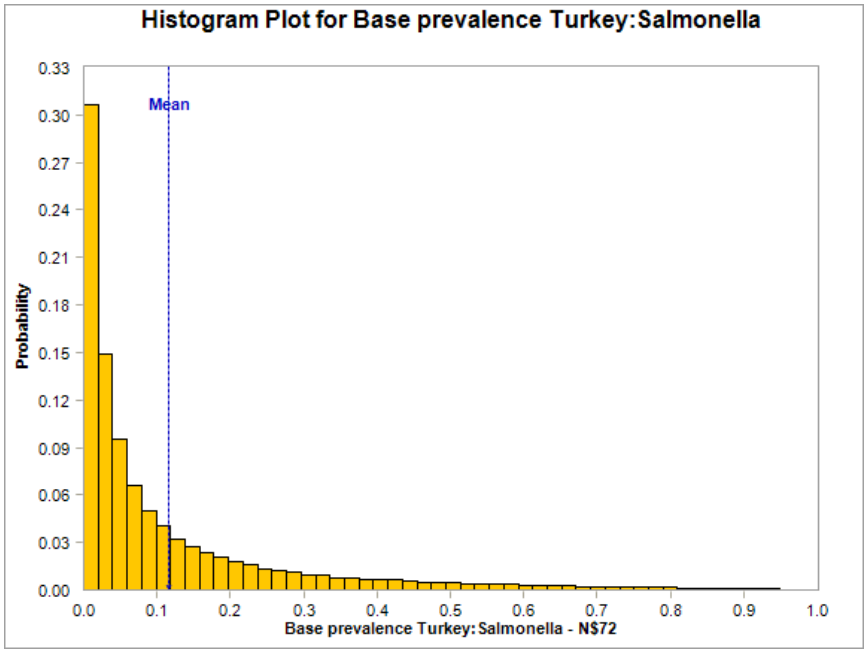




813



814
815
816



Bearing in mind that, as I understand the Report, all turkey samples are taken at rehang and for chicken between 50% (*Campylobacter*) and 85% (*Salmonella*) are taken at rehang. The above plots intuitively give a far greater uncertainty than would be generated in translating the remaining pre-chill data to rehang 'data' to produce current estimates of prevalence. ***FSIS Response: To the contrary the turkey baseline samples consist of 50% from rehang and 50% from post-chill. Therefore the reviewer's comment concerning uncertainty is questionable where this assumption is applied.***

The following table compares several prevalence estimates. *Observed* is the fraction of the relevant observations where a carcass is contaminated. *Deterministic* refers to the prevalence estimate using the Model if no uncertainty is included in the coefficient estimates (I have left the rehang parameters at the values provided in the Model, as there is some confusion which values should be used). *Sim Mean* and *Sim Median* refer to the mean and median respectively of the simulated distribution using the Model when the uncertainty in the parameters is included.

Base run prevalence	Observed*	Deterministic	Sim Mean	Sim Median
Chicken:Campy	73.33%	62.84%	52.88%	62.22%
Chicken:Salmonella	12.31%	8.39%	9.10%	8.41%
Turkey:Campy	11.89%	0.97%	20.15%	0.97%
Turkey:Salmonella	7.29%	4.75%	11.67%	4.78%

* The observed prevalence calculations are based on the following values taken from the SAS Code: Campy in chickens: 1:0 = 4809:1749 = 73.33%, *Salmonella* in chickens: 1:0 = 2790:19881 = 12.31%, Campy in turkeys: 1:0 = 343:2541 = 11.89%, *Salmonella* in Turkeys: 1:0 = 638:8111 = 7.29%

It is notable that there is a significant difference between the observed prevalence and any of the three estimates derived from the Model. I would have expected that Deterministic values to be either consistently more, or consistently less, than the Observed value due to the effect of changing the rehang parameter to 1, but that isn't the case. The order of magnitude difference for Turkey:Campy is particularly noteworthy. The *Deterministic* and *Sim Median* estimates are similar, as one might expect, but can be very different from the Sim Mean value.

Covariance structure of the error terms in the regression analysis has been ignored

Note 3 on page 19 of the Report states:

> We assume independence in the errors among the independent variables (i.e., we do not include covariance terms between these variables). The calculated standard error from the regression is somewhat smaller than the value as we have simulated it; this result suggests that the aggregate effect of any non-zero covariance terms is to reduce uncertainty in modeled forecasts. Therefore, our simple treatment increases uncertainty and is deemed conservative for that reason.

The Model describes an option to simulate the current state by replacing with a value of 1 all of the Pert distributions simulating the variation of the variables modeling the effect of policy change. Doing this, one gets the results plotted above.

Note 3 states that ignoring the covariance structure of the uncertainty in the estimates of the parameters increases the overall uncertainty and is therefore conservative. This is not, however, a general statement, since if the observed prevalence from historic data is greater than 50% the effect of exaggerating the uncertainty is to decrease the mean prevalence estimate.

More importantly, there should be very little uncertainty in the observed prevalence estimate when one simulates the current state because it is a known value. The great uncertainty actually shown occurs because the regression analysis looks backwards from

a data set to estimate the influence of different factors that produced the observed prevalence. If those factors remain constant a simulation model predicting what the observed prevalence would be 'next year' should only show the level of random variation that would occur in a binomial process and the translation of some samples from pre- to post-chiller.

Factors used in the logistic regression model

The factors used differ between each pathogen:poultry type combination as show in the following table (* denotes used and blank denotes not used, C=Chicken, T = Turkey, Ca = *Campylobacter*, Sa = *Salmonella*):

Factor	**C-**	**C-**	**T-**	**T-**
Intercept	*	*	*	*
rehang 0	*	*	*	*
loglinespeed	*	*	*	*
logemployees	*	*	*	*
lines	*	*	*	*
Himp	*	*		
month0 1	*	*		*
month0 2	*	*		*
month0 3	*	*		*
month0 4	*	*		*
month0 5	*	*		*
month0 6	*	*		*
month0 7	*	*		*
month0 8	*	*		*
month0 9	*	*		*
month0 10	*	*		*
month0 11	*	*		*
month0 12		*		*
month0 13		*		*
month0 14		*	*	*
month0 15		*	*	*
month0 16		*	*	*
month0 17		*	*	*
month0 18		*	*	*
month0 19		*	*	*
month0 20		*	*	*
month0 21		*	*	*

month0	22		*	*	*
month0	23		*	*	*
month0	24		*	*	*
month0	25		*	*	*
month0	26		*		*
month0	27		*		*
month0	28		*		*
month0	29		*		*
month0	30		*		*
month0	31		*		*
month0	32		*		*
month0	33		*		*
month0	34		*		*
month0	35		*		*
month0	36		*		*
month0	37		*		*
month0	38		*		*
District	5	*	*	*	*
District	15	*	*	*	*
District	20	*	*	*	*
District	25	*	*	*	*
District	30	*	*	*	*
District	35	*	*	*	*
District	40	*	*	*	*
District	45	*	*	*	*
District	50	*	*	*	*
District	60	*	*	*	*
District	65	*	*	*	*
District	75	*	*	*	*
District	80	*	*		*
District	85	*	*		
InspectionSystem	MAESTRO	*	*		
InspectionSystem	MAESTRO,Nu-Tech	*	*		
InspectionSystem	MAESTRO,Religious	*	*		
InspectionSystem	MAESTRO-SIS	*	*		
InspectionSystem	NELS	*	*		
InspectionSystem	NELS,MAESTRO	*	*		
InspectionSystem	NELS,NTIS,MAESTRO	*	*		
InspectionSystem	NELS,Nu-Tech	*	*		

InspectionSystem NELS,Nu-Tech,Relig	*	*		
InspectionSystem NELS,Religious Sla	*	*		
InspectionSystem NELS,SIS	*	*		
InspectionSystem NELS,SIS,Religious	*	*		
InspectionSystem Nu-Ova		*		
InspectionSystem Nu-Tech	*	*		
InspectionSystem Nu-Tech,Religious	*	*		
InspectionSystem SIS	*	*		
InspectionSystem SIS,MAESTRO	*	*		
InspectionSystem SIS,MAESTRO,Religi	*	*		
InspectionSystem SIS,Religious Slau	*	*		
InspectionSystem SIS-Nu-Tech	*	*		
InspectionSystem SIS-NuOva	*	*		
InspectionSystem HIMP			*	*
InspectionSystem NTIS			*	*
InspectionSystem OtherNTIS			*	*
Sep_Tox	*	*	*	*
Contam	*	*	*	*
AirSac	*	*	*	*
synovitis			*	*
sum_SP	*	*	*	*
sum_SNP	*	*	*	*
sum_U	*	*	*	*
sum_NC	*	*	*	*

There should always be a reasoning to suggest a possible causal relationship behind the selection of factors to be tested for use in a logistic regression, as with any other type of regression model. Without such reasoning, one risks finding a statistically significant but meaningless, relationship to completely unrelated variables.

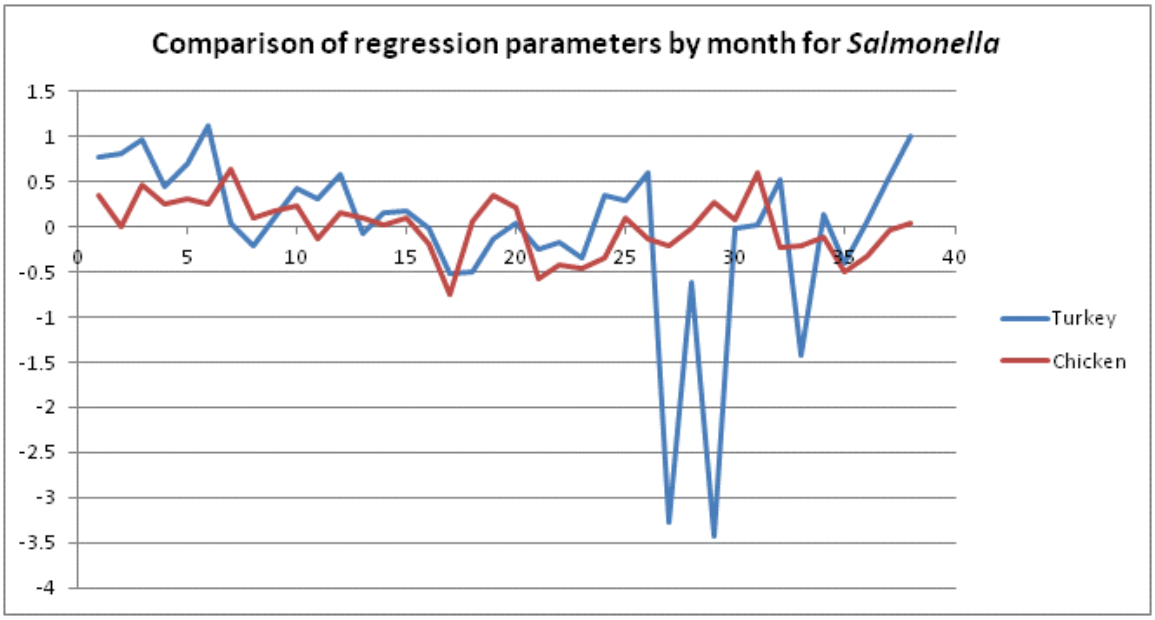
In this case, I can see no reasoning behind the use of the month factors. Up to 38 of these are used, depleting the degrees of freedom of the regression. If the idea was to account for seasonal variation, then four factors at most would probably have been sufficient (spring, summer, fall, winter).

The estimated coefficients for each month also vary significantly, as shown in the following plot for the larger set used for *Salmonella*. Months 27, 29 and 33 for turkey have very large negative values in comparison with others. Unless there is a good explanation for why these negative values could occur for turkeys, and not for chickens,

the month factors should not be used because they amount to an over-parameterizing of the model.

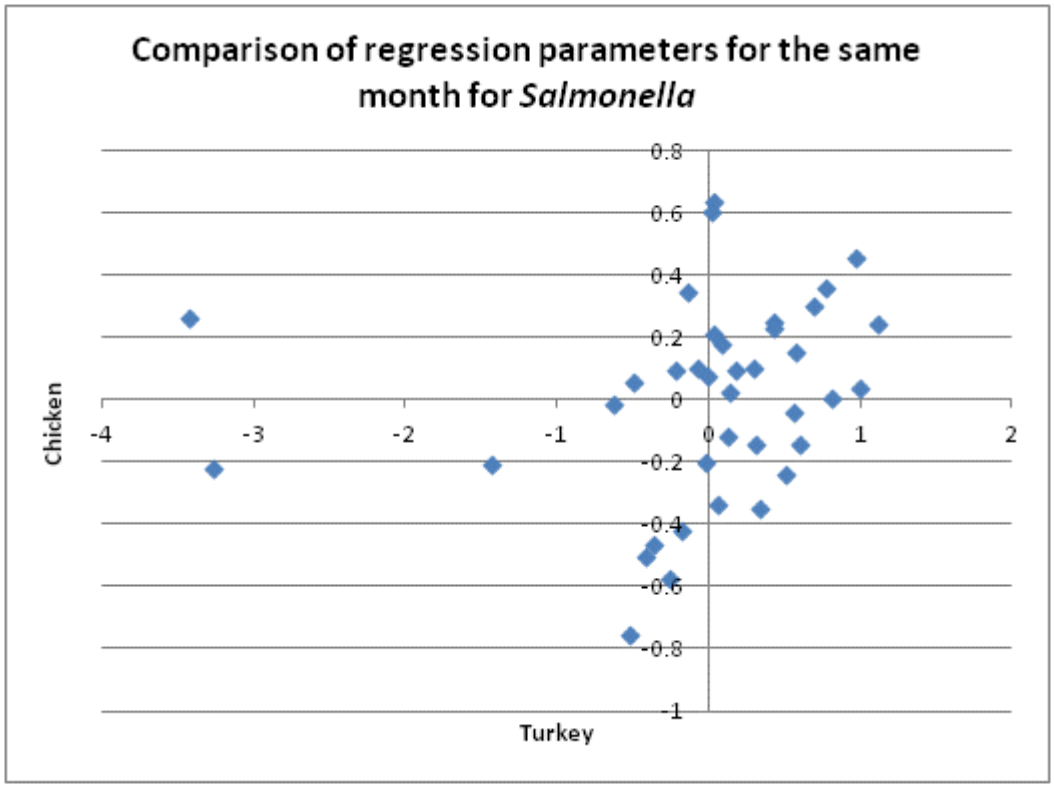
FSIS Response: As stated above the rationale for including a categorical monthly parameter in the model is primarily due to the data having been biased by the economic recession over most of the study period. The effect is obvious in the turkey Salmonella data where the negative parameters for months 27, 29, and 33 correspond with the 2009 holiday period where sales fell markedly and multiple economic factors resulted in decreased Salmonella prevalence. If the model were over-parameterized, one might expect it to perform poorly out of sample.

Due to the complex construction of the datasets, the non-stationary four-quarter seasonal time series, and the irregular seasonal patterns over the study period more monthly parameters were included in the models. The models did not seem to be overparameterized when the balance between the increasing amount of variance explained by each model (that increases with increasing number of parameters) and the validation error does not increase with increasing numbers of parameters. Because the amount of variance explained by each finalized model was not accompanied by an increase in validation error the models were not considered to be over-parameterized.



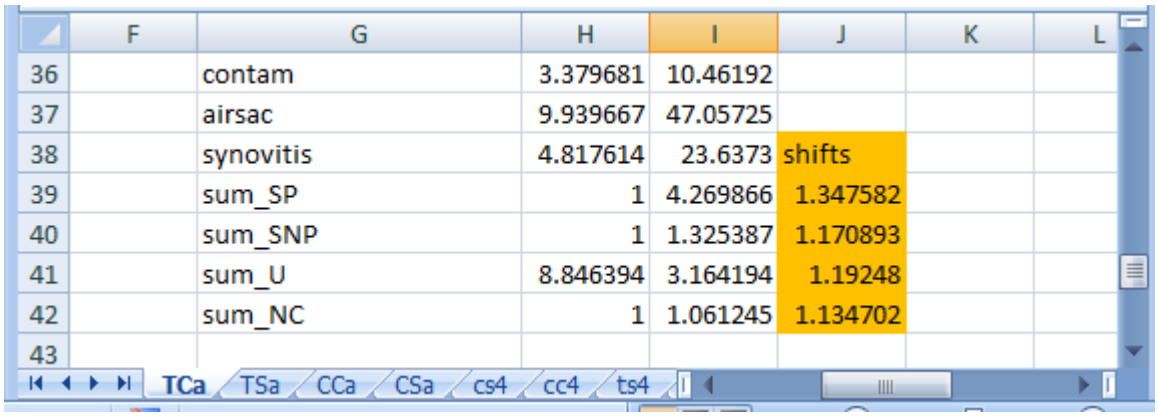
The following plot shows that there is little if any relationship between the monthly coefficients for chickens and turkeys (Spearman correlation of 0.2). This provides further evidence to suggest that the monthly factors have no intrinsic predictive value.

FSIS Response: The correlation plot is for the parameter estimates which are average estimates of the regression adjusted monthly categories relative to the September 2010 reference category. This plot and associated rank correlation demonstrate no strong correlation between the turkey and chicken monthly parameters. However, we do not assume that the Salmonella prevalence in chicken and turkey should show the same monthly pattern because two separate grower industries are examined which occur and different locations and with different management practices. Additionally, the markets for chicken and turkey products are different and follow different seasonal patterns. Especially with the market forces operating during the observation period it is not expected that there should be a strong correlation between these parameters.



Small Model inconsistencies

In the Model file I was given, the variables sum_SP, sum_SNP, sum_U and sum_NC were all set to 1. This is inconsistent with the other sheets, and I wonder if it is in error (see following screen capture):

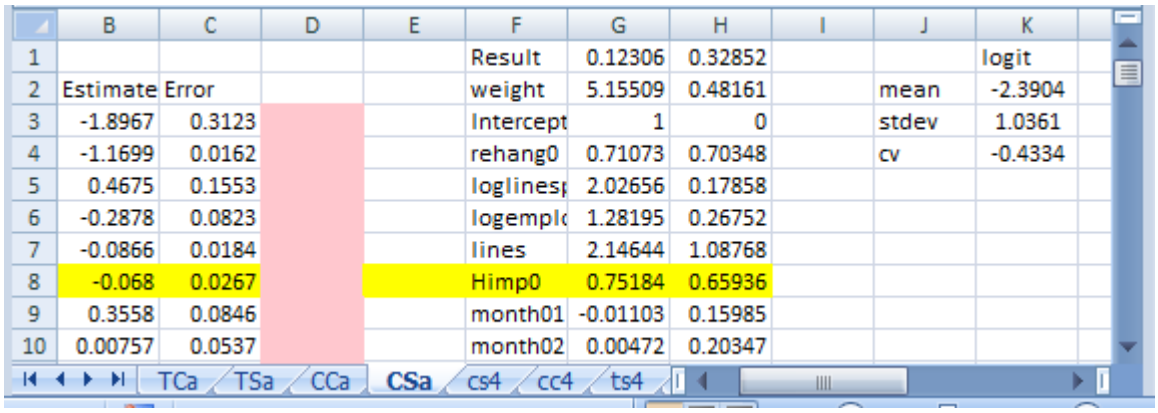


	F	G	H	I	J	K	L
36		contam	3.379681	10.46192			
37		airsac	9.939667	47.05725			
38		synovitis	4.817614	23.6373	shifts		
39		sum_SP	1	4.269866	1.347582		
40		sum_SNP	1	1.325387	1.170893		
41		sum_U	8.846394	3.164194	1.19248		
42		sum_NC	1	1.061245	1.134702		
43							

TCa | TSa | CCa | CSa | cs4 | cc4 | ts4

FSIS Response: the reviewer comments on data given for review- the data given were not fully explained. This provided the impression in the reviewer appendix screen shots that the shifts for the S, SP, and NC variable had not been set to 1.0 in the indiscriminate scenario when in fact they had been set to 1.0. And additionally, that the rehang variable had been set to 1.0 or -1.0 for post-chill or rehang respectively and not to the mean value indicated in the screen shot.

Sheet CSa of the Model file I was given also had the Rehang variable set to 0.71073 as shown in the following screen capture, though the Report set this value was set to 1 (pages 18 and 38 of the Report):



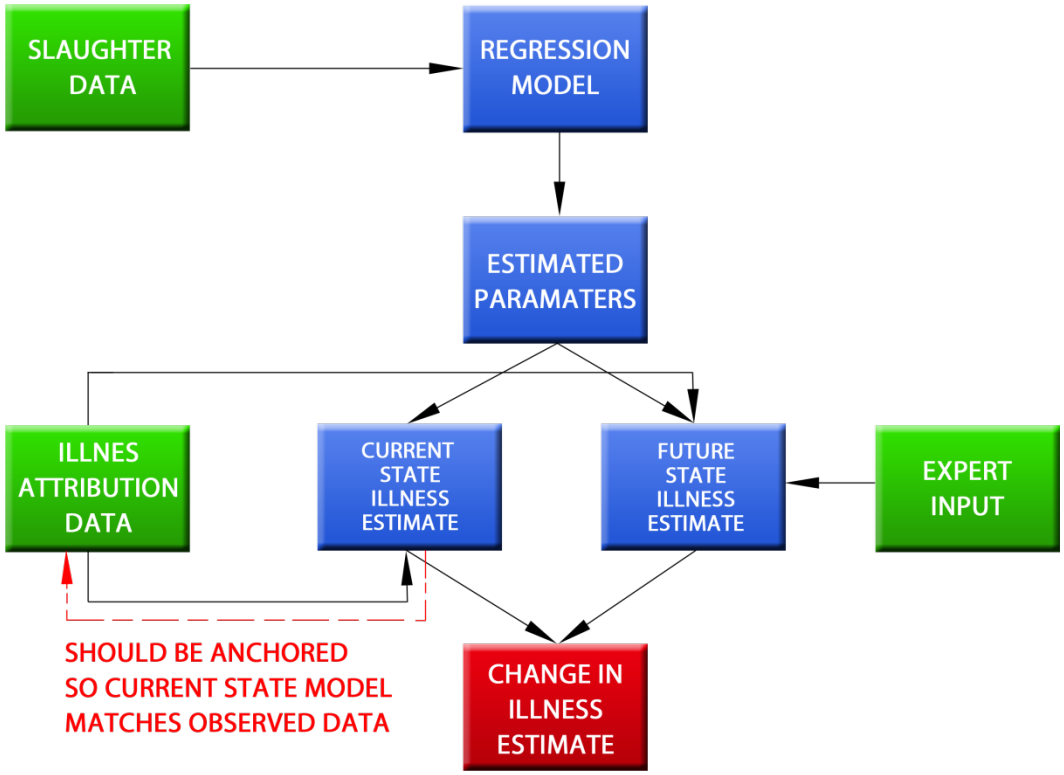
	B	C	D	E	F	G	H	I	J	K
1					Result	0.12306	0.32852			logit
2	Estimate	Error			weight	5.15509	0.48161		mean	-2.3904
3	-1.8967	0.3123			Intercept	1	0		stdev	1.0361
4	-1.1699	0.0162			rehang0	0.71073	0.70348		cv	-0.4334
5	0.4675	0.1553			loglinesp	2.02656	0.17858			
6	-0.2878	0.0823			logemplc	1.28195	0.26752			
7	-0.0866	0.0184			lines	2.14644	1.08768			
8	-0.068	0.0267			Himp0	0.75184	0.65936			
9	0.3558	0.0846			month01	-0.01103	0.15985			
10	0.00757	0.0537			month02	0.00472	0.20347			

TCa | TSa | CCa | CSa | cs4 | cc4 | ts4

FSIS Response: We apologize if the model provided was confusing. The reviewer's understanding is correct on both counts. Occasional runs of the model were completed without modeling changes to the decision variables (hence the values could be set to one on those occasions). The rehang variable should always be set to 1 to model the probability of a positive sample at post-chill. In the revised model, this amounts to setting this variable to 1 for all available data for the purposes of predicting the probability of a positive sample at post-chill.

Solution to the Model issues

The following flow diagram illustrates the general logical flow of the Model as it is currently implemented.



The red arrow illustrates the component that is missing in the analysis. The arrows between Illness Attribution Data and Current State Illness Estimate go both ways, since the current state is partially known from the illness attribution data and must therefore anchor to that data. Monte Carlo simulation only allows for a logical flow in one direction. However, Markov Chain Monte Carlo (MCMC) models can accommodate this anchoring. MCMC models will also automatically account for the correlation structure between the uncertainty distributions for the fitted parameters and allow one to mix pre- and post-chiller data.

If the attributable fraction data are reevaluated, the Pert-distributed expert estimates revisited, and the varying number of month indices replaced with four seasonal factors I think that the MCMC method would address all of the Model's behavioral issues I have discussed in this Appendix.

FSIS Response: Practically it is extremely difficult, if not infeasible, to move this complex problem into an MCMC framework. Reviewer 1 suggests that the MCMC

modeling approach would be more logically correct and result in narrower uncertainty. However, the current FSIS model is simulating the baseline and alternative scenarios in parallel, so it is not clear that the FSIS modeling approach is logically incorrect. We were able to correct the error of propagating the mean through a nonlinear model. as described by the reviewer, without changing modeling platforms. And, finally, the recommended anchoring approach can be useful, but a decision not to use an anchoring approach is not fundamentally incorrect. Furthermore, it seems unlikely that the results would be substantially different given the findings of an MCMC model that examined the effects of HACCP on poultry-associated Salmonella illnesses in the United States (Williams and Ebel, 2012). That analysis – using public health surveillance data and modestly-informed prior distributions for attribution – did not generate substantially better informed estimates of model inputs. Ultimately, the development of an MCMC model is not necessarily preferable in this case and, at a minimum, is not a necessary replacement for the Monte Carlo model developed here.

Specific editorial comments

There is no line numbering in the report, so I refer to page.

Formula for the regression equation

On Page 18 the regression equation is show as:

$$Prev(policy) = \frac{e^{\alpha+\beta_1 X_1+\ldots+\beta_i X_i A_i+\ldots+\beta_n X_n+\varepsilon}}{1+e^{\alpha+\beta_1 X_1+\ldots+\beta_i X_i A_i+\ldots+\beta_n X_n+\varepsilon}}$$

On Page 35 it is written as:

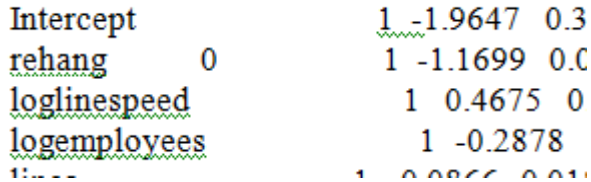
$$p = \exp(b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n) / (1 + \exp(b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n))$$

The two equations are equivalent except for the term in the first equation. This describes a latent (unobserved) variable that would follow a logistic distribution by convention (if it followed a Normal distribution, it would be impossible to distinguish it from the uncertainty distribution for a). The regression results in SAS Code provided to me suggest that the variable was not used, so the former equation on Page 18 should be edited to remove , which makes it equivalent to the latter equation on Page 35. The equations should also share the same convention (b or , or or).

FSIS Response: We agree and have revised the equations in the report.

Inconsistent regression coefficient value

The regression results in the SAS Code page give an intercept value of -1.9647:



The table of results at page 9, the values that are in the Model, gives a different intercept value:

FSIS Response: this error has been corrected.

Appendix Table 2. Parameter Estimates for Young Chicken *Salmonella* Model Used in Scenario Analysis

Parameter	Estimate	Std Error	p-value	Mean	Std Dev
Intercept	-1.8967	0.3123	<0.0001*	1.0000	0.0000
rehang	-1.1699	0.0162	<0.0001*	0.7107	0.7035
loglinespeed	0.4675	0.1553	0.0013*	2.0266	0.1786
logInspectors	-0.2878	0.0823	0.0002*	1.2820	0.2675

Estimates of current levels of illness

Page 19 of the Report states:

A previous analysis estimated that the fractions of total Salmonella and Campylobacter illnesses per year attributable to young chicken as 16.33% (167,831/1,027,561) and 19.71% (168,291/845,024), respectively (FSIS, 2011, 1). That analysis also estimated the fraction of total Salmonella and Campylobacter illnesses per year attributable to young turkeys as 0.67% (6855/1,027,561) and 0.08% (714/845,024), respectively.

The calculation performed in the 2011 FSIS reference is duplicated below:

Table 1. Data source and methods used for estimating illnesses from *Campylobacter* and *Salmonella* attributed to young chickens and turkeys in the U.S. population. Minor discrepancies in final estimates are due to rounding of intermediate values.

Step	Input	*Campylobacter*	*Salmonella*	Data Source & Time Period / Estimation
1	Foodborne illnesses	845,024	1,027,561	Scallan *et al.*, 2011
2	Poultry attribution fraction	0.20	0.17	CDC outbreak data, 2001-2007 and Pires *et al.*, 2009[2]
3	Young chicken volume adjusted proportion positive	0.467	0.075	FSIS Young Chicken Baseline Study (2007-2008)
4	Turkey volume adjusted percent positive	0.011	0.017	FSIS Young Turkey Baseline Study (2008-2009)
5	Young chicken production fraction	0.838	0.838	ERS (2003-2008)
6	Young turkey production fraction	0.151	0.151	ERS (2003-2008)
7	Contaminated young chicken fraction	0.996	0.961	Step =(3 x 5)/ ((3 x 5) + (4 x 6))
8	Contaminated young turkey fraction	0.004	0.039	Step = (4 x 6)/ ((3 x 5) + (4 x 6))
9	Young chicken attribution fraction	0.199	0.163	Step = 2 x 7
10	Young turkey attribution fraction	0.001	0.007	Step = 2 x 8
11	Total foodborne illnesses from young chickens	168,291	167,831	Step = 1 x 9
12	Total foodborne illnesses from young turkeys	714	6,855	Step = 1 x 10

There is an assumption in the analysis of this table that a contaminated turkey carcass causes on average the same number of illnesses as a contaminated chicken carcass. I have no evidence either way: intuitively a turkey is a lot more meat and can infect a lot more people as a result, but on the other hand would be expected to be cooked a lot longer and perhaps killing more bacteria. In any event, there is better information available, described below.

Step 2 in this table states that the poultry attribution fraction comes from data gained from investigating outbreaks. The footnote to that table states:

[2] An assessment of publically available Centers for Disease Control and Prevention (CDC) outbreak cases, 2001 – 2007 for *Salmonella* gave poultry attribution estimates of 16.9%. This estimate was rounded up to 17%. For *Campylobacter* poultry attribution, a CDC case-control study suggested the "population attributable fraction (PAF) of 24% was related to consumption of chicken prepared at a restaurant" (Friedman *et al.*, 2004). This estimate was rounded down to 20%.

I cannot see why the chicken-attributable fraction for *Campylobacter* 24% was rounded down to 20%. In any event, the 20% source attribution is far too low. This is just consumption of chicken in a restaurant. The figure is more likely to be 40% or more, though it is difficult to estimate because of the sporadic nature of infections. Two epidemiological events give some indication: It was at least 40% in Belgium (Vellinga, A. and Van Lock, F. (2002) The dioxin crisis as experiment to determine poultry related campylobacter enteritis. Emerg. Infect. Dis. 8, 19-22.) and 70% in Iceland (Stern, N.J., Hiett, K.L., Alfredsson, G.A., Kristinsson, K.G., Reiersen, J., Hardardottir, H., Briem, H., Gunnarsson, E., Georgsson, F., Lowman, R., Berndtson, E., Lammerding, A.M., Paoli, G.M. and Musgrove, M,T. (2003) *Campylobacter* spp. in Icelandic poultry operations and human disease. Epidemiol Infect., 130(1),23-32.). I don't know of any robust data for turkey-attributable fractions for *Campylobacter*.

The poultry-attributable fractions for *Salmonella* are also far too low: for chicken it should be around 48%, and for turkey around 17% (see, for example, the far more robust analysis using serovar pattern matching rather than case-control studies in http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3123837/)

The effect of making corrections to the attributable fractions would be to increase the magnitude of the predicted reduction in illnesses.

FSIS Response: Our justification for the attribution fractions used has been explained in other responses. We can simply add that the product attribution estimates are from published U.S. data. We can also explain that the production fractions used in the table (reproduced above) do, in fact, account for the different masses of chicken and turkey carcasses because those fractions reflect mass of products produced in the United States.

Lognormal fit to illness rate estimates

Page 20 of the Report describes how a Lognormal (to base e) distribution is fit to the uncertainty around the illness rates. The method described fits to confidence interval values but cannot match the mean.

There are two better approaches:

1. Fit to a shifted Lognormal distribution. This gives a three parameter distribution which can be matched to the confidence intervals and the mean (Table 2 of the Report) for all incidence rates except turkey-*Campylobacter*, which is negatively skewed.

2. Much better: Go back to the original data, which was a total estimate of illness rates combined with an uncertainty distribution for the attributable fraction. This needs changing anyway, see other comments, but one can use a Beta distribution to give the fraction, which is more likely to match the data. Better still, a Dirichlet distribution would allow one to model turkey and chicken attributable fractions together, since the uncertainties are necessarily jointly distributed.

FSIS Response: We agree that alternative approaches could be used. Due to the fact that the Scallan et al (2011) uncertainty distributions are only approximately lognormally, there are slight discrepancies between the estimated annual illnesses summary statistics and the associated lognormal distributions. For example, Lognormal (mu = 12.043, sigma = 0.291) has mean = 177,254 (not 167,831). Thus, the model actually assumes that the young chicken attribution fraction for Salmonella is 0.172 rather than the nominal 0.163. The difference appears inconsequentially small.

The approach taken here was simple to explain and fit for this purpose. More elaborate techniques would be needed if uncertainty about the attribution fraction was included. As the Sensitivity Analysis section of the Nov 2012 revision explains, the influence of this input on changes to human illness attributable to poultry slaughter inspection decisions is less than other model inputs. Therefore, more precision about the illness rate estimate (through more complex fitting methods) does not seem necessary.

Expert estimates of fractions for which control changes would apply

Page 21 and 22 of the Report provide estimates of the factor to multiply current use by so that, for example, 1.6 represents a 60% increase, 1 represents no change, and 0 represents a complete cessation of that activity. The expert estimates are:

Activity code	**Min**	**Mode**	**Max**
SP	1	1.25	1.6
SNP	0	0.9	Hanley JA

U	1	1.25	1.6
NC	0	.74	1

The three values are interpreted as a PERT distribution. Note that SP and U take the same parameter values, which makes me wonder whether they share some common assumptions about their range, in which case they should be correlated.

There is a lot of opportunity to produce a mismatch in interpretation when eliciting expert estimates between what the expert is thinking and how the estimate is used in the model. It strikes me that these ranges may be extremely wide (particularly SNP) when one considers that the same value is applied universally across the Model, i.e. that it is assuming *on average* all plants will adopt the simulated value. I recommend that this be revisited with the experts.

FSIS Response: The available evidence about these adjustments is limited to the analysis of HIMP establishments explained in the HIMP Report. Because that report does not delineate between Scheduled and Unscheduled Procedures performed, the effects are assumed to be similar for each type of completed procedure, although their future effects are considered independent of each other. Our sensitivity analysis suggests that the uncertainty about these inputs is an important contributor to the uncertainty about the model outputs. Nevertheless, the characterizations of the Pert distributions is reasonable in the absence of more evidence. In fact, these parameterizations intend to account for a population-level effect; this is why the most likely values for SP and U are assumed to be 1.25 rather than the 1.6 estimated from comparing HIMP and non-HIMP establishments in the HIMP report (i.e., this more conservative most likely value assumes that the population effect will be less than what is observed among the volunteer participants in the HIMP study).

Placement of mode, median and mean

Page 23 attempts to interpret the reasoning for mean > median > mode. The reason is simply that whenever a distribution is right skewed, this is the order in ascending value in which the statistics will occur.

FSIS Response: We agree; that was our point.

Placement of mode, median and mean

Page 24 mentions two variables being ‘perfectly correlated’. I think the author(s) mean that they took the same random value in any particular sample of the Model.

FSIS Response: that is correct.

Small editorials

sows = shows

a average = an average

FSIS Response: fixed

Itemized FSIS Replies to Reviewer #2

Reviewer #2's comments:

The logistic regression models and scenario models are well documented, and rationale is provided for assumptions. However, the soundness of the overall approach cannot be determined due to the lack of transparency for data and models used for estimating reductions in annual human illness rates. A manuscript is cited for a 'simple prevalence-based method' that reports modeling annual illness as a Poisson process. However, no biological data or rationale is provided in the report. Available datasets for human salmonellosis and campylobacteriosis from experimental and epidemiologic studies, as well as extensive analyses of these datasets, are not provided or referenced in the report.

FSIS Response: FSIS respectfully disagrees. The approach is reasonable both because it has the potential to answer the set of posed questions, and because it is practical given the limited amount of available data. In fact, it makes relatively few assumptions – and the data used is well documented in the report. This seems appropriate, given the limited knowledge available about the steps between the slaughter plant and consumer illness (e.g. preparation, storage, cooking, dose-response by sub-population, etc.). Other risk assessments in both Europe and the USA have shown a linear relationship between prevalence and expected incidence, so the method used here skips these steps without any loss of accuracy.

A comprehensive and transparent synthesis is needed that rigorously assesses the strengths and weaknesses of the data and models used, with supporting scientific rationale for applications of the data and models, in order to evaluate robustness of the approach for estimating reductions in annual human illness as a 'simple prevalence-based risk assessment' for campylobacteriosis and salmonellosis. This synthesis is essential for improving transparency of this report as a 'stand-alone' analysis. The manuscript cited in the report for the 'simple prevalence-based risk assessment' noted the importance for 'analysts to convey how the outputs of a risk model will change with alternative assumptions' in the manuscript discussion section describing determination of the robustness of this approach as an area for future research. Further documentation of the analyses, including more comprehensive sensitivity analysis and exploration of alternative assumptions about dose-dependencies for likelihood and severity of disease, would strengthen the report, even if validity of the approach cannot be verified with data presently available.

FSIS Response: please see our response just above this.

The material in the report body, appendices, and supporting SAS code and Excel sheets is accurate and consistently reported. Editorial corrections are needed as follows.

Page Number	Section or Paragraph	Correction
3	Title for Table 5	Insert space after 'for'
11	2nd paragraph from bottom	Delete 'might' on second line
22	2nd paragraph	Replace 'sows' with 'shows'
32	Last paragraph	Replace 'testin' with 'testing' and delete second '.' In last sentence
37-38	Numbered list	Punctuation is inconsistent
39-40	Last sentence	Awkward construction 'farther away the curve is away from'
40	1st full paragraph	1st sentence awkward
43-60	Appendix Tables	Need header rows for tables continuing on multiple pages

FSIS Response: these corrections have been addressed in the report.

1. *Evaluate if the overall approach for modeling the public health benefits potentially realized from the change in inspection system examined is fundamentally sound.*

    a. *Is the overall approach used in the analysis to evaluate the linkage between inspection activities and potential reductions in annual human illnesses fundamentally sound? The regression model used to estimate changes in establishment prevalence should be addressed separately from the model used to estimate reductions in annual human illness.*

    The logistic regression models for four decision variables (and groups of inspection system procedure codes) and simulation models for indiscriminate and discriminating scenarios across decision variables were used to estimate changes in establishment prevalence. These models are described with sufficient transparency to support the soundness of the approach. In general, the basis of assumptions or rationale for inferences is provided or referenced. However, the authors noted ambiguous effects across pathogens and products. Further, the impact of the policy changes on linespeed and worker safety are not discussed, though linespeed is a parameter estimate for scenario analysis.

    FSIS Response: linespeed in poultry slaughter establishments is not recorded as a part on ongoing FSIS surveillance activities. Instead the structural variable used in the model was simply the rated line speed for the type of inspection-evisceration system operating in that establishment at the time the observational data were collected. As indicated by Tables 4-7

in the appendix of the risk assessment, the significant parameter estimates for the line speed variable do have the expected sign (positive), indicating that in the absence of compensating measures, increased nominal line speed is predicted to result in higher prevalence of poultry carcasses. The FSIS 2011 Evaluation of HACCP Inspection Models Project (HIMP) indicates that with adoption of compensating control measures, equivalent pathogen performance can be achieved at higher line speeds.

Worker Safety, although an important issue for human welfare, is not a food safety issue.

The soundness of the overall approach cannot be determined due to the lack of transparency for data and models used for estimating reductions in annual human illness. In particular, no data are provided or referenced in the report for salmonellosis or campylobacteriosis dose-response datasets and models considered and used.

FSIS Response: please see our previous response. Our approach is well documented – and we have strengthened the documentation in the Nov 2012 version. The FSIS risk assessment is transparent that it does not use dose-response analysis to estimate reductions in illness.

b. If not fundamentally sound, in each case, what problems exist and how should they be addressed?

For the models estimating changes in establishment prevalence, scenario results could report on the impact of increasing linespeed on prevalence of microbial contamination, and perhaps on worker safety. It is unclear why the number of months varies (11, 25, and 38 months in appendix tables), though sources were reported for 12-month baseline studies (*Campylobacter* and *Salmonella*) and 38-month PR/HACCP *Salmonella* verification program.

FSIS Response: Ideally, estimated relationships between line speed in poultry establishments and prevalence in those establishments would use "pair-wise" data – data collected on observed values of both at the same point in time. The proxy of "rated maximum line speed" as a substitute for actual linespeed is a limitation of the analysis and is duly noted. In this case line speed is an independent variable in the model – albeit a left-centered independent variable due to a lack of disaggregated data. Regression based on grouped data is sometimes unavoidable, and generally contributes to less efficient parameter estimates but improved fit of the regression.(Greene, W. 1997. Econometric Analysis).

FSIS Response: The number of months varied because the number of months observed was not the same for each data set. Also, it should be recalled that the number of parameters for each categorical time period cited as 11, 25, and 38 are actually the number of months in the data set minus one due to there being one month used as reference.

For the models estimating reduction in annual human illness, the authors state an assumption that a 'simple prevalence-based risk assessment method' (Williams et al., 2011) that models annual illness as a Poisson process. In the referenced manuscript, no example depicting the behavior of the model for salmonellosis is provided. The example of campylobacteriosis that is provided applies one set of beta-Poisson model parameters reported by Medema et al. (1996) for the dataset of human campylobacteriosis from a volunteer study (Black et al., 1988). The authors do not state the strains and endpoints represented by the beta-Poisson parameters for campylobacteriosis or provide a description and rationale for their treatment of strain variability and model uncertainty for either campylobacteriosis or salmonellosis.

FSIS Response: in the Nov 2012 revision, we have added/amended the following explanatory paragraphs:

"The modeling framework stems from the three primary determinants of adverse human health outcomes from foodborne pathogens; 1) the frequency of exposure to the pathogen; 2) the distribution of pathogens in a random exposure event on a per serving basis; and 3) the probability that a random exposure event causes the adverse human health outcome (Cox, 2006; Haas, 1996). In microbial food safety, sporadic exposure events are considered independent events and chronic exposures to pathogens are historically not considered. These characteristics support modeling the occurrence of human illnesses as a Poisson process.

A prevalence-based model estimates changes in annual illness counts based on changes in the frequency of occurrence among food commodities (Williams et al., 2011). The basic model is:

$$P(ill) = P(ill \mid exp)P(exp)$$

where $P(ill)$ ***is the probability of illness from a product-pathogen pairing across a population,*** $P(ill \mid exp)$ ***is the probability that exposure to a random contaminated serving will produce illness[1] and*** $P(exp)$ ***is the frequency of exposure to the pathogen on a per serving basis[2].***"

---

[1] $P(ill \mid exp)$ is the solution to the integral $\int_{>0}^{\infty} R(D) f(D) dD$ where $R(D)$ is the dose-response function and the exposure distribution of doses (D > 0 organisms) is the probability density $f(D)$.

[2] Exposure to a contaminated serving can be defined at any point in the farm-to-table continuum assuming that $P(exp)$ is proportional to the percent of positive units observed at some point prior to consumption (i.e., these measures of occurrence differ by a multiplicative constant). In food safety applications, the best data for measuring frequency is usually at the point of commercial production (e.g., retail-ready raw chicken carcasses).

"The advantage of this modeling approach is that it prevents the need to estimate an exposure distribution or a dose-response relationship. The critical assumption needed to apply a prevalence-based approach is that dose levels at consumption are independent of the frequency of contamination. This assumption asserts that $P(ill \mid exp)$ ***is constant regardless of changes in*** $P(exp)$***. There is empiric evidence that supports the independence of prevalence and contamination levels at the end of the production of raw poultry carcasses. For example, in rinse samples of young chicken carcasses that test positive, the average concentration of Salmonella per ml of sample rinsate was 0.16 and 0.14 colony forming units (cfu) in the 1995 and 2007 baseline surveys, respectively (FSIS, 1996; FSIS, 2009). Yet, the prevalence of positive carcasses was demonstrably different (20% vs. 7.5%) in those surveys. Similarly, those same surveys found the average concentration of Campylobacter per ml of sample rinsate was 21 and 9.1 cfu in 1995 and 2007, respectively; despite a dramatic reduction in the prevalence of positive carcasses from 88% to 11%. Other studies have drawn similar conclusions with respect to other product-pathogen pairs (Crouch et al., 2009; Withee et al., 2009)."***

FSIS Response: The available evidence about the effect of changing inspection activities is limited to the data used in this assessment. That data will not support more detailed assessment with respect to bacterial strain variability. Furthermore, it is difficult to imagine how any model would examine strain-type effects from changes to inspection activities that do not target specific bacterial strains.

FSIS Response: With respect to model uncertainty, the Nov 2012 revision includes more sensitivity analysis. Nevertheless, the simple modeling approach used here is intended to generate more conservative estimates of illnesses avoided relative to more complex process modeling approaches. In general, it is reasonable to assume that a reduction in proportion of positive samples will correlate with a reduction in pathogen levels on carcasses that remain contaminated (and vice versa). This modeling approach does not account for any change in pathogen levels on contaminated carcasses. Therefore, it is possible that our model outputs under-estimate the effects of modeled inspection changes for those results that predict a reduction in proportion of positive samples. Because the model results suggest a high confidence that prevalence will decrease, an assertion that the current results are conservative seems reasonable.

In addition, for salmonellosis, the authors do not acknowledge uncertainty regarding alignment of serotypes causing human outbreaks and sporadic illness (MMWR June 20, 2011 60(22)748-755; MMWR September 9, 2011, 60(35):1197-1202) and those reported in baseline studies for young chickens and young turkeys. Of an unspecified number of isolates serotyped in the baseline studies for *Salmonella* in young turkeys, the three serotypes reported by FSIS (Heidelberg, Saint Paul, Hadar) are indeed listed by CDC as associated

with human illness, but accounted for less than 6.3% of human cases reported by CDC in 2009. For more than 1,000 isolates from young chickens, the major serotypes isolated by FSIS (Kentucky, Heidelburg, and Typhimurium) accounted for less than 20% of human cases reported by CDC in 2009. Though uncertainty in attribution of human cases is high, the extensive literature on this issue, and its potential impact on predicting reductions of human cases, is largely ignored in this report.

FSIS Response: We agree that uncertainty about attribution fractions is high and has been largely ignored in this analysis. As explained above, our choice of attributions was based on the principles of consistency with previous analyses and transparency in their development. We are aware of a pending publication from the Centers for Disease Control that will explain the state of the art with respect to estimating attribution, as well as provide attribution estimates from many product-pathogen pairs. Although it is reasonable to consider distributions of serotypes between human illnesses and poultry in estimating attribution fractions, our approach based on outbreaks has been used commonly in the past. Furthermore, although Reviewer 2 correctly notes that the uncertainty in attribution of human cases is largely ignored, this uncertainty does not impact the probability of increased illness.

A comprehensive and transparent synthesis is needed that rigorously assesses the strengths and weaknesses of the data and models used, with supporting scientific rationale for applications of the data and models, in order to evaluate robustness of the approach for estimating reductions in annual human illness as a 'simple prevalence-based risk assessment' for campylobacteriosis and salmonellosis. This synthesis is essential for improving transparency of this report as a 'stand-alone' analysis. The manuscript cited in the report for the modeling framework (Williams et al., 2011) noted the importance for 'analysts to convey how the outputs of a risk model will change with alternative assumptions' in the manuscript discussion section describing determination of the robustness of this approach as an area for future research. Further documentation of the analyses, including more comprehensive sensitivity analysis and exploration of alternative assumptions about dose-dependencies for likelihood and severity of disease, would strengthen the report, even if validity of the approach cannot be verified with data presently available.

FSIS Response: Support for the prevalence-based approach used in this risk assessment is provided by three sources. 1) Discussion of the results of the FSIS baseline studies have been added to the document. These studies demonstrate, particularly for Salmonella, that while the prevalence of test-positive carcasses has decreased, the levels of the pathogen on carcasses is low (e.g., ~ 0.15 cfu/ml) and are essentially unchanged between the FSIS baseline studies. Given that the levels on test-positive carcasses are unchanged, the servings derived from these carcasses have the same P(ill|exp) (under the assumption of

similar consumer handling). The Williams et al. (2011) study provides a derivation that demonstrates that the more complex dose-dependent model simplifies to the prevalence-based model. 2) The prevalence-based approach was used to back-calculate the number of cases of salmonellosis caused by the consumption of broiler chickens prior to the implementation of the HACCP program (Williams and Ebel 2010). Estimates derived from this analysis match a similar analysis of the FoodNet data performed by CDC, which suggests the simpler prevalence-based model provides reasonable estimates of illnesses avoided. 3) While FSIS has more limited data for Campylobacter on poultry, two additional studies show the appropriateness of the prevalence based model for this pathogen. These being the Vose/FDA model Vose et al. 2000 and Rosenquist et al, 2003. This latter study used a dose-dependent model, but validation of the model demonstrated the linear relationship between prevalence and illnesses that one expects to see under the prevalence based model. To quote the Rosenquist article "The simulations showed a linear relationship … between the flock prevalence and the incidence of campylobacteriosis. … The simulations indicated that if the flock prevalence was reduced for example two times then the number of cases associated with consumption of chicken meat would also be reduced approximately two times. This is because there is a one-to-one relationship between the two parameters".

Williams, M.S. and Ebel, E.D. 2012. Estimating Changes in Public Health Following Implementation of Hazard Analysis and Critical Control Point in the United States Broiler Slaughter Industry. Foodborne Pathogens and Disease. 9(1):59-67

Rosenquist, H., Nielsen,N.L., Sommer,H.M., Nørrung,B., and Christensen, B.B. 2003. Quantitative risk assessment of human campylobacteriosis associated with thermophilic Campylobacter species in chickens. International Journal of Food Microbiology, 83:87–103.

2. Evaluate the complexity of the model in areas where the reviewer identifies limitations, weaknesses, or inadequacies; the reviewer must provide alternative data, data analysis, and/or modeling approaches.

a. Is the model too complex, or not complex enough, to adequately address the risk management questions?

Unless FSIS conducted additional analyses that were not included in the body of the report or appendices, the model is not complex enough to address the impact of alternative

assumptions for the 'simple prevalence-based risk assessment', as noted above. The uncertain alignment of serotype prevalence between poultry baselines and human salmonellosis cases merits mention, even if this complexity is impractical to include variability in the 2,500 *Salmonella* serotypes (or even the top 20 serotypes) in the risk models. If prevalence maps poorly, the framework may not be appropriate to judge the relative benefits and costs of procedural changes that rely on estimates of pathogen prevalence.

FSIS Response: we recognize the concern of the reviewer here. The relative simplicity of the model, making few assumptions and focusing on the specific problem, is a key positive attribute of the Model.

b. Is the model over- or under-parameterized?

As a microbiologist, the model as presented is under-parameterized in the sense that biologically meaningful parameters are not considered, or at least not explained.

FSIS Response: please see our previous responses. Other reviewers are of the opposite opinion.

c. Does the model adequately characterize the uncertainty present?

The model adequately addresses parameter uncertainty, not model uncertainty or errors in model structure for alternative assumptions.

FSIS Response: we have strengthened the discussion of model uncertainty and alternative assumptions by including a section on sensitivity analysis in the Nov 2012 report.

d. Is variability sufficiently addressed?

No, particularly strain variability is ignored in the model and is influential for predicting risk of campylobacteriosis and salmonellosis.

FSIS Response: Reviewer 2 does not provide any such data or analysis to back up this contention. We are unclear how strain variability might influence risk in the context of the analysis conducted here. Although we can understand how the relative frequencies of different strains among humans and food products might be insightful for examining attribution fractions for various product-pathogen pairs, the complexity of incorporating bacterial strain variability into an assessment of the effect of inspection changes on bacterial occurrence on carcasses does not seem warranted. The data available for inferring the influence of various allocations of inspection resources on carcass contamination would be

stretched too thin if such inferences were targeted to specific bacterial strains. Furthermore, the lack of available attributions for specific strains would require more extensive assumptions than used here.

3. *Evaluate whether the model source code and mathematics are correct. If not, the reviewer must provide alternative modeling techniques.*

   a. *Are the modeling techniques (model mathematics and equations) appropriate?*

   The logistic regression approach and equations are well documented and appropriate. The SAS code is consistent with tables and text descriptions of methodology and results. The simulations for indiscriminate and discriminating scenarios are well described, in text and appendices. The modeling of the linkage between prevalence in poultry baselines and prevalence of human cases is not well-characterized or validated with available data, as noted above.

   FSIS Response: please see our previous response to this comment.

   b. *Are the methodologies used in the risk assessment for estimating parameters from the data appropriate (i.e., follow scientifically accepted methodologies)?*

   Procedures appear appropriate with the exception of the 'simple prevalence-based risk assessment' approach for modeling dose-dependent relationships for campylobacteriosis and salmonellosis.

   FSIS Response: please see our previous response to this comment.

   c. *Are the data analyses and source code accurate?*

   The SAS code for regression modeling and Excel sheets for simulation modeling are consistent with tables and text descriptions of methodology and results. No data or analyses are provided for modeling dose-dependencies.

   FSIS Response: please see our previous response to this comment.

4. *Evaluate whether adequate sensitivity analysis has been provided. If not, the reviewer must provide an alternative approach or application for sensitivity analysis and/or identify those parameters that should have been included.*

   a. *Have the most important variables in the model been identified?*

The authors assume a 'simple prevalence-based risk assessment' based on Williams et al. (2011) with no biological rational or synthesis of data on dose-dependencies for salmonellosis and campylobacteriosis. No alternatives to this assumption appear to have been tested, nor were results of sensitivity analyses provided to assess the impact on relative risks for procedural changes.

FSIS Response: please see our previous response to this first comment. In the Nov 2012 version of the report, we include a sensitivity analysis section that addresses the reviewer's concern.

   b. *Has an important variable been left out?*

As noted above, the authors chose to assume a 'simple prevalence-based risk assessment' based on Williams et al. (2011). It is unclear how important dose-dependency is to predicting relative risks for procedural changes.

   b. *Has the impact of including or excluding scientific studies or other data been adequately explored?*

No.

FSIS Response: given the limited knowledge available about the steps between the slaughter plant and consumer illness (e.g. preparation, storage, cooking, dose-response by sub-population, etc.), this approach is reasonable. Other risk assessments in both Europe and the USA have shown a linear relationship between prevalence and expected incidence, so the method used here skips these steps without any loss of accuracy.

5. *Evaluate the available data and the underlying assumptions used in this risk assessment. Are they complete and correctly analyzed and interpreted? If not, the reviewer must provide additional data sources and citations (where appropriate) or provide alternative interpretations, analysis, or suggested use of the data.*

a. Have all key studies and data been identified?

The transparency of the report would be improved by inclusion of a table or section that provides more structured information on the available data and underlying assumptions. For a regulatory decision as important as this proposed rule, very little available scientific evidence is provided or referenced.

FSIS Response: we have tried to strengthen the documentation in the Nov 2012 report

None of the studies linking dose and response from clinical and epidemiologic studies of salmonellosis and campylobacteriosis were cited or analyzed systematically. For campylobacteriosis, a series of human volunteer studies were conducted by Tribble et al. (2002, 2004, 2007, 2009, 2010) and a single study by Black et al. (1988). For salmonellosis, a series of human volunteer studies were conducted by McCullough and Eisele (1951a,b,c,d). The journal *Risk Analysis* published 63 manuscripts on *Campylobacter* dose response and 98 on *Salmonella* dose response. Reviews by WHO/FAO could also be cited and included in a synthesis of available datasets and models.

FSIS Response: although entirely appropriate in an academic publication, we do not agree that this was a necessary ingredient in the background material for this RA – given the simplicity of the methodology employed.

c. Have the data been correctly interpreted, analyzed, and used in the risk assessment?

No for studies linking exposure to cases of human illness. It is unclear how consistent these studies are to the forecasts of the 'simple prevalence-based model'. The description provided is not transparent biologically, nor does the manuscript cited provide biologically meaningful interpretation and rationale for comparing and selecting approaches for estimating influence on rates of human illness. It is not clear that changes in pathogen prevalence 'mathematically map' as a Poisson process (or another process) to changes in rates of annual human illness, particularly when serotype prevalence in poultry do not appear to map to observed rates of human salmonellosis cases.

FSIS Response: please see our previous discussion above.

1555
1556

6. *Evaluate the regression analysis used to estimate baseline and scenario aggregate establishment prevalence.*

   a. *Is the technique accurately described, utilized, and appropriate for its intended use?*

The authors provide helpful descriptions of the proposed logistic regression with four decision variables each representing a grouping of off-line inspection procedures for each product-pathogen pair. Some inconsistencies were noted across product-pathogen pairs, as each product-pathogen pair included two significant decision variables, but no two the same variable or same direction (+ or -). The models with significant SNP and NC were chicken-*Campylobacter* and turkey *Salmonella*, but signs of estimated coefficients differed. The decision variable U was significant and in the same direction for chicken- *Salmonella* and turkey- *Campylobacter*, but SNP and SP respectively, were the second significant variables. Expanding Table 1 to list the average number of procedures used as explanatory variables would be more helpful to the reader attempting to interpret these mixed results, rather than the current examples included in the text.

FSIS Response: more detail is provided in extensive appendices included with the November 2012 report

In addition, the authors briefly describe more complex regressions with 43 and 21 decision variables representing specific inspection system procedure codes, as well as previous versions including submodels that demanded a more complex and difficult weighting scheme. Including summary tables of these additional analyses considered, along with reasons for rejection, would be helpful for transparency and completeness.

FSIS Response: more detail is provided in extensive appendices included with the November 2012 report.

The body of the report does not cross-reference the key sections of the appendix that describe rationale/criteria for selection of the 4 decision variable model. The appendix does provide helpful detail on the use of split datasets to demonstrate stability of the aggregate establishment prevalence estimates. The analysis appears to be appropriate for estimating baseline and scenario aggregate establishment prevalence.

   b. *If not, reviewer must provide rationale for why not and detail better alternatives.*

   c. *Are the conclusions drawn from the regression analysis appropriate?*

The conclusions of the regression analysis are well supported, and uncertainties are acknowledged appropriately.

d. If not, reviewer must provide alternative interpretation of the results derived from this analysis.

7. *Evaluate the scenario approach taken to quantify changes in establishment prevalence due to additional off-line inspection activities.*

   a. *Is this scenario approach reasonable, given the limited amount of data available?*

   The scenario approach testing an indiscriminate scenario and the alternative scenario increasing U is reasonable, based on inferences from the HIMP report, descriptions of inspection procedures, and assumptions developed in the body of the report.

   b. *If not, what flaws do you perceive in the rationale and what information is lacking to make the case as proposed?*

   c. *What alternatives exist and how could they be incorporated?*

8. *Evaluate whether the documentation, discussion, and interpretation of results is appropriate. If not, the reviewer must provide an alternative outline and/or approach for adequately and clearly documenting this risk assessment.*

   a. *Is the report clearly written?*

   The report is not clear and transparent or well cross-referenced between the body of the report and the appendices. The gap in analysis of data and models of dose-dependency is puzzling.

   FSIS Response: First, the readability of Nov 2012 report has been improved. Second, because FSIS policies are targeted towards reducing human health risk in young chicken and young turkey slaughter establishments – that is the focus of the risk assessment. The number of human illnesses is the means by which we measure the effectiveness of our policies –but in effect what we are really trying to regulate is the probability of human illness as the product leaves the establishment. Over-emphasis on the uncertainties associated with dose-response tends to envelop the effects of the policy at the point where we have some influence. By simplifying that portion of the farm-to-table risk continuum, we are able to focus on that portion where we have more influence on the outcome.

   b. *Is it complete?*

   No. The report is not transparent or complete as a stand-alone document. The authors assume a 'simple prevalence-based risk assessment' based on Williams et al. (2011) with no biological rational or synthesis of data on dose-dependencies for salmonellosis and

campylobacteriosis. No alternatives to this assumption appear to have been tested, nor were results of sensitivity analyses provided to assess the impact on relative risks for procedural changes.

FSIS Response: please see our previous responses to this reviewer.

c. Does it follow a logical structure and layout?

The organization of material on the regression modeling in the body of the report and the appendix is fragmented and difficult to follow. Combining or cross-referencing this material would be helpful to the reader.

FSIS Response: we have improved the readability of the document with the Nov 2012 version.

d. Is it useful?

Yes, as a proposed framework; no as a regulatory analysis due to incompleteness.
FSIS Response: We respectfully disagree (along with the majority of peer reviewers).

e. Does the risk assessment support the conclusions reached?

Not at present.
FSIS Response: We respectfully disagree (along with the majority of peer reviewers).

Itemized FSIS Replies to Reviewer #3

Reviewer #3's comments:

A marked up copy of the PDF is being provided along with this review, but here is a point-by-point listing of all comments based on the order in which they appear in the document. All of my comments follow the same format first the page number is listed, then the word "content" appears followed by the text of the report that I am commenting on, finally the word "comment" appears, and is followed by my comments relative to the text listed under "content".

Page 7:

Content: "If this efficiency either reduces (or does not change) the occurrence of foodborne pathogens such as Salmonella and Campylobacter on finished poultry products, then a net public health benefit may result."

Comment: It's not clear how no change in the occurrence of a food borne pathogen will result in net public health benefits. I agree that if there's no change in the occurrence of food borne pathogens there may be benefits just not public *health* benefit.

FSIS Response: this wording has been changed in the November 2012 report.

Page 7:

Content: "The original risk management questions were:"

Comment: Please clarify these were the original risk management questions but they are still the current risk management questions as well, correct?

FSIS Response: original and current –clarified in Nov 2012 version.

Page 8:

Content: "As Agency guidance has heretofore been unspecific about procedures that could improve from the new inspection system, an “indiscriminate” scenario is propagated in which all 4 categories of decision variables are randomly changed."

Comment: Not clear what this sentence means. Does “indiscriminate” mean the same as random? If so, why not just call it the random scenario?

FSIS Response: Indiscriminate is not random, but simply refers to the lack of acting on prior information or beliefs that would lead to the targeting of any additional resources towards specific inspection activities. We have clarified this description in the Nov 2012 report.

Page 9:

Content: "These results describe estimated changes in both poultry slaughter establishment prevalence”

Comment: It's not the prevalence of establishments; it's the prevalence of pathogens, right? You should search for "establishment prevalence" throughout the document, and correct as needed

FSIS Response: this has been corrected.

Page 9:

Content: "indiscriminately changed"

Comment: Is "indiscriminately changed" the best phrasing here? It's a bit pejorative sounding, and not completely clear. See also the same comment above.

FSIS Response: see our explanation above.

Page 9:

Content: "(.005, .04)"

Comment: Here and throughout the document fractional decimals are presented without the leading zero. I think the standard method of presenting these types of numbers is with a leading zero. Therefore it should read 0.005 rather than .005.

FSIS Response: this has correctly been identified and corrected.

Page 10:

Content: "This decision variable is poorly understood"

Comment: Why is it poorly understood? What does this mean in layman's terms? Please expand.

FSIS Response: we have rewritten this explanation in the text.

Page 11:

Content: "human Salmonella and Campylobacter illness attributable to poultry."

Comment: Italics needed for pathogen names

FSIS Response; done.

Page 13:

Content: "This should result in the efficient production of poultry products."

Comment: Perhaps you mean this should result in *more* efficient production. They are somewhat efficient already, one would assume.

FSIS Response: we agree.

Page 13:

Content: "either reduces (or does not change) the occurrence of food borne pathogens such as Salmonella and Campylobacter on finished poultry products, then a net public health benefit may result."

Comment: If pathogen prevalence does not change then there should be no net public health benefit. There might however be a cost savings or other benefits. Please clarify.

FSIS Response: this wording has been changed in the Nov 2012 report.

Page 14:

Content: "The four decision variables are Scheduled and Performed procedures (SP), Scheduled and Not Performed procedures (SNP), unscheduled procedures (U), and Non-Compliances (NC)."

Comment: A clear description of what the decision variables mean would be helpful to the lay reader. For example an SNP is a procedure that was planned, but never occurs, correct? When would an unscheduled procedure occur? Is this when the inspector notices that something is wrong? When does non-compliance occur? Could it result from an SP or from a U?

FSIS Response: The definitions for Scheduled and Performed (SP), Scheduled and Not Performed (SNP), Unscheduled (U), and Non-Compliances have been clarified in the main body of the report.

Page 15:

Content: "potentially invalid. ." ***FSIS Response: fixed***

Comment: Extra "."

FSIS Response: typo fixed

Page 16:

Content: "Nevertheless, the sign of the turkey-Salmonella model suggests that reducing SNP will actually increase Salmonella prevalence in turkey."

Comment: But this is a nonsensical finding, correct? Please comment.

FSIS Response: this language has changed in the November 2012 report.

Page 21:

Content: "most likely10%,"

Comment: Typo adds space after “likely”.

FSIS Response: fixed

Page 21:

Content: "An alternative scenario (Increase U) considers how human illness forecasts might change by emphasizing changes to the unscheduled procedures (U) decision variable while leaving other decision variables unchanged."

Comment: Given that inspector time is constrained within a given establishment, is this a valid assumption?

FSIS Response: this is the point. Freeing up additional inspection resources (time and personnel) allows the completion of more of these procedures that are shown to correlate with human health risk.

Page 22:

Content: "Table 4 sows"

Comment: Typo should be "shows".

FSIS Response: fixed

Page 22:

Content: "percent(.021, 32)"

FSIS Response: fixed Comment: Typo should be .32, I assume.

FSIS Response: fixed

Page 22:

Content: "predicts a average "

FSIS Response: fixed Comment: Typo should be an average.

FSIS Response: fixed

Page 23:

Content: "turkey-Campylobacter models, respectively"

FSIS Response: fixed Comment: Typo missing period.

FSIS Response: fixed

Page 23:

Content: " 0..01 "

FSIS Response: fixed Comment: Typo extra "."

FSIS Response: fixed

Page 27:

Content: "Figure 1. "

Comment: Should the cumulative probability not sum to one? Please explain why it does not sum to 1 if this is correct. Also why use cumulative probability instead of a probability distribution function?

FSIS Response: this figure has changed in the November 2012 report. The intent of the Figures 1 – 4 was to provide the reader some insight regarding the cumulative probability around the "no change" (i.e., illnesses neither decrease nor increase) value for Annual Illnesses Avoided. To provide sufficient resolution and balance for these graphs, the graphs were sometimes truncated at larger/smaller values for Annual Illnesses Avoided such that the cumulative probability for values shown did not reach 0 or 1at the left or right extremes, respectively.

Page 28:

Content: "Figure 2. "

FSIS Response: fixed Comment: Same comment as figure 1. ***FSIS Response: fixed***

Page 31:

Content: "This decision variable is poorly understood"

Comment: Again more details are needed. What makes this variable poorly understood?

FSIS Response: see our previous response to this.

Page 32:

Content: "The most reliable implication from the regression models is that increasing unscheduled procedures seems to reduce pathogen occurrence on carcasses."

Comment: Again more details are needed on unscheduled processes. How does an inspector decide to conduct an unscheduled process?

FSIS Response: see our previous response.

Page 32:

Content: "for FSIS.."

FSIS Response: fixed.

Comment: Typo. Extra "."

FSIS Response: fixed

Page 32:

Content: "testin data"

FSIS Response: fixed

Comment: Typo should be "testing".

FSIS Response: fixed

Page 34:

Content: "Regression Modeling Methods and Observational Datasets"

Comment: Understanding the regression modeling is <u>essential</u> to understanding the risk assessment. I would suggest at least a portion of this go in the main document, or a justification be made for relegating the regression models to the appendix.

FSIS Response: We agree. We have rewritten the methodological section of the main body of the Nov 2012 document, including more discussion on the regression analysis.

Page 34:

Content: "Each model evaluates pathogen prevalence in relation to four off-line inspection procedure categories; (i) scheduled and performed, (ii) scheduled but not performed, (iii) unscheduled, and (iv) non-compliances."

Comment: As noted above, these need to be further explained in layman's terms. Clearly there are activities that are being scheduled, and most of the time they are performed, and some times they are not performed. When do unscheduled activities occur? When the inspector feels like it? Are inspectors expected to perform a certain number of unscheduled activities? What event occurs that triggers a "non-compliance"? Can a non-compliance occur from an SP or a U?

FSIS Response: see our previous response.

Page 35:

Content: "increased availability of off-line inspectors should increase unscheduled procedures"

Comment: Why? What triggers an unscheduled inspection? Will increased availability of off-line inspectors increase NC's or at least the chance of an NC?

FSIS Response: As stated in the text unscheduled procedures occur as the result of inspector availability to perform them. Also, as stated in the text, given the observation that there are fewer scheduled but not performed procedures and more unscheduled procedures performed when establishments are fully staffed and off-line inspectors are not required to fill line positions- it may be assumed that increased inspection scrutiny will result in more non-compliances that were not detected previously because of lack of man power. And, it may be expected that continued increased scrutiny will result in a decrease in non-compliances finally resulting in a fully compliant establishment.

Page 35:

Content: "We also assume that – in the long-run – reported non-compliances will decrease with more off-line inspectors in slaughter establishments because such establishments will attain appropriate process control."

Comment: It is very important to emphasize that this is the long run. If there are currently undetected non-compliances, increasing off-line inspectors in the short run will find these non-compliances, and detected non-compliances will go up. Eventually the root-causes should be addressed and the NC's will go down.

FSIS Response: basically what we said in the previous response.

Page 35:

Content: " a random variable that summarized HACCP procedures would need to increase scheduled and performed procedures (and unscheduled procedures) but also decrease scheduled but not performed procedures (and non-compliances)."

Comment: This sentence is unclear, even after repeated reading. Please expand and further explain.

FSIS Response: this sentence is not relevant in the Nov 2012 report.

Page 36:

Content: "There are six general inspection system procedure (ISP) code activity categories captured in the FSIS database (Table 1)."

Comment: It would be most helpful to have the tables and figures embedded in the appendix text in approximately the location where they are first referenced.

FSIS Response: we have improved the readability of the Nov 2012 report.

Page 36:

Content: "Unscheduled procedures are performed according to in-establishment inspector needs;"

Comment: What does this mean in plain English? i.e. what is “in-establishment inspector needs”? Does it mean that inspectors do these when they have time? ***FSIS Response: This is a misstatement. This has been corrected in the document. Inspector needs is changed to inspector availability. The reviewer is correct in stating that unscheduled procedures are performed when all other duties are performed or when there is an obvious non-compliance that needs to be addressed.***

Page 36:

Content: "performed in response to unforeseen hazards,"

Comment: Please give an example of an unforeseen hazard. ***FSIS Response: Unforeseen hazard has been defined in the document.***

Page 36:

Content: "SNP = scheduled not performed procedures for sanitation(01),"

Comment: In fact, this is the IDENTICAL LIST as for SP, correct? If true, why not just say that? Also it would be very helpful if the entire list of procedures for all 4 categories could be explain in plain English. For example, what is a “sanitation(01)”? What is a “fecals (03J)"?

FSIS Response; these lists have been more clearly defined in the document text and appendix.

Page 37:

Content: "U = unscheduled procedures performed for sanitation(01),"

Comment: This appears to be the same as the list for SP and SNP with the addition of emergency procedures. If this is the case why not just say this?

FSIS Response: The reviewer's observation is correct. The lists have been clarified in the document.

Page 37:

Content: "fecals (03J),"

Comment: This is called "fecal" and "fecals". Be consistent.

FSIS Response: The terminology is changed to "fecal check" in the document to more accurately reflect the procedure.

Page 37:

Content: "NC = non-compliant procedures for sanitation(01),"

Comment: As above this appears to be just a minor modification to the same list. Isn't there an easier way and clearer way to explain this same information rather than just repeating the same list?

FSIS Response: No, because this is actually a simplification of the total data analyzed. We decided to err on the side of repetitive simplicity rather than exhaustive complexity.

Page 37:

Content: "The re-hang variable distinguishes between locations of sample collection (where 1 signifies post-chill samples and 0 signifies re-hang samples)."

Comment: Why call this variable rehanging when it refers to the location? Wouldn't location be a better variable name?

FSIS Response: Actually, in the slaughter establishment the re-hanging activity is accomplished at the rehang location which is a specific location identified in each establishment.

Page 37:

Content: "The categorical month variable breaks down the time dependency into 39 consecutive months."

Comment: Why consider months at all in the model? What would happen if you ran the model ignoring the month variable?

FSIS Response: Please refer to the response to this question for reviewer 1.

Page 37:

Content: "District 90 is used as the reference."

Comment: Is this arbitrary? Des it matter?

FSIS Response: ***District 90 is an arbitrary selection. Any other reference would have yielded difference numerical estimates for each parameter but the prevalence estimate would be the same.***

Page 37:

Content: "The categorical district variable differentiates the 15 districts."

Comment: As above with respect to months, why use district at all as a variable? What would happen to the model if this variable was not used?

FSIS Response: The district variable was found to be important to the model because omitting it resulted in a significantly decreased amount of variance explained by the model. This was a good categorical variable because of the high degree of variability between districts.

Page 37:

Content: " Line-speed,"

Comment: Explain. What are the units? How is line speed measured? Does it change throughout the day or day to day?

.

FSIS Response: Line speed has been defined in the document.

Page 37:

Content: "Number of establishment inspectors,"

Comment: As above, explain. Does this vary? Is this an average?

.

FSIS Response: The number of establishment inspectors variable has been defined in the document.

Page 37:

Content: "Line count"

Comment: Is this the number of processing lines in the plant?

FSIS Response: The number of processing lines definition has been made explicit in the document as the number of slaughter lines in the establishment.

Page 37:

Content: "(MAESTRO, NELS, Nu-Tech, Nuova, SIS, HIMP, Traditional, and Religious Slaughter)."

Comment: These all need to be explained somewhere.

FSIS Response: The inspection system abbreviations have been defined in Tables 4 and 5 in the appendix.

Page 37:

Content: "HACCP size,"

Comment: What is "HACCP size"?

FSIS Response: The definition of HACCP size has been made explicit in the document as the same as the Small Business Administration definition of business size.

Page 37:

Content: "inspector positions,"

Comment: How is this different from number of establishment inspectors?

FSIS Response: The definition of inspector positions has been clarified in the document to mean the number of supervisors, on-line inspectors, and off-line inspectors for each establishment as separate variables.

Page 37:

Content: "time in weeks (52), time in months (12), time in quarters (4 and 12), time in years (4),"

Comment: Explain how these are different from the categorical dates used.

FSIS Response: These five types of categorical time variables have been defined in Table 2 of the appendix.

Page 38:

Content: "septicemia-toxemia condemnations of carcasses,"

Comment: More details. Is this the number of carcasses, percent, or something else?

FSIS Response: This variable refers to the daily number of carcasses condemned in the septicemia-toxemia category for each establishment.

Page 38:

Content: "contamination (fecal, ingesta, body fluids, etc.) of carcasses,"

Comment: As above, number, percent, etc.

FSIS Response: The contamination variable has been clearly redefined in the document.

Page 38:

Content: "Some coefficients have non-significant contributions according to a 0.05 significance assumption but were retained in the model for consistency across all four models."

Comment: Were any of the coefficients non-significant across all four models?

FSIS Response: No.

Page 38:

Content: "Among structural variables, a common finding was the (statistically significant) negative coefficient for HIMP participation across all four models. The HIMP participation variable is a separate structural variable in the chicken models, but it is incorporated into an inspection system variable in the turkey models. "

Comment: I'm not sure how you can make this statement across all four models since the HIMP variable is confounded within the turkey model. Please explain why it is incorporated into an inspection system variable in the turkey models.

FSIS Response: In the turkey models, when the "coded" categorical variables relative to a base system for establishment inspection system are decoded to produce the "decoded" main effects models, the same significance relationships hold for HIMP establishments as when the establishment inspection system variables were in relative form.

Page 39:

Content: "The BX element in Table 9 is the sum of cross products of the B regression parameter"

Comment: What is the B regression parameter?

FSIS Response: The scalar quantity, η, is defined in the text as equal to the coefficient-wise multiplication and summation (linear form) of the vectors B and X and further explained in Appendix Tables 9, 11, 13, and 15.

Page 40:

Content: "100% sensitivity and 0% 1-Specifity corner point."

Comment: Typo the word "specificity" is misspelled

FSIS Response: fixed

Page 40:

Content: "The predictive order of c coefficients across the four models is 0.702, 0.710, 0.792, and 0.852,"

Comment: Please tell us which coefficient corresponds to which model.

FSIS Response: The predictive order of c coefficients across the four models is 0.702, 0.710, 0.792, and 0.852 respectively for young chicken Campylobacter the least predictive, young turkey Salmonella somewhat more predictive, young chicken Salmonella still more predictive, and the young turkey Campylobacter model the most predictive. This was an oversight that is corrected in the risk assessment text.

Page 40:

Content: "The 03, 04, and 06 procedure elements have this characteristic in the chicken-Salmonella model and the 04 and 05"

Comment: Please tell us what these procedure element numbers correspond to in words.

FSIS Response: The wording has been changed to be more explicit in the document.

Page 40:

Content: "The turkey-Campylobacter model has the 03 and 06 elements significant. It is not clear why the 05 and 06 coefficients have significant positive signs in the chicken models. Table 15 shows the results for further disaggregated models. It becomes clear that the 03J procedures are the drivers decreasing prevalence for HACCP in the chicken-Campylobacter model and the 06D01 procedures are drivers"

Comment: As above please use words not numbers to describe the coefficients.

FSIS Response: The wording has been changed to be more explicit in the document.

Page 40:

Content: "Table16"

Comment: Typo missing space. ***FSIS Response: fixed***

Page 40:

Content: "Because the original observational dataset used to develop the four models for scenario analysis excluded some of the establishments that are predicted to adopt the new inspection system requiring a shift of the majority of on-line inspectors to off-line inspection duties while leaving one inspector on-line for final carcass inspection according to the Preliminary Regulatory Impact Analysis (PRIA) of the proposed poultry slaughter rule, we decided to create a simulated dataset corresponding to all establishments expected to adopt the new inspection system."

Comment: This is an incredibly long sentence. Please break it into shorter sentences.

FSIS Response: The sentence has been simplified in the document.

Page 40:

Content: "none of the very small establishments in the observational dataset are expected to adopt the new inspection system."

Comment: Why are the very small establishments not expected to adopt the new system? Please explain. ***FSIS Response: This is an assumption from the PRIA that is now made explicit in the document text.***

Page 41:

Content: "The 19 establishments in the “other” category were placed in either the chicken or the turkey datasets according to size and predominant production characteristics."

Comment: Please explain what these establishments are. Are they establishments that process both turkey and chicken? Or something else?

Page 41:

Content: "1-Specificity"

Comment: Please explain what "1-specificity" means.

FSIS Response: This use of this term has been made clear in the document.

Page 43:

Content: "Appendix Table 1. "

Comment: What is the purpose of the two columns that don't have column headers that start with the number one and the number 24?

FSIS Response: The absent column heading has been changed to number (No.).

Page 43:

Content: "Code Sum "

Comment: Does this column tell the reader anything useful?

FSIS Response: The heading now has been explicitly defined in the table.

Page 43:

Content: "Other Sum "

Comment: Likewise for this column. Is any information being communicated to the reader?

FSIS Response: The heading was misleading and has been changed to “detail sum” and is now fully explained in the table.

Page 45:

Content: "Appendix Table 1."

Comment: What is the purpose of breaking this table into a separate table when the first table 1 above is already split across a page break?

FSIS Response: please see our earlier comments with respect to Appendices material.

Page 46:

Content: "loglinespeed "

Comment: Is this the logarithm of the linespeed?

FSIS Response: yes it is the base ten logarithm..

Page 46:

Content: "logInspectors "

Comment: Is this the logarithm of the number of inspectors? ***FSIS Response: yes, it is the base ten logarithm.***

Page 56:

Content: "BX (rehang= mean)"

Comment: I understand what this table is trying to say but these descriptions are very hard to interpret. They could be rewritten in plain English. ***FSIS Response: please see our earlier comments with respect to Appendices material.***

Page 61:

Content: "sum01_U "

Comment: Please use English here rather than variable names. ***FSIS Response: please see our earlier comments with respect to Appendices material.***

Page 62:

Content: "sum01B_U"

Comment: Please do not use IST code here. Please write in English.

FSIS Response: All tables have been annotated to make the ISP code jargon clear as to its meaning.

Page 63:

Content: "Number of Establishments Expected to adopt the New Inspection"

Comment: What information is used to calculate this expectation?

FSIS Response:The language in the risk assessment has been corrected to distinguish between the expected number of establishments to adopt the new inspection system given in the PRIA for all poultry slaughter establishments and the expectation for the number of establishments to adopt the new system based on our observational study. The reviewer is referring to the latter expectation. The expectations for large, small, and very small establishments based on the observed dataset were estimated. These expectations are the distribution averages of a Monte Carlo process of repeated random selection of establishments with known establishment characteristics that we had data for and for those establishments for which we only had incomplete data because they were not in our observed dataset. The assumptions used to calculate the expectations are now clarified in the text.

Page 63:

Content: "switch"

Comment: What does switch mean?

FSIS Response: the number of establishments expected to adopt the new inspection system This term has been annotated in the tables.

1. Evaluate if the overall approach for modeling the public health benefits potentially realized from the change in inspection system examined is fundamentally sound.

a. Is the overall approach used in the analysis to evaluate the linkage between inspection activities and potential reductions in annual human illnesses fundamentally sound? The regression model used to estimate changes in establishment prevalence should be addressed separately from the model used to estimate reductions in annual human illness.

Both the risk assessment and the regression model appear to be fundamentally sound. The regression model description in the appendix contains a great deal of jargon and otherwise unexplained information. It would benefit the reader if the jargon could be eliminated or explained.

FSIS Response: We have attempted to clarify the jargon used in the risk assessment with more complete explanation of individual jargon items.

b. If not fundamentally sound, in each case, what problems exist and how should they be addressed?

As noted above, I believe the analysis is fundamentally sound however the presentation is unclear. Readers of the report would benefit from a clarified presentation.
.

FSIS Response: We have attempted a more clarified presentation in the November 2012 version which we think is much improved.

2. Evaluate the complexity of the model in areas where the reviewer identifies limitations, weaknesses, or inadequacies; the reviewer must provide alternative data, data analysis, and/or modeling approaches.

   a. Is the model too complex, or not complex enough, to adequately address the risk management questions?

   The model appears to have the correct degree of complexity to adequately address risk management questions. As noted in my main comments above, I question the need to include the months as variables and the districts as variables.

   FSIS Response: This has been explained in the comments to another reviewer.

   b. Is the model over- or under-parameterized?

   The parameterization of the model appears adequate.

   c. Does the model adequately characterize the uncertainty present?

   Yes.

   d. Is variability sufficiently addressed?

   Yes.

3. Evaluate whether the model source code and mathematics are correct. If not, the reviewer must provide alternative modeling techniques.

   a. Are the modeling techniques (model mathematics and equations) appropriate?

   The modeling techniques both math equations appear appropriate. As noted above I question the need to include some of the variables in the regression model. The authors should justify the inclusion of these variables.

   FSIS Response: please see our response above.

   b. Are the methodologies used in the risk assessment for estimating parameters from the data appropriate (i.e., follow scientifically accepted methodologies)?

The methodologies used are scientifically accepted.

c. Are the data analyses and source code accurate?

The analyses and the source code appear to be accurate.

4. Evaluate whether adequate sensitivity analysis has been provided. If not, the reviewer must provide an alternative approach or application for sensitivity analysis and/or identify those parameters that should have been included.

a. Have the most important variables in the model been identified?

The most important variables in the model do appear to have been identified.

b. Has an important variable been left out?

No important variables appear to have been left out.

c. Has the impact of including or excluding scientific studies or other data been adequately explored?

The document contains very few scientific studies. This is largely appropriate however because the studies that are referenced are generally federal reports that informed the risk assessment. The small number of studies published in the scientific literature that are cited are appropriate.

5. Evaluate the available data and the underlying assumptions used in this risk assessment. Are they complete and correctly analyzed and interpreted? If not, the reviewer must provide additional data sources and citations (where appropriate) or provide alternative interpretations, analysis, or suggested use of the data.

a. Have all key studies and data been identified?

Yes.

b. Have the data been correctly interpreted, analyzed, and used in the risk assessment?

Yes, the data appear to have been correctly interpreted and analyzed.

6. Evaluate the regression analysis used to estimate baseline and scenario aggregate establishment prevalence.

   a. Is the technique accurately described, utilized, and appropriate for its intended use?

   The regression analysis appears to be appropriate for its intended use. There are a number of places where the description of the variables could be significantly improved. These places have been indicated in my general comments above.

   FSIS Response: the November 2012 report improves readability of the document and we have made changes where indicated in the reviewer's general comments.

   b. If not, reviewer must provide rationale for why not and detail better alternatives.

   The single biggest problem that the report suffers from is its lack of intelligibility to an informed lay reader. The report assumes that the reader understands all of the phrasing and jargon used within the context of FSI S inspections of chicken and turkey slaughter facilities. While much of the definitions can be inferred from context, the reader should not have to work that hard. Once all of my comments listed in the general comments section above are addressed the document should have a much-improved readability to informed lay reader.

   FSIS Response: We have improved the readability of the document in the November 2012 report by explaining difficult to understand jargon and more fully explaining the model assumptions and results.

   c. Are the conclusions drawn from the regression analysis appropriate?

   The conclusions drawn appear to be appropriate.

   d. If not, reviewer must provide alternative interpretation of the results derived from this analysis.

7. Evaluate the scenario approach taken to quantify changes in establishment prevalence due to additional off-line inspection activities.

   a. Is this scenario approach reasonable, given the limited amount of data available?

   The scenario approach appears reasonable. As noted above however, it is difficult to understand in many cases exactly what is meant by the different terms used in the scenarios.

   b. If not, what flaws do you perceive in the rationale and what information is lacking to make the case as proposed?

   The document could be improved by providing additional information and definitions as noted the general comments section above.

   FSIS Response: We have included additional tables, annotation, textual information, jargon definitions as indicated above.

   c. What alternatives exist and how could they be incorporated?

   See detailed comments above.

   FSIS Response: See our comments above.

8. Evaluate whether the documentation, discussion, and interpretation of results is appropriate. If not, the reviewer must provide an alternative outline and/or approach for adequately and clearly documenting this risk assessment.

a. Is the report clearly written?

Single biggest issue with the report is its lack of clarity in some places. If all of my comments noted in the general section above are addressed, this should significantly improve the intelligibility and the clarity of the report.

FSIS Response: As stated above, we have improved the readability, intelligibility, and clarity of the report through additional text, tables, definitions, and annotations.

b. Is it complete?

Definitions of important terms are missing. Details are provided above.

FSIS Response: please see or response above.

c. Does it follow a logical structure and layout?

The report is generally logically structured. I think relegating the regression analysis to an appendix diminishes its importance. As noted above, understanding the regression analysis is central to understanding the risk assessment upon which it is based. Additionally as noted above including the figures and the tables at the end of the appendix distracts the reader and reduces readability.

FSIS Response: we have expanded discussion of the regression analysis in the methodology section of the main report. See our previous response on the Appendices.

d. Is it useful?

The report is highly readable. It appears to sufficiently support the case for the implementation of a new inspection system.

e. Does the risk assessment support the conclusions reached?

Yes.

2367

Itemized FSIS Replies to Reviewer #4

Reviewer #4's comments:

This 2011 version of the risk assessment is an updated version of a previous 2008 risk assessment, with new data and a modified modeling approach. The main goal of the risk assessment was to evaluate the change in the prevalence of both, *Salmonella* and *Campylobacter,* on chicken and turkey and, subsequently, attributable human illnesses as a result of changes in off-line inspection procedures in FSIS poultry slaughter facilities. Overall, given the scope of the risk assessment, the approach undertaken to assess the relationship between inspection activities and potential changes in annual human illnesses seems logical and appropriate. The modeling techniques and methods, data and results analyses appear appropriate. It seems relevant studies and data were used in this risk assessment. Nonetheless, the report is not well written and needs additional proof reading.

FSIS Response: we have revised the document. The Nov 2012 version is more readable and has been proofed.

Please find below the responses to each charge question.

1. Evaluate if the overall approach for modeling the public health benefits potentially realized from the change in inspection system examined is fundamentally sound.

a. Is the overall approach used in the analysis to evaluate the linkage between inspection activities and potential reductions in annual human illnesses fundamentally sound? T The regression model used to estimate changes in establishment prevalence should be addressed separately from the model used to estimate reductions in annual human illness.

FSIS Response: we have revised the document – along with the model used to estimate changes in human illness. The regression model is described separately from the description of the simulation model used to predict changes in attributable human illnesses.

b. If not fundamentally sound, in each case, what problems exist and how should they be addressed?

Comment: The objective of this risk assessment is to evaluate the change in the prevalence of both, *Salmonella* and *Campylobacter*, on chicken and turkey and, subsequently, in the attributable human illnesses as a result of changes in inspection procedures in FSIS poultry slaughter facilities. A logistic regression analysis was performed to estimate the relationship between the prevalence of *Salmonella* or *Campylobacter* on carcasses and off-line inspection procedures, followed by a stochastic simulation to predict the effect of changes in off-line inspection procedures on changes in human *Salmonella* or *Campylobacter* illnesses attributable to the consumption of chicken and turkey. Overall, given the scope of the risk assessment, the approach undertaken to assess the relationship between inspection activities and potential changes in annual human illnesses seems logical and appropriate.

The change in the number of illnesses by a proposed inspection procedure was estimated by a simple prevalence-based calculation based on a published paper by Williams et al. 2011. This prevalence-based method is simply a linear relationship between contaminated carcasses prevalence and human illnesses, which suggests that number of illnesses avoided by a policy aims at reducing prevalence, is a simple proportion of the number of illnesses for baseline scenario, i.e., that occurred prior to implementing the policy. However, estimation of human illnesses is not a simple process as reflected by this approach. In addition to the existence of variability among strains of pathogens, among population groups of different susceptibility, there are many steps involve after carcasses leave the primary processing facilities to arrive at consumer's table, which may change the contamination status and microbial level in the food used for consumption. Although these factors along with dose-response modeling were

not considered, because of the scope of the risk assessment that focused on inspection procedures at primary processing facilities, the approach undertaken to estimate change in human illnesses is reasonable and seems appropriate.

FSIS Response: we agree with the reviewer, but as the reviewer points out, appropriately chose to focus on that aspect of the farm-to-table continuum for which the Agency is attempting to influence.

2. Evaluate the complexity of the model in areas where the reviewer identifies limitations, weaknesses, or inadequacies; the reviewer must provide alternative data, data analysis, and/or modeling approaches.

   a. Is the model too complex, or not complex enough, to adequately address the risk management questions?

   b. Is the model over- or under-parameterized?

   c. Does the model adequately characterize the uncertainty present?

   d. Is variability sufficiently addressed?

Comment: This reviewer appreciates the efforts of carefully considering several alternative sets of decision variables and finally choosing four defined categories (decision variables) such as Scheduled and Performed procedures (SP), Scheduled and Not Performed procedures (SNP), Unscheduled procedures (U), and Non-Compliances (NC) in the analyses. Four decision variables represent the sum of activities across the various Inspection System Procedure (ISP) codes into mutually exclusive classes. Although the whole spectrum of variability and uncertainty in the data set may not be captured by such aggregation, this approach seems provide meaningful results. The authors indicated that this approach also avoids over-interpretation of specific procedures that might simply reflect random associations that can occur with over-parameterized models. While inclusion of many variables in a model appear adequate and add complexity in the analysis, the model-generated results may be intractable and very difficult to interpret.

FSIS Response: we agree with the reviewer, and this is a primary reason for aggregation across procedure types in the final regression analysis used.

In the model, while estimating the change in human illnesses that could occur as a result of implementation of the new inspection system, uncertainty were incorporated for the regression coefficients, change in off-line inspection activities with the new inspection system, in the current estimate of human illnesses using probability distributions. Overall, the characterization of uncertainty appears reasonable.

The uncertainty in the current annual rate of product-pathogen illness (λ_{ill}) was characterized as a lognormal distribution with mean (μ) and standard deviation (σ). The mean and standard deviation values for the lognormal distributions were estimated using a percentile fitting algorithm (described in page 20)
[FSIS Response: note in the Nov 2012 report these page #s have changed]

and then used in the lognormal distributions as parameter values. The authors mentioned that this approach is a reasonable approximation of the intended uncertainty distribution. Instead of this approximations, the authors could define the lognormal distribution in @Risk with percentile values (e.g., 5^{th}, 50^{th}, and 95^{th}) to get the better representation of the actual distribution. This could be done by selecting "Alternate Parameters" instead of "Standard" Parameters while using "Define Distribution" menu. There may not be any changes in results either way one defines the uncertainty distributions, as both distributions seem approximately the same.

FSIS Response: we agree that alternative methods would have produced similar uncertainty distributions for attributable human illnesses. Nevertheless, the published credibility bounds from Scallan et al. (2011) were used here because these were available transparently. .

3. Evaluate whether the model source code and mathematics are correct. If not, the reviewer must provide alternative modeling techniques.

   a. Are the modeling techniques (model mathematics and equations) appropriate?

   b. Are the methodologies used in the risk assessment for estimating parameters from the data appropriate (i.e., follow scientifically accepted methodologies)?

   c. Are the data analyses and source code accurate?

Comment: It would have been better if the authors could have presented information about different model variables, equations, etc. in the excel sheet in a clear way. It is difficult to quickly locate and follow the models and results as presented in the excel sheets provided. The modeling techniques seem appropriate and the model source codes and mathematics are correct.

FSIS Response: Because there are 2 primary components on top of this "model", it is difficult to glean everything from the excel spreadsheet. All equations used in the simulation analysis are clearly available in the excel spreadsheets. One must, however, refer to the text and appendices for more information on the equations used in the regression analyses. Note: the results of the regression analyses are incorporated into the simulation analyses in a slightly different way in the Nov 2012 report – and the corresponding spreadsheets have changed as well.

On Page 46: Appendix Table 2, the estimate for "Intercept" was mentioned as "-1.8967" whereas in SAS code file this value is "-1.9647". This reviewer is wondering about this discrepancy**.**

FSIS Response: the correct intercept is cited here, however, in the appendix 2 table of the risk assessment document - the incorrect intercept is given. This has been corrected.

On Page 20, it is mentioned that "Scheduled and performed and unscheduled procedures in an establishment could either increase, *decrease*, or stay the same, once an establishment adopts the new inspection system in the proposed rule." However, for the SP and U decision variables the authors represented A_i as Pert distribution with values 1.0, 1.25, and 1.6, which implies the decision variables did not change, increased by 25%, and increased by 60%, respectively. I was wondering why not any other values were tested for to take into account any *decrease* in the scheduled and performed and unscheduled procedures in an establishment.

FSIS Response: the reviewer's point is correct. However, as "decisional" variables, we were not concerned with a decrease in scheduled and performed or unscheduled procedures – only the uncertain potential for allocating more resources to these inspection activities.

In the second paragraph on page 24, the authors mentioned "The combined illnesses avoided results suggest the probability that illnesses associated with both young chicken and turkey establishments might increase is ~0.13. This result suggests with approximately 87% confidence that aggregate human illnesses will be unchanged or decrease following an indiscriminate implementation of the proposed poultry rule." And on the last paragraph on the same page, "…. These results suggest that aggregate human illnesses will be unchanged - or decrease - with approximately 100% and 94% confidence among young chicken and young turkey establishments, respectively, if increasing unscheduled procedures is emphasized in the proposed rule. This reviewer suggests changing the word "confidence" as this is mere a proportion or percentage

FSIS Response: the results have changed in response to changes made to the model and the language has been modified in the Nov 2012 report.

Although described in texts on page 24, please provide these numbers 0.13, 0.0009, and 0.0603 for combined illnesses avoided for chicken and turkey, in Tables 5 and 6. These numbers (0.1281 for chicken, 0.1293 for turkey, for indiscriminate scenario; and 0.0009 for chicken and 0.0603 for turkey, for alternative scenario (increased unscheduled) were found in the excel file "PSRA RA 2012 (supplemental) - NEW RUN SCEARIO w Agg illness". In excel file, this reviewer could not find the numbers for combined illnesses avoided for *Salmonella* and *Campylobacter* such as the values 0.0407 & .40, and 0.0058 & 0.0501, for the probability of increased illnesses, as mentioned in Tables 5 and 6, respectively.

FSIS Response: the results have changed in response to changes made to the simulation model and the Nov 2012 report reflects these changes.

4. Evaluate whether adequate sensitivity analysis has been provided. If not, the reviewer must provide an alternative approach or application for sensitivity analysis and/or identify those parameters that should have been included.

   a. Have the most important variables in the model been identified?

   b. Has an important variable been left out?

   c. Has the impact of including or excluding scientific studies or other data been adequately explored?

   Comment: In the report, this reviewer could not find any explicit section on sensitivity analysis (if any). On pages 20-21, the authors only mentioned, they tested the sensitivity of the assumptions for values of the adjustment parameter (A_i) for SNP and NC variables by changing the minimum value of the Pert distribution but the results were not significantly altered.

   FSIS Response: No explicit section was included on 'sensitivity analysis' in the Nov 2011 report, but, by design, this type of modeling framework incorporates fairly extensive implicit sensitivity analysis. For example, The out-of-sample regression model evaluation reported in the original appendix, as well as the Nov. 2012 update is an important element of the implicit sensitivity analysis. We have modified the Nov 2012 report to explicitly include a section in the results on sensitivity analysis for the appropriate input variables mentioned.

5. Evaluate the available data and the underlying assumptions used in this risk assessment. Are they complete and correctly analyzed and interpreted? If not, the reviewer must provide additional data sources and citations (where appropriate) or provide alternative interpretations, analysis, or suggested use of the data.

   a. Have all key studies and data been identified?

   b. Have the data been correctly interpreted, analyzed, and used in the risk assessment?

   Comment: The microbiological contamination data for this risk assessment were obtained from different surveys conducted by FSIS: Young Chicken Baseline study (July 2007 through September 2008), Young Turkey Baseline study (August 2008 through July 2009), and from PR/HACCP *Salmonella* verification program (from July 2007 to September 2010). Based on the total numbers of samples and establishments from which microbial data were collected implies good and quality data. From FSIS's PBIS database, corresponding inspection activities data were taken for *Salmonella* and *Campylobacter* prevalence data for the same establishments and timeframes. Estimates for the number of human illnesses due to *Salmonella* and *Campylobacter* attributable to young chicken and turkey consumptions were based on the annual domestically acquired foodborne illnesses recently estimated by the CDC (Scallan et al., 2011). It appears that relevant data were identified and used in this risk assessment.

6. Evaluate the regression analysis used to estimate baseline and scenario aggregate establishment prevalence.

   a. Is the technique accurately described, utilized, and appropriate for its intended use?

   b. If not, reviewer must provide rationale for why not and detail better alternatives.

   c. Are the conclusions drawn from the regression analysis appropriate?

   d. If not, reviewer must provide alternative interpretation of the results derived from this analysis.

   Comment: This reviewer thinks it is appropriate to use the logistic regression analysis to estimate the prevalence of *Salmonella* and *Campylobacter* on poultry carcasses both at the baseline and alternative scenarios (based on four categories of decision variables). The outcome of interest is a binary variable, i.e., either *Salmonella* or *Campylobacter* positive on carcass and as such the relationship between the outcome and variables was estimated using logistic regression (with logit link). Separate logistic regressions were performed for each product-pathogen pair (i.e., young chicken-*Salmonella*, young chicken-*Campylobacter*, young turkey-*Salmonella*, and young turkey-*Campylobacter*). Overall, the regression model, analysis and interpretation of results seem logical and appropriate.

In the regression model, "MONTH" was included as a categorical variable. This reviewer was wondering if "MONTH" could be changed to 12 categories as January, February …instead of coding each month as a unique category. For example, in the Chicken-*Salmonella* model, data were from the samples collected at 39 different months and used as such. I understand that for *Campylobacter* data are available for 12 months. However, I was wondering if we could get a sense of seasonal variation in contamination prevalence by coding "MONTH" to 12 categories (such as for *Salmonella* here).

FSIS Response: The reasons for using the month categories as defined in the text and the consideration of seasonal categories have been given in response to other reviewers.

On Page 39 of the report, the authors mentioned "For model evaluation and validation, we randomly split the datasets used in model development, re-estimated the regression coefficients for each subset of data and assessed the stability of the prevalence estimates." The process of randomly splitting the dataset needs to be mentioned. Also this reviewer is thinking how about doing this validation with an independent data set, if available. Basically, same data that were

used for model development were used for validation and checking the model stability. How about splitting the data into two halves and use one half for model development and the other half for model validation?

FSIS Response: The split dataset procedure reported seemed adequate to prove model stability. However, the SAS logistic procedure used also includes delete-one validation. This procedure indicated sufficient model validity had been achieved for each of the four models in addition to the data splitting validation reported in the risk assessment. In a sense the model has also been validated using alternative data over time. A 2008 version of the analyses used data from an earlier time frame, and came to similar results.

On Page 40, in the first paragraph the authors mentioned "However, all models are sufficiently predictive with areas under the curve all greater than 0.7." Please provide a reference for this.

FSIS Response: References have been provided in the risk assessment table of references Hanley et al, 1982)***. In addition it has been noted that the recommended statistical test from the reference was done for each model to show that each AUC was significant at the 95% level of confidence.***

7. Evaluate the scenario approach taken to quantify changes in establishment prevalence due to additional off-line inspection activities.

   a. Is this scenario approach reasonable, given the limited amount of data available?

   b. If not, what flaws do you perceive in the rationale and what information is lacking to make the case as proposed?

   c. What alternatives exist and how could they be incorporated?

   Comment: The scenario approach to predict how prevalence of both *Salmonella* and *Campylobacter* on poultry carcasses and ultimately annual human illnesses might change based on four categories of decision variables (SP, U, SNP, and NC) seems reasonable. The authors mainly evaluated an indiscriminate scenario, where there would be an indiscriminate change across all four decision variables and an alternative scenario, which considered the effect of only increasing unscheduled procedures (discriminative scenario).

8. Evaluate whether the documentation, discussion, and interpretation of results is appropriate. If not, the reviewer must provide an alternative outline and/or approach for adequately and clearly documenting this risk assessment.

   a. Is the report clearly written?

   b. Is it complete?

   c. Does it follow a logical structure and layout?

   d. Is it useful?

   e. Does the risk assessment support the conclusions reached?

Comment: In this reviewer's opinion, overall, the report is not well written and this report needs additional proof reading. Some of the suggestions are given below:

FSIS Response: the Nov 2012 Report has been re-written and proofed.

Page 10, in the last paragraph, the authors mentioned "In general, the probability that indiscriminate changes in off-line inspection procedures will increase the annual rate of human illnesses is small, and there is a greater probability that such changes would contribute to no net change or even reductions in human illnesses." I was wondering greater than what? The authors need to provide information on what they are comparing.

FSIS Response: the Nov 2012 Report has been re-written and proofed.

Page 11, for the answer to the risk management question (Q3) Where within the establishment can relocated inspection activities have the most impact toward reducing microbial prevalence and corresponding human illness?, the authors replied "The most reliable implication from the regression models is that increasing unscheduled procedures seems to reduce pathogen occurrence on carcasses." Although this statement appears correct, this statement is equally valid for the indiscriminate scenario, based on the reported results. On Pages 9-10, in *Model Results* section the authors mentioned that when off-line procedures are indiscriminately changed, for chickens, the estimated mean of decrease in prevalence is 2% for *Salmonella*, and 0.02% increase in prevalence for *Campylobacter*. On the other hand, for unscheduled inspection procedures the decrease in prevalence values was 2% for *Salmonella* and 0.5% for *Campylobacter*. For turkey, the corresponding decrease in

prevalence value, for indiscriminate scenario was 4% for *Salmonella* and 17% for *Campylobacter* and for unscheduled scenario was 3% and 17%.

FSIS Response: the Nov 2012 Report has been re-written and proofed.

Page 14; Lines 1-2: "Logistic regression analysis is performed to estimate the relationship between off-line inspection procedures and contamination of carcasses with either *Salmonella* or *Campylobacter*." It is not apparent from the sentence, which are the off-line procedures. This reviewer recommends including the information about four decision variables here: Scheduled and Performed procedures (SP), Scheduled and Not Performed procedures (SNP), Unscheduled procedures (U), and Non-Compliances (NC).

FSIS Response: the November 2012 Report has been re-written and proofed.

Page 22-23: Results section: It seems there is a major error in presenting the results for young chicken establishments and young turkey establishments from Tables 3 and 4. Two paragraphs on top of Table 3 are exactly the same as on bottom of Table 3. It appears that the authors forgot to edit the text appropriately. This reviewer also suggests combining Tables 3 and 4 to one table for better comparison of results.

FSIS Response: this error has been corrected in the November 2012 Report.

Page 27-30: Figures 1-4: In figure captions, authors should clearly mention which description they are referring to for the figure legends inside the figure; it is not clear. For example, this reviewer is wondering what is "No change"?

FSIS Response: these figures have changed in the November 2012 report.

Page 23; the last paragraph and Page 32; the first paragraph, is it Table 5 instead of Table 3?

FSIS Response: fixed.

Page 32: last sentence: spelling error "testing".

FSIS Response: fixed.

In Appendix, when referring to any Table, it would be better to write it as Appendix Table # in the text. Otherwise, if only Table # is written, it is confusing to readers whether the authors are referring to tables inside the main report or in the appendix.

FSIS Response: these references have been changed as the reviewer suggests.

Pages 38-39: Last paragraph of Page 38 and first paragraph of Page 39: This reviewer suggests a table with all three statistics for all four product-pathogen models.

Pages 61-62: Appendix Tables 14 and 15: Please provide information on why the authors have not included results for Young Turkey-Salmonella; need to provide information from page 40, "Because the turkey-*Salmonella* model does not have a significant aggregate coefficient only the three remaining models were considered."

FSIS Response: This omission has been corrected in the risk assessment tables referred to.

"Forecast" is used throughout the report. Is there any specific reason for such use? This reviewer suggests considering replacing that with "predict".

FSIS Response: this language has been changed.

Appendix #1: List of Peer Reviewers with Brief Biographical Sketches

NOTE: Reviewers were blinded until the Reviews were completed – and remain blinded as to who submitted which review.

Peg Coleman

Ms. Peg Coleman is a Senior Scientist and sole proprietor of Coleman Scientific Consulting, and serves as a Medical Microbiologist for ICF International. She is a risk assessor with thirty years of experience in regulatory, consulting, and academic environments synthesizing bodies of scientific data and technical information to support risk assessments for chemical, physical, and microbial hazards in air, food, and water. She has been invited to serve as an expert reviewer on projects with National Academies of Science committees and multiple government agencies who seek to develop comprehensive guidance for microbial risk assessments and improve their practice in support of policy decisions. Ms. Coleman delivers briefings and lectures on microbial risk for organizations including the American Association for the Advancement of Science, Society for Risk Analysis (SRA), and Interagency Risk Assessment Consortium. She serves on the editorial board for the SRA journal Risk Analysis, and is a reviewer for other scientific journals and the National Academy of Science. She received her M.S. in Biology/Biochemistry from Utah State University, and a second M.S. in Medical Microbiology from the University of Georgia.

Abani Pradhan, Ph.D.

Dr. Abani Pradhan is an Assistant Professor in the Department of Nutrition and Food Science & the Center for Food Safety and Security Systems (CFS3) at the University of Maryland (UMD), College Park. Prior to joining UMD, Dr. Pradhan was working as a Research Associate at Cornell University in Ithaca, New York, where he also received his post-doctoral training. He received his Ph.D. in Biological Engineering from the University of Arkansas. His research interests include food safety, quantitative microbial risk assessment, predictive microbiology, food safety engineering, and molecular epidemiology. Some of his recent research projects focused on quantitative risk assessments for *Listeria monocytogenes* contamination in foods, and molecular epidemiology and dynamics of endemic infectious diseases on dairy farms. Dr. Pradhan is a member of numerous professional organizations, including the Society for Risk Analysis (SRA) and the International Association for Food Protection (IAFP). He has presented his research work at a number of professional meetings and conferences, and has published in refereed journals such as the Journal of Food Protection, Applied and Environmental Microbiology, the Journal of Dairy Science, and Poultry Science.

Donald Schaffner, Ph.D.

Dr. Donald Schaffner is an Extension Specialist in Food Science and a Professor at Rutgers University. He also serves as the Director of the Center for Advanced Food Technology. His research interests include quantitative microbial risk assessment and predictive food microbiology. Dr. Schaffner has authored more than 100 peer-reviewed publications, book chapters and abstracts. Dr. Schaffner is the recipient of multiple awards, including the International Association for Food Protection (IAFP) Elmer Marth Educator Award in 2009 and the Sustained Research and Impact Award in 2008 from the Rutgers School of Environmental and Biological Sciences and NJ Agricultural Experiment Station. Dr. Schaffner has served on a variety of national and international expert committees, including service to US National Academy of Sciences and the World Health Organization (WHO) and Food and Agriculture Organization (FAO) of the United Nations, the Institute of Food Technologist (IFT) and US National Advisory Committee on Microbial Criteria for Foods (NACMCF). Dr. Schaffner is active in several scientific or associations including the IAFP, IFT, Society for Risk Analysis (SRA), the American Society for Microbiology (ASM), and the Conference for Food Protection (CFP). Dr. Schaffner was elected a Fellow of the IFT in 2010 and is an Editor for the ASM journal Applied and Environmental Microbiology. Dr. Schaffner was elected the Secretary of the IAFP in 2010, a five-year commitment ending with his service of the President of the organization. He holds a Ph.D. in Food Science and Technology from the University of Georgia.

David Vose

Mr. David Vose is the Director of Vose Software, based in Belgium. He has twenty three years of experience in risk analysis modeling and decision support. He has written the textbook *Risk Analysis*, published by John Wiley and Sons, now in its third edition. He is also the author of the ModelAssist risk training software and the designer and key mathematician for the development of the ModelRisk software product. Mr. Vose maintains a large focus on animal imports and microbial and antimicrobial food safety issues, and has been a member of various committees charged with the development of international guidelines in these fields. Mr. Vose has provided training on microbial food safety risk analysis to government agencies in over 35 countries in a span of 12 years. He has performed food safety risk assessments for a wide variety of pathogens and food sources for the Danish Veterinary and Food Administration, the European Food Safety Authority, the World Health Organization (WHO), and the US Food and Drug Administration (FDA). He is an active member of the Society for Risk Analysis and ORMS. He holds an M.S. in Physical Oceanography from Southampton University.

Appendix #2: Charge to Peer Reviewers

The "charge to peer reviewers", as defined in the OMB's Peer Review Guidelines, are the issues and areas reviewers are expected to focus on in their evaluation of the risk assessment. The charge to the peer reviewers for this risk assessment evaluation included the following questions:

1. Evaluate if the overall approach for modeling the public health benefits potentially realized from the change in inspection system examined is fundamentally sound.
    a. Is the overall approach used in the analysis to evaluate the linkage between inspection activities and potential reductions in annual human illnesses fundamentally sound? The regression model used to estimate changes in establishment prevalence should be addressed separately from the model use to estimate reductions in annual human illness.
    b. If not fundamentally sound, in each case, what problems exist and how should they be addressed?

2. Evaluate the complexity of the model. In areas where the reviewer identifies limitations, weaknesses, or inadequacies, the reviewer must provide alternative data, data analysis, and/or modeling approaches.
    a. Is the model too complex, or not complex enough, to adequately address the risk management questions?
    b. Is the model over-or under-parameterized?
    c. Does the model adequately characterize the uncertainty present?
    d. Is variability sufficiently addressed?

3. Evaluate whether the model source code and mathematics are correct. If not, the reviewer must provide alternative modeling techniques.
    a. Are the modeling techniques (model mathematics and equations) appropriate?
    b. Are the methodologies used in the risk assessment for estimating parameters from the data appropriate (*i.e.*, follow scientifically accepted methodologies)?
    c. Are the data analyses and source code accurate?

4. Evaluate whether adequate sensitivity analysis has been provided. If not, the reviewer must provide an alternative approach or application for sensitivity analysis and/or identify those parameters that should have been included.
    a. Have the most important variables in the model been identified?
    b. Has an important variable been left out?
    c. Has the impact of including or excluding scientific studies or other data been adequately explored?

5. Evaluate the available data and the underlying assumptions used in this risk assessment. Are they complete and correctly analyzed and interpreted? If not, the reviewer must provide additional data sources and citations (where appropriate) or provide alternative interpretations, analysis, or suggested use of the data.
    a. Have all key studies and data been identified?
    b. Have the data been correctly interpreted, analyzed, and used in the risk assessment?

6. Evaluate the regression analysis used to estimate baseline and scenario aggregate establishment prevalence.
    a. Is the technique accurately described, utilized, and appropriate for its intended use?
    b. If not, reviewer must provide rationale for why not, and detail better alternatives.
    c. Are the conclusions drawn from the regression analysis appropriate?
    d. If not, reviewer must provide alternative interpretation of the results derived from this analysis.

7. Evaluate the scenario approach taken to quantify changes in establishment prevalence due to additional off-line inspection activities.
    a. Is this scenario approach reasonable, given the limited amount of data available?
    b. If not, what flaws do you perceive in the rationale and what information is lacking to make the case as proposed?
    c. What alternatives exist and how could they be incorporated?

8. Evaluate whether the documentation, discussion and interpretation of results is appropriate. If not, the reviewer must provide an alternative outline and/or approach for adequately and clearly documenting this risk assessment.
    a. Is the report clearly written?

b. Is it complete?
c. Does it follow a logical structure and layout?
d. Is it useful?
e. Does the risk assessment support the conclusions reached?