Response to Peer Review Comments on:

Assessment of the Potential Change in Human Risk of Salmonella Illnesses Associated with Modernizing Inspection of Market Hog Slaughter Establishments

August 2018

Table of Contents

Introduction

The U.S. Department of Agriculture (USDA), Food Safety and Inspection Service (FSIS) Office of Public Health Science addresses food safety hazards through prevention and control activities. Risk assessments are used to evaluate intervention strategies to reduce foodborne risks and to guide, support, and enhance the Agency's overall decision-making process, risk management policies, outreach efforts, data collection initiatives, and research priorities. The risk assessment project being considered focuses on the public health impact associated with the potential reallocation of USDA inspection personnel inside market swine slaughter plants. Specifically, what would be the risk to human health as a result of reallocating inspection resources as a result of this rule?

Currently, inspector assignments vary from plant to plant across several areas of responsibility, including online inspection assignments and offline inspection responsibilities such as Hazard Analysis and Critical Control Point (HACCP) verification procedures, sanitation verification procedures, and other consumer protection (OCP) procedures. Under the new proposed rule, certain online procedures will be consolidated, allowing more inspector time to be focused on offline procedures. Specifically, more resources could potentially be available to conduct verification procedures designed to protect human health (HACCP and sanitation verification procedures) or humane handling verification procedures.

FSIS initiated the pilot study, HACCP-Based Inspection Models Project (HIMP), in 1997. Establishment-directed sorting for food safety and non-food safety defects occurs in the "pilot plants," while nonparticipating establishments continue to operate on the standard inspection system. In the pilot plants, some on-line inspectors are made available to conduct a host of off-line verification activities, (e.g., they can check for more necessary HACCP and sanitation verification procedures, or they can conduct more humane handling verification tasks). In addition, data have been collected to allow further quantitative analysis of the potential human health impact, if any, due to the changes initiated by implementing HIMP.

This risk assessment addresses three risk management questions:

(1) What predicted effects will various models for increasing the number of off-line inspection tasks in non-HIMP establishments have on human salmonellosis numbers?
(2) Where within a hog slaughter establishment can relocated inspectors have the most impact toward reducing *Salmonella* prevalence and corresponding numbers of human illness?
(3) What is the magnitude of uncertainty about the predicted prevalence and number of illnesses?

This effort involved 154 hog plants spanning a broad range of production volumes. Five of these 154 plants participated in HIMP. Data regarding Inspection System Procedure (ISP) checks, inspector assignment profiles, and sampling results for *Salmonella* were obtained and used in the risk assessment. A subset of 35 plants currently not adopting a HIMP system but expected

to convert to the new HIMP inspection system in the near future were the focus of the final analysis. The overall risk modeling effort consisted of four submodels: (1) post-chill location in non-HIMP plants, (2) post-chill location in HIMP plants, (3) pre-evisceration location in non-HIMP plants, and (4) pre-evisceration location in HIMP plants. Each of the four submodels was applied using both (1) baseline market hog and (2) regulatory *Salmonella* sampling and inspection data as input. While the project focused on non-HIMP plants at the post-chill location, the four submodels were developed to leverage all data available for the time frame considered in the study. Additional scenarios were included in the modeling effort, namely the (1) impact of minimizing on-line inspectors and reallocating current inspectors to off-line inspections, (2) impact of reducing the number of inspectors in a plant, (3) impact of increasing the number of scheduled and unscheduled ISP procedures, and (4) impact of decreasing the number of scheduled procedures not performed.

The risk assessment was conducted in two main stages:

- Stage 1: Investigate the relationship between inspection procedures and *Salmonella* prevalence. A regression analysis approach was used to assess if different inspection protocols were significantly correlated with the *Salmonella* percent-positive samples by establishment.
- Stage 2: Calculate the prevalence ratio and correction of the baseline salmonellosis distribution for market hogs using this ratio. A prevalence ratio was calculated by dividing the estimated prevalence with HIMP-like system implementation by the baseline prevalence in non-HIMP establishments. Subsequently, the baseline attributable human distribution of salmonellosis was multiplied by the prevalence ratio (Stage 1) to estimate the illness distribution associated with implementing HIMP-like systems.

Peer Review Charge Questions

The selected peer reviewers were asked to address the following questions while conducting their review:

1. Please evaluate the available data and the underlying assumptions used in this risk assessment.
    a. Would the results of this simplified prevalence model be similarly reliable to a more complicated, traditional risk assessment model of the market hog slaughter inspection process?
    b. Have all key studies and data been identified, correctly analyzed and properly interpreted? If not, please provide additional data sources and citations (where appropriate) or alternative interpretations or analyses.
    c. Have the strengths and limitations of the data been transparently explained?
    d. Given the differences in the data, is the overall modeling approach for hog slaughter, including its differences from the poultry slaughter risk assessment, appropriate?
    e. Are there additional differences between hog slaughter or poultry slaughter, in the data or modeling approaches, that should be articulated? If so, would they affect the modeling approach?
    f. Is the model's sample size adequate to estimate prevalence change?
    g. Are the illness distribution calculations adequate to describe market hog-attributable illnesses?

2. Please identify limitations, weaknesses, or inadequacies of the logistic regression techniques and data (Stage 1); the reviewer must provide alternative data, data analysis, and/or modeling approaches.
    a. Is the technique accurately described, utilized, and appropriate for its intended use?
    b. Are the data analyses and source code accurate?
    c. Does the low Salmonella prevalence affect the validity of the model?
    d. Is the use of logarithmic weighting appropriate?
    e. Is the model over- or under-parameterized?
    f. Does the model clearly characterize the uncertainty present?
    g. Is variability in contamination prevalence and establishment characteristics sufficiently addressed?
    h. Has confounding been adequately controlled?
    i. Has multicollinearity been addressed?
    j. Are the decision variables modeled correctly?
    k. Is evaluation of non-HIMP establishment results at post-chill adequate as a sub-model?
    l. Are the conclusions drawn from the regression analysis appropriate?

3. Please identify limitations, weaknesses, or inadequacies of the Monte Carlo simulation techniques and data (Stage 2); the reviewer must provide alternative data, data analysis, and/or modeling approaches.
    a. Are the data analyses and source code accurate?
    b. Is the final model based on data from thirty-five establishments sufficient to estimate prevalence and changes in illness number from the scenarios that were included?
    c. Are the scenarios with altered decision variable profiles well-characterized?
    d. Is the applicability of the prevalence-illness relationship adequately demonstrated?
    4. Evaluate whether adequate sensitivity analysis has been provided.
    a. Have the most important variables in the model been identified? If not, which have been left out?
5. Evaluate whether adequate sensitivity analysis has been provided.
    a. Is the report clearly written and complete?
    b. Does the report follow a logical structure and layout?
    c. Are the conclusions supported by the risk assessment?

Selection of Peer Reviewers

RTI International (Research Triangle Park, NC, USA) was selected as the contracting agency which would carry out the peer review logistics and blinding. The process for selecting reviewers is detailed below, as described by RTI International:

RTI identified potential peer reviewers with overlapping and complementary expertise in key technical areas such as risk assessment, food safety, epidemiology, predictive microbiology, infectious disease, veterinary science, and food production as per discussion with FSIS. The initial list was composed of 19 experts.

We then contacted 10 reviewers to determine their availability and interest in participating. We asked them to submit their curriculum vitae as well as a form detailing their expertise and potential conflicts of interest (see form in Appendix A). This information ensured we recruited reviewers with the appropriate scientific stature and experience with related projects who were also independent from FSIS.

Based on the information collected in the second step, the RTI team finalized a list of five experts. RTI submitted blinded biographies of the peer reviewers to the Risk Assessment and Analytics staff. Reviewer identities were kept confidential for the duration of the review process. Biographies of the identified peer reviewers are listed in Section 2. Reviewers signed a nondisclosure agreement as part of establishing a consulting agreement.

RTI provided reviewers with the risk assessment report; model; cited references; link to the March 22, 2018 webinar; the charge to peer reviewers detailing the objectives to address

during the review; and a template for them to record their answers. RTI answered clarifying questions as needed. Upon receiving each review, the RTI's risk assessment expert reviewed each report and communicated as needed with the reviewers to address any gaps or ambiguities in the reviews. RTI notified FSIS once the reviewers began their review and submitted their unedited responses as they were completed.

Selected Peer Reviewers' Biographies

The following peer reviewers were selected to address the charge questions provided by FSIS:

David Vose is the Director and Founder of Vose Software. He has a background in risk analysis research and has extensive expertise in Monte Carlo simulation, risk assessment, quantitative modeling and has extensive knowledge of *Salmonella*. He has extensive experience in developing quantitative risk assessments for a variety of fields, including food safety and animal health, and has contributed to over 25 publications related to food safety risk assessments. He is knowledgeable about FSIS inspection data and demonstrates expertise in both @RISK and R software.

Donald Schaffner is a Distinguished Professor and Extension Specialist in the Food Science Department at Rutgers University. He has a background in food science and technology with research that focuses on microbial risk assessment, food microbiology, and predictive models. He has extensive expertise in food safety risk analysis, as well as a broad understanding of the food continuum and how models such as the swine slaughter model fit into it; he is familiar with the swine slaughter industry in the United States, as well as FSIS inspection data. Further, he has developed several quantitative models using Monte Carlo simulation and has been refereed in over 160 publications. He demonstrates expertise in both @RISK and SAS software.

Renata Ivanek is an Associate Professor of Epidemiology in the Department of Population Medicine and Diagnostic Sciences within the College of Veterinary Medicine at Cornell University. She has a background in epidemiology and veterinary medicine with research related to disease transmission, ecology, and epidemiology. In particular, she is familiar with the swine slaughter industry in the United States. She has a strong background in statistical methods and extensive experience using Monte Carlo simulation to develop quantitative risk assessment models. She has contributed to over 60 peer-reviewed publications in the field and demonstrates expertise in @RISK, R, and SAS software.

Ursula Gonzales Barron is a Food Engineer and Principal Investigator at the Polytechnic Institute of Braganza, Portugal. She has a background in food microbiology and risk assessment of foodborne pathogens with research focusing on various types of predictive modeling. She has extensive experience in applying Monte Carlo simulation to develop quantitative models. *Salmonella* in pork meat has been a focus of her research over the past years. She has contributed to over 60 peer-reviewed publications and multiple book chapters. She demonstrates expertise in @RISK, R, and SAS software.

Beatriz Martinez Lopez is an Associate Professor and the Director of the Center for Animal Disease Modeling and Surveillance within the Department of Medicine and Epidemiology of the College of

Veterinary Medicine at University of California, Davis. She has a background in epidemiology and veterinary medicine and has conducted disease transmission and surveillance research relating to swine and other livestock. She is familiar with the swine slaughter industry in the United States, as well as FSIS inspection data. She has extensive experience in Monte Carlo simulation and has developed several quantitative models to quantify the impact of several zoonotic diseases on both human and animal health. She has contributed to 70 peer-reviewed publications. She has a strong background and training in statistical methods and demonstrates expertise in @RISK, R, and SAS software.

Reviewer responses can be found in Appendix B. In some cases, reviewers commented directly in the risk assessment document. Those files were shared with FSIS before submission of this report.

Individual Reviewer Comments and FSIS' Responses

Reviewer A

A1. Please evaluate the available data and the underlying assumptions used in this risk assessment.

A1a. Would the results of this simplified prevalence model be similarly reliable to a more complicated, traditional risk assessment model of the market hog slaughter inspection process?

The linear model approach is now quite well established. It was first used by me/CVM for the fluoroquinolone-resistant *Campylobacter* risk assessment. The same basic idea is also used in Salmonellosis-attribution models developed by Tine Hald and me, and this model has been highly consistent as a predictor. Risk assessment models that perform entire farm to fork pathway models reproduce the same linearity, but with many additional assumptions. Given the huge uncertainty and variability modelling pathway probabilities from a carcass at the slaughter plant to the contracted illness of a consumer, the linear model is definitely the best approach. Most importantly, it allows anchoring of the illness rates to estimates from CDC. Doing whole pathway models, in which the illness rate is the final predicted variable, cannot anchor to 'observed' illness rates.

No response is necessary.

A1b. Have all key studies and data been identified, correctly analyzed and properly interpreted? If not, please provide additional data sources and citations (where appropriate) or alternative interpretations or analyses.

I cannot comment on whether all data have been identified. However, the data that have been used in this analysis have not been correctly interpreted – see comments at end

Additional analyses have been conducted in response to the comments at the end of this reviewer's comments. Please see response below and the new Appendix H, which addresses the reviewer's concerns. No additional response is necessary.

A1c. Have the strengths and limitations of the data been transparently explained?

Yes, except I think that they have missed the underlying pattern – see below.

No response is necessary. Please see Appendix H and response below.

A1d. Given the differences in the data, is the overall modeling approach for hog slaughter, including its differences from the poultry slaughter risk assessment, appropriate?

Yes. The modeling approach makes general sense, but I don't believe the regression assumption is correct – see below. I haven't been asked to review the poultry risk assessment.

No response is necessary. Please see Appendix H and response below.

A1e. Are there additional differences between hog slaughter and poultry slaughter, in the data or modeling approaches, that should be articulated? If so, would they affect the modeling approach?

I have not been asked to review the poultry risk assessment.

I have some experience with observing how poultry are processed, but not for swine, so these comments are predicated on some assumptions.

There are differences in the slaughtering process. The primary one, I believe, is to do with chilling. For poultry, carcasses are chilled in baths which creates a great deal of cross-contamination. The result is that whether a poultry carcass is contaminated or not strongly depends on whether there is salmonella in the batch of birds that is being processed. Pig carcasses are, I believe, chilled by hanging them individually in a cool room.

A secondary difference is the method of evisceration. For poultry, this is done with a de-cloacing device automatically. It can split the gut and spread fecal material from one carcass to another, especially if the poultry carcass varies from the ideal size for which the machine is designed. For pigs, I believe, the carcasses are gutted individually with less risk of contamination.

No response is necessary.

A1f. Is the model's sample size adequate to estimate prevalence change?

I don't believe so. In effect, there are just five observations – one each for the five different slaughter plants that adopted HIMP. One might have a lot of data within each plant, but there are only five plants from which one is attempting to extrapolate experience to the remaining many plants. If other non-measured factors play a part (like discipline, building work conditions) then the model does not have enough data points to have such unknown factors average out. These plants have been using HIMP for many years, so there are no data that could be used to compare the same plant's contamination post-chilling with and without HIMP – which would have been ideal.

The five plants are also large. It is not easy to see how one could support the idea that their experience could be translated to smaller plants.

Additional analyses have been conducted to explore the power of the analyses given the sample size; those analyses indicate that the sample size is adequate to detect differences. Please see Appendix H for a discussion of those analyses.

It was beyond the scope of this report to address factors such as working conditions and behavior change.

A1g. Are the illness distribution calculations adequate to describe market hog-attributable illnesses?

Yes. There is a small bias (a 5% overestimation) created in the Lognormal distribution used to match the P05, P95 from Scallan et al but this is of no great importance.

No response is necessary.

A2. Please identify limitations, weaknesses, or inadequacies of the logistic regression techniques and data (Stage 1); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

A2a. Is the technique accurately described, utilized, and appropriate for its intended use?

The technique is accurately described. I do not believe that it is appropriate, however – see below.

No response is necessary specifically for this comment. Other concerns are addressed below.

A2b. Are the data analyses and source code accurate?

Yes.

No response is necessary.

A2c. Does the low *Salmonella* prevalence affect the validity of the model?

Yes, very much. Not just the low prevalence but also the low number of samples overall. It opens up a debate about how one should treat the data. I describe at the end what I did to try to compensate for the low numbers.

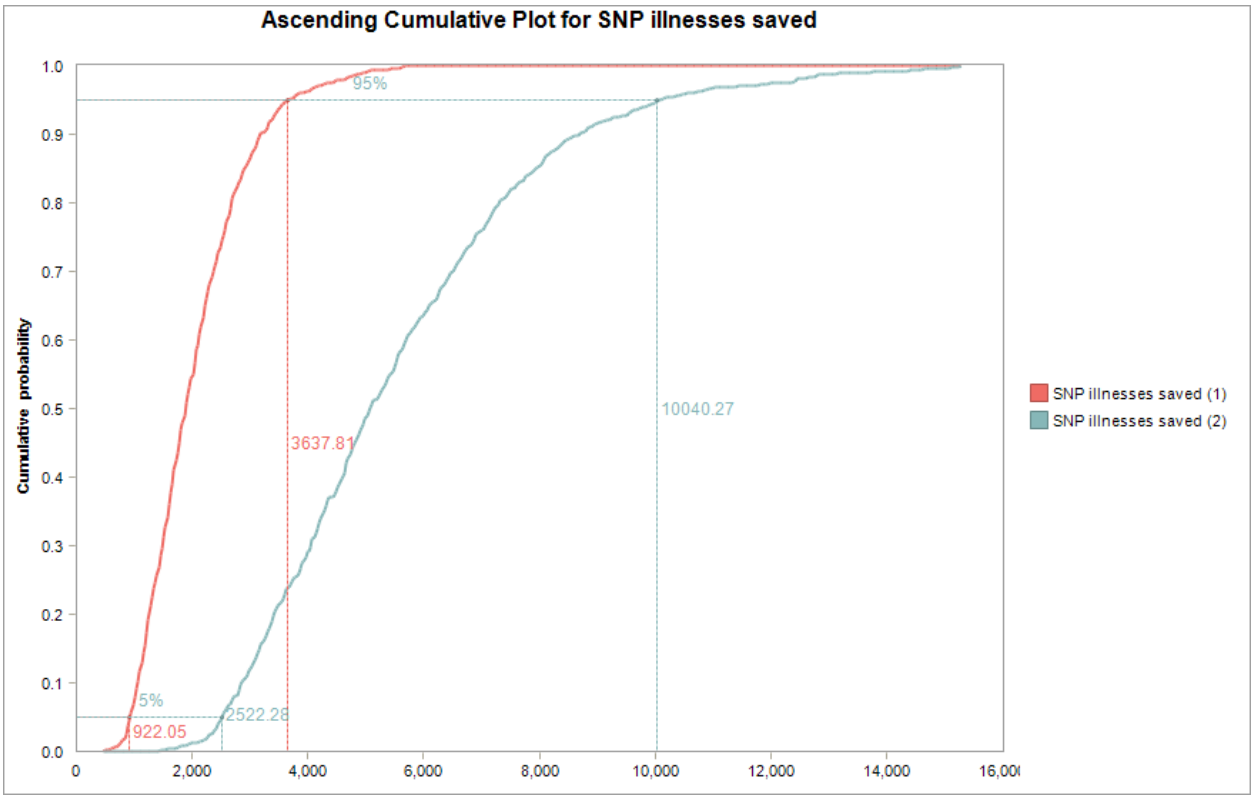
Please see response below.

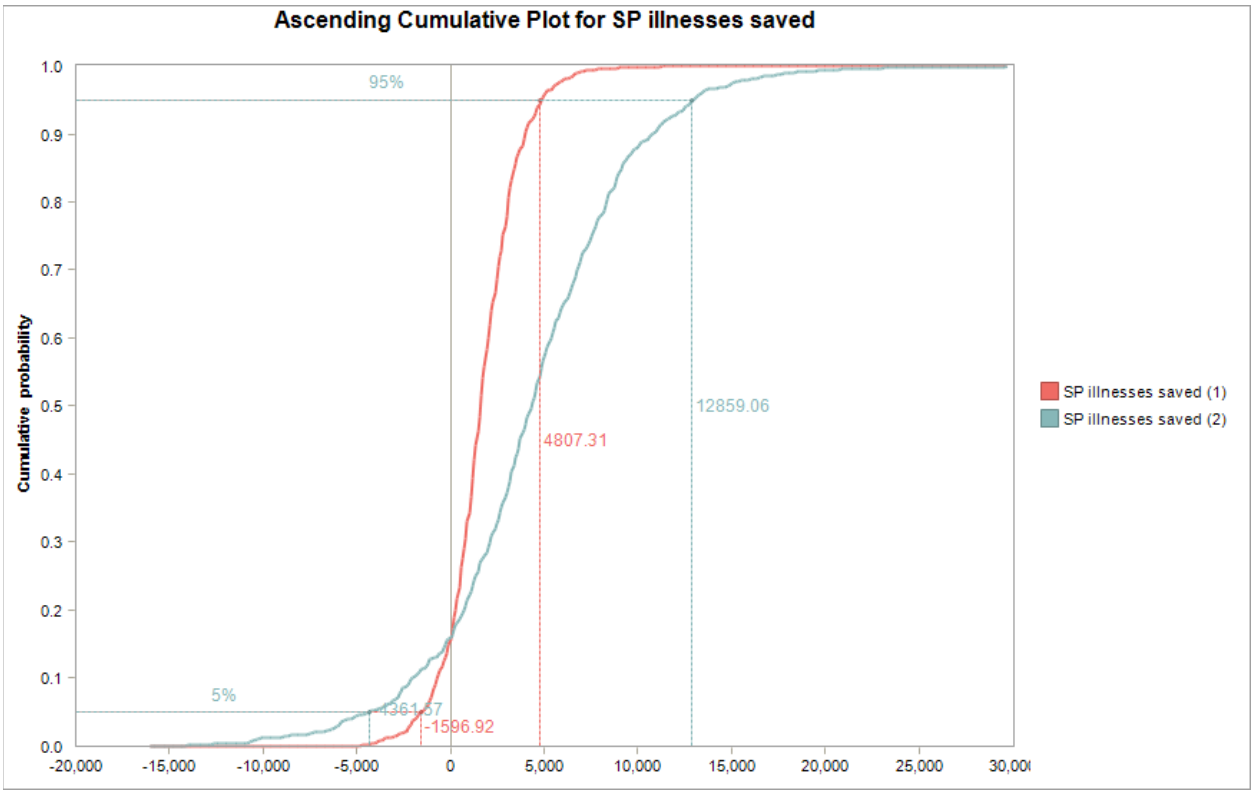
A2d. Is the use of logarithmic weighting appropriate?

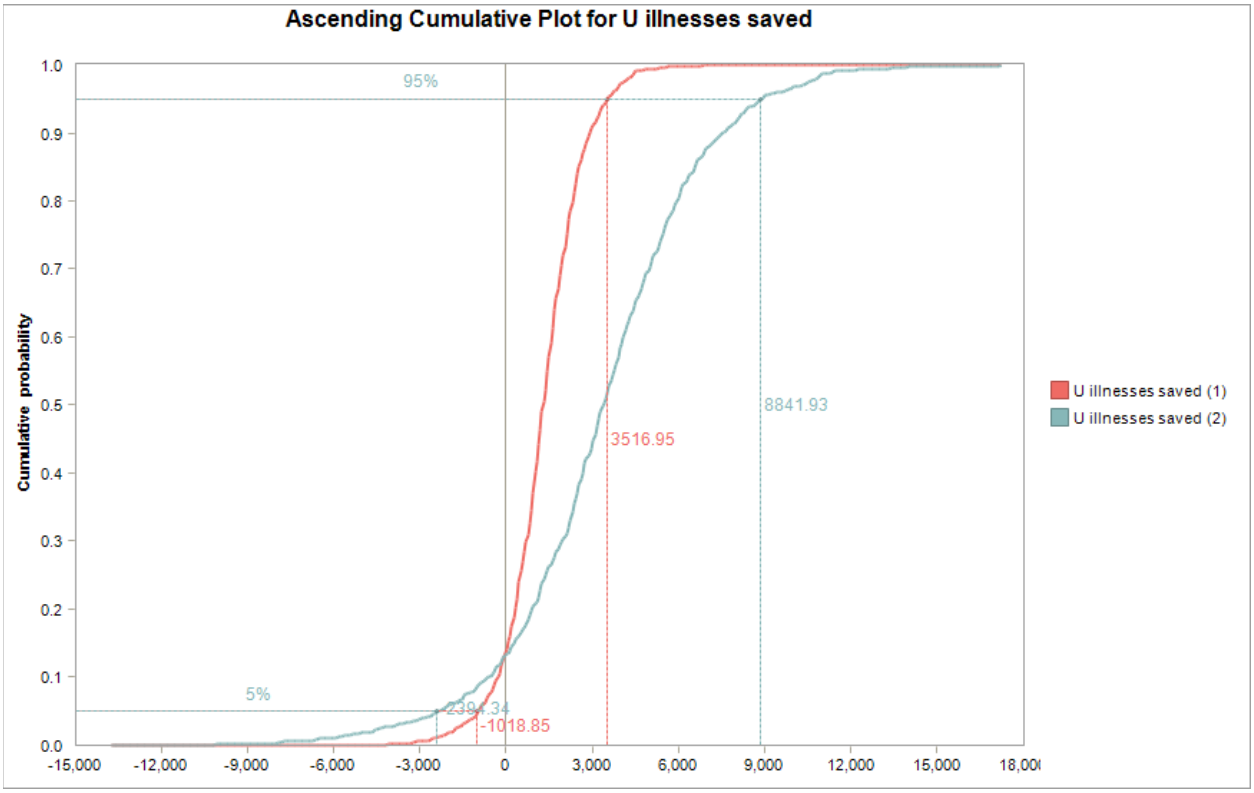
Not at all. The weighting is incorrect for three reasons:

(1) The weighting should be by actual volume (not log). The reason is that the model is trying to establish a nationwide prevalence, which should be done by multiplying the probability of contamination of a random post-chill carcass from a specific plant by the fraction of the pork production coming from that plant, which is standard epidemiological modeling.

Error 1 has a large impact. The estimates of illnesses avoided roughly triples. See graphs (1=original, 2 = corrected):







(2) The second weighting error is that the contribution of each plant is determined also by the number of samples that were taken from each plant. The weighting should actually be Volume/(k*TotalVolume), where k is the number of samples recorded for that plant. This produces a relatively small error in practice given the data set.

To illustrate – 127 samples were taken from plant #[redacted], and so 127 probabilities of contamination were estimated. For 110 or so plants, only one sample was taken and thus 150 or so probabilities of contamination – one for each plant – were calculated. By taking the weighted sum of these probabilities, one is assigning 127 times more weight to plant #[redacted] than these others. One would expect that more samples are taken for larger plants, so it is a matter of 'luck' that the error of under-weighting plants by the log of their production, rather than the production itself, countered the over-weighting by averaging over all probabilities.

If the model is deemed still to be of value given the other comments I have made, this error warrants correction since it is an incorrect methodology. If the data were different (e.g., few samples taken for some large producers, many samples for some small producers) a much larger error in the results could have been created. It is important that, in published risk assessments, correct methodologies are described as other practitioners will consider applying the same methodology to their own problems. Moreover, a third party seeking to challenge a risk assessment could point out these more obvious errors, claim that the whole risk assessment is therefore invalid, and question what other unknown mistakes there may be.

(3) The weighting should have the total domestically-consumed production volume as its denominator, not the sum of the plants selected. However (see discussion below) the plants selected provide about 95% of the total US production so the error is small.

Again, this error should be corrected for the same reasons as error 2, namely that one should publish correct methodologies since others may adopt them for a different problem in which the error could then be large, or because errors in simple parts of a model call into question the correctness of the more complex parts.

In response to this comment, additional analyses are included in Appendix H to explore the effect of not log transforming the volume data. Although the reviewer is correct that not log-transforming the data would approximately triple the estimates of the illnesses avoided, we respectfully disagree that using such untransformed data is appropriate, as it risk weighting (logVolume) gives weight to all establishment's contamination while non-transformed Volume weighting gives risk to the largest volume establishments. Furthermore, using the log transformed volume data results in a more conservative estimate of illnesses avoided. Please see Appendix H for a detailed discussion of weighting and model considerations that were used to arrive at the final model included in the risk assessment report.

A2e. Is the model over- or under-parameterized?

Given the regression routine, it is correctly parameterized. However, I believe that the fundamental assumption about the effect of HIMP may be wrong. See below.

Please see response below.

A2f. Does the model clearly characterize the uncertainty present?

Yes

No response is necessary.

A2g. Is variability in contamination prevalence and establishment characteristics sufficiently addressed?

No. The prevalence variation appears to be a function (see below) of the production volume of the plant, which is not in the regression equation. I find this surprising. For example, the number of workers/inspectors is included, but their effectiveness is dependent on how many hogs need checking each – which comes back to the production volume. There is a relationship between workers/inspectors and production volume, but it is highly non-linear – see graph at end.

Please see response below.

A2h. Has confounding been adequately controlled?

Not relevant, I think, because the regression assumption does not seem valid to me – see below

No response is necessary. Please see below.

A2i. Has multicollinearity been addressed?

Not relevant, I think, because the regression assumption does not seem valid to me – see below

No response is necessary. Please see below.

A2j. Are the decision variables modeled correctly?

Yes

No response is necessary.

A2k. Is evaluation of non-HIMP establishment results at post-chill adequate as a sub-model?

Yes, but again, not relevant, I think, because the regression assumption does not seem valid to me – see below

No response is necessary. Please see below.

A2l. Are the conclusions drawn from the regression analysis appropriate?

I do not think so – see below

No response is necessary. Please see below.

A3. Please identify limitations, weaknesses, or inadequacies of the Monte Carlo simulation techniques and data (Stage 2); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

A3a. Are the data analyses and source code accurate?

Accurate, but based on what I think is a flawed assumption

No response is necessary. Please see response below.

A3b. Is the final model based on data from thirty-five establishments sufficient to estimate prevalence and changes in illness number from the scenarios that were included?

Not relevant, I think, because the regression assumption does not seem valid to me – see below

No response is necessary. Please see response below.

A3c. Are the scenarios with altered decision variable profiles well-characterized?

They are correct under the given assumptions

No response is necessary.

A3d. Is the applicability of the prevalence-illness relationship adequately demonstrated?

It is not demonstrated. However, it refers to references that have already made this argument well.

No response is necessary.

A4. Evaluate whether adequate sensitivity analysis has been provided.

A4a. Have the most important variables in the model been identified? If not, which have been left out?

No. The production volume is not used as a causal factor for contamination levels – see below.

No response is necessary. Please see response below.

A5. Evaluate other characteristics of the report.

A5a. Is the report clearly written and complete?

Pretty clearly. I think too much emphasis was placed on the technical details of the regression, and too little on exploring the original data. In the end, stepwise logistical regression is a standard statistical procedure and we should be able to know it was well-carried out with just a couple of pages.

No response is necessary.

A5b. Does the report follow a logical structure and layout?

Yes

No response is necessary.

A5c. Are the conclusions supported by the risk assessment?

The conclusions are supported by the risk assessment in that - given the risk assessment results - one could arrive at the stated conclusions. However, the risk assessment is not correct in my view – see below.

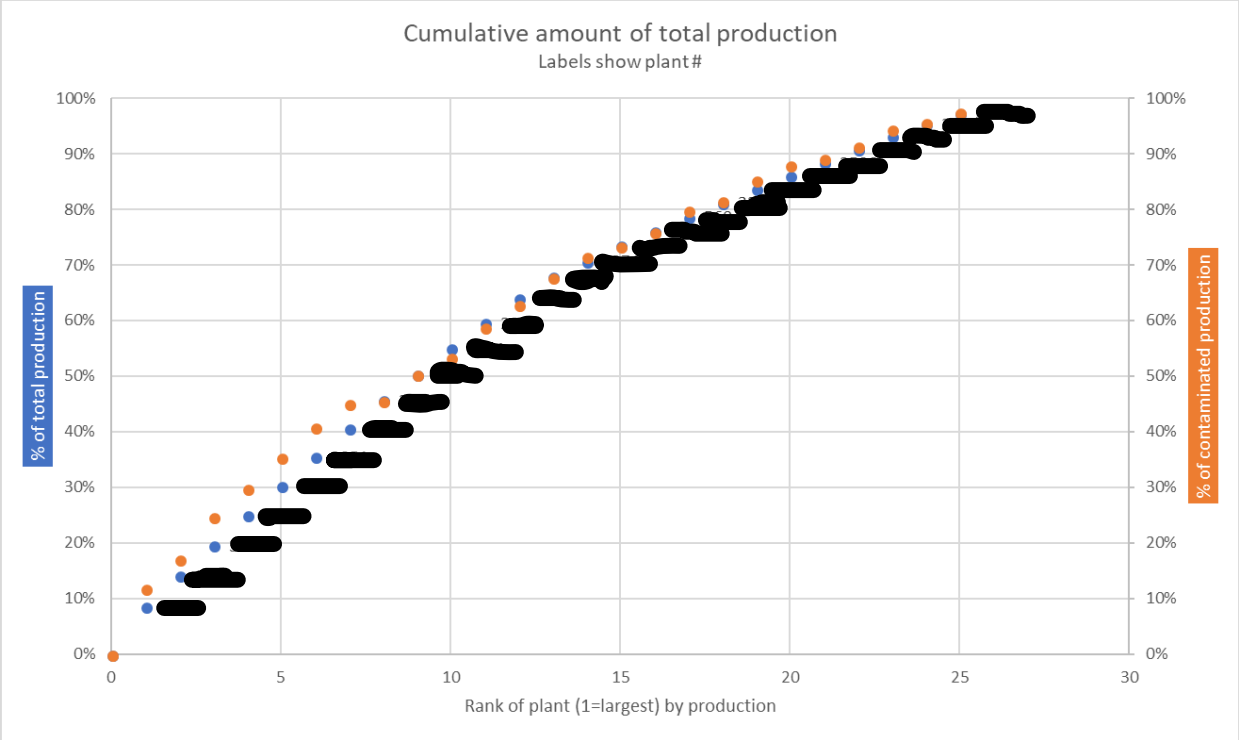
My doubts about the basic assumption that HIMP changes post-chill contamination prevalence

I took all the data from hogs1c.csv and analyzed it using pivot tables in Excel. The Excel file is sent along with this report to see how I performed the analysis. The plants that use HIMP are shown with orange dots in all the following plots, and the non-HIMP plants are shown with blue dots.

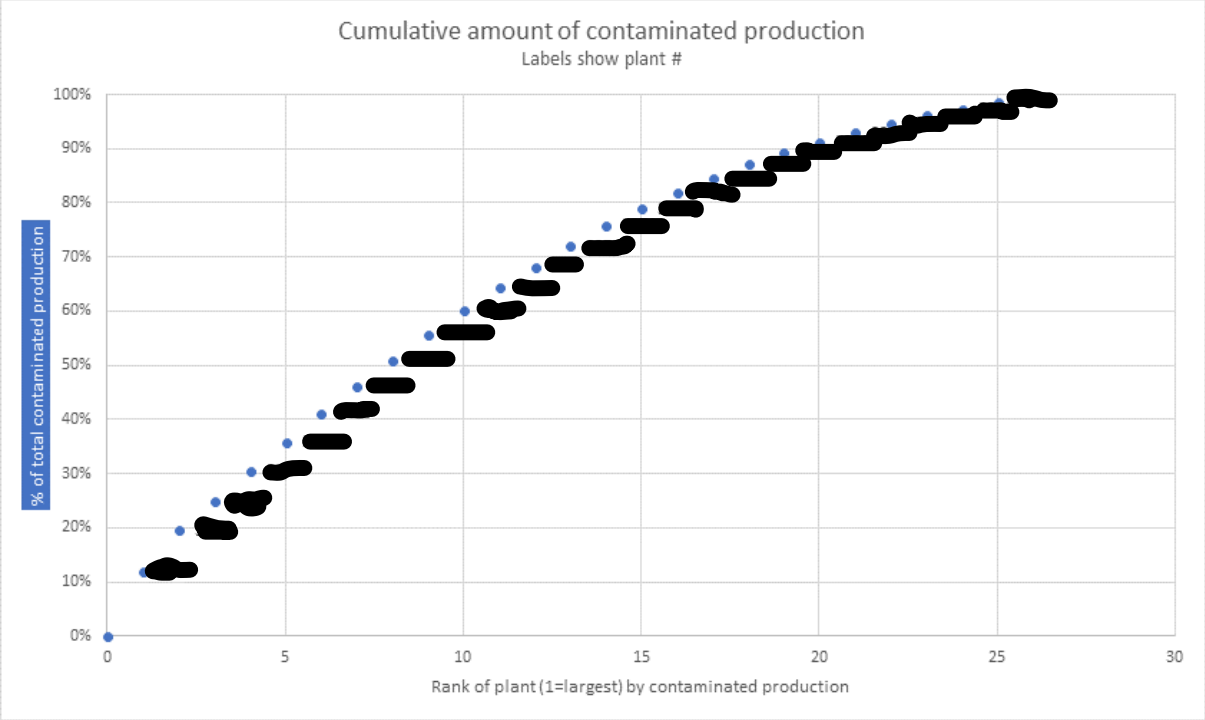
Distribution of plant production volume

A plant with a 10% prevalence and a production of 100,000 hogs should contribute much more to the illness rate than another plant with 50% prevalence and 1,000 hog production. So, the first step must be to analyze the volume data.

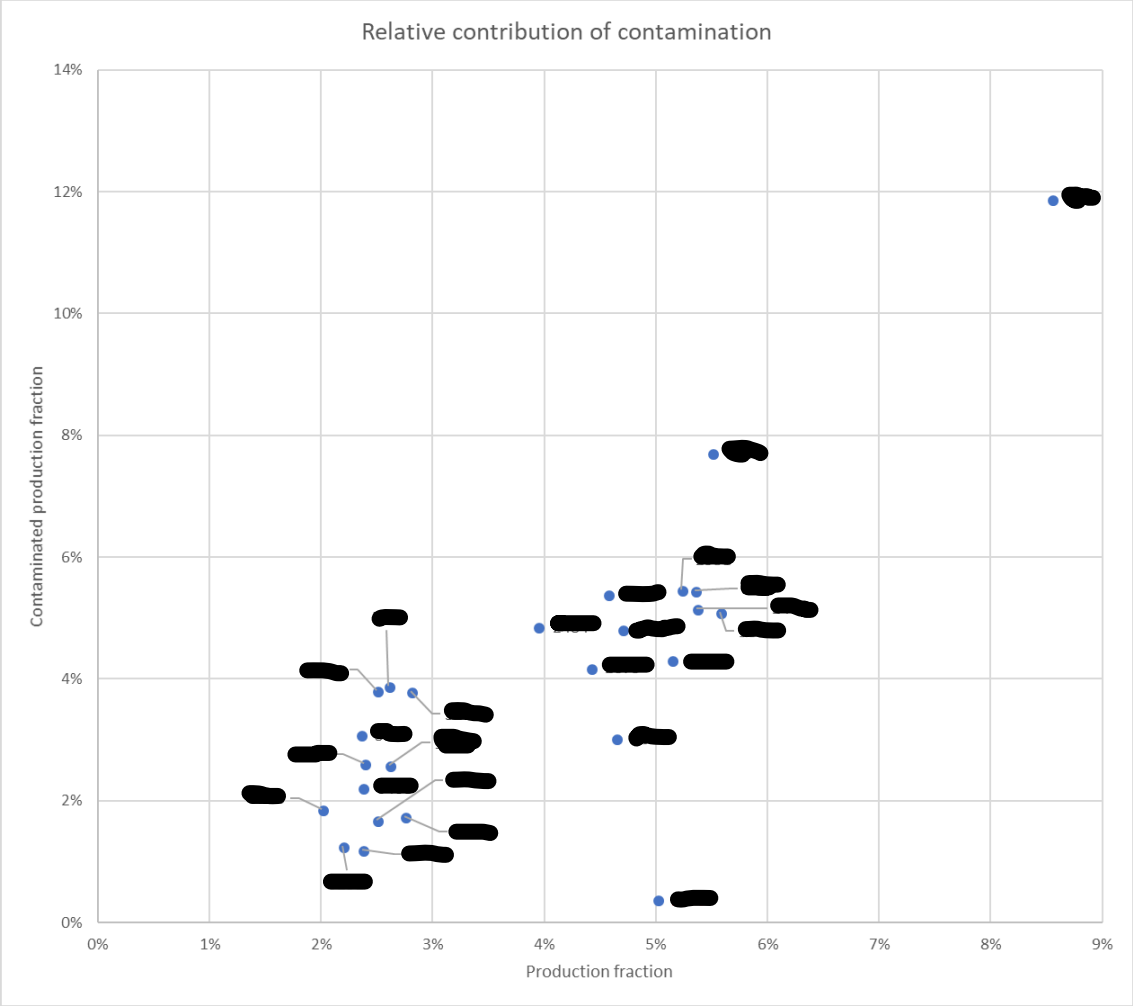
There are 164 plants in the hogs1c dataset with a total annual production of [redacted] hogs. Plotting the cumulative fraction of the total production starting with the largest producer and working down the list one gets the following plot:



So, for example, plant [redacted] produces [redacted] hogs annually, or [redacted]% of the total production (blue dots) but has a higher than average post-chill prevalence so contributes a bit more of the total contaminated hogs (orange dots). The next largest plant ([redacted]) produce [redacted] hogs annually, giving a running total of [redacted] hogs or [redacted]% of total production, and so the chart builds for the largest 25 plants by production. We see that, of the 164 plants, 25 contribute 97% of the production and 98% of the contaminated production.

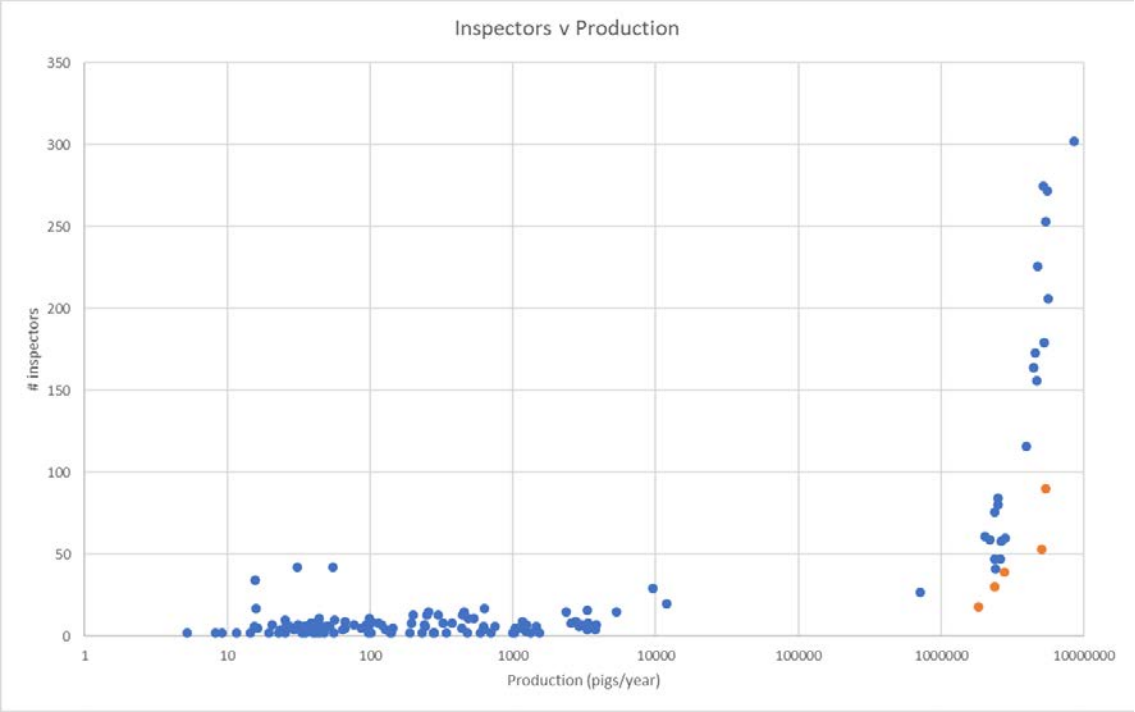


The same chart, ordered by the number of contaminated carcasses produced, gives essentially the same list of plants. These 25 plants contribute 98% of the total contaminated hogs. Plotting the fractional contribution to total and contaminated, we roughly see the expected y=x linear relationship:



All of the top 25 plants described above are included in the base model with the exception of plant [redacted] labelled in red above, presumably because its very low prevalence already means that HIMP would not likely add any benefit and potentially mess about with a system that seems to be working perfectly.

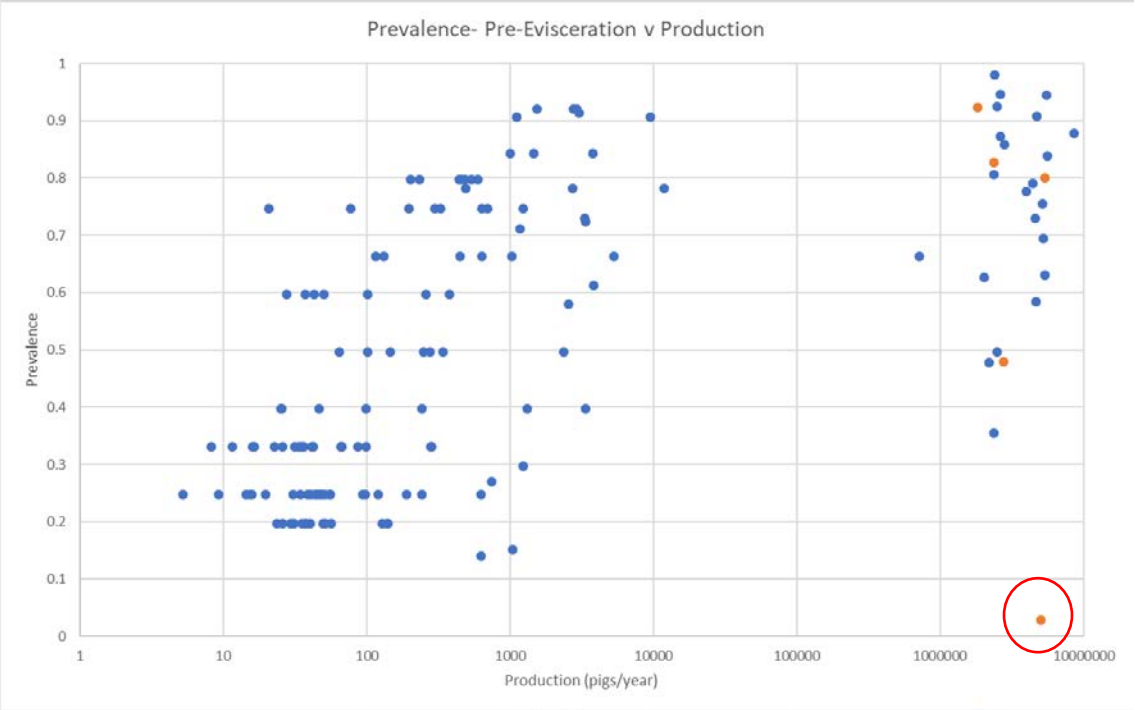
Number of inspectors plotted against the production volume

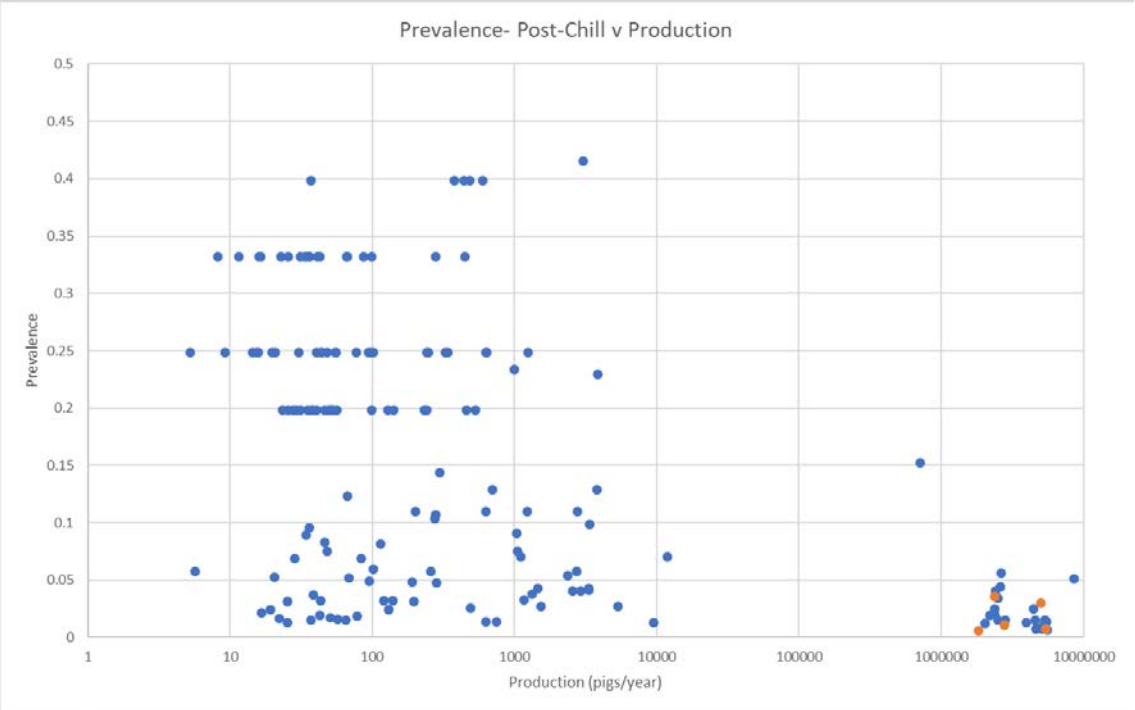


This says that to me the number of inspectors for the HIMP plants is much lower than for others of the same size (I don't know why). It also shows that the number of inspectors is a non-linear function of production volume for the non-HIMP plants.

Prevalence pattern as a function of production volumes

In the following charts, the HIMP prevalence pre-and post-chill are plotted as a function of the production volume by plant:





The prevalence of HIMP and non-HIMP plants match well when one accounts for the production volume, i.e. the location of the HIMP data points lies within the scatter of the non-HIMP data points. The exception is one plant which – for pre-evisceration – shows a very low prevalence (ringed) but this prevalence is unrelated to HIMP inspection activities. Why this plant has such low pre-evisceration prevalence would perhaps yield a more interesting insight into how to control salmonella contamination.

I therefore conclude that the regression model assumption that there is a relationship between HIMP inspection activities and post-chill prevalence of Salmonella contamination has not been established. In my view, this makes the risk assessment model invalid.

Additional point on prevalence estimates

If a plant has n samples, and s are positive for salmonella, the natural estimate of prevalence is s/n. However, this estimate is biased away from 0.5. For example, if s=0 the estimate would give a prevalence of 0 irrespective of the size of n. That is relevant particularly when one has a small number of samples. Three of the HIMP plants give s=0. A way to compensate for this is to use the quite standard description of the uncertainty of prevalence using the distribution Beta(s+1, n-s+1), which has a mean of (s+1)/(n+2). I have translated the observed prevalence using this formula in the charts above. Many of the non-HIMP plants have very few samples – hence the visible horizontal lines in the scatter plots above at 0.2 (n=3, s=0), 0.25 (n=2, s=0), 0.333 (n=1, s=0), 0.4 (n=3, s=1).

In response to Reviewer A's comments and concerns, additional analyses were conducted to ensure that the approach in the risk assessment is not affected by the Reviewer A's concerns. Please see Appendix H for a detailed discussion of weighting and model considerations that were used to arrive at the final model included in the risk assessment report.

Small errors

Incorrect parameters for fhog and fmarkethog given at p42 and in Table 6.

We are unsure of what the reviewer is referring to. Fhog and fmarkethog are not actually parameters, but are Pert uncertainty distributions which were calculated using a mode as the measure of most likely value for the center of the Pert distribution. The reviewer may have expected that such calculations would be based on a mean, however we believe that using mode is the proper statistical convention for such distributions. No change has been made to the document.

Reference error on pp 117, 125

***These modifications have been incorporated in the updated version of the risk assessment report*.**

Reviewer B

B1. Please evaluate the available data and the underlying assumptions used in this risk assessment.

B1a. Would the results of this simplified prevalence model be similarly reliable to a more complicated, traditional risk assessment model of the market hog slaughter inspection process?

The results of this simplified prevalence-based risk assessment rely on the assumption that a change in *Salmonella* prevalence in a slaughterhouse does not affect the concentration of *Salmonella* at the moment of human exposure. In other words, the model accounts for the observed reduction of *Salmonella* prevalence of contaminated carcasses (based on statistical analysis of available data) due to change in the inspection system but then assumes that the concentration of *Salmonella* on these contaminated carcasses is not affected by the change in the inspection system; it assumes that inspection activities are equally likely to eliminate/prevent carcasses with high and low *Salmonella* concentrations.

In the risk assessment, this assumption was justified by internal analysis of *Salmonella* concentrations from high and low prevalence settings (pre-evisceration vs post-chill), which showed no differences between the contamination levels. However, to support this assumption, more information needs to be provided in the risk assessment about the sampling approach and *Salmonella* quantification method used (including how and where were samples collected, which enumeration method was used and the detection limit).

On a more fundamental level it is important to consider that this assumption would not hold under two possible scenarios. (1) The inspection activities may be more likely to eliminate carcasses with a higher contamination level. If one considers that the new inspection system (HIMP) is more likely to eliminate/prevent carcasses visibly contaminated with fecal matter, which may include high levels of *Salmonella*, then it would be feasible to assume that the new inspection system would be more likely to prevent carcass contamination at a high level. Neglecting that in the risk model (stage 2) would result in underestimation of the public health benefits of the new inspection system. (2) On the other hand, if the inspection activities are more likely to eliminate carcasses with low levels of contamination per carcass while carcasses with high *Salmonella* concentrations remain uneliminated, that would reduce the benefits of the inspection system and in the model would result in underestimation of the human health risk.

While to my understanding there is no data to conclusively support/reject either of the two scenarios, the scenario (1) seems more plausible, meaning that it is more plausible that the risk assessment has underestimated than overestimated the public health benefits of the new inspection system. The risk assessment should discuss the limitations of data and this assumption further because that will provide for a more effective risk communication. Additional analyses of the risk could be conducted to relax this assumption (i.e., evaluate the risk under each of the two scenarios). However, unless new data are identified/collected, this would likely have a limited value because the scenarios would have to be based on hypothetical parameter values. Also, if one accepts that scenario (1) is indeed more plausible the final conclusions of the risk assessment would remain the same. This too should be discussed to allow for a more effective risk communication.

Another assumption underlying this risk assessment is that prevalence of *Salmonella* contamination at post-chill is causally dependent on the inspection procedures in the facility and therefore that any change in the inspection decision variables is predictive of *Salmonella* prevalence. This assumption is based on statistical analysis of observational data, which measured association that is not evidence of a causal effect. While the causal effect may be true, it is also possible that inspection variables are just a proxy for some other unmeasured variables that affect *Salmonella* prevalence (e.g., variables that define inspector behavior and designing of scheduled activities). It is also possible that the observed association is a false positive result related to selection bias or random chance. For example, in FSIS, 2011 Evaluation of HIMP, the association between decision variables and *Salmonella* prevalence was identified only in some years. However, despite these challenges, there currently is no a mechanistic and better way to model this relationship and therefore the approach used seems acceptable.

We appreciate the reviewer's thoughtful comments. We agree that the causal link between inspection procedures and contamination is difficult to prove given the data and the limitations thereof, but we feel that the relationship between contamination and illnesses is well supported. In addition, like in the case of the Poultry Slaughter Modernization risk assessment, this relationship provides the best framework for an analysis describing the impact of an inspection program change. Additional information is provided in the updated risk assessment report describing the data collection, sample method, and results of the Market Hog Baseline study that produced the data for our internal analysis of contamination levels.

B1b. Have all key studies and data been identified, correctly analyzed and properly interpreted? If not, please provide additional data sources and citations (where appropriate) or alternative interpretations or analyses.

All key studies and data seem to have been identified and properly analyzed/interpreted. The only recommendation is to provide more information about the characteristics of enrolled slaughterhouses (to be able to better judge generalizability of findings in stage 1) and the *Salmonella* enumeration approach.

Providing additional details about the enrolled slaughter facilities is beyond the scope of this risk assessment. The 2014 report,* Evaluation of HIMP for Market Hogs*, may provide more information for readers and is cited in the risk assessment report[1].

B1c. Have the strengths and limitations of the data been transparently explained?

The characteristics of the enrolled establishments (particularly the 5 HIMP slaughterhouses) need to be provided to be able to assess representativeness and generalizability of data obtained from those plants. Other than that, limitations of used data have been transparently explained.

[1] Accessible as of July 12, 2018 at https://www.fsis.usda.gov/wps/wcm/connect/f7be3e74-552f-4239-ac4c-59a024fd0ec2/Evaluation-HIMP-Market-Hogs.pdf?MOD=AJPERES

See response to Comment B1b.

B1d. Given the differences in the data, is the overall modeling approach for hog slaughter, including its differences from the poultry slaughter risk assessment, appropriate?

The overall modeling approach seems appropriate. The poultry risk assessment was based on 20 HIMP establishments and thus provided stronger statistical support for modeling.

No response is necessary.

B1e. Are there additional differences between hog slaughter and poultry slaughter, in the data or modeling approaches, that should be articulated? If so, would they affect the modeling approach?

This reviewer is unaware of additional differences between hog slaughter or poultry slaughter that would be relevant for consideration in this risk assessment.

No response is necessary.

B1f. Is the model's sample size adequate to estimate prevalence change?

With > 5,000 observations in the non-HIMP post-chill model the sample size seems to be sufficient. My only concern was the presence of clustering due to collection of multiple datapoints within the same establishments over time, and related to that the potential for underestimated standard errors. To evaluate this, I considered a random effect model with Locid as a random effect, however, the model did not converge when all variables from the reported final model were considered. The HIMP and inspection procedures were significant in a more parsimonious random effect model but could not all be fitted simultaneously into the model. Considering the challenges with random effect modeling (and the potential for incomplete understanding of data structure), this reviewer considers as appropriate the approach taken (with application of weights to correct standard errors for clustering).

We appreciate the reviewer's thoughts on an alternative model, and agree with the reviewer's conclusion regarding the random effect model not being as appropriate. In response to this and other comments, however, evaluations of other potential models, is now included in the risk assessment. Please see the discussion of alternative modeling approaches, including clustering, in Appendix H section 2.

B1g. Are the illness distribution calculations adequate to describe market hog-attributable illnesses?

A linear relationship was assumed between post-chill prevalence of *Salmonella* and the associated human illness. If one accepts that prevalence and load of *Salmonella* are independent (see response to charge 1a), this relationship is adequate.

No response is necessary.

B2. Please identify limitations, weaknesses, or inadequacies of the logistic regression techniques and data (Stage 1); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

B2a. Is the technique accurately described, utilized, and appropriate for its intended use?

The applied approach seems adequate and is well described. The only recommendation is to provide more information about the logic and effect of dummy variables to represent different interactions in the model.

A new appendix, Appendix H, has been added and provides more information, generally, about variables and interactions.

B2b. Are the data analyses and source code accurate?

R source code was limited to a few lines of R code that read the data into R and produced the final model. This code was reviewed and found adequate. The description of statistical analyses was evaluated from the risk assessment text and seemed feasible. A strong recommendation would be to include the R codes for all performed analyses, not just the final model, into the Appendix.

The risk assessment was primarily constructed using SAS 9.4 and @Risk 7.5; the R code for the final model was provided to reviewers as a courtesy.

B2c. Does the low *Salmonella* prevalence affect the validity of the model?

The low *Salmonella* prevalence did not affect the model validity.

No response is necessary.

B2d. Is the use of logarithmic weighting appropriate?

I agree that some way of weighting was necessary. Logarithmic weighting produced weights that penalized large production volumes while relative to that may have overestimated the importance of low production volumes. I considered a version of the model with absolute values of volume of production as weights but the model didn't converge. Conversions other than logarithmic could have been considered but would distort the production volume data as well. Therefore, the approach taken seems adequate and no additional analysis regarding logarithmic weighting is deemed necessary.

No response is necessary.

B2e. Is the model over- or under-parameterized?

Considering that all variables in the final model are significant at the 5% level the model seems well parameterized.

No response is necessary.

B2f. Does the model clearly characterize the uncertainty present?

To my understanding, the uncertainty in the context of stage 1 model arises from the potential sampling bias, i.e., the fact that the enrolled establishments may not have been representative of the general hog

slaughterhouses. Related to that, additional information should be provided about enrolled establishments. Additional information could include age of the facility, sourcing of pigs for slaughter (e.g., preferential sourcing based on *Salmonell*a infection status or implemented control strategies at the farm level), slaughter line characteristics, any *Salmonella* control strategies in place at the facility (e.g., hot water decontamination of carcasses or decontamination with organic acids, and application of logistic slaughtering (i.e., separation of pigs at slaughtering based on their risk of *Salmonella* infection). The characteristics of the enrolled facilities should be compared with the rest of the industry. From this information the reader should be able to infer whether the enrolled establishments were representative of the variation in the industry and any non-enrolled facility should be able to judge whether the results apply to their facility or not. It is impossible to control for the presence of selection bias at the analysis stage (stage 1); this bias is controlled at the study design stage (i.e., enrollment of facilities). However, the stage 1 results should be discussed in the context of any evidence for or against the presence of selection bias. If the characteristics of the enrolled facilities are representative of the industry that would reduce the uncertainty around the results. If there is evidence of selection bias, that limitation and the likely direction of bias (i.e., under- or over-estimation of the measures of association) should be discussed.

Although having such detailed information about each establishment might provide more information on potential biases, such information was not available. Because this risk assessment was not carried out as an experiment but rather an observational study and so the representativeness of the sample could not be controlled. Additional details about the data can be found in Appendix H, section 4.

B2g. Is variability in contamination prevalence and establishment characteristics sufficiently addressed?

Variability in contamination prevalence and measured establishment characteristics have been statistically captured, with confidence intervals reflecting the sample size and variability.

No response is necessary.

B2h. Has confounding been adequately controlled?

Multivariable modeling measured the effect of inspection activities on *Salmonella* prevalence while controlling for the effect of other significant factors in the final model. Additionally, the effect modifying effects have been accounted for through dummy variables that modeled interactions at the multiplicative level. However, strictly speaking, statistical analysis in stage 1 did not evaluate whether controlling of factors other than inspection activities has produced meaningful confounding or interaction effects. Considering that the purpose of statistical modeling is to use the regression data in stage 2, evaluation of the magnitude of confounding is not necessary but could be done for completeness.

To control for potential confounding repeated measures modeling was considered but did not provide a substantively better fit for the data than the final model chosen. See Appendix H where alternative models are discussed for additional details.

B2i. Has multicollinearity been addressed?

The descriptions of statistical treatment of multicollinearity seem adequate.

No response is necessary.

B2j. Are the decision variables modeled correctly?

The decision variables have been modeled adequately. I agree with the approach to place emphasis on the scenario PS+PNS+U (i.e., without NC) as that seems to be the most meaningful multivariable treatment.

The 3-ways interaction between U*HIMP*COLL is interesting and should be discussed to help understand the model results.

The three-way interaction was mistakenly included from an earlier draft and is not part of the updated report. We have corrected that typographical error.

B2k. Is evaluation of non-HIMP establishment results at post-chill adequate as a sub-model?

The evaluation of non-HIMP establishment at post-chill seems to be an adequate sub-model.

No response is necessary.

B2l. Are the conclusions drawn from the regression analysis appropriate?

The conclusions drawn from the regression model are adequate from the statistics point of view, but there are concerns related to the study design (selection of participating facilities and whether they are representative).

This comment has been addressed in previous responses. Please see response to B2f.

B3. Please identify limitations, weaknesses, or inadequacies of the Monte Carlo simulation techniques and data (Stage 2); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

B3a. Are the data analyses and source code accurate?

The described data and the @Risk model setup seem adequate.

The only suggestion is to potentially consider additional factors that may mechanistically explain inspectors' behavior in the establishments and inspection procedures scheduled by the establishments (though it is not clear if such data exist).

Although it could be interesting to consider the factors mentioned by the review, such data do not exist.

B3b. Is the final model based on data from thirty-five establishments sufficient to estimate prevalence and changes in illness number from the scenarios that were included?

The use of the results from the final model with thirty-five establishments supported with the modeled uncertainty about the change in the inspection procedures seem adequate.

No response is necessary.

B3c. Are the scenarios with altered decision variable profiles well-characterized?

It seems that the inspection procedures were varied simultaneously and independently of each other, meaning that in a particular iteration the total number of performed inspection activities could have been unrealistically low or high. This needs to be clarified in the model description and if indeed the inspection procedures were allowed to vary simultaneously and independently, this should be corrected either (i) by sampling the number of inspection procedures from a total number fixed for the establishment and iteration or (ii) by eliminating iterations with unrealistic total numbers of inspection procedures.

We agree with the reviewer's comment and, in fact, scenarios with unrealistic uncertainty bounds were not included in the final model, and the data were screened for outliers that appeared more likely to reflect errors in notation than true variability in the measurement. SAS 9.4 software was used to apply regression diagnostics and validity checks as part of this screening.

B3d. Is the applicability of the prevalence-illness relationship adequately demonstrated?

The applicability of the relationship between carcass *Salmonella* prevalence and human illnesses has been based on empirical evidence in FSIS chicken carcasses, which indicated that the average concentration of *Salmonella* per milliliter of rinsate did not change over time but the prevalence of positive carcasses changed. As discussed in response to charge 1a, this relationship is based on the assumption that the dose levels at consumption are independent of the frequency of contamination. There is some support for this assumption also based on the described internal comparison of *Salmonella* concentrations in areas with different prevalence (described under charge 1a).

No response is necessary.

B4. Evaluate whether adequate sensitivity analysis has been provided.

B4a. Have the most important variables in the model been identified? If not, which have been left out?

The sensitivity analysis was based on parameters identified from stage 1 and additional parameters in stage 2 and it seems adequate.

No response is necessary.

B5. Evaluate other characteristics of the report.

B5a. Is the report clearly written and complete?

The report seems complete and for the most part it is clearly written. Below are listed specific comments to improve clarity:

- Page 17, paragraph 1: clarify what is "adverse effect" in this context.
  The text has been clarified.
- Page 30, the FSIS reference seems to be FSIS, 2016 (not 2014).
  The final report, "Evaluation of HIMP for Market Hogs" was published in November 2014.
- Page 39, 3rd bullet point: do you mean 3 vs. 1 average bacteria per sample or 3 log10 vs 1 log10?
  This value is based on an analysis of the FSIS Market Hog Baseline data and the report has been updated with the correct units and citation.
- Figures 8, 9 and 10 seem to be included too early compared their mentioning in the text.
  This has been corrected in the updated version of the risk assessment report.
- Page 88, 2nd sentence: "The deletion of" is redundant, delete.
  This has been corrected in the updated version of the risk assessment report.
- Page 88, mid paragraph, "This plot indicates that model tends to have more outliers …" add 1 after model.
  This has been corrected in the updated version of the risk assessment report.
- Page 94: 3rd paragraph: For clarity indicate that "Data Sources and Structure" refers to Table 5.
  This has been corrected in the updated version of the risk assessment report.
- Page 103, end of paragraphs 2 and 3: I assume that Figure A should eb corrected into Figure A4.
- Page 104, text at the bottom, sentence 1: correct Figure A into Figure A5?
- Page 115, 2nd paragraph, 1st sentence: you may refer the reader to Figure A7.
- Page 117: figure number shows as error, but I assume that Figure A7 is meant.
- Page 119, paragraph 2, 1st sentence: correct figures to A8 and A10? Also, mid-paragraph, figures seem mislabeled (A8 should be A11 and A9 should be A12? Last sentence, it is unclear what figures the text refers to.
  All of these figure labels have been corrected in the updated version of the risk assessment report.
- Page 121: mismatch between figure caption (U) and text title in the figure (SNP).
- Page 122: mismatch between figure caption (SNP) and text title in the figure (U).
  Both of these caption errors have been corrected in the updated version of the risk assessment report.
- Page 125: figure A11 should be Figure A14? Paragraph starts with error in place of figure.
  This has been corrected in the updated version of the risk assessment report.

- Page 129: table is unclear, consider switching columns “After” and “Intervention” and changing “Intervention” into “Attributable to intervention”

  This has been corrected in the updated version of the risk assessment report.
- “Additional details…” document: difficult to follow because several tables seem to be mislabeled; Figures not labeled; Figure on page 3 x-axis label in lower panel seems incorrect.

  Please see the response below to Comment B5b.

B5b. Does the report follow a logical structure and layout?

The risk assessment report follows logical structure. However, in its present form the “Additional details…” document is confusing.

> ***This document was created for reviewers to facilitate their running of models, etc. However, the elements of the “Additional details” document that provide necessary corrections or clarifications to the original risk assessment report have been incorporated into the updated version of the report.***

B5c. Are the conclusions supported by the risk assessment?

The main conclusion of this risk assessment is that acceptance of the new inspection system (HIMP) is very unlikely to result in adverse effects on public health with respect to salmonellosis attributable to market hog carcasses. This conclusion is supported by risk assessment. The uncertainties around this conclusion have been evaluated and described.

> ***No response is necessary.***

Reviewer C

C1. Please evaluate the available data and the underlying assumptions used in this risk assessment.

C1a. Would the results of this simplified prevalence model be similarly reliable to a more complicated, traditional risk assessment model of the market hog slaughter inspection process?

The level of complexity of a model depends on the amount and structure of data available. On the other hand, we must take into account that the complexity introduced into a model does not guarantee its reliability.
I believe the results of this QRA would be reliable inasmuch as the underlying model (the logistic one) could appropriately describe the data.
Although the logistic model chosen is very simple, it is not the most parsimonious one. There are a number of inadequacies that must be solved in the logistic modelling itself before the results from the simulations can be assessed.

In response to this and other reviewer comments, we have evaluated a number or alternative models, including some hierarchical modeling. Those models are now addressed in Appendix H, in the Alternative Models section. However, because those approaches do not substantively change the report's conclusions, and are much more complex, the main part of the risk assessment retains the simpler, logistic model.

C1b. Have all key studies and data been identified, correctly analyzed and properly interpreted? If not, please provide additional data sources and citations (where appropriate) or alternative interpretations or analyses.

There is a lack of information on:

- The contamination force imparted by the groups of slaughter pigs entering the abattoirs. For sure, in some of the herds, pigs would be more infected than in others, and such information is not taken into account to model the *Salmonella* prevalence on pre-eviscerated carcasses, which necessarily impacts on the prevalence after chilling.
- The effects of processing stages such as scalding, polishing, application of sanitizers/carcass washing, or even the type of chilling in the abattoirs; in other words, important variables that can be useful inputs for simulation of risk management strategies.

I believe that this risk assessment study can be underpinned by much more data from the literature, either from predictive microbiology experiments or from meta-analysis of relevant published studies. A very careful systematic review of published studies on Salmonella in the productive process of pork meat should be undertaken, and this can be limited to research conducted only in the USA.

As mentioned on page 12 of the original risk assessment, this report aims to estimate potential reductions in illness or risks from modifying the allocation of FSIS inspectors in market hog slaughter establishments. Hence, we use the total number of Inspection System Procedures (ISP) and a published empirical contamination prevalence-illness relationship to estimate how those procedures would impact illness rates. This relationship implicitly takes into account the processing stage effects that the reviewer mentions and takes advantage of FSIS data, in line with previously published and peer-reviewed methodology (see Table 5 for details). This prevalence-based model differs from a risk

assessment that looks at the changes in prevalence with different interventions in the slaughter establishments, circumventing the need to look at changes in specific points in the slaughter process with different interventions, data which are not available. Estimating the effect of upstream changes, such as* Salmonella *contamination load or prevalence on hogs entering the slaughter process, is beyond the scope of this assessment.

C1c. Have the strengths and limitations of the data been transparently explained?

The data have not been fully explained. For instance:

- Both, "Districts" and "Regions" are contemplated in the model. However, there might be more HIMP or non-HIMP might abattoirs in some Districts/Regions than others. However, this information was not given to the Reviewer.
  The information about Districts and Regions are given in the data that was shared with reviewers. The model does not improve by adding the interaction between District and Region with HIMP. For instance, the AUC for the original is 0.9351 and adding the interaction of Region*HIMP to the model, the AUC shifts only slightly to 0.940.

- The difference and/or relationship between Districts and Regions have not been explained either. For the "South" Region, there are three levels "-1", "0" and "1", which is not clear to this reviewer. The variable "District" has also the same levels. Did you attempt to standardize the Regions? If so, class variables should not be standardized; that can only be done with continuous variables.
  There is no need to standardize the variable District; the descriptions are given in Appendix E: Structural Variables.

- Linked to the previous remark, why is it that instead you did not use three levels MidWest, NorthEast and South for the variable called "Region", instead of using "-1", "0" and "1"? This is very confusing to the reviewer.
  This is a preference and does not substantively affect the model outputs.

Why are the variables "U", "S", "SP", "SNP" and "NC" standardized?

The variables are not standardized. Typographical errors where "NC" was included instead of "NR" have been corrected in the updated version of the report. If the reviewer is referring to Table 9, it is common practice to give standardized regression coefficients in regression analysis results as this method removes the unit of measurement for predictor and outcome variables.

C1d. Given the differences in the data, is the overall modeling approach for hog slaughter, including its differences from the poultry slaughter risk assessment, appropriate?

This reviewer believes that the simple logistic model is completely inadequate to describe the hog slaughter data. The rationale will be exposed in detail in the next section.

Please see Appendix H for justification of simple logistic model.

C1e. Are there additional differences between hog slaughter and poultry slaughter, in the data or modeling approaches, that should be articulated? If so, would they affect the modeling approach?

For both models, it is necessary to find the most parsimonious model. However, in the case of the hog slaughter regression model, there are a lot of non-detections (zeros) particularly for the prevalence after

chilling. If this fact causes significant over-dispersion of the data, it is something that has to be taken care of, which can be done by a random-effects model, or by adding zero-inflation to the underlying distribution.

In response to the reviewer's comments, we have conducted further analysis to evaluate whether there is over-dispersion of the data. Please see Appendix H which addresses reviewers' concerns and explains how overdispersion has been contained.

Campylobacter is not important in pork meat, as in poultry.

No response necessary. The poultry report considered* Campylobacter *a significant hazard while the market hog report did not.

C1f. Is the model's sample size adequate to estimate prevalence change?

This question can be addressed at the second step of modelling through a Monte Carlo simulation.

Sample size calculation for logistic regression is a complex problem, but based on the work of Peduzzi et al. (1996) the following guideline for a minimum number of cases to include in your study can be suggested. Let *p* be the smallest of the proportions of negative or positive cases in the population and *k* the number of covariates (the number of independent variables), then the minimum number of cases to include is: N = 10k/p.

```
# sample size
table(hogs1$Sal)/sum(5984+1487)
#     0         1
#  0.8009637 0.1990363
# lower proportion
p=0.19903
#number of covariates
k=18
# Minimum number of cases to include is:
N = 10*k/p
N
# The minimum number of cases required is:
[1] 904.386
```

Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology 49:1373-1379.

Additional analyses have been conducted to explore the power of the analyses given the sample size; those analyses indicate that the sample size is adequate to detect differences. Please see Appendix H for a discussion of those analyses.

C1g. Are the illness distribution calculations adequate to describe market hog-attributable illnesses?

The calculations are simple; they assume direct relationships, as in a simple rule of three; yet they can be considered as adequate.

No response is necessary.

C2. Please identify limitations, weaknesses, or inadequacies of the logistic regression techniques and data (Stage 1); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

C2a. Is the technique accurately described, utilized, and appropriate for its intended use?

The technique is acceptably described in the report; however the simple fixed-effects logistic regression is largely inappropriate. There are two aspects that the logistic regression must take care of:

(i) there is an underlying dependency of variables, which the error structure must account for (see Figure 1 at the end of this document); and

(ii) overdispersion (the binary data you have is highly overdispersed (variance>>>mean), arising from the fact, that in one single model, you model prevalences pre-evisceration (high prevalences ~70%) and post-chill prevalences (low prevalences ~2.5%)). A test for overdispersion showed that the data is significantly overdispersed.

If variable dependencies and over-dispersion are left unattended, the model results will be inaccurate, and the conclusions drawn for the simulations simply wrong.

Please see Appendix H for information about the model chosen and the alternative models considered – including a section where overdispersion containment is addressed.

C2b. Are the data analyses and source code accurate?

The error structure of the model seems very inadequate. In the model, "District" does not appear to depend on Region. Is this really true? If not, a nested design must be used. Likewise, the HIMP/non-HIMP abattoirs might be linked to the Region, or even maybe non-HIMP abattoirs are more likely to arise in certain Districts than others. Since these dependencies have not been adequately taken into account, multicollinearity can be present in the model.

In addition, in the data analysis, there is no effort in finding out whether the model "can appropriately characterise" the data. Before you carry out Monte Carlo simulation using the logistic model, you must ensure the model closely describes the data. This has to be evaluated by at least these two graphs:

- Fitted values versus observations

Errors versus fitted values

Please see Appendix H for information about the model chosen and the alternative models considered. As show there, incorporating a nested design would not substantively change the risk assessment report's conclusions.

C2c. Does the low *Salmonella* prevalence affect the validity of the model?

A short answer for this is "yes". As mentioned in Query 2(a), modelling together the prevalences pre-evisceration and post-chill is not that convenient, because you cause the "joint" data to become highly over-dispersed (i.e., variance >>>> mean). Furthermore, you are assuming that both prevalences (pre-

evisceration and post-chill) are affected by the same variables. This is not necessarily true, the pre-evisceration prevalence and the post-chill prevalence may not be affected by the same variables (!).

In order to test this, I split the data into two groups, and ran a comparable code: Doing this, I found that **pre-evisceration prevalence** was affected by: HIMP, logNBrEmp, Fall, MidWest, NorthEast, South, District2, District3, District4, U and NC. However, the **post-chill prevalence** was affected by a lower number of variables, namely, HIMP, logNBrEmp, MidWest, South, District4 and SP. This seems reasonable – and was in fact what I expected – because the prevalence of Salmonella on pre-evisceration carcasses depends on many more factors – from the on-farm herd controls up to the good manufacturing practices during slaughter, for instance; whereas the Salmonella prevalence measured on the chilled carcasses would be mostly dependent on the level of contamination of carcasses at the point of evisceration and the effect of chilling.

A way to address this would be to define "Salmonella status on post-chill carcasses" (**y**) as the response variable. "Salmonella status on pre-chill carcasses" (**x1**) would be then an explanatory variable, which is at the same time affected by the other variables. In this way, the other variables might enter the model in interaction with **x1**. If by doing this way, the model still does not sufficiently handle over-dispersion, the next step would be to incorporate random-effects. The clustering variable for the random effects, for instance, can be the abattoir/establishment. If doing this way, the over-dispersion has not been sufficiently accounted for, then more complex models using zero-inflation must be contemplated.

Additional analyses have been conducted and presented to evaluate the effect of the reviewer's suggestion. Please see Appendix H for information about the model chosen and the alternative models considered, including zero-inflated versions as suggested by the reviewer. Incorporating these different approaches would not substantively change the conclusions of the risk assessment.

C2d. Is the use of logarithmic weighting appropriate?

The idea seems to be appropriate, and can be kept if a (more appropriate) random-effects logistic model is to be developed.

As discussed in response to previous comments, please see Appendix H for information about the model chosen and the alternative models considered.

C2e. Is the model over- or under-parameterized?

The model, as is, is over-parameterised or overfitted because, although, some variables are not significant, according to the chi-square test (use the anova() statement in R to retrieve these results), they were still kept in the full model. For instance, "Summer" is not significant, and when this variable is removed from the full model, the reduced model is not significantly different from the full model. The same was observed for "District1" and "Spring". In addition, the interaction "HIMP*COLL" was not significant (Pr(chi)=0.323). When using the fixed-effects logistic regression (the model presented in this QRA), only the variables "HIMP", "COLL", "LogNbrEmp", "Fall", "MidWest", "NorthEast", "South", "District2", "District3", "District4" and "NC" were significant. Therefore, the current model is over-parameterised and non-parsimonious – apart from being inappropriate from the point of view of the lack of random effects and the lack of a correlated error structure.

We respectfully disagree with the reviewer, and note that other reviewers did not think the model is over-parameterized. As indicated in Appendix A: Model Validation, the model's parameterization scheme is appropriate for the final model. In addition, the interaction HIMP*COLL was significant (Pr(chi)=0.0039). As discussed in response to other comments by this reviewer, the use of random effects models did not substantively change the conclusions (see Appendix H).

C2f. Does the model clearly characterize the uncertainty present?

The model does not extract the data over-dispersion by placing appropriate random effects; therefore, it does not characterise uncertainty at all. The only source of uncertainty the model has is the sampling error. To account for uncertainty in contamination, you can extract measures of the variability between abattoirs, the variability between production batches within an abattoir, or even the variability between districts within a region.

Please see Appendix H for detailed uncertainty analysis. The suggested variability analyses would constitute a random effects model which has been explored in Appendix H and did not substantively change the model's conclusions.

C2g. Is variability in contamination prevalence and establishment characteristics sufficiently addressed?

The variability in prevalence is not well addressed because the present model is built upon a simple binomial distribution (variance < mean), when in fact the data (overdispersed) calls for the utilization of random effects. As mentioned previously, in the response to query 2(f), from the current data it is possible to characterise the variability in prevalence due to abattoir, due to batches of production, or other clustering variable considered as strategic by the US risk managers.

Please see Appendix H for a detailed explanation of models considered and the final model chosen.

C2h. Has confounding been adequately controlled?

I do not believe confounding has been adequately controlled. To deal with confounders, there are two options: Stratification and Multivariate methods. Confounding can be a problem in the association **HIMP versus Region**

```
#table(hogs1$South, hogs1$HIMP)
#
#          -1    1
#    -1   190  700
#     0   597 4391
#     1     0 1593
```

Please see Appendix H for a detailed explanation of models considered and the final model chosen. As now shown in Appendix H, neither stratification nor multivariate (mixed models) methods substantively changed the model's conclusions. Additionally, the observational nature of the risk assessment results in a model in which confounding cannot be completely eliminated.

C2i. Has multicollinearity been addressed?

Since the dependency between variables has not been addressed utilising an adequate error correlation matrix, confounding has not been controlled at all. As mentioned in 2(b), nested variables are a good option for this. In figure 1, you will find a representation of the correlation of variables.

Multicollinearity was only tested a posteriori (after the simple model without an adequate error correlation structure was fitted) by means of the VIF of the independent variables. VIF values between 5 and 10 are worrisome because they denote multicollinearity of parameters and instability of their estimates. VIF should be lower than 3.

In response to this comment additional tests of multicollinearity have been conducted and demonstate that the model chosen better controls for multicollinearity than other models. Please see Appendix H for a detailed discussion of multicollinearity and the results of our analyses evaluating it.

C2j. Are the decision variables modeled correctly?

According to the current (fixed-effects) logistic model, the variables "SNP" and "U" are non-significant (please retrieve these results with the anova() statement in R); therefore their presence in this model does not contribute to a parsimonious model, but, on the contrary, makes the model overfitted.

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)	
NULL			5545	4696.7		
HIMP	1	41.684	5544	4655.0	0.0015244	**
COLL	0	0.000	5544	4655.0		
HIMPCOLL	0	0.000	5544	4655.0		
logNbrEmp	1	219.750	5543	4435.3	3.376e-13	***
Fall	1	0.273	5542	4435.0	0.7974160	
Spring	1	0.016	5541	4435.0	0.9505497	
Summer	1	2.553	5540	4432.4	0.4327711	
MidWest	1	49.213	5539	4383.2	0.0005723	***
NorthEast	1	2.592	5538	4380.6	0.4292846	
South	1	93.089	5537	4287.5	2.167e-06	***
District1	1	0.505	5536	4287.0	0.7272215	
District2	1	3.560	5535	4283.5	0.3542533	
District3	1	0.445	5534	4283.0	0.7432160	
District4	1	20.372	5533	4262.7	0.0266853	*
SP	1	51.335	5532	4211.3	0.0004350	***
SNP	1	0.291	5531	4211.0	0.7911570	
U	1	7.829	5530	4203.2	0.1694866	
NC	1	0.729	5529	4202.5	0.6751153	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

It would be interesting to test whether the decision variables are significant under the random-effects logistic regression this reviewer proposes.

Please see Appendix H for explanation of the models considered and justification of the model chosen. As mentioned previously, incorporating random effects did not substantively change the model's conclusion.

C2k. Is evaluation of non-HIMP establishment results at post-chill adequate as a sub-model?

In the way the logistic model is currently formulated, such evaluation is not appropriate, once again, because of the large over-dispersion (generated by modelling both the pre-evisceration prevalence and the post-chill prevalence) not adequately handled. This reviewer suggests that you model "post-chill prevalence" as a function of "pre-chill prevalence" and the other variables in interaction or not with pre-chill prevalence, placing random effects in a variable likely to be affected by a high heterogeneity in prevalence, such as establishment.

As discussed in response to this reviewer's other comments, we have now explored a random effects models, and it did not substantively change the outcome and conclusions. Please see Appendix H for an explanation of the models considered and justification of the model chosen.

C2l. Are the conclusions drawn from the regression analysis appropriate?

Given all the facts discussed above, the conclusions drawn from the logistic regression are largely inaccurate. It is important to know that, when data are over-dispersed, and over-dispersion is not accounted for, the variables will tend to show up as significant, when in fact, they may be non-significant.

As discussed in response to Comment C2a, additional analyses have been conducted to account for overdispersion. Please see Appendix H for a discussion of over-dispersion.

C3. Please identify limitations, weaknesses, or inadequacies of the Monte Carlo simulation techniques and data (Stage 2); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

C3a. Are the data analyses and source code accurate?

The Monte Carlo simulation has been carried out upon a fixed-effects logistic regression. This fixed-effects logistic regression is largely unsuitable for the data; therefore, the results from the simulation stage cannot be expected to be accurate. What the authors should fit is a **random-effects** logistic model, taking into account nested variables and an adequate error structure, therefore capable of extracting the variability in prevalence due to abattoir, due to product batch and/or due to region.

As discussed in response to other comments by this reviewer, incorporating random effects did not substantively change the model's conclusion.

C3b. Is the final model based on data from thirty-five establishments sufficient to estimate prevalence and changes in illness number from the scenarios that were included?

With a proper model, I believe the data from 35 establishments should suffice. Bear in mind that when a fixed-effects model is fitted (viz. the type of model fitted here), the results can only be generalizable to the 35 establishments. Said otherwise, it only represents what happens in the 35 establishments. However, if a random-effects model is fitted (and if the clustering variable is the establishment/abattoir), the results are, at least theoretically, generalizable to all the population of establishments. Therefore, this is an advantage gained when adjusting a random effects model.

As discussed in response to other comments by this reviewer, incorporating random effects did not substantively change the model's conclusion.

C3c. Are the scenarios with altered decision variable profiles well-characterized?

The scenarios with altered decision variable profiles do not seem incorrect, and they can be kept, however simulation should be done on the basis of an adequate model (this is, a random effects logistic regression).

As discussed in response to other comments by this reviewer, incorporating random effects did not substantively change the model's conclusion.

C3d. Is the applicability of the prevalence-illness relationship adequately demonstrated?

The prevalence-illness relationship is a very simple one, which can be used in absence of better data.

No response is necessary.

C4. Evaluate whether adequate sensitivity analysis has been provided.

C4a. Have the most important variables in the model been identified? If not, which have been left out?

The variables chosen for the sensitivity analysis could as well be kept. However, the results of the sensitivity analysis are very likely to be biased. The simulation (and hence the sensitivity analysis) must be re-run using an appropriate random-effects logistic regression. Thus, I cannot stress enough that the main thing to correct is the logistic regression.

As discussed in response to other comments by this reviewer, incorporating random effects did not substantively change the model's conclusion.

C5. Evaluate other characteristics of the report.

C5a. Is the report clearly written and complete?

The report is written in a way that, at least for me, was difficult to read and interpret. It is also incomplete because it does not present the model fitting in an unambiguous way. Nowhere in the report can be found an assessment of the fitting quality of the logistic model by simple charts, such as "Fitted values versus observations" and "Errors versus fitted values".

Please see Appendix H for explanation of the models considered and justification of the model chosen. Incorporating random effects did not substantively change the model's conclusion. The requested simple chart has also been added.

C5b. Does the report follow a logical structure and layout?

The report should have this layout:

1) Modelling of Salmonella prevalence on post-chill carcasses
   Introduction

Methodology
Results

2) Modelling the relationship between Salmonella prevalence on post-chill carcasses and probability of infection
Literature sources
Model development
3) Monte-Carlo simulation of the effects of FSIS interventions on Salmonella post-chill prevalence and illness
Modelling of decision variables
Monte-Carlo simulation results

Conclusions

We appreciate the reviewer's suggestion, however this question of formatting seems more a matter of preference and so we have structured it to be consistent with the previous risk assessments.

C5c. Are the conclusions supported by the risk assessment?

No, the conclusions are largely unreliable, because the core of the QRA (viz. the logistic regression) was performed incorrectly, disregarding data overdispersion. **Random effects**, **nested variables** and a **correct error covariance structure** must be necessarily introduced to the logistic regression to cope with the significant over-dispersion of the data. After this, the Monte-Carlo simulations must be conducted based on the right logistic model.

As discussed in response to the reviewer's comments above, we have now considered the suggested models and they did not substantively change the outcome and conclusions. We have also addressed the potential for over-dispersion. Please see Appendix H for an explanation of the models considered and justification of the model chosen.

Reviewer D

D1. Please evaluate the available data and the underlying assumptions used in this risk assessment.

D1a. Would the results of this simplified prevalence model be similarly reliable to a more complicated, traditional risk assessment model of the market hog slaughter inspection process?

Yes, based on model results and the information provided in the Appendix C (i.e., H-L test and the AUC of the ROC curve results), this simplified prevalence model seems adequate and reliable.

No response is necessary.

D1b. Have all key studies and data been identified, correctly analyzed and properly interpreted? If not, please provide additional data sources and citations (where appropriate) or alternative interpretations or analyses.

Yes, key studies and data have been identified, analyzed and interpreted.

No response is necessary.

D1c. Have the strengths and limitations of the data been transparently explained?

The approach and assumptions are well explained at the beginning of the document (e.g., Table 3) and throughout the text. However, a specific section or an appendix with the main limitations and strengths would have been valuable.

In response to this and other comments, we have added an appendix, Appendix H, that goes into greater depth about the models and data sources chosen, which includes a discussion of strengths and limitations.

D1d. Given the differences in the data, is the overall modeling approach for hog slaughter, including its differences from the poultry slaughter risk assessment, appropriate?

Yes, it seems appropriate.

No response is necessary.

D1e. Are there additional differences between hog slaughter and poultry slaughter, in the data or modeling approaches, that should be articulated? If so, would they affect the modeling approach?

I don't see the need to articulate any additional differences in the data or modeling approach between hog slaughter or poultry slaughter.

No response is necessary.

D1f. Is the model's sample size adequate to estimate prevalence change?

Yes, 5 establishments is not a lot, but should provide adequate numbers, given the volume of samples taken in all of them (n=787). The only concern is how the selection of the 5 establishments for the implementation of the HIMP was conducted. Were those 5 HIMP establishments significantly different from the other 159 non-HIMP establishments **before** the FSIS initiated the voluntary HACCP-based Inspection Models Project (HIMP) in each of them? Certainly in Appendix B we can see that (as expected) there are differences in the prevalence of Salmonella in the HIMP vs non-HIMP establishments. Now, the questions is, is that prevalence difference due to the implementation of the HIMP or is this something that may have been observed even before the implementation of HIMP just due to the nature and specific characteristics of those particular 5 establishments. It will be great (if possible) to present some tables similar to the ones in Appendix B of the historical info of those 5 establishments **before** the actual implementation of the HIMP protocols to actually address this question/concern.

Additional details about the enrolled slaughter facilities are included in the 2014 report,* Evaluation of HIMP for Market Hogs*, which is referenced in the risk assessment. Including such details, therefore, is beyond the scope of this risk assessment[2].

D1g. Are the illness distribution calculations adequate to describe market hog-attributable illnesses?

Yes, the illness distribution calculations (and sensitivity analysis) seem OK as presented in the document (particularly in Table 6 and Appendix G)

No response is necessary.

D2. Please identify limitations, weaknesses, or inadequacies of the logistic regression techniques and data (Stage 1); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

D2a. Is the technique accurately described, utilized, and appropriate for its intended use?

Yes, the technique seems appropriate. I would have thought in the first place to use a multi-level logistic regression model using "establishment" as random effect. However, most of this clustering is taken into account by including the dummy variables "Region" and "District", as well as including the weighting at establishment level.

We agree with the reviewer and, as discussed in response to comments by Reviewer C, we have now evaluated the random effects model and it did not differ substantially from the chosen model (see Appendix H).

[2] Accessible as of July 12, 2018 at https://www.fsis.usda.gov/wps/wcm/connect/f7be3e74-552f-4239-ac4c-59a024fd0ec2/Evaluation-HIMP-Market-Hogs.pdf?MOD=AJPERES

D2b. Are the data analyses and source code accurate?

Yes. However, there are several inconsistencies all over the document as well as in the raw hogs1.csv data (see specific comments in the text, but for example, in the hogs1.csv we have "Midwest" "NorthEast" and "South" but in the APPENDIX E the categories mentioned are “North-East”, “North-West”, “South” and “West”. So not sure here if Midwest corresponds to just West or to North-West?; the non-compliance records is coded as NC in the hogs1.csv data but as NR in the text (2018-05-16 SAS&atRisk Data Dictionary.docx document mentions about this NC vs NR use, but why just don't using NR consistently everywhere??); etc.

The inconsistencies with NR and NC have been corrected in the updated version of the report. The model in R encodes the regions at three levels relative to the model in SAS as follows:

Region\Variable	MidWest	NorthEast	South
MidWest	1	0	0
NorthEast	0	1	0
South	0	0	1
West	-1	-1	-1

D2c. Does the low Salmonella prevalence affect the validity of the model?

Certainly, when the prevalence is low and the sample size is large, Poisson approach can be used to approximate probabilities from the binomial distribution. However, the use of this approach may lead to low power and large standard deviations. Previous studies like the one presented by Petersen and Deddens (2008) (available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2292207/) suggested that for most of the cases, particularly with moderate prevalence and moderate sample sizes, the log-binomial method is slightly less bias than the Robust Poisson method and provided higher power and smaller standard errors in most of the cases. Therefore, they suggested that the log-binomial method is superior in most common situations, even in those with very low or very high prevalence.

As discussed in response to Comment A1f, additional analyses have been conducted to explore the power of the analyses given the sample size; those analyses indicate that the sample size is adequate to detect differences. Please see Appendix H for a discussion of those analyses.

D2d. Is the use of logarithmic weighting appropriate?

Yes, it is ok.

No response is necessary.

D2e. Is the model over- or under-parameterized?

No, it looks OK.

No response is necessary.

D2f. Does the model clearly characterize the uncertainty present?

Yes. The uncertainty presented is clearly described, evaluated and discussed

No response is necessary.

D2g. Is variability in contamination prevalence and establishment characteristics sufficiently addressed?

Yes, the variability in contamination prevalence is sufficiently addressed

No response is necessary.

D2h. Has confounding been adequately controlled?

There are some vague mentions about the control for confounding in the text (page 31, 76, 77, 97); however, the specific approach has not been described. For example, I wonder if variables 7, 8, 9, 10 and 11 as described in page 98 (which were not included in the final model) were evaluated as potential confounders. Particularly, I would have checked for the categorical HACCP size of establishment (large, small and very small). Some discussion will be valuable in this regard.

The stepwise regression process detailed in the original report is now expanded in Appendix H, providings details about the approach to control confounding. HACCP size was considered for inclusion but, as it was not statistically significant, was not included in the final model.

D2i. Has multicollinearity been addressed?

Yes. Appendix C presents a section about “Multiple Collinearity Analysis” presenting all VIF (and Tolerance) values reflecting that there are no multicollinearity problems in the model

No response is necessary.

D2j. Are the decision variables modeled correctly?

I believe the coding in the HIMP and COLL variables is not the most appropriate (see my comment in the text: “ I just don’t see convenient to use the coding "-1" and "1" for HIMP and COLL columns. It turns out that for your HIMPCOLL (=HIMP*COLL) you won’t be able to distinguish "-1"*"-1" vs "1"*"1". In other words, the coding for "collection in market hog baseline pre-evisceration" in a "HIMP establishment" will be the same (HIMPCOLL code = 1*1= 1) than the collection in "market hog baseline and PR/HACCP post-chill" in a "non-HIMP establishment" (code = -1*-1 = 1). Similarly issue if we apply this code for the

other combinations: -1*1 vs code 1*-1. This coding definitely can significantly affect the results. I would suggest using other coding that allows a clear distinction between all the HIMPCOLL categories

Although the reviewer's concern would be valid if any 3-way interactions were included in the model. However, the final model did not include any 3-way interactions so the reviewers concern is not applicable. The lack of 3-way interactions has been made clearer in the post-review version of the document.

D2k. Is evaluation of non-HIMP establishment results at post-chill adequate as a sub-model?

Yes

No response is necessary.

D2l. Are the conclusions drawn from the regression analysis appropriate?

Yes, although I would double check the model results after re-coding HIMP, COLL and HIMPCOLL variables as suggested in section 2.j.

As stated in the response to 2j, because there were no 3-way interactions, this distinction does not substantively impact the outcome of the analysis.

D3. Please identify limitations, weaknesses, or inadequacies of the Monte Carlo simulation techniques and data (Stage 2); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

D3a. Are the data analyses and source code accurate?

Yes

No response is necessary.

D3b. Is the final model based on data from thirty-five establishments sufficient to estimate prevalence and changes in illness number from the scenarios that were included?

Yes

No response is necessary.

D3c. Are the scenarios with altered decision variable profiles well-characterized?

Yes, the scenarios with altered decision variable are well described and characterized

No response is necessary.

D3d. Is the applicability of the prevalence-illness relationship adequately demonstrated?

Yes

No response is necessary.

D4. Evaluate whether adequate sensitivity analysis has been provided.

D4a. Have the most important variables in the model been identified? If not, which have been left out?

Yes, the most important variables have been identified. The sensitivity analysis for illnesses avoided and product attribution is described in detail in the text and Appendix G. However, instead of the Figure 7 or the Figure A12, I would suggest using tornado graphs that are more informative (e.g. Spearman rank correlation coefficients or the one with regression coefficients – see Figure 1 and 2 below with the specific comments). Based on those graphs is easier to identify the inputs which value change may have significant impact in model outputs. The choice of the values for the min, most likely and max of the Pert distributions is critical to have accurate and more reliable risk estimation.

We appreciate the reviewer's suggestion and, in response, have included a number of additional tornado graphs. Please see Appendix H for the additional sensitivity analysis including Spearman correlation coefficients and tornado graphs.

D5. Evaluate other characteristics of the report.

D5a. Is the report clearly written and complete?

Yes, although some edits/suggestions have been provided (in the comments thought the text) to fix some errors and for overall consistency/clarification.

These edits have been incorporated into the updated version of the risk assessment report.

D5b. Does the report follow a logical structure and layout?

Yes, overall is a well written report

No response is necessary.

D5c. Are the conclusions supported by the risk assessment?

Yes, the conclusions are adequate and supported by the risk assessment

No response is necessary.

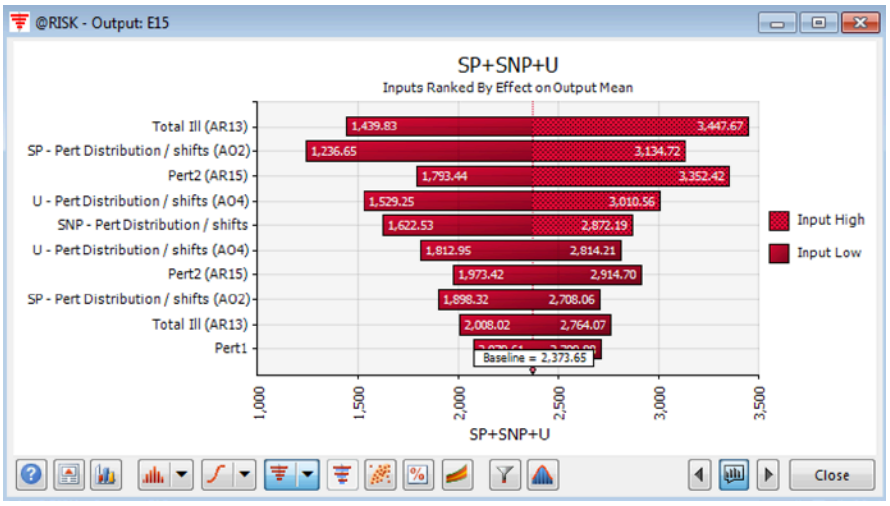


Figure 1. Tornado graph using the inputs ranked by effect on output mean. The use of this type of tornado graph (which I believe is similar to the one presented in Figure A12) does not necessarily indicate any influence of the inputs in the output and, in general, with more iterations you will see less variability in the change in output mean.

Figure 2. Tornado graphs using the Spearman's rank correlation coefficient (left) and the regression coefficient (right). These two types of tornado graphs provide a better representation of the impact that particular input variables have on the model outputs.

As discussed above, additional details of the sensitivity analyses are now included in Appendix H.

Reviewer E

E1. Please evaluate the available data and the underlying assumptions used in this risk assessment.

E1a. Would the results of this simplified prevalence model be similarly reliable to a more complicated, traditional risk assessment model of the market hog slaughter inspection process?

Yes. In fact, I'm not sure a more complicated, traditional risk assessment model would even be possible given lack of available data.

I don't think the question about "similarly reliable" can reasonably be answered without also conducting a complicated traditional risk assessment model, and then comparing the two results side-by-side.

I think the current approach is defensible on the grounds that I noted a traditional risk assessment model is not in fact possible.

No response is necessary.

E1b. Have all key studies and data been identified, correctly analyzed and properly interpreted? If not, please provide additional data sources and citations (where appropriate) or alternative interpretations or analyses.

Yes.

No response is necessary.

E1c. Have the strengths and limitations of the data been transparently explained?

Yes.

No response is necessary.

E1d. Given the differences in the data, is the overall modeling approach for hog slaughter, including its differences from the poultry slaughter risk assessment, appropriate?

Yes.

No response is necessary.

E1e. Are there additional differences between hog slaughter and poultry slaughter, in the data or modeling approaches, that should be articulated? If so, would they affect the modeling approach?

No.

No response is necessary.

E1f. Is the model's sample size adequate to estimate prevalence change?

Yes, however as noted in my detailed comments below, I am not sure that the authors approach of reducing uncertainty by increasing the sample size to 22,631 is statistically appropriate.

Reviewer E questions, in a number of places, whether it is statistically appropriate to include all inspection data, rather than solely the inspection data collected on the same day as sample collection, in the risk assessment. (In addition to this comment, Comments E2F and E3a, and when discussing pages 52, 71, 72, 119 and 126.) We have considered this reviewer's comments at length, however, believe that the use of all inspection data is appropriate because of, among other things, the major sources of variability and uncertainty in the two different stages of the risk assessment. This is now explained and evaluated in detail in Appendix H (in the "Rationale for Decreased Uncertainty with Expanded Sample Size). It should also be noted that the main risk assessment presents results using both the smaller and larger sample size, and the central tendency estimate from both data sets is the same.

E1g. Are the illness distribution calculations adequate to describe market hog-attributable illnesses?

Yes.

No response is necessary.

E2. Please identify limitations, weaknesses, or inadequacies of the logistic regression techniques and data (Stage 1); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

E2a. Is the technique accurately described, utilized, and appropriate for its intended use?

Yes, except as noted below where I have asked questions or raise concerns regarding report clarity.

These edits have been incorporated into the updated version of the risk assessment report.

E2b. Are the data analyses and source code accurate?

To the extent that I was able to determine, yes, the data analysis and source code are appropriate, except as noted below.

No response is necessary.

E2c. Does the low Salmonella prevalence affect the validity of the model?

No, it does not appear to affect model validity.

I suppose the authors could re-run the model, and assume higher prevalence values, and see if it makes a difference in the overall conclusions.

No response is necessary.

E2d. Is the use of logarithmic weighting appropriate?

I'm not sure. I was never able to find a location in the document where the authors actually explained and justified their reasons for choosing to weight prevalence by logarithm of production volume. Note that this choice dramatically over-emphasizes the weight of smaller facilities.

As discussed in response to Comment A2d, additional analyses are included in Appendix H to explore the effect of not log transforming the volume data. Although the reviewer is correct that not log-transforming the data would approximately triple the estimates of the illnesses avoided, we respectfully disagree that using such untransformed data is appropriate, as it risk weighting (logVolume) gives weight to all establishment's contamination while non-transformed Volume weighting gives risk to the largest volume establishments. Furthermore, using the log transformed volume data results in a more conservative estimate of illnesses avoided. Please see Appendix H for a detailed discussion of weighting and model considerations that were used to arrive at the final model included in the risk assessment report.

E2e. Is the model over- or under-parameterized?

The model appears to be more or less correctly parameterized.

No response is necessary.

E2f. Does the model clearly characterize the uncertainty present?

As noted in my comments below, I am not sure that the approach the authors used to reduce uncertainty is valid. In the first version of the model the authors assumed a sample size of 2230 days. They then decided to increase this number to 22,632 days by including all days of inspection recorded whether salmonella samples were taken or not. I believe that this approach may artificially and inappropriately reduce uncertainty.

Please see response to Comment E1f.

E2g. Is variability in contamination prevalence and establishment characteristics sufficiently addressed?

Yes.

No response is necessary.

E2h. Has confounding been adequately controlled?

Yes.

No response is necessary.

E2i. Has multicollinearity been addressed?

Yes.

No response is necessary.

E2j. Are the decision variables modeled correctly?

Yes.

No response is necessary.

E2k. Is evaluation of non-HIMP establishment results at post-chill adequate as a sub-model?

Yes.

No response is necessary.

E2l. Are the conclusions drawn from the regression analysis appropriate?

Yes.

No response is necessary.

E3. Please identify limitations, weaknesses, or inadequacies of the Monte Carlo simulation techniques and data (Stage 2); the reviewer must provide alternative data, data analysis, and/or modeling approaches.

E3a. Are the data analyses and source code accurate?

To the extent that I was able to determine, yes, the data analysis and source code are appropriate, except as noted below.

Let me clarify that what I'm talking about here are the technique used to reduce uncertainty, which is mentioned below, and which I believe is not appropriate. This is the analysis presented in table 14.

Please see response to Comment E1f.

E3b. Is the final model based on data from thirty-five establishments sufficient to estimate prevalence and changes in illness number from the scenarios that were included?

Yes.

No response is necessary.

E3c. Are the scenarios with altered decision variable profiles well-characterized?

Yes.

No response is necessary.

E3d. Is the applicability of the prevalence-illness relationship adequately demonstrated?

Yes.

No response is necessary.

E4. Evaluate whether adequate sensitivity analysis has been provided.

E4a. Have the most important variables in the model been identified? If not, which have been left out?

Yes.

No response is necessary.

E5. Evaluate other characteristics of the report.

E5a. Is the report clearly written and complete?

Yes, except as noted below.

No response is necessary. See below.

E5b. Does the report follow a logical structure and layout?

Yes, except as noted below.

No response is necessary. See below.

E5c. Are the conclusions supported by the risk assessment?

Yes, except as noted below.

Detailed comments appear below. The page numbers below correspond to the physical page numbers in the PDF document, which also matches the numbering scheme used. In each case I have flagged a section of text preceded by the word "content", which refers to text appearing in the report. Following this, my comments appear following the word "comment all learning".

Page 11:

Content: "a variety of online and offline duties."

Comment: Online and offline need to be defined.

Examples have been added to this paragraph in the updated risk assessment report.

Page 11:

Content: "Consistent with FSIS' focus on Salmonella outlined in the Agency's1996 implementation of the Hazard Analysis and Critical Control Point inspection system."

Comment: Not a complete sentence.

This has been corrected in the updated version of the risk assessment report.

Page 11:

Content: "Agency's1996"

Comment: Space needed before date

This has been corrected in the updated version of the risk assessment report.

Page 11:

Content: "(4.5%)1."

Comment: Footnote here is superscript 1, but actual footnote has 11. Clarify numbering.

This has been corrected in the updated version of the risk assessment report.

Page 12:

Content: "estimate potential reductions in illness or risks from modifying the allocation of FSIS"

Comment: Could be an increase too, correct?

This has been corrected in the updated version of the risk assessment report.

Page 13:

Content: "Salmonella."

Comment: Delete underline below "."

This has been corrected in the updated version of the risk assessment report.

Page 14:

Content: "as sanitation, HAACP,"

Comment: Should be HACCP

This has been corrected in the updated version of the risk assessment report.

Page 14:

Content: "For this assessment,"

Comment: This = swine or this = poultry. Rewrite to clarify.

This has been corrected in the updated version of the risk assessment report.

Page 14:

Content: "is strictly not an FSIS"

Comment: “Strictly not” or “not strictly”? I understand that the establishment does something that results in an NR, but there is some inspector-to-inspector difference. Some inspectors may miss problems and other may be overly aggressive in finding them.

This has been corrected in the updated version of the risk assessment report.

Page 14:

Content: "However, because future NR rates depend on the behavior of establishments,"

Comment: Mostly depend (see above)

This has been corrected in the updated version of the risk assessment report.

Page 19:

Content: "perform hands-on online inspection tasks"

Comment: Give examples

This has been corrected in the updated version of the risk assessment report.

Page 19:

Content: "offline inspection tasks"

Comment: Give examples

Examples of offline inspection tasks are given in the next paragraph, so no correction was necessary.

Page 19:

Content: "procedures, HAACP verification,"

Comment: HACCP

This has been corrected in the updated version of the risk assessment report.

Page 19:

Content: "Therefore, this risk assessment is designed using weighted regression modeling and"

Comment: Why *therefore* use weighted regression? Not clear why this logically follows.

The wording in this sentence has been edited for clarity in the updated version of the risk assessment report.

Page 22:

Content: "1278 1276"

Comment: Formatting not consistent with above (i.e. add “,” at thousands place.

This has been corrected in the updated version of the risk assessment report.

Page 22:

Content: "the PR-HAACP study."

Comment: Someone really needs to do a search and replace for HAACP to HACCP.

The authors have made this correction as suggested.

Page 23:

Content: "HAACP,"

Comment: HACCP

This has been corrected in the updated version of the risk assessment report.

Page 35:

Content: "The regression coefficients for all continuous variables in the first stage of the model are considered multivariate normal distributed."

Comment: Any evidence for this assumption?

These distributions did not contribute substantively to the outcome of the model as most of the variables are categorical. The true variances were typically smaller than would have been the case

with these distributions, so the assumption is conservative with respect to eventual public health conclusions. Appendix H also discusses overdispersion containment, which reduces error introduced by this assumption. Because of this, the text has been deleted from the body of the risk assessment.

Page 36:

Content: "Salmonella prevalence is estimated using these coefficients in log production volume weighted"

Comment: Why log production volume?

Please see Appendix H for a justification of this weighting scheme.

Page 37:

Content: "Contaminated carcass population prevalence estimates are derived from the average annual production log-volume weighted average prevalence estimates for individual non- HIMP establishments."

Comment: Why log weighted?

Please see response to Comment e2b.

Page 40:

Content: "The production volume grouping roughly corresponds to Very Small, Small, and Large HACCP establishment sizes."

Comment: Figure 2 shows 2 clear clouds of points. Not clear how this corresponds to 3 sizes.

This has been clarified with an explanation and demonstration of the clustering in the updated version of the risk assessment report.

Page 50:

Content: "NRs are not only a function of how frequently FSIS conducts inspection tasks but also indicate the effectiveness of the establishment's food safety practices."

Comment: May also be a function of inspector will of the parties irrevocably clearance personality/attention to detail, etc.

My point here is that there are other factors that influence when Inspector might write an NR. For example, it is well known that some inspectors are what is commonly referred to as "a pain in the butt". These inspectors tend to write a higher number of NRs than inspectors that don't fall into this category.

There also may be inspectors that are simply more observant, these inspectors may write a higher number of NRs.

This is not a suggestion to include additional factors into the analysis, but only to point out that ultimately human beings conduct inspections, and there are differences from person to person that should be acknowledged.

This has been acknowledged in the updated version of the risk assessment report.

Page 50:

Content: "Decreasing the number of NRs, according to the regression analysis, could theoretically reduce Salmonella prevalence (βNR = 0.0978, $p < 0.0001$) as a result of a higher number of inspections targeting food safety procedures."

Comment: I'm not sure I follow/agree with the logic here. If an establishment has fewer NRs they are theoretically being "safer" so of course they should be producing product with less Salmonella.

This sentence describes the findings of the authors' regression model. One explanation for fewer NRs associated with reduced* Salmonella *contamination is "safer" behaviors and procedures, generally, as indicated by the reviewer. The wording of this sentence has been edited for clarity in the updated version of the risk assessment report.

Page 50:

Content: "These products are: SP (-0.03424); SNP (0.0085); U (- 0.0158); NR (0.0137)."

Comment: Might be useful to include in table 9 above.

This is a matter of preference and we have chosen to keep the original format for Table 9.

Page 52:

Content: "Table 12and"

Comment: Add space before "and".

This has been corrected in the updated version of the risk assessment report.

Page 52:

Content: "Table 14 summarizes the expected change in human illness for the 35 establishment subsample using a larger sample size to better estimate the uncertainty distributions for each scenario."

Comment: I'm not sure the authors logic is correct here. They have decreased uncertainty by increasing the simulated number of days.

The impact of this step is that the authors have created an unrealistically precise estimate of the risk.

The only way to support the validity of this step would be to actually go out and collect 10 times as many samples and see if that also reduces the uncertainty.

I don't think there is a way to solve this deficiency using math. Put bluntly, the authors are stuck with the level of uncertainty inherent in the original data set. No amount of math or statistics is going to make that go away. The only thing that can make it go away is to collect more data.

Please see response to Comment E1f.

Page 53:

Content: "Reduction Cases "

Comment: This is a bit odd. The value for baseline seems to be total cases, but for other scenarios it's reduction. For consistency I suggest baseline be changed to 0 cases. If the authors want to include cases, make a separate column, or just include cases for baseline in a footnote.

We respectfully disagree that this is odd. This stylistic change has not been made.

Page 53:

Content: "35 Large and Small Market Hog Establishments (1)"

Comment: Unclear from the table what (1) and (2) mean in the table.

These numbers each refer to versions of the model where the sample included either procedures from days on which* Salmonella *samples were drawn (version 1) or all procedures in the establishments of interest (version 2). The two versions differ in sample size and that has been clarified throughout the document wherever necessary in the updated version of the risk assessment report.

Page 56:

Content: "Table 11 for all estimates."

Comment: Move to join with paragraph on prior page.

The formatting has been corrected in the updated version of the risk assessment report.

Page 57:

Content: "aThe indiscriminate scenarios show the range of illnesses avoided if any combination of inspection activity category is increased."

Comment: Superscript not found in table above.

Page 57:

Content: "bThis percentage represents the probability that an increase in illness of any size, even one illness, will occur. In other words, it is the likelihood that the decrease in illnesses will be negative."

Comment: Same comment

These two errors in superscripts have been corrected on each relevant table in the updated version of the risk assessment report.

Page 58:

Content: "Table 13: Estimated Illness Reduction Scenario Uncertainty-35 Selected Establishments (1)"

Comment: What Des (1) mean?

As stated regarding page 35, this has been corrected in the updated version of the risk assessment report.

Page 58:

Content: "aThe indiscriminate scenarios show the range of illnesses avoided if any combination of inspection activity category is increased. bThis percentage represents the probability that an increase in illness of any size, even one illness, will occur. In other words, it is the likelihood that the decrease in illnesses will be negative."

Comment: Same comments, no superscripts in tables.

These two errors in superscripts have been corrected on each relevant table in the updated version of the risk assessment report.

Page 58:

Content: "Table 14: Estimated Illness Reduction Scenario Uncertainty-35 Selected Establishments (2)"

Comment: What does (2) mean?

As stated above regarding page 35, this has been corrected in the updated version of the risk assessment report.

Page 58:

Content: "aThe indiscriminate scenarios show the range of illnesses avoided if any combination of inspection activity category is increased. bThis percentage represents the probability that an increase in illness of any size, even one illness, will occur. In other words, it is the likelihood that the decrease in illnesses will be negative."

Comment: Same comment, missing superscripts.

These two errors in superscripts have been corrected on each relevant table in the updated version of the risk assessment report.

Page 61:

Typo in text above Fig 4 “Establisments”

Page 63:

Same typo in Fig 6

These errors in figure titles have been corrected in the updated version of the risk assessment report.

Page 64:

Content: "and scheduled not preformed"

Comment: “Preformed” is a typo.

This has been corrected in the updated version of the risk assessment report.

Page 67:

Content: "Figure 11: Relative Contributions to Uncertainty in Illnesses Avoided (λavoided) Estimate for 35 Market Hog Establishments (1)"

Comment: A, beta, lambda, total and none are explained in the text, but also need to be explained in the figure legend.

An explanation has been added to the figure legend.

Page 71:

Content: "The sample size was increased from 2,230 to 22,632 by using all inspection data from 2010 through 2011 which included all days of inspection recorded whether Salmonella samples were taken or not."

Comment: As noted above, I’m not sure that this is a statistically valid approach.

Please see response to Comment E1f.

Page 71:

Content: "Using a larger dataset of 22,631 inspection days the feasible scenario (SP+SNP+ U) has an expected reduction in market hog-attributable salmonellosis cases"

Comment: Why expected reduction? Clarify.

Please see response to Comment E1f.

Page 72:

Content: "The magnitude of the uncertainty is such that while the mean of the estimated uncertainty distribution suggests a reduction in illnesses under all scenarios considered, the estimated probability of increased illnesses exceeds 5% in the SP+U scenario using the 22,631 sample size. The feasible SP+SNP+U scenario has the lowest probability of increased illnesses at 4.0% while reducing illnesses an average of 2,533. However, only targeting the SNP decision variable has a probability of increased illnesses of less than 0.01% while reducing illnesses an average of 1,257."

Comment: This is all dependent on the validity of the assumption that increased sample size will decrease uncertainty.

Please see response to Comment E1f.

Page 80:

Content: "Table A7through"

Comment: Space needed after 7.

This has been corrected in the updated version of the risk assessment report.

Page 81:

Content: "Table A3: Number of Salmonella Positive Samples Used in Model"

Comment: Numbers in this table should be decimal aligned.

Percentages in this table have been corrected to have three significant figures.

Page 81:

Content: "Pre-Eviscerationa"

Comment: Fix inconsistent bolding here and below.

This has been corrected in the updated version of the risk assessment report.

Page 81:

Content: "PREVPOST"

Comment: These terms need to be defined in the table.

The terms are now written out in the table.

Page 82:

Content: "PREVaPOSTa "

Comment: These terms need to be defined in the table.

The terms are now written out in the table..

Page 99:

Content: "The weighting scheme does not seem to unduly bias the percent positive estimates (in plant prevalence for this sample of establishments) because the crude (unweighted) percent positive values are reasonably close to the model estimates as evidenced by the standard errors of the crude estimates."

Comment: Yes, but I still do not see a justification for the weighting in the first place.

Please see response to Comment E2b.

Page 103:

Content: "Figure A for the market hog Salmonella full data set."

Comment: Figure is not numbered correctly.

This has been corrected in the updated version of the risk assessment report.

Page 103:

Content: "Figure A shows that stability of R-Square"

Comment: Figure not numbered correctly.

This has been corrected in the updated version of the risk assessment report.

Page 104:

Content: "Figure A shows the Receiver Operating Characteristic (ROC) "

Comment: Figure not numbered correctly

This has been corrected in the updated version of the risk assessment report.

Page 114:

Content: "Figure A"

Comment: Figure not identified correctly

This has been corrected in the updated version of the risk assessment report.

Page 115:

Content: "of λavoided Probability "

Comment: Is this a typo? Is there supposed to be a lambda before avoided?

This has been corrected in the figure.

Page 117:

Content: "Also shown in Error! Reference source not found."

Comment: Fix this missing reference

This has been corrected in the updated version of the risk assessment report.

Page 119:

Content: "This increased the sample size from 2, 330 to 22,631."

Comment: I am not convinced this is a statistically legitimate way I'm increasing the sample size. This is just resampling repeatedly from the original data set.

Unfortunately, I don't see any recommended course of action other than (A) living with the level of uncertainty inherent in the original analysis (B) spending the resources needed to collect additional data.

Please see response to Comment E1f.

Page 125:

Content: "Error! Reference source not found. shows the same relative uncertainty component"

Comment: Text missing reference

This has been corrected in the updated version of the risk assessment report.

Page 126:

Content: "The second dataset for the 35 establishments used inspection data over the same time period incorporating results from a total of 22,631 days of inspection whether Salmonella samples were taken or not."

Comment: As noted above, this is not necessarily a statistically legitimate approach.

Please see response to Comment E1f.