

**Summary Report (Draft)  
External Peer Review of FDA's**

***Draft Risk Ranking Model for Product Tracing as Required by Section  
204 of FSMA – Supporting Data Review***

Contract No. HHSF223201210011B  
BPA No. 06

March 8, 2016

**Prepared for:**

Lori Papadakis & Yuhuan Chen, Ph.D.  
U.S. Food and Drug Administration  
Center for Food Safety and Applied Nutrition  
5100 Paint Branch Parkway  
College Park, MD 20740

**Prepared by:**

Versar, Inc.  
6850 Versar Center  
Springfield, VA 22151

**Peer Reviewers:**

Jean-Louis Cordier, Ph.D.  
Thomas Montville, Ph.D.  
Sara Monteiro Pires, Ph.D.  
Thomas Ross, Ph.D.

## TABLE OF CONTENTS

I.	INTRODUCTION .....	1
II.	CHARGE TO REVIEWERS .....	2
III.	INDIVIDUAL REVIEWER COMMENTS .....	4
	Reviewer #1 .....	5
	Reviewer #2 .....	16
	Reviewer #3 .....	25
	Reviewer #4 .....	44
IV.	PEER REVIEWER COMMENT TABLE.....	68
	I. General Impressions .....	69
	II. Response to Charge Questions.....	74
	III. Specific Observations on <i>Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report)</i> within the context of the supporting data.....	118
	IV. Specific Observations on Appendices to the <i>Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report)</i> within the context of the model itself.....	131

## I. INTRODUCTION

Section 204(d) (2) of the FDA Food Safety Modernization Act (FSMA) requires FDA to designate high-risk foods (HRFs) for which additional recordkeeping requirements are necessary to protect public health. Under FSMA, FDA's designation of HRFs is based on six factors. Although FSMA section 204 requires FDA to designate "high-risk foods," in order to apply the FSMA factors, it is necessary to first take into account characteristics of both foods and hazards, i.e., food-hazard pairs. To address the requirements of FSMA Section 204, CFSAN has developed a data-driven model, the *Draft Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA*, that uses seven explicit criteria related to public health risk. Both microbial and chemical hazards are considered in the model as required by FSMA in the HRF designation.

For this peer review, four experts were selected to answer 12 charge questions and to evaluate and provide written comments on the HRF Model Report, underlying data, and risk scores. The peer review focused on the data used to populate the HRF model, including data from the literature, FDA, and expert elicitation, and the scientific rationale and justification for the scoring of the food-hazard combinations in the model.

### **Peer Reviewers:**

#### **Jean-Louis Cordier, Ph.D.**

CH-1806 Saint-Légier-La Chiésaz  
Switzerland

#### **Thomas Montville, Ph.D.**

Environmental Laboratory  
School of Environmental and Biological Sciences  
Rutgers, the State University of New Jersey  
New Brunswick, NJ 08901-18520

#### **Sara Monteiro Pires, Ph.D.**

Research Group for Risk-Benefit  
National Food Institute, Technical University of Denmark  
1601 Copenhagen V, Denmark

#### **Thomas Ross, Ph.D.**

School of Land and Food – Tasmanian Institute of Agriculture  
University of Tasmania  
University of Tasmania, Australia

## II. CHARGE TO REVIEWERS

FDA has developed a draft risk ranking model to inform the designation of high-risk foods for which additional recordkeeping requirements are appropriate and necessary to assist product tracing, as required under section 204(d)(2) of the FDA Food Safety Modernization Act (FSMA). The draft model was developed through an iterative process that involved, among other things, using the FSMA statutory factors to define criteria and scoring functions of the criteria, and collecting data relevant to the scoring criteria for food-hazard pairs to identify those foods which should be designated as high-risk for future consideration in policy decision.

The focus of this review is on the underlying data used in the model, within the context of the overall risk ranking approach, criteria and results. Note: a separate panel is reviewing the model.

### Charge Questions:

1. In order to apply the FSMA factors it is necessary to first take into account both the characteristics of foods and known or reasonably foreseeable hazards, i.e., food-hazard pairs. The food categorization scheme involves a list of 1,286 food-hazard pairs (candidates) and 335 foods (secondary commodities) linked to approximately 50 food categories (primary commodities).
  - a. Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?
  - b. Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?
  - c. Are the method and data used to identify the food-hazard pairs adequate?
  - d. If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.
2. Please evaluate the scoring definitions for the seven criteria:
  - a. Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.
  - b. For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).
3. The draft model integrates a substantial amount of public health and commodity-specific data about food-hazard pairs, as well as information regarding manufacturing and processing of different foods.
  - a. Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.
  - b. Are there data sources not yet used but that should be considered? If so, please provide specific examples of data source(s) for each criterion, and explain why the additional data sources might enhance the criteria scoring.

4. Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.
5. Consumption data from the NHANES What We Eat in America database were used and, for certain commodities (e.g., ice) where no data or limited data were available, expert opinion was used in scoring Criterion 6. Is the timeframe used (2011-2012 or 2009-2010 survey cycle) appropriate? Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.
6. Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate? Are there any underlying data not appropriately used? Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.
7. The draft model relies on expert elicitation to fill data gaps, in particular for the scoring of Criterion 5 and for the scoring of criteria for chemical hazards and undeclared allergens. Is the expert elicitation process (which involves external panels and FDA subject matter experts) adequate to address data gaps? Is the scoring logic used to combine data and expert opinions and the order of preference in section 4 (e.g., Figure 4.3) appropriate? If not, please explain what changes might be considered and why.
8. Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.
9. Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?
10. How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?
11. Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.
12. Do you have any additional comments? Please share them in your review.

### **III. INDIVIDUAL REVIEWER COMMENTS**

**Reviewer #1**

**Peer Review Comments on FDA's Draft Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA:**  
HRF Model Report, Underlying Data, and Risk Scores

Reviewer #1

---

**I. GENERAL IMPRESSIONS**

This report describes an impressive effort to rank foodborne diseases and foods on the basis of an extensive list of criteria. The report is very clear, well written and structured, and the authors undertook a notable task of data gathering and analysis. Overall, the seven criteria chosen to score food-hazard pairs are adequate and comprehensive.

The purpose of the project is to identify high-risk foods to “inform proposed rule-making”, and “establish recordkeeping requirements for the designated high-risk foods to prevent and mitigate outbreaks”. In this context, the model focuses largely on foods at the end of the food chain. Still, I have concerns about the general reliance of the model on outbreak data (specifically for microbial hazards). Outbreak events are random and can have multiple causes, from contaminated primary foods to some type of failure in the production or distribution process (e.g., environmental contamination, disrespect of optimal time-temperature). Even though these factors are potentially accounted for in the model, I believe that prevention of outbreaks (and of foodborne disease in general) requires the control and prevention of contamination of primary ingredients as well. Because random failures in the chain are difficult to predict and prevent, reduction of hazards at a point in the food chain that is closer to the original production of the food may actually be the most effective way to prevent many types of events. In this context, I have concerns about the use outbreak-related data for criteria 1 and 7. Outbreak cases correspond only to a fraction of the total cases of infection by most pathogens, and this proportion varies from pathogen to pathogen (e.g., 6% for *Salmonella* in the U.S. in 2013; 7% for *Listeria m.*; 15% for STEC (CDC data, see link below)). On the other hand, these data were also used for C7 (economic impact), but here outbreak cases were corrected for underreporting based on Scallan *et al.* (2011). However, Scallan's estimates were derived for correcting sporadic cases for underreporting, which are less likely to be reported than outbreak cases. Furthermore, the relative contribution of different foods for outbreak cases is not necessarily representative of the causes of all cases (i.e., of sporadic cases also), because some foods are more prone to cause large-scale outbreaks than others. This can lead to an overrating of the economic impact of disease by some foods. Lastly, and as mentioned in the discussion of the report, not all hazards cause outbreaks (or at least not frequently).

It is apparent that the authors have focused on outbreak data because the essential piece in the model is the food-hazard pair, but I believe that an alternative approach would be to start by ranking foodborne hazards on the basis of their disease burden in the population (e.g., by applying health metrics), and then estimating the proportion of this burden that is caused by different foods (using source attribution, which can also use outbreak data).



## II. RESPONSE TO CHARGE QUESTIONS

1. *In order to apply the FSMA factors it is necessary to first take into account both the characteristics of foods and known or reasonably foreseeable hazards, i.e., food-hazard pairs. The food categorization scheme involves a list of 1,286 food-hazard pairs (candidates) and 335 foods (secondary commodities) linked to approximately 50 food categories (primary commodities).*
  - a. *Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?*

The authors justify the choice of the categorization scheme well. The described arguments for the use of RFR-adapted primary categories [a) is in line with preventive control measures by FDA, b) considers processing, (c) no need for re-categorization] are reasonable, and the definition of secondary commodities under these primary ones facilitates the aggregation of information at different levels. The list is representative of (and more detailed than) the FDA's RFR commodities definitions.

I have one comment: some secondary commodities seem too unspecific (in the sense that they correspond to multi-ingredient foods that can potentially include almost any type of single food), and their link with hazards (to generate food-hazard pairs) somewhat arbitrary. For example, "sandwiches" can include nearly any type of food commodity (in addition to bread), which means that many of these foods can be contaminated with a wide range of hazards (which is reflected in the number of food-hazard pairs with sandwiches: 15). However, information on this food-hazard pair will not give any information on the contaminated original ingredient or point of contamination in the processing chain. Another example is "RTE dinners", which can be constituted by an even wider variety of food commodities. The fact that these are paired with hazards which have been associated with outbreaks or have been recalled due to mislabeling (undeclared allergens) at some point in time, does not ensure that all potential hazards associated with these type of foods are listed, or that the relative occurrence of different hazards in these foods is well-established.

That said, I would suggest that all secondary commodities have some kind of description of the basic food (ingredient) that can be linked to the hazard in the pair. In my view, this would allow for generating food-hazard pairs that are more useful for regulation purposes.

- b. *Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?*

Both primary and secondary commodities appear appropriate, and data to populate the model at the secondary-commodity level seems to be available in most cases. However, because of the point in the food chain that most foods are classified (i.e., prepared foods), contamination data for some seem to be lacking. Following the same example as above, contamination data from sandwiches seem to be sparse for most hazards. Categorizing foods on the basis of their ingredients (like foods that constitute a sandwich) may allow for collecting contamination data.

***c. Are the method and data used to identify the food-hazard pairs adequate?***

After establishing the categorization scheme, the authors listed all hazards that historically have been associated with each food. As stated in the report, this food list has been associated with known or reasonably foreseeable hazards. Even though this methodology is reasonable and potentially allows for the inclusion of most combinations of foods-hazards, I have one concern: it does not allow for the flexibility of allowing considering new, unseen events. Acknowledging that food contamination and consequently foodborne outbreaks are to a wide extent random and a consequence of so many different factors (e.g., environmental contamination, cross-contamination, carrier food-handlers), I think it would be important to identify food-hazard pairs not only on the basis of outbreaks and recalls, but also on other types of epidemiological evidence on the potential sources of foodborne hazards (see “general impressions”).

***d. If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.***

The authors explain in the report that they have investigated alternative categorization schemes (like CDC’s scheme by Painter *et al.*), and justify well why they have chosen to adopt the here presented one. Another option would be some kind of hybrid, which could allow having the advantages of the currently used scheme (i.e., in line with preventive control measures by FDA, and considering processing), but also the inclusion of evidence other than outbreak data in the risk ranking model. More specifically, including “meat and poultry” primary and secondary commodities would allow for listing a wide variety of microbial hazards (which are very common causes of foodborne disease, such as *Campylobacter*-chicken), as well as chemical hazards (such as PAHs and meats (grilled, fried, etc.)). Such an approach could also allow for classifying ingredients of unspecific composite foods (like sandwiches and RTE foods).

***2. Please evaluate the scoring definitions for the seven criteria:***

***a. Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.***

The seven criteria chosen to score food-risk pairs are adequate and comprehensive. In general, the scoring criteria are well defined. When explaining data sources for the criteria, it is extremely helpful that the same 3 (at least) examples (1 micro, 1 chemical, 1 allergen) are followed.

I have some specific comments on the description of some of the criteria (in addition to the responses given below):

**Criterion 1:**

- It is stated that this criterion is applicable to microbial and chemical hazards, and that the epidemiological link is based on the frequency and number of outbreak-related cases. However, it is also mentioned that chemicals and undeclared allergens have not been involved in outbreaks, so evidence that these hazards cause disease, and how much disease they cause in the population, is based solely on experts. It is of general agreement

within the scientific community that the epidemiological evidence for the casual effect of exposure to most chemicals and disease is frequently inconsistent, and the estimation of the incidence of disease that can actually be attributed to these chemicals complicated. Still, epidemiological and animal studies assessing these relationships exist, and an assessment of all available evidence seems crucial to make inferences on the relative occurrence of foodborne disease due to exposure to chemicals.

- It is described that “both the frequency of reported outbreaks and the occurrence of illnesses” (page 14 of the main report) (i.e., the number of reported outbreaks AND sporadic cases per year) are used in scoring. However, it is mentioned in the following page (page 15, line 10) that only outbreak data were used for this criterion (i.e., data on reported sporadic cases or estimates of total illnesses corrected for underreporting – burden of illness estimates) were not included. Aside from my concerns about the exclusion of sporadic disease in this criterion, the initial statement should be corrected.

Criterion 2:

- I have concerns about the indicators used to describe the severity of disease by the different hazards. Specifically:
  - For hazards that cause acute disease and for which quantitative data were available, only hospitalization and mortality rates are used. However, some foodborne diseases may lead to severe sequelae that do not necessarily reflect on a high mortality or hospitalization rate. It is described that these were not considered, but they can be relevant to describe severity. For example, Irritable Bowel Disease and Guillian-Barré Syndrome (potential sequelae of *Salmonella* and *Campylobacter* infection, respectively) can be very severe in the sense that they have large negative influence on the quality of life of the individual, but do not necessarily reflect in high rates of hospitalization (and certainly not mortality rates). One way of taking these differences into account and to be able to compare across hazards would be to use the disability weights (DWs) generated by the Global Burden of Disease (GBD) study as a basis for the indicator of severity. DWs reflect the impact of a health condition in terms of health-related quality of life estimated based on preferences obtained from a panel of judges. Preferences are defined as quantitative expressions or valuations for certain health states, which reflect the relative desirability of the health states. (See Salomont *et al.*, 2013). DWs for a wide range of conditions (which I expect include the vast majority of adverse health effects of foodborne hazards) have been estimated, and using DWs would allow for being much more comparative across hazards.
  - I feel that the semi-quantitative scale used for Criterion 2 does not adequately characterize the severity of different hazards. As described, the severity score for *Salmonella* and *Listeria* is the same (9), and – somewhat surprisingly - would be the same for adverse effects caused by exposure to some chemicals, such as cancer. However, the severity of disease by these hazards is very different. The authors argue that this will be compensated by Criterion 7, but the example of *Salmonella* and *Listeria* illustrates that this is not necessarily the case – a 3 times higher cost of disease does not necessarily compensate for the difference in severity. It may however balance if (true) incidence is taken into account (which is the highest for *Salmonella*), but it seems to me that this would require using corrected incidence estimates (BoI) in Criterion 1.

- b. For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).**

If Criterion 5 is defined for different food-hazard pairs taking into account if a food is RTE or not – i.e., if the indicator “steps taken to reduce contamination” for pairs with RTE foods are classified as “weak”, and thus Criterion 5’s score would be minimum 3 for that pair, I think it is reasonable to have the same definition for RTE and non-RTE foods in Criterion 3.

**3. The draft model integrates a substantial amount of public health and commodity-specific data about food-hazard pairs, as well as information regarding manufacturing and processing of different foods.**

- a. Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.**

- Criterion 1: The report uses official data for outbreak data (CDC and FDA), which are the appropriate data sources. Still (as described earlier), I have concerns about the use of outbreaks (only) as an indicator of the occurrence of disease in the population (i.e., for the epi link). On one hand, outbreaks caused by foodborne pathogens are recognized as a small fraction of cases by foodborne pathogens (this proportion varies, and it’s particularly small for some pathogens, like *Campylobacter* (<1%); for *Salmonella* was 6% in 2013: <http://www.cdc.gov/foodnet/reports/annual-reports-2013.html>). Even though foodborne illnesses by microbial hazards are known to be highly underdiagnosed and underreported, several studies have made efforts to estimate the “true” incidence of a wide variety of foodborne diseases. In the U.S., Scallan and co-authors have estimated the burden of illness of 31 pathogens. This study’s estimated multiplication factors are used in this report in Criterion 7 (see comments below), and I would think that it would make sense to use them in this criterion as well. In other words, I believe the overall incidence of disease (sporadic + outbreak cases) should be used as an indicator of the epidemiological link. On the other hand, and as well described in the report, chemicals and undeclared allergens seldom or never cause outbreaks, and therefore this type of evidence cannot be used as an indicator for disease. In addition, because many chemical hazards cause chronic disease, which is difficult to link with exposure, incidence data are lacking, and I believe experts would also have difficulties comparing disease by different hazards in the absence of incidence estimates. This task is particularly difficult because many of these chemical hazards are (considered potentially) carcinogenic, but the probability of a given individual developing disease is determined by many other factors. I recognized that these are difficult to derive, and there are few studies available so far, but believe that it is an important piece of information and a research gap that would be valuable to address, particularly in a risk ranking effort like this one. I am aware of ongoing research efforts that would be worth including in the analysis. Also worth mentioning is that the WHO’s initiative to estimate the global burden of foodborne diseases (FERG, [http://www.who.int/foodsafety/areas\\_work/foodborne-diseases/ferg/en/](http://www.who.int/foodsafety/areas_work/foodborne-diseases/ferg/en/)) also looked into the burden of some chemical hazards (at the regional level), and results will be publicly available latest in December 2015.

- Criterion 2: To describe severity of different diseases, hospitalization and mortality rates from Scallan *et al.* (for microbial hazards), and expert opinions (for chemicals and undeclared allergens) were used. It seems that the data source is appropriate for these rates, but the fact that Scallan did not consider potential sequelae of foodborne diseases hampers a complete picture of the severity of the different pathogens, and thus comparability of diseases. As examples, in addition to acute diarrhea, which is common to many of the considered hazards, *Salmonella* infections can lead to reactive arthritis, irritable bowel disease, irritable bowel syndrome, *Campylobacter* to these and to Guillian-Barré Syndrome, VTEC to hemolytic uremic syndrome, etc. Even though these diseases are rare and may also result in hospitalization and even death (in some cases), most will probably not be linked with the foodborne infection, because they occur relatively long after exposure and acute disease (HUS by VTEC and abortion by *Listeria* are examples of exceptions). Methodologies that allow for the inclusion of sequelae and their severity are available (health metrics like disability adjusted life years - DALYs), and would allow for a better comparability between diseases. In addition, this measure of severity (DW) would allow for a better comparability with disease by chemical hazards. As examples, the DW for cancer as estimated by the GBD 2010 is around 0.5, for diarrhea from 0.061 (for mild cases, >80% for most pathogens) to 0.281 (for severe, ~3%), and for HUS 0.21 (Salomon *et al.*, 2013). I believe that the fact that the final score for Criterion 2 (which is 9, the maximum, for ~50% of considered hazards) is for example the same for *Salmonella* spp. and *Salmonella Typhi*, and the same for *Yersinia* and acrylamide (which can cause cancer) illustrates the limitations of the current data for comparing between diseases.
- Criterion 7: I have some concerns about the approach and data used for this criterion. Like in Criterion 1, the authors use outbreak cases, but here these data are corrected for underreporting using Scallan's multipliers. However, these multipliers were derived to correct all illnesses captured by different laboratory surveillance systems, i.e., including sporadic illnesses. As mentioned before, sporadic cases are, for many pathogens, the majority of all reported cases. It is recognized that outbreak-related cases are more likely to be reported, and thus applying an overall underdiagnosis/underreporting factor to these is not necessarily appropriate. Even though this could be irrelevant in terms of disease burden in general – because the estimated burden of a disease would still be lower if outbreak cases are used as a basis for estimations, as opposed to all reported cases –, it may make a difference when the risk ranking model has as a starting point the food-hazard pair. In other words, as I understand it, in this criterion the model is correcting the (outbreak) cases caused by a food-hazard pair using the UR factors derived for that disease, which could lead to an overestimation of the relative contribution of that specific source for disease with that pathogen. The limitations of using outbreak cases for source attribution (which I think also apply here) have been widely discussed (e.g., Pires *et al.*, 2009; Hald *et al.*, 2004, Batz *et al.*, 2004, etc.).

I also have trouble understanding why UR factors of 1 were used for pathogens for which multipliers have not been estimated. For example, even though Scallan did not publish a multiplier for norovirus (because a top down approach was used, as opposed to a bottom-up/pyramid approach), that study did estimate burden of illness for norovirus (i.e., unreported cases), and the gap between reported cases and estimated burden was extremely large. This study estimated a total of ~21 million cases (which corresponds to 5,461,731 foodborne cases) in the

study year (U.S. Census population data from 2006 was used, just for reference). This can be compared with the number of outbreak-related reported cases used in this project. Data from 2009-2010 shows 69,145 norovirus outbreak-related reported cases ([http://wwwnc.cdc.gov/eid/article/19/8/13-0482\\_article](http://wwwnc.cdc.gov/eid/article/19/8/13-0482_article)). (From appendix D1 I can see 2,701 for 2006?).

In addition, the model used surrogates for the cost of illness (CoI) for some pathogens; it would be helpful to understand why specific surrogate data were chosen. For example, the cost for norovirus illness was used for parasites. Were these defined on the basis of similarities between symptoms and duration of disease?

It was also not clear how the experts provided input for chemicals and allergens. For e.g., some chemicals' health effect is cancer, and the costs of disease have been estimated. The problem with these hazards is estimating the incidence of disease – for which there are no data and few studies. Was there some type of scientific basis for the experts' input?

- b. Are there data sources not yet used but that should be considered? If so, please provide specific examples of data source(s) for each criterion, and explain why the additional data sources might enhance the criteria scoring.***

Please see detailed answers above.

- 4. Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.***

The weighting factors seem appropriate.

- 5. Consumption data from the NHANES What We Eat in America database were used and, for certain commodities (e.g., ice) where no data or limited data were available, expert opinion was used in scoring Criterion 6. Is the timeframe used (2011-2012 or 2009-2010 survey cycle) appropriate? Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.***

The timeframe used seems appropriate, and, because the purpose is to estimate the proportion for the population that is exposed (i.e., consumes) each food, I agree with using the rate from the day with highest consumption.

- 6. Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?***
- a. Are there any underlying data not appropriately used? Is any score assigned to a criterion not appropriate?***
- b. If so, please provide suggestions on how the process of data collection and documentation might be refined.***

My expertise is within microbial hazards, and so I will mainly review in detail pairs with some pathogens.

*Salmonella* appears in 4 of the first 20 pairs in the ranking (in this order):

- Tomatoes-*Salmonella*
- Sandwiches-*Salmonella*
- Shell eggs-*Salmonella*
- Fresh salsa-*Salmonella*

Assessing the epidemiological evidence for all pairs (Criterion 1), it is evident that the number of outbreaks and outbreak-related illnesses varies substantially (from 492 (weighted) outbreak cases from fresh salsa to 2,678 from eggs); still, they all fit into scoring 9 according to the scoring bins defined, with the exception of salsa, which had <13 outbreaks reported in the time period). Data sources seem appropriate.

The criteria that drive these different rankings are mainly prevalence and economic impact. For sandwiches, contamination data were very poor. The score was based on one single study (from Greece). Was this assumed to be representative of “sandwiches” to which the U.S. population is exposed? I also did not understand the data – e.g., total samples appear in decimals. (Also see my comments on the use of composite food categories in the ranking). For fresh salsa, I did not understand why the score for Criterion 1 is 3, while for Criterion 7 is 9. The inverse happens for sandwiches.

I was also interested in how *Campylobacter* pairs scored in the model. It is evident that all food-hazard pairs with this pathogen scored very low, and I believe there are two important reasons for this: a) *Campylobacter* seldom causes outbreaks, and this is recognized worldwide, and b) no meat and poultry products were considered in the model, and these are the most important sources of *Campylobacter* (which has been proven by a wide variety of source attribution studies). However, *Campylobacter* is the most reported foodborne illness in the U.S., and estimated to be the fourth most imported foodborne pathogen in terms of total (corrected) incidence (Scallan *et al.*). I would think that foods that typically cause *Campylobacter* infections would be worth labeling as “high risk”, along with the others included in this report. I could see that fresh salsa appears highlighted at some points in the results table (appendix L), and was interested in knowing why. I noted it myself for the same reasons I have pointed for sandwiches – the fact that salsa is a composite food, potentially constituted by a wide variety of ingredients (commodities) makes me question the utility of including it in such a risk ranking exercise.

Looking into a chemical hazard-pair in more detail, I can see that arsenic in rice is the first on the ranking. It is widely accepted that the link between exposure to arsenic and potential onset of cancer is one of the better established scientifically. Still, I wondered about the score for Criterion 3 (which is 9). I could not find the data used to support this score. Was the prevalence above the accepted level >1%? Or have there been recalls? This sounds very high, especially for the U.S. (i.e., maybe it would be more likely in high producers and highly exposed populations).

**7. The draft model relies on expert elicitation to fill data gaps, in particular for the scoring of Criterion 5 and for the scoring of criteria for chemical hazards and undeclared allergens.**

***Is the expert elicitation process (which involves external panels and FDA subject matter experts) adequate to address data gaps?***

***a. Is the scoring logic used to combine data and expert opinions and the order of preference in section 4 (e.g., Figure 4.3) appropriate? If not, please explain what changes might be considered and why.***

Expert elicitations are a widely used and well accepted method to address knowledge gaps and produce evidence when basic data are lacking. Within food safety, they have been successfully used for instance with source attribution, for which they are often the only method available to answer certain public health questions. In this project, I feel they were the best approach to address gaps, and the panels seemed to be varied in the sense that they included experts with experience in different fields within food safety. They were however small (maximum 4 experts in each panel), and I understood that they were not provided with an answer sheet, but rather discussed questions and reached agreement. I wondered if, in this process, they did not influence each other? Also, in structured EEs the experts' knowledge is often validated (through initial validating questions) and eventually weighted, and the answers are analyzed to derive uncertainty around the final estimates. I understand that this would not be possible with such panels, but these limitations should be discussed.

The scoring method used to combine available data and expert input seems reasonable.

***8. Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.***

I think the approach taken is appropriate, and would not agree with scaling down consumption: even if the foods would only be of high risk for certain subgroups in the population, there would be no way of protecting these populations if foods of this risk were not considered in the same way as foods/hazards that can cause disease to the wider population.

***9. Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?***

I am not sure I understand this question. I believe that risk scores should be compared, taking into account that the epidemiological link that constitutes the basis of the identification of the food-hazard pairs, and which contributes to the overall scores as well, is informed by outbreak data, which may not allow for the flexibility of considering new events, or for measuring the relative contribution of specific foods for disease by a given hazard if this pair has not caused (many) outbreaks before but may have caused sporadic illnesses (e.g., *Campylobacter* pairs).

***10. How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?***

The model uses an impressive amount of data, which were retrieved from numerous data sources and collected over more than ten years. I could see that more recent data from e.g., surveillance programs (namely outbreak data, used for criteria 1 and 7) are publicly available, and this could



now be included. However, for most other criteria new data would probably come slowly (e.g., scientific articles), and thus updating the model would make sense only every few years. That said, recent efforts to estimate the burden of foodborne diseases, particularly caused by chemical hazards, would be worth evaluating and including.

***11. Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.***

As mentioned in the general comments, the report is very clear, well written and well structured. The level of detail is adequate in the sense that it allows for understanding the approach and decisions, and refers to very detailed, complete and transparent appendices. I would perhaps suggest having a chapter on the general approach on uncertainty earlier on the report. Even though the current description is clear, I did wonder about it while I was reading the earlier chapters.

***12. Do you have any additional comments? Please share them in your review.***  
 No further comments.

**III. SPECIFIC OBSERVATIONS ON DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Page	Paragraph/Line	Comment
		No comments

**IV. SPECIFIC OBSERVATIONS ON APPENDICES TO THE DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Appendix	Page/Row	Paragraph/Line/Column	Comment
			No comments

**Reviewer #2**

**Peer Review Comments on FDA's Draft Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA:**

HRF Model Report, Underlying Data, and Risk Scores

Reviewer #2

---

**I. GENERAL IMPRESSIONS**

The FDA has done an admirable job in building a science-based model for risk ranking and in developing a report that makes it accessible to a reader with no previous experience in risk ranking or modeling. The targeting of biological, chemical, and undeclared allergen hazards is appropriate and need not be expanded to include other potential hazards (e.g., physical hazards that are addressed in GMPs). The fifty Primary Commodities proposed appear to encompass all foods regulated by FDA, but because some high-risk and low-risk foods are combined in the Secondary Commodity groupings, some of these secondary groupings may require further refinement in the next iteration of the model. (For the current model, >1200 food pairs is enough. One has to stop somewhere.) The seven criteria incorporate the six factors required by FSMA. The scoring and the underlying data appear to be accurate for each of the criteria. However, because low risk ranking the finfish:methyl mercury pair provided in the report is inconsistent with previous strong FDA advisories on methyl mercury in fish, the finfish:methyl mercury pair and its underlying data should be further scrutinized. This might be done in the broader context of discussing the issues associated with sensitive populations and the scoring of chronic versus acute illnesses. By the end of the report, readers have a clear understanding of the risk ranking process and the procedures which implement it. However, the beginning of the document is not user friendly; the incorporation of the relatively few editorial suggestions given in Section III of this review would help address this issue. A glossary would also be very helpful. The conclusions of each document follow from the underlying data. In summary, FDA has done an impressive job of tackling the very complex task of developing a semi-quantitative method which will allow foods to be regulated according to their risk as required by FSMA.

**II. RESPONSE TO CHARGE QUESTIONS**

1. *In order to apply the FSMA factors it is necessary to first take into account both the characteristics of foods and known or reasonably foreseeable hazards, i.e., food-hazard pairs. The food categorization scheme involves a list of 1,286 food-hazard pairs (candidates) and 335 foods (secondary commodities) linked to approximately 50 food categories (primary commodities).*
  - a. *Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?*

It appears that all FDA-regulated food products can be placed in this classification scheme.

- b. *Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?*

As noted in the public comments, many “risky” and “nonrisky” food products are aggregated into the same food-hazard pair. This is problematic and should be addressed in the next iteration of the model.

There is one example of aggregating that is of particular concern; treating finfish as a single group for the secondary commodity, secondary food-hazard combination “finfish:methyl mercury.” A secondary commodity, “small fish” would not present a methyl mercury hazard in this category, but a secondary commodity “large fish”, e.g., swordfish, tile fish, king mackerel and shark), (i.e., a “large fish”:methyl mercury pair) present such a large risk that the FDA/EPA Joint Advisory (2004, 2014) states that pregnant women, nursing mothers, and young children “*should not eat*” swordfish, tile fish, king mackerel and shark. There should be intra-agency agreement regarding the risk of consuming large fish.

***c. Are the method and data used to identify the food-hazard pairs adequate?***

Yes, the method and data build from the ground up, starting with the primary data from the literature or other sources. This provides a sound scientific basis for all of the food-hazard pairs.

***d. If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.***

See comment 1b.

***2. Please evaluate the scoring definitions for the seven criteria:***

***a. Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.***

C1. The scoring is appropriate for biological, undeclared allergens, and chemical hazards.

C2. Table 2-1 is unclear. Are hospitalization rates and hazards being scored separately, or do hospitalization rates follow from the severity of the hazard, in which case they refer to the same thing? That is, severe hazards result in high hospitalization rates. It's difficult to imagine a severe hazard with a low hospitalization rate or a moderate hazard with a high hospitalization rate. Thus, I think the scoring double counts for the same characteristic.

C3. The scoring is appropriate if the percentiles refer to % of samples.

C4. The scoring is only for “the temperature at which the food is *intended* to be held and stored.” The potential for growth under abuse conditions should also be considered. The growth potential for e.g., potatoes salad would be high under abuse conditions, whereas foods with  $a_w < 0.85$ , pH  $< 4.6$  would have low growth potential. Use of three factors would, of course, require presentation as a cube rather than grid. This may be beyond the scope of the program.

C5. The scoring is appropriate.

C6a. The scoring for consumption is appropriate.

C6b. This may be off-topic but should be noted somewhere: increased consumption does not equate to increased risk. Consider “Food A” which has 50 cases and 100,000 units of consumption versus “Food B” which has 5 cases and 1,000 units of consumption. Under the present scheme, based on the number of cases without normalization for consumption, Food A would appear to have greater risk, when in fact Food B is ten-fold more risky. Because of this, the score for C6 should be weighted, perhaps as 7 rather than 10.

C7. The scoring is appropriate, as is a weighting of 5.

- b. For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).***

It is generally accepted that RTE foods present a greater risk than NRTE foods. This should be taken into account in the model. The suggested C3 weight of 15 for RTE food and 10 for NRTE foods is appropriate.

- 3. The draft model integrates a substantial amount of public health and commodity-specific data about food-hazard pairs, as well as information regarding manufacturing and processing of different foods.***
- a. Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.***

The data sources are comprehensive and appropriate.

- b. Are there data sources not yet used but that should be considered? If so, please provide specific examples of data source(s) for each criterion, and explain why the additional data sources might enhance the criteria scoring.***

The underlying data used for the FDA/EPA advisory on methyl mercury in fish would be relevant in determining the risk score for the finfish:methyl mercury pair.

- 4. Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.***

If “age” of the data is being used as a proxy for “quality” of data, weighting is not appropriate since the best science available may not be the most recent data. If, however, the weighting is meant to reflect the current situation in a progressive fashion (rather than retrospective) in keeping with the intent of the model, the weighting is appropriate.

- 5. Consumption data from the NHANES What We Eat in America database were used and, for certain commodities (e.g., ice) where no data or limited data were available, expert***

***opinion was used in scoring Criterion 6. Is the timeframe used (2011-2012 or 2009-2010 survey cycle) appropriate?***

Yes, the time frame is appropriate.

***a. Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.***

The use of highest consumption is appropriate since it will produce the most conservative model.

***6. Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?***

All food:botulism pairs were evaluated and compared, seeking to understand both the scoring of each pair and why the scoring of some pairs differed for some of criteria in a given food:risk pair. The data for each pair were consistent and the differences among them transparent.

The finfish:methyl mercury pair was also evaluated in greater depth, and for chemical hazards in general, the scoring was found to be more subjective. There are several issues that make the finfish:methyl mercury pair difficult to evaluate: 1. There is high risk for the sensitive population (pregnant women, nursing mothers, and young children), but low risk for the general population. 2. High-risk fish are aggregated with low-risk fish into a single category of “finfish.” 3. There is a huge data gap. There is compelling evidence that consumption of methyl mercury can lead to deficits in neurological and cognitive development. The PTWI has been determined to be 1.6 µg/kg body weight and a guideline for methyl mercury levels in fish set at 0.1-1.0 mg/kg (ref 47). However, the distribution of methyl mercury concentrations in fish is unknown, so there cannot be a direct connection between fish consumption and the PTWI.

n.b.

The FDA/EPA Joint Advisory on Methyl Mercury in Fish (2004 and 2014) states that pregnant women, nursing mothers, and young children “*should not eat*” swordfish, tile fish, king mackerel and shark. This implies a large risk even though the risk ranking score is low. There should be intra-agency agreement regarding the risk of consuming large fish. (A large volume of supporting data was provided to the FDA Food Advisory Committee in the 2004 advisory but only one (Choy, 2009) was cited in the 2014 document).

Although I was furnished with the documents which were cited as references for the underlying data, it would not be possible for an end-user to determine the “correctness” of the scoring if the references were not readily available. In many cases it is impossible to access the underlying data because the relevant documents are not available online (i.e., the paper had to be purchased if one was not a subscriber). This issue could be overcome if there were hot-links from the score to the reference and from the reference to the actual document. Some of references are incomplete, e.g., a reader may not know that JFP refers to the Journal of Food Protection. At the very least, the full citation, including the title of the document, should be provided in the reference list.

C1 is scored as a “1” based on refs 4, 42, 52, and 60, which as noted below, are inappropriate. Based on my knowledge and experience, a C1 score of 9 would be appropriate because there is

compelling evidence that consumption of methyl mercury causes adverse health effects in the U.S. There is wide consensus that consumption of methyl mercury can lead to deficits in neurological and cognitive development in sensitive populations.

C5 is scored as a “3” based on ref 6041, input from subject matter expert. This reference is difficult to evaluate, but C5 could be rated as a 9 because there is a high incidence of contamination and weak control. (Granted that the distribution of methyl mercury in fish is unknown and if available, data for the distribution in large fish could justify a score of 3).

The differences between risk scores generated by the program (190) and the alternate scoring proposed here (250) could be enough to move the finfish:methyl mercury pair from a low-risk to a high-risk food. Whatever the final risk score is, it should be consistent with the risk presented in the FDA/EPA joint advisory.

***a. Are there any underlying data not appropriately used?***

Re: the finfish:methyl mercury pair:

Ref 4 gives data for outbreaks, but not the chronic exposure data which are required for the scoring.

Ref 47 sets the PTWI as 1.6 µg/kg body weight, but does not relate this to consumption of fish. The report notes that if there were a known distribution of methyl mercury for the aggregated “finfish,” it would be dominated by species that do not have a high concentration of methyl mercury.

Ref 52 provides information similar to ref 47, but focuses on the neurodevelopmental effects. It does not provide the consumption or incidence data needed for the scoring.

Ref 60 grapples with the issue of risk v. benefits and sensitive populations. It concludes that for finfish consumption the benefit of polyunsaturated fats outweighs the risk from methyl mercury in women of childbearing age. However, this report aggregates high-risk and low-risk species of fish. This reference notes that FDA advises the sensitive population not to eat the four species of fish that are high in mercury.

At the end of the day, the scoring of both C1 and C5 of the finfish:methyl mercury pair might come down to expert opinions. If this is the case, the opinions which gave rise to the FDA/EPA joint advisory should be given considerable weight.

***b. Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.***

See note above.

***7. The draft model relies on expert elicitation to fill data gaps, in particular for the scoring of Criterion 5 and for the scoring of criteria for chemical hazards and undeclared allergens. Is the expert elicitation process (which involves external panels and FDA subject matter experts) adequate to address data gaps?***

Yes, the process is adequate, but it is unclear how heavily FDA subject matter experts (versus expert panels) are used and the number of subject experts involved for each determination. If, for

example, the scoring reflects the opinion of a single subject matter expert, that scoring might be open to question.

- a. ***Is the scoring logic used to combine data and expert opinions and the order of preference in section 4 (e.g., Figure 4.3) appropriate? If not, please explain what changes might be considered and why.***

The hierarchy is appropriate.

8. ***Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.***

Scaling down would present a slippery slope. If allergens are scaled, shouldn't the sensitive populations for *Listeria monocytogenes*, and methyl mercury also be scaled? Would it be possible to give two separate risk scores, one for the general population and one for susceptible populations, perhaps with an explanatory note for each case?

9. ***Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?***

If this question is asking where the line between high-risk and non-high-risk foods should be drawn, it's subjective and akin to scoring the cutoff number for a student's average % grade on an exam with specific letter grades. One might default to the traditional A =  $\geq 90\%$ , B = 80 -89%, C = 70-79% etc. But this may be inappropriate for a given class, and if the grade distribution curve is smooth, even these cut-offs would be arbitrary (i.e., is there really a difference between a student who gets an 89.9 and one who gets a 90?).

That having been said, there has to be some sound scientific basis for making the cut-off. One might consider making some % of all food-hazard pairs the cut-off (e.g., defining the top quadrille of risk scores as the cut-off). How this can be done remains to be determined, but clearly a cut-off score that places 90% of the food as high risk has too low a cut-off and one that places 90% of foods as low risk wouldn't be very useful either.

Perhaps the cut-off could be approached statistically by looking at the risk score distribution and finding the 95% (or 90%, or whatever) confidence interval (i.e., the risk ranking score that is significantly different from the population) and basing the cut-off on that.

10. ***How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?***

Once the model is in use, areas for improvement and new data will become apparent, so it would be useful to update the model after three years of usage.

11. ***Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.***



6.1 The risk ranking approach is clear. Figure 6-1 is especially useful.

**12. Do you have any additional comments? Please share them in your review.**

No additional comments, except for the editorial ones in section III. Many of these are very important for clarity in the presentation of the report.

**III. SPECIFIC OBSERVATIONS ON DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Page	Paragraph/Line	Comment
General		In general, all of the scoring grids should be presented in the same format. For example, scores are given within the grid in Figure 2-2, but as column labels in Table 2-1. Placing the scores consistently in the grid would have greater transparency and make them easier to use.
General		A glossary of technical terms used (e.g., criteria, factors, bin, risk, hazard, acute, chronic, foodborne disease, granularity, etc.) would greatly improve the report’s utility for the nonexpert.
12	33 and others	“Granularity” can be considered jargon. It might be helpful to define it here as, e.g., “level of detail.” The definition could go in the glossary.
8	20-29	How the 7 criteria relate to the 6 factors was unclear in the first reading. Figure 2.6 is a great explanation. Could it be moved up to follow the initial paragraphs on criteria and factors? The figure on Slide 33 from the webinar was very instructive in putting everything together and should be included in the report, perhaps after Figure 2.1.
C1	Figure 2.2	The figure would be clearer if the column headings were above rather than below the columns. As labeled, the difference between “frequency of outbreaks” “and occurrence of illness” is unclear. (Is this per year or over the whole period?) Does the latter refer to the number of cases since 1998 or the number of cases per outbreak?
Table 2.1		The rows should be labeled, e.g., “hospitalization data,” “severity data.” % relative to what should be defined. Presumably it’s % of cases. It is unclear whether score is based on “hospitalization data,” and/or, “severity data.”
Table 2.3		Same concerns as Table 2.1.
40		This refers to two methods of scoring. It was unclear what these are.
73	10-15	It would be useful to give a sample calculation here.
73	27	“PAG” is not included in list of abbreviations.
75	25	This paragraph is very difficult to follow, especially without having the model in hand. For example, line 35 refers to Risk Scores, Food Rank, and FRRS. It took several readings to understand that the Risk Score gives rise to the Food Rank, which is the FRRS. FRRS is not included in list of abbreviations.
85	1	This sentence needs a verb. Should it read “Appendix L <i>includes</i> the food risk score and ...”?
General		Biological hazards, chemical hazards, and undeclared hazards are not defined until page 27-28 of the report. They should be defined at the beginning of the report or in the glossary.

**IV. SPECIFIC OBSERVATIONS ON APPENDICES TO THE DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

<b>Appendix</b>	<b>Page/Row</b>	<b>Paragraph/Line/Column</b>	<b>Comment</b>
			None to report.

**Reviewer #3**

**Peer Review Comments on FDA's *Draft Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA*:  
HRF Model Report, Underlying Data, and Risk Scores**

Reviewer #3

---

**I. GENERAL IMPRESSIONS**

The aim of the document to determine the risk of food-hazard pairs is clear and as such it is a useful and valuable approach which could benefit public health authorities as well as other stakeholders of the food chain. Although in certain cases and for certain food-hazards pairs, the need for an assessment may be questioned and whether the fact that undeclared allergens appear systematically in any type of food is justified and also, generally speaking, the outcome provides this and as such the goal is achieved.

The document is not very clear in terms of language and structure – contains a lot of repeats between different sections/chapters and even within single paragraphs. While section 2 seems designed to provide the background and details to all criteria used in the model, often it is fragmentary and part of the information/explanations are provided much later in the document, e.g., certain calculations (equations are provided in section 2, others in section 4, in text or figures), aspects such as confidence, uncertainty appear in section 4, discussions on certain foods in section 4, while a clear understanding would be needed earlier on, e.g., cascade from primary, secondary and tertiary foods which makes it difficult to understand why certain foods are classified in the tertiary list (e.g., chicken sandwich) but other sandwich types are only kept at secondary level. Terminology is also not consistent or new terms are introduced later in the document. Clear definitions would be useful for example for terms such as cases, illness, outbreaks, sporadic cases – generally speaking a section for definitions and acronyms would be helpful.

There are quite a number of inconsistencies between Appendices – e.g., numbering of hazards in Annex B (in principle the master) is not always reflected in other appendices.

The rationale for assigning certain “chemicals” as “biological” due to growth leading to their formation is not applied consistently – mycotoxins are typically associated with growth and the step of the food chain is also a key factor in their formation but they are handled as “inert” chemicals.

It is also not completely clear what is included in the assessment, for example of C5 – which steps of the food chain are exactly considered is unclear as the text is very vague.

Weighing is not always understandable – C3 is about contamination of a food – i.e., an event occurring before consumption and is therefore not linked. Nevertheless in the model consideration is given to importance of the food in the diet with a weighing factor. However, for the determination of C6, the importance of the food in the diet is again used for the scoring, which is the appropriate way to consider the impact a contamination will have on consumers.

Section 4 – it is not clear why in any case the whole process is run for the different criteria. While from the text it seems that if information is available to determine a score, then the process stops, whether additional information exists or not. In the case of the figures and examples – it seems however that the whole process needs to be run, e.g., even if available quantitative data are sufficient to score “0”, scoring process continues. In other words – processes do not seem to follow a Y/N decision tree, with an end at the level where the score should be considered as substantiated.

## II. RESPONSE TO CHARGE QUESTIONS

**1. In order to apply the FSMA factors it is necessary to first take into account both the characteristics of foods and known or reasonably foreseeable hazards, i.e., food-hazard pairs. The food categorization scheme involves a list of 1,286 food-hazard pairs (candidates) and 335 foods (secondary commodities) linked to approximately 50 food categories (primary commodities).**

**a. Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?**

At the level of primary commodities it is helpful to categorize broad food categories allowing to fine tune at secondary and then at tertiary level.

**b. Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?**

A clear scheme or diagram on the three levels would be helpful to understand – the challenge is to determine on how far a food-hazard pair is valid applicable for the higher (or lower) level – for example what is valid for a particular cheese might be valid for others, but not for cream; what is valid for infant formula is not valid for milk powder due to the different sensitivity of consumers; what is valid for a particular food is valid for others with a very similar ecology and manufacturing conditions, although not involved so far in outbreaks.

The food classification does not always appear appropriate and the difference between a secondary and a tertiary commodity is not always understandable.

Example: Secondary – Acidified vegetables and beans: these are in my opinion two different levels. If beans are not considered vegetables, then a secondary commodity would be “pulses”, encompassing more than beans. Aside of acidified vegetables and beans there is also a category canned vegetables and fruits – seems that they are overlaps making it difficult to either identify a food-hazard pair or to “forget” some in case a food is considered by individuals in one of the three possible options.

Another example – shelf-stable milk as secondary commodity to LACF (assuming this means low acid canned food) considering that there are other types of shelf-stable milk such as condensed milk (usually also canned) which is however part of “Dairy” as first commodity.

Frozen vegetables as a secondary commodity, frozen vegetables (beans, potatoes) as well – in my opinion beans and potatoes are subsets and sometimes this gives the impression that the assignment to a category of commodity is driven by the hazard (in this example different hazards for the three).

Salads – certain of the categories seem to describe products at a tertiary level – what is the difference between a mixed RTE salad and specific ones such as taco salad, fish salad...why not chicken salad, meat salad, tomato salad. Avocados are listed under dressings – why not salads? Most salads are “mixed” – and if there is a specificity then this would fit better at tertiary level. As an analogical example – sandwich is dealt with as secondary commodity, apparently irrespective of the type of sandwich.

Grouping completely different forms at secondary level, e.g., dry and liquid gravies does not seem to allow for an appropriate assessment of certain of the criteria such as 4 and 5. Hence, it would probably be more appropriate to differentiate between dry and wet (low and high water activity products), also to be able to take into consideration elements pertaining to the rating of individual criteria, in this case “supporting growth or not”. By having a single commodity – which one should be used to answer these questions.

***c. Are the method and data used to identify the food-hazard pairs adequate?***

To some extent it seems driven by publications which do not use the same classification of foods or do not precisely describe the food. Example – *Cronobacter* spp. (different from *Cronobacter sakazakii* as per annex B is listed for a few commodities – with the exception of infant formula, none of them has ever been involved in an outbreak. If the purpose is to list all commodities for which isolation of *Cronobacter* spp. has been described, then the list of commodities is far from being complete.

In the report mention is made on the inclusion of “potential” hazards – from my understanding, the possibility to assign a hazard to a commodity based on expert knowledge. One element could be the microbial ecology, manufacturing processes, behavior of microorganisms – while this option seems to exist – it does not seem to have been applied systematically: for example 1249 identifies *L. monocytogenes* as microbial hazard, 1251 *Salmonella*. The products are probably very similar by nature, process, etc. – why should there be a difference in type of contamination ...even if not published? *C. botulinum* and ETEC are hazards for Tofu but not for Tofu products – does not seem consistent in terms of ecology and with the assessment of the process contamination (Tofu as raw material for tofu products).

The same question could be raised for other categories – e.g., seafood with completely different microbial hazards for octopus, squids, although quite similar in terms of origin, processing etc., *C. botulinum* and histamine in smoked finfish but not in dried.

***d. If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.***

Not sure there is a “magic” way to establish for the classification other than establishing rules on what needs to be considered which includes as well a cross-check between categories – for example if a hazard is identified in a commodity, then it must be assessed whether another commodity including this has a similar profile – see Tofu example.

**2. Please evaluate the scoring definitions for the seven criteria:**

- a. Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.**

The rationale for biological toxins is not completely clear and consistent – on one side some microbiological toxins are missing, such as staphylococcal enterotoxins which can be present in the absence of *S. aureus* (die-off), on the other side the element “ability of growth” if used to classify metabolites such as histamine. For mycotoxins however, where elements such as ability to grow, the step at which contamination takes place, and shelf-life are important to the formation of mycotoxins, it is considered as “inert” as would be arsenic or lead.

In terms of allergens, 8 are considered and it is understood that “undeclared allergens” is considered a “worst case” that covers all of them, i.e., without taking into account whether it is likely to occur or not (e.g., fish in dairy products...). It is however not clear why hazard sub-types 91–95 have been defined, what about eggs, soybean, tree nuts, for which no sub-types exist. What is the rationale for this difference?

- b. For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).**

The likelihood of contamination is, in my opinion, not dependent on whether it's an RTE or NRTE food but is the outcome of food supply chain and the control measures applied during the manufacture up to consumption. Hence the data gathered reflect the percent contamination rate and should already reflect these differences. Criteria 4 and in particular 5 are also linked to the differentiation of the two categories. Different weighing would also require to take into consideration the likelihood of misuse – i.e., how likely is it that a NRTE food is not cooked before consumption – should it then be weighed as a RTE food?

**3. The draft model integrates a substantial amount of public health and commodity-specific data about food-hazard pairs, as well as information regarding manufacturing and processing of different foods.**

- a. Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.**

In principle all the sources of information are appropriate – the only question is whether, taking the amount of documents, publications, etc., everything has been taken into account. From a few examples it seems that there could be a bias – e.g., *Cronobacter sakazakii* which has been isolated from numerous commodities but only involved in outbreaks with infant formula. In terms of criteria, there is probably much more information available than what is indicated – one example is ICMSF Book 8 which discussed the relative role and impact of contamination along the food chain. This does not seem to have been taken into consideration – compared to other ICMSF books. At least it is not mentioned in chapter 2.2.5 or 4.2.5.

It seems that only U.S. data have been used for outbreaks (appendix D1) while other references are listed in appendix E. Some product and product categories have an extensive listing of data – while for others, relevant outbreaks (not in the U.S.) do not seem to have been taken into consideration. But it is very difficult to judge from appendix E what has been included (sometimes no title) and what not as the list is valid for all criteria, includes reports based or including themselves on reviews and compilations. What can be commented is that for certain publications or books, probably much more references are used.

- b. Are there data sources not yet used but that should be considered? If so, please provide specific examples of data source(s) for each criterion, and explain why the additional data sources might enhance the criteria scoring.***

See comments on C5.

- 4. Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.***

One can argue that outbreak is outbreak, whether it occurred today or in the past. The weighing is based on an assumption that control measures were not as effective in the past which is probably questionable for quite a number of food commodities – e.g., control measures for numerous food commodities are much older than 1998 – e.g., heat processes such as pasteurization, sterilization, hygiene control measures related to environmental contamination have been described, or included in guidance documents, e.g., for chocolate or dairy products before 1998.

However if this is accepted as assumption – then the question is whether or how far this is not taken care in the determination of C5.

- 5. Consumption data from the NHANES What We Eat in America database were used and, for certain commodities (e.g., ice) where no data or limited data were available, expert opinion was used in scoring Criterion 6. Is the timeframe used (2011-2012 or 2009-2010 survey cycle) appropriate?***

Yes.

- a. Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.***

Seems Ok – could possibly (if not already done) be compared with the approach applied in FAO/WHO risk assessments to ensure consistency on the way it's done.



**6. Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?**

Scones – why are microbial pathogens included? As indicated somewhere else *Cronobacter* spp. are ubiquitous, and finding it in any type of commodity is not too difficult – relevance is however very limited.

Beverage bases – relevance of *C. botulinum* is questionable, no reference on why it's included (even if rated 0).

Coffee – virus – without more details on what type of coffee, it is not very useful – liquid coffee, roasted coffee?? Same hold true for *Salmonella* – “0” but nevertheless included and no reference.

Tea – *Salmonella*, there have been 2-3 outbreaks in Europe (herbal tea) which do not seem to have been considered

Chocolate – *Salmonella*, several outbreaks (e.g., Germany) and recalls (UK, Germany). On the other hand never heard about STEC.

Confections – *Cronobacter* spp., see earlier comments – not relevant.

Dried milk – *Cronobacter* spp., irrelevant as only infants sensitive. On the other hand *Salmonella* outbreak occurred in France 2005.

Frozen fruits (berries) – Several outbreaks – at least one with several hundreds of cases – in Europe.

Dry instant breakfast – *Cronobacter* spp. Same comment as above.

Infant formula – *Cronobacter* spp. Strange to see a C1 = 0 considering that WHO/FAO classifies this a high-risk food and that a number of outbreaks have involved more than one baby. Same comment for *Salmonella* – there have been several outbreaks since 1998 but they do not seem to have been taken into consideration.

**a. Are there any underlying data not appropriately used?**

[Reviewer did not comment.]

**b. Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.**

[Reviewer did not comment.]

**7. The draft model relies on expert elicitation to fill data gaps, in particular for the scoring of Criterion 5 and for the scoring of criteria for chemical hazards and undeclared allergens. Is the expert elicitation process (which involves external panels and FDA subject matter experts) adequate to address data gaps?**

If experts are chosen correctly, then this is certainly an appropriate procedure. What would need however to be captured is the rationale for certain decisions, rating. In some of the examples (chapter 4) it is often only mentioned that this was done based on expert elicitation – the rationale is however not given, also not when a calculated score is changed.

- a. Is the scoring logic used to combine data and expert opinions and the order of preference in section 4 (e.g., Figure 4.3) appropriate? If not, please explain what changes might be considered and why.**

What does not look very logical is that the process from top to the bottom in the different processes illustrated in figures chapter 4 is not stopped as soon as there is a quantitative “zero” with a high confidence – e.g., no outbreaks ever registered, no recalls, no contamination. Otherwise the process seems quite heavy – e.g., to assess C4, C5 for foods which have never been implicated in outbreaks, e.g., scones and *Cronobacter* spp. It does somehow not fit with other provisions, for example the fact that a food is not or only marginally consumed.

- 8. Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.**

It seems inappropriate to rate a pathogen for a whole population but which has a specific “target”, e.g., *Cronobacter* only causes severe disease in infants below 6 months – which will not be consuming some of the foods which have been associated to, such as chocolate, scones, foods for adults consuming this type of food severity is “0”...and they would probably also not be consuming formula. In such a case the score 9 should only be used for that particular population.

In terms of undeclared allergens – one could approach it in a similar way and consider only the fraction of the population which is susceptible. This population is however probably less well defined than in the above example. Considering that for undeclared allergens the worst case scenario is taken, i.e., (a) just “presence” and not a threshold below which it would no longer represent a severe hazard, and (b) significant percentage of mortality in case of presence, the risk is 9 and weighing in terms of consumer population is not really going to contribute to a change in the rating.

If a weighing is to be considered, then this would, in my opinion, be better done with respect to the likelihood of finding an undeclared allergen above a certain threshold in a certain food – i.e., high likelihood to find a dairy ingredient or residues in a culinary preparation or nuts in confectionery, than finding shellfish in milk. The fact that no differentiation is made between allergens, excludes however this option. A possibility of weighing is C5 – which however would imply a very good knowledge on processing lines and processes (One could even argue on why for certain commodities “undeclared allergens” are not mentioned – as in annex C; for example butter and buttermilk, while it is included for cream and cheese, both used to manufacture these. A further example is nuts, where undeclared allergens is found for some but not for others or wrongly undeclared allergens other than nuts in the case of hummus which is (at least in the rest of the world) manufactured with chick peas and not nuts and contains other ingredients.

- 9. Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?**

To have a cross-check to see how far it makes sense and whether there is a bias due for example to the structure primary, secondary, and tertiary commodity, impact of a single report,

contamination not related directly with the food as could happen in a home (see comments under IV for annex C). What would also be important is to check for consistency – it seems illogical to have hazards associated with certain commodities but not with others which contain the first as an ingredient – or to make distinctions such as mycotoxins (overall) for apple and apple products and specifically patulin for apple juice concentrate; acrylamide in peanuts but no longer in peanut butter – unless there is a clear rationale. A further example is nuts, where undeclared allergens are found for some but not for others or wrongly undeclared allergens other than nuts in the case of hummus which is (at least in the rest of the world) manufactured with chick peas and not nuts and contains other ingredients.

***10. How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?***

It is difficult to give an answer but this should probably be based on changes of criteria which could evolve such as C1. C6 may be the most critical as in today's world eating habits may change rapidly to follow new social trends and globalization (see comment on C6, page 19 in the specific observations).

***11. Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.***

Overall the purpose of the report is apparent after repeated reading. However, aside of more technical question as the previous charges, the report would, as discussed above as well in the following section III, greatly benefit from serious editorial review, to streamline the structure, avoiding numerous repetitions, grouping of related text currently spread across different sections/chapters, introduction of new terms without explaining them (e.g., score used in relation with criteria and then later for uncertainty/confidence – is it really a score?), one equation in section 2, one in section 4 and another one in one of the figures + others somewhere else. The figures in chapter are very difficult to read and not always aligned with text (or contain new elements such as equations not explained before).

Simple elements such as sub-sections on the individual hazards would help better understanding of “what is what” without having to read several times the text, list of definitions, list of acronyms, etc., changes in terminology. Expert conclusions not always understandable – e.g., in the examples...”has been decided or changed because of expert ...” does not help in transparency and understanding of the rationale.

It is also quite cumbersome to find associated information across the numerous annexes and excel-spread sheets.

***12. Do you have any additional comments? Please share them in your review.***

[Reviewer did not comment.]

**III. SPECIFIC OBSERVATIONS ON DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Page	Paragraph/Line	Comment
8	25/26 (iii)	The term processing step is probably more appropriate than point of manufacturing process
8	27 (iv)	Normally one refers to control or preventive measures rather than “steps taken”, in addition “step” is used in (iii) with a different meaning.
8	27 (iv)	The reference to “manufacturing process” does not seem to be aligned with further sections dealing with the subject – e.g., 2.2.5 talking about “the entire food supply chain” which encompasses more than the manufacturing process.
9	1	What are “ <u>the</u> statutory factors”?
9	14	What is RTI – acronyms should be spelled out, at least the first time it’s used. Valid throughout the document. A table summarizing them would be useful.
9	34	What is meant by “additional analysis” – information, data?
10	5	Since animal food/feed is not covered – should this category not be deleted from Appendix A nos. 3 and 4, or at least a comment provided?
11	19	“...and adapted it to account...” to which model does this refer?
12	11	See section Expert Elicitations below...where below?
12	37	An annex with secondary commodities would be helpful
13	14	Assume you are referring to factors listed on page 8 – cross-reference would be helpful
13	22	Section 2.2.5 speaks of “manufacturing process” as one of the elements of the whole supply chain – should Criterion 5 not also reflect this?
13	26	Different for undeclared allergens which are not mentioned here?
14	11	A numerical value <b>from 0 to 9</b> – this is not really correct as only agreed upon values are used, 0, 1, 3 and 9 and not any value between these two limits.
14	16	Since this section covers the three hazards – it would certainly help the reader to have a further sub-division: 2.2.1.1, 2.2.1.2. and 2.2.1.3 for the individual hazards.
14	23	What about an anaphylactic shock caused by undeclared allergens such as peanuts- is this not part of the acute effects?
15	2 / 3	Included or include? In the rest of the document undeclared allergens are handled separately (text, comments, examples), why then saying they are included in chemical hazards. Should the “hazards” not be described/defined up front as they are generally valid and not just for 2.2.1
15	Figure	Should read low <b>tens</b>
15	9	“food hazard pair” – is this valid for all three hazards or just micro (see my comment on a further page)
15	12	“Sporadic cases” – how is this defined?
15	26	What does “including the outbreak itself” mean?
15	27	Why is “timing of outbreak data” defined by (C1)?
15	27	Is C1 definition correct – not rather scoring?
15	6, 11-13	Are marine biotoxins such as algal toxins not chemical hazards – why not include them after row 24 which refers broadly to “chemical hazards”?

Page	Paragraph/Line	Comment
		See comment page 14/Line 16 on sub-division – this would help avoiding confusions.
16	8 / 9	Fig 2.2 “Have the potential to be involved – what does this mean? Unclear since, as understood, Figure 2.2 shows “real” outbreaks. How could potential be classified as low, medium or high?
16	15	Why e.g., undeclared allergens and chemicals – does this imply also microbiological hazards which are the only ones also included in the model?
16	26	Reference is made in this part on “U.S. food consumption” – does this mean this aspect is only valid for chemical hazards, not for other hazards? Does this also apply to the data used – only U.S. data for chemical contaminants? Is this not an element which is anyways considered for C6?
17	5 – 10	It is ICMSF (2002) – since the chemical hazards are mycotoxins and not any other “traditional” chemical hazard (e.g., arsenic used in later examples, it would be good to specify to avoid confusion).
17	15	Also here it is confusing as here the chemical hazards referred to are not of microbial origin (which include mycotoxin).
17	Tables 2-1	Title –hazard – all of them? 1 <sup>st</sup> column/3 <sup>rd</sup> row should probably read “not a hazard” to be consistent with the rest of the row. 2 <sup>nd</sup> column/3 <sup>rd</sup> row – with little or <b>no</b> medical 3 <sup>rd</sup> column/3 <sup>rd</sup> row – is “not life threatening” aligned with 0.5% mortality?
18	Table 2-2	1 <sup>st</sup> column/3 <sup>rd</sup> row Should probably read “not a hazard” to be consistent with the rest of the row. Since the descriptions of the hazards are identical to those in Table 2-1 – why making a difference? The text itself is more descriptive of acute symptoms short duration, “self-limiting” effects, little medical intervention, life threatening – does not sound like effects of chronic exposure leading at some stage to effects. What is IQ reduction considering that sequelae are infrequent?
19	5	Is food supply system – the food supply chain as described under 2.2.5? Suggest to use same terminology throughout report.
19	21	Unclear why the geographical origin is used to weight as this chapter deals with the likelihood of presence in a finished product (irrespective of where it’s produced) and not with the exposure of the consumer. The fact that a food is not consumed in the U.S. or only in limited quantities is considered under 2.2.6. What if the consumption of a certain food is increasing in the U.S. or completely new – does it mean you will have to change C3 and C6? Why then not consider outbreaks related to those foods which have happened outside of the U.S. differently than those within the U.S.?
19	36	Is a scoring bin a definition?
19/20	35 – 40/1 – 8	Suggest to move discussion around what is a RTE/NRTE to a separate section –e.g., along with discussion on food categories. It is not specific to C3 and it also detracts from the content of the section – likelihood of contamination.

Page	Paragraph/Line	Comment
20	Table 2-3	<p>1<sup>st</sup> column/1<sup>st</sup> row – why not just “0” since n (positive samples) = 0 or no reported occurrence? “Known” sounds “qualitative”.</p> <p>1<sup>st</sup> column/2<sup>nd</sup> row – what are “indicators” – defined or explained somewhere before?</p> <p>2<sup>nd</sup> column/1<sup>st</sup> row – is <math>\leq 0.1\%</math> which includes 0 correct?</p> <p>3<sup>rd</sup> column/1<sup>st</sup> row – should probably read <math>&gt;0.1</math>.</p> <p>3<sup>rd</sup> column/1<sup>st</sup> row – should probably read <math>&gt;1</math>.</p> <p>How are the two rows connected? Is it one <b>and</b> the other? For example what if 0.1% occurrence but no recalls?</p>
21	1 – 6	Why is the same parameter “includes conflicting studies....” valid to assign to both Moderate and Low?
21	21	C5 refers to manufacturing process, while the text below explains that this is only one of the elements. C5 is, according to this text the entire food supply chain – from the farm to the consumer (as per line 27). Believe this needs clarification.
21	24	All three hazards included here- would be good to specify here to avoid confusion.
21/22	28 / 1-6	Would certainly be clearer if sentences on C5 are grouped and not mixed with C3. C3 is in principle the outcome of C5.
22	16	<p>According to the introductory part of 2.2.5, manufacturing is just one element – wording should be consistent as it is also a possibility to establish the contribution of individual steps in the food chain, e.g., before and after a kill step.</p> <p>As a consequence what is assessed in terms of contamination probability – the occurrence in finished products, of raw materials, intermediate product or processing environments leading to contamination of product? Believe it would be important to understand the rationale to understand Figure 2-4, as to some extent it seems that it is very much focused on finished product and hence not very different to C3.</p>
22	18/19	<p>Moderate: Sporadic; Low: Infrequent detection of contamination – what is the difference?</p> <p>What means “contamination introduced post manufacturing” – post-kill step, during the distribution, during preparation by consumer? Why should this type of contamination be of a low probability?</p>
22	25/26	<p>Control measure is probably a more usual term in food processing than “step”.</p> <p>Would Effective not be more accurate than “strong”?</p> <p>Is the assessment and rating done for a whole industry or for individual branches or factories?</p>
22	34	Seems that discussion around manufacturing should be moved up front at the beginning of this section for a more logical structure.
23	1	What are activities? Unclear why the physical location is not taken into account as it is an essential part of the hygiene control measures – e.g., related to post-process (kill) step contamination.
23	Figure 2-4	<p>What is the contamination potential – same as probability used before?</p> <p>Scoring not completely clear – if control measures are effective, why is the contamination potential rated 3 for “high”, which would rather correspond to “not completely effective”? The same for the combination weak/low – in the case of post-kill step contamination even low levels of, for example</p>

Page	Paragraph/Line	Comment
		<i>Salmonella</i> , will invariably lead over time to a high probability of contamination.
23	5	What does this mean – negative results for all elements contributing to contamination, e.g., raw materials, processing environment etc.? It would seem more appropriate to have the first column scored 1, the second scored 3 and the last scored 9. ICMSF Volumes 6 and 8 provide estimates on the relative impact of different steps in the food chain in terms of contamination – does not seem to have been taken into consideration.
24	10/11	What is distinct from C1 and C2?
24	15	“Where appropriate....” – according to Figure 4-7 it’s done in any case...
24	16/17	For other criteria details (e.g., values) are provided on factors to weight scores as well as calculations (Equation) – why not in this section and only in Figure 4-7?
24	Table 2-5	How are the qualitative ratings established, lower or higher than what?
25	8	Should the rationale to decide on “reasonably foreseeable hazards” not be explained somewhere? E.g., same ecology, same behavior, same sources....
28	3	The rationale for dealing with histamine is clear – why was not the same rationale used for mycotoxins whose formation is also dependent on growth of molds and therefore elements considered for bacterial growth equally valid?
31	5	Should probably read scored instead of defined
31	14	Term attribute introduced here – along with indicator, underlying indicator (line 19), data indicator in Table 4-2. None of them seems to have been defined and explained in the introductory part of the document.
32	Table 4-2	Title – what is the difference between ranking criteria and model criteria as per 2.2? Is this table not a summary of elements used to determine individual criteria 1 to 7? 5 <sup>th</sup> column/1 <sup>st</sup> and 3 <sup>rd</sup> row: would probably be useful to qualify the chemical hazard (acute, chronic) which are handled differently. 4 <sup>th</sup> column/4 <sup>th</sup> – 6 <sup>th</sup> row – also here difference acute/chronic +Likelihood of contamination” – “Average number/year – seems to be the number of recalls or reports per year – hence an absolute number and not an average. C4 – Is it not time rather than days as for certain products shelf life can be months or even years?
35	Figure 4-1	First step – why not just referring to Table 4-1, rather citing a few elements of info (not complete – rationale)? Second box/right – 2. Is only if several data sets are available. Third box/right – why introducing this table here and not page 15 where the subject is dealt with in details? Fourth box/right – discrepancies between table and text as text page 15/16 do not speak about “experts” just “0” or does not refer to “the occurrence of illnesses includes outbreak-associated cases <b>only</b> ” – Figure 4.1 does not seem to be aligned.
37	2	Why e.g., in brackets? – This gives the impression there are others. However considering that only few hazards are dealt with, would it not be clearer just to cite the ones for which the statement is valid? – Otherwise there is ambiguity.

Page	Paragraph/Line	Comment
37	11	Why not introducing and explaining the concept of confidence and uncertainty scores (is it really a score not a level?), e.g., in chapter 2?
37	35	Why not using the example Cantaloupe- <i>Salmonella</i> as for the other criteria?
39	28	Estimates of estimates – or should it read CDC reports?
40	5	Any particular reason to write – including the definitions for microbiological and chemical hazards?
41	Figure 4-2	First boxes right – examples of references of the boxes point to chemical/allergen hazards for the “Quantitative” one and to micro for the “Qualitative”, while text on page 40/line 4 – 17 says something different. 2 <sup>nd</sup> box/left – should it not read “the most instead of “more”?”
42	24	Wrong reference – ICMSF (2002) does not provide any information on arsenic.
43	6	Same comment – Book 7 does not address allergens.
44	Figure 4-3	2 <sup>nd</sup> box/right – why referring to ICMSF 2002 considering that Books 6 and 8 are much better references for information on contamination? 4 <sup>th</sup> boxes/right – should probably read “study weight” 6 <sup>th</sup> box/right – Equation shows $n_p$ text below $n$ . Why does this equation appear here and not earlier in the text as well $n_p$ as in the case of Equation 1?
45	Figure 4-3	Several indexes not explained 1 <sup>st</sup> box/right For score = should it not read 0%, and then $>0 - <1\%$ ? For score 3 it should probably read $>0.1$ as the 0.1% is already included in score 2. Are the qualitative qualifiers needed since quantitative % are given? 2 <sup>nd</sup> box /left – if weighted prevalence is 0, what is the purpose for moving it to the next step? Does it mean a calculated and weighted can be changed further down? 3 <sup>rd</sup> box/right – should score 1 not read 1 – 5, considering that score 0 is “0” and $<5$ is then undefined (could also be 0.....)? 5 <sup>th</sup> box/left – would eLEXNET not be used in the previous step by an expert anyways?
46	27	Why is this equation (probably not 2, considering the one on weighted positive samples) not introduced and explained earlier – chapter 2?
48	11	Unclear – what if they have a quantitative “0”? Should then the scoring not stop at this level.
48	15	Unclear – according to Figure 4-3 it’s a number of reports per year and not an average (not sure anyway to understand how one can do an average of reports per year)?
48	34	Not sure to understand why eLEXNET is needed if recall $>0$
48	2	According to Figure 4-3 expert opinion comes before eLEXNET. Unclear why a process is continued for scores determined as score “0”



Page	Paragraph/Line	Comment
48/49	Example C3.1	It is said that if no “quantitative”, then RFR – in this example there is however a quantitative prevalence, hence a score of 3 based on the 0.246% -What is the aim - to continue with the process with other elements? Seems that whether data are available or not the process goes on... In terms of example, it would certainly be more illustrative to have one where certain elements are missing and would really require to go to a next step.
49	34/35	Seems to be inverted – according to Figure 4.3 RFR comes before recall – but also here there is a weighted positive rate, so why pursuing the process? Seems that is principle of the two options – response available = “stop at this level” and response not available = “continue to next level” is not applied.
51	9	Does seem as if Figure 4-3 is not followed – in principle with a positive rate of > 1% then the score should be 9 (as indicated in line 27 of previous page). Does C3 = 3, now mean that scoring is changed afterwards...or the rationale is not aligned with the example? Clarification would be needed.
51	Example C3.3	Same comment – according to Figure 4-3 RFR is considered first: C = 1. What is the purpose of looking at recalls which provides a C = 3...which is then ignored? Or is the purpose to determine confidence/uncertainty? If this is the case, then this should be explained better in the text.
52	13	Growth potential – has a reference rate been established to assess growth potential, e.g., expressed as minimum log increase to differentiate between foods supporting or not growth? This would also provide experts a common reference and avoid different interpretations on what “potential for growth” means.
52	16	Microbial or bacterial pathogen – two different things and putting one in brackets does not help a lot.
52	17	Also undeclared allergens do not grow
52	17 – 21	Unclear – the shelf-life will have an impact on the levels reached if growth is supported. The term the “extent” seems not appropriate – also because in the bracket there is a need to explain further and then again in another sentence (line 23). To be logical the determination on whether a food can support growth comes first and then the shelf-life which will determine the levels. In this text the shelf-life seems to be considered first...at least according to the sequence of the discussion and numerous repetitions of the same. These two indicators – which one, as the text before is quite general?
52	22	Unclear what is meant by taking into account the point where contamination takes place. How does this relate to potential for growth and shelf-life and how is this taken into consideration?
52	30	What about molds which can grow at lower water activities? – The potential for growth and associated potential mycotoxins formation does not seem to have been considered in this model as mycotoxins are handled as “pure chemicals” unlike histamine, enterotoxins.
52	34	Unclear – microbial hazards is wider than bacterial hazards, e.g., molds. Is a pathogen not a hazard per definition?
53	Figure 4-4	Why assigning a shelf-life for foods which do not support growth?

Page	Paragraph/Line	Comment
		4 <sup>th</sup> box/right – is it here necessary to indicate days which are only an example for a specific commodity? Why not providing more guidance (introductory part) on how to proceed with other commodities? The table does not include foods not supporting growth, while they seem to have to undergo the whole process.
54	Example C4.1	Is it whole cantaloupe, fresh cut? Why assigning confidence/uncertainty levels for both elements – not as in Figure 4-3? What means moderate – not exceeding a certain level?
55	Examples C4.2 & C.4.3	Why considering shelf-life since there is anyways no potential for growth – or were other aspects of arsenic considered which would modify toxicity over time? Same comment applies for allergens – is it not a bit of a waste of time to look for shelf-life in this case. While the confidence /uncertainty is 9/1 for both elements, the overall is 1/9. What is the rationale for this change?
57	7	Book 8 ICMSF provides lot of information of contamination in the food chain – does not seem to have been used.
58	Figure 4-5	3 <sup>rd</sup> box/right – the wording seems very quantitative, while most is based on a qualitative assessment.
59	Example C5.1	The scoring will very much depend on the type of product – whole fruit, fresh-cut. Since there are several “1s” in the table – which one is it? Unclear how two scores of “1”, both on expert advice can then give a final score of “3”, also on expert advice (different ones?). – Would be good to explain the rationale. Same for example C5.2 the final score of 1 is just a fact – what is the rationale for different expert judgements as this is what is expect in a section entitled “rationale behind scoring”?
60	Example C5.3	Not sure I understand why a high score of “9” has a very low confidence level – this seems quite contradictory. These two elements would certainly benefit from more explanations in the text as to their determination, role, etc.
62	Figure 4-6	4 <sup>th</sup> and 5 <sup>th</sup> box/right there is overlap for scores 0 and 1 in terms of %, since the 1 in 1-5% belongs also to $\leq 1$ of score 0. Should probably read $> 1-5\%$ .
66	Figure 4-7	2 <sup>nd</sup> box/right – Equation is missing on page 24, the location where one would expect explanations on calculations.
67	3	Cases – illnesses, outbreaks, sporadic cases? Would be helpful to have definitions somewhere clarifying terminologies used throughout the document (and which are not always used consistently).
75	28	(See section below) – where below, there is no further section in this chapter – it would be clearer if cross-references within the report – see above or see below would be identified with the relevant reference (e.g., see 2.2.3).

**IV. SPECIFIC OBSERVATIONS ON APPENDICES TO THE DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Appendix	Page/Row	Paragraph/Line/Column	Comment
A	1/no 1		LACF – abbreviation should be explained somewhere.
A	1/no 1		The definition” of baby is quite vague. Infant is normally defined as 0 – 12 months. Junior does also not seem to be an official definition – to what age does this correspond – 1 – 6 years?
A	1/no.2		What is N.E.C?
A	1/ nos. 3 & 4		See comments page 10 of main document
A	1/ no.7		What means “Waters” – flavored water? If the meaning is mineral or bottled water, then this would be quite a different category from soft-drinks.
A	1/ no. 9		Liquid coffee and tea or dry?
A	1/ no. 17		Does this include all types of products: raw, pasteurized, ESL, UHT?
A	1/ no. 25		Are these prepared dishes or also frozen vegetables
A	1/ no. 28		Are simulated meats an example? – If the case, then “e.g.,” would be appropriate as there are other example such as soya or bean flours.
A	2 /no. 37		Are refrigerated RTE salads an example of refrigerated RTE foods such as delis, cold cuts refrigerated pasta or prepared dishes etc., or are those included in NEC (whatever it means)? What is the rationale to single out salads?
A	2/ no. 45		What is “soup” – dry, refrigerated, canned? No secondary category found.
B	1	<i>Aeromonas</i>	To our knowledge it is <i>Aeromonas hydrophila</i> and should be specifically mentioned to be consistent with other pathogens such as <i>L. monocytogenes</i> , not just <i>Listeria</i> .
B	1	Hazards 4, 20, 21, 22, 23, 34	Hazards are listed but have not been considered further – e.g., absent in appendix D2.
B	1	Hazard 14	<i>Enterococcus faecalis</i> – absent in appendix D2, but ETEC listed as hazard 14.
B	1	Hazard 18	Norovirus – but in appendix D2, it’s <i>Listeria monocytogenes</i> .
B	1	Hazard 19	Parasite – but absent in appendix D2 and becomes Norovirus.  These are three examples but there are more where there is a discrepancy between Annex B

Appendix	Page/Row	Paragraph/Line/Column	Comment
			and others. – In principle B should be the master and the others aligned.
C	1	15	Wonder why <i>Cryptosporidium parvum</i> appears – have tried to find information in the numerous spreadsheets and annexes to understand– considering the type of product, preparation, etc. Believe it is this type of example which would benefit from a review to decide whether it is really beneficial for public health to perform the whole risk ranking – while beans are probably just a vehicle and not a systematic carrier of this parasite.
C	2	23	Same problem with <i>Listeria</i> in canned foods – realistic or not, what is the cause? Poor hygiene in handling after opening but not a systematic problem.
C	5	112	Muffins and <i>norovirus</i> – is muffin really the issue and not, for example one of the ingredients such as berries? The assessment would certainly be more beneficial if a specific ingredient could be targeted – would also allow improvement – rather than “all muffins” irrespective of their recipe. If a specific ingredient is identified as high risk, this would allow to focus preventive measures to that one which may also be used in other products.
H	1		The text indicates that food/feed for animals is out of the scope – no. 1- 3 refer pet food. If this is due to the occurrence of human cases, then these are not the only outbreaks. Should be clarified in the text.
File 09b (I2)			File name does not match content – comments on Criteria rather than info on control measures.
L		Undeclared allergens	Rating, e.g., of C1 does not seem to be consistent – e.g., condiments is scored 1, while ingredients which are used to manufacture them such as flavorings, flour, spices are scored higher. Bakery products with scores of 1 up to 9 for C1 and C5 – basically manufactured with the same ingredients, frequently on the same lines and some being ingredients of others.
L		<i>Cronobacter</i>	C1 rated as “0” including for infant formula despite around 50 outbreaks and sporadic cases since 1998. The C2 score corresponds to the sensitivity of infants up 6 months, while at this age they will never consume chocolate. Does C6 correspond to the infant population

Appendix	Page/Row	Paragraph/Line/Column	Comment
			which is consuming 100% of the products, C6 for coatings – not relevant for infants?
L		Beverages	<i>C. botulinum</i> is mentioned as C2 = 9 and C4 = 0, hence no growth and therefore C2 is questionable as toxin formation depends on growth (only exception is infant botulism, but then it would be, if at all, a very defined category of beverages). The same is valid for some commodities rated as C4 = 0 – no growth = no toxin. The only exception is honey causing infant botulism.
L		<i>S. aureus</i>	Similar problem – microorganism as such is not causing any illness, only its toxin. This, however requires growth and in several cases there is a discrepancy between scores – e.g., scores of 1 or 3 for C1, C3 but C4 of 0 or 1 or C5 is wrongly scored.
L		<i>B. cereus</i>	Not sure why C3 was rated 0 and C4 as 1 – number of outbreaks are associated with cooked rice and growth is needed to permit toxin formation.

**Reviewer #4**

**Peer Review Comments on FDA's Draft Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA:**  
HRF Model Report, Underlying Data, and Risk Scores

Reviewer #4

---

**I. GENERAL IMPRESSIONS**

The ultimate aim of this model/tool appears to be to identify foods under USFDA oversight that represent a 'high' risk of causing food-borne illness and which, consequently, may warrant extra record keeping, e.g., to facilitate identification of outbreaks and hasten recalls of implicated products to minimize public health risk. As the model authors indicate in the report, they do not attempt to identify the 'cut-off' in the risk ranking that would identify foods that *do* require the additional record keeping.

There are potentially two main approaches to identifying and ranking the relative risk of foodborne illness from specific foods. One is to rely on the available epidemiological evidence for foodborne illness to identify "problem" foods and hazards. There is a large amount of relevant data, however, not all foodborne diseases are reported, or even required to be reported. As such, this empirical approach alone is not sufficiently robust to meet the objective.

A second approach is based on synthesis of relevant knowledge about factors, and their interactions, that influence risks of foodborne illness. To do so requires, however, requires characterization of the kind and magnitude of those factors for each product, process, intended consumer population, intended end-use, etc. Additionally, it would be ideal to be able to identify all hazards that are, or might be, associated with that product. Using some form of formalized risk assessment, these data could be used to estimate relative risks from different foods, processes, hazards, etc. This approach, being more fundamentally based on an *understanding* of risk-affecting factors than the empirical epidemiological data, might also be applied to estimate risks from foods that represent new formulations and processes, and for which there are currently no epidemiological data or, conversely, a log record of safety. This approach, however, could be used proactively to *predict* relative risk, rather than a reactive approach based on epidemiological data.

There are, however, a large number of variables that influence food safety risk, even on a relative scale. Assessing all of these factors for each food-hazard pair creates a significant challenge to the achievement of the ambition of risk ranking. Many of these variables have potential to profoundly affect the risk from a single category of product, e.g., the addition, or not, of preservative; the time and temperature of storage of the product; whether the product is consistently manufactured hygienically and under GMP; the intended consumer group and whether some groups are predisposed to specific hazards that may be present in the product. In many cases data needed to estimate risk will not be available and the data can change over time, e.g., as new technologies and products are introduced to the market, when processes or suppliers change, and if products or ingredients begin to be sourced from regions outside of the USFDA's regulatory oversight.

Thus, such (inferential) ranking schemes are likely to involve compromises and assumptions that potentially lead to over- or under-estimation of the relative risks from specific products, under specific processes, for specific populations. The challenge is to produce a decision tool that is simple enough to be workable but that is scientifically/logically robust enough to achieve the correct relative risk for most products and circumstances, but also to be able to recognize when the answer is inappropriate and to recognize that special considerations, not explicitly considered in the model/decision tool, might sometimes have to be applied to achieve a credible relative risk ranking. (Recognition and accommodation of such ‘anomalies’ should also help to improve the logic of the model and its robustness).

The two approaches described above are not mutually exclusive and each has its own strengths and weaknesses. **However, it seems reasonable that, if a risk model were correctly formulated, it should produce predictions consistent with the observed epidemiological data (at least where such data are available). If so, it would give confidence that the predictions of the model for other products that have not (yet) been linked with food-borne illness outbreaks would also be reliable.**

Upon reading the report and working through the model and its output what seemed to be missing were ‘reality checks’ to test whether the model, and the data it is based on, produce relative risks estimates that are close to, or at least consistent with, what is observed from the available epidemiological data. However, it is noted that the development of the model has received oversight, comment and insight from industry and scientific experts, and that there has been public consultation enabling stakeholders to share their knowledge, insights and concerns about the relative risks from relevant product:hazard pairs. If after those opportunities the stakeholders agree that the model is sufficiently credible for its intended use, there may be no need for further ‘reality checks’, i.e., if consensus/acceptance has been achieved.

In looking closely at some of the data it also seems there are some inconsistencies in the approaches taken and decisions made, and some potentially flawed logic in the scoring scheme, particularly because some of the criteria are confounded, e.g., as suggested above, the epidemiological data would be expected to be a reflection of other risk-influencing factors (Criteria 3 – 6), so that effectively a “double-counting” of some factors is occurring in the logic of the model. It is also noted, however, that the model is based on an EFSA food risk model which used a similar approach of including both epidemiological data and data on products and processes to characterize risk.

Nonetheless for reason discussed above, it seems that it would be a valuable experiment to remove the epidemiological data from the scoring, and use it instead to assess the reliability of the relative risk estimates of the risk assessment model based on Criteria 2 – 7 only, for example. Many minor typographical and grammatical errors were also evident, which are described in the accompanying reports (Sections III and IV on the model and report).

Principle 6 of the Codex Alimentarius Guidelines<sup>1</sup> for the conduct of microbial food safety risk assessment, of which the report under review is an example, states that: “Any constraints that impact on the Risk Assessment such as cost, resources or time, should be identified and their

---

<sup>1</sup> Codex Alimentarius Commission (1999). Principles and Guidelines for the Conduct of Microbiological Risk Assessment CAC/GL-30 (1999)



possible consequences described”. Given some of the potential shortcomings described it may be appropriate to include a relevant statement concerning potential constraints in the report itself. Detailed comments against the Charge Questions are presented below and elaborate on comments made above.

## II. RESPONSE TO CHARGE QUESTIONS

**1. *In order to apply the FSMA factors it is necessary to first take into account both the characteristics of foods and known or reasonably foreseeable hazards, i.e., food-hazard pairs. The food categorization scheme involves a list of 1,286 food-hazard pairs (candidates) and 335 foods (secondary commodities) linked to approximately 50 food categories (primary commodities).***

**a. *Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?***

The list of foods seems appropriate having been based on commodity groups that fall within USFDA jurisdiction. It might be useful however to accentuate foods that do not fall under USFDA regulatory oversight, e.g., the reasons for not considering red meats and chicken products (risks of STECs, *Salmonella* and *Campylobacter*) as high-risk food:product pairs. The further refinement of the main categories into secondary categories is also appropriate because the 49 primary commodities groups are very broad and encompass products that would be expected to represent vastly different risk due to differences in product formulation, processing, packaging, shelf-life, intended end-use, that affect potential for pathogen growth in particular, but also likelihood of contamination.

(*n.b.*, Table A-1 still lists pet foods/feed even though the text says they were excluded from consideration).

**b. *Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?***

As noted above, risk assessments are often a compromise between i) the desire to include all relevant details and differences to be able to account for all important variability, and ii) the availability of relevant data to support those distinctions in the risk assessment. In risk assessment terminology this might be considered the paradox of disaggregation.

The food groups selected seem to be “fine” enough to enable identification and differentiation of relevant classes of product, particularly as they relate to potential for pathogen growth, although details of packaging and storage conditions (as well as shelf life) might also be relevant for correct characterization of potential for growth. The other consideration, however, is the availability of relevant data to support assessment of relative risk of each of those sub-divisions.

From the text of the report it seems that the consumption data from NHANES did support the level of ‘granularity’ adopted in the model. However, the data to assess relative outbreak incidence and hazard prevalence seem less complete. From Appendix 4b (Prevalence data) only ~220 of the 1286 nominated product:hazard pairs appear to have published (refereed or otherwise) data to support the prevalence estimates needed. Similarly, for outbreak data, only

~411 of the 1286 product:hazard pairs have data to support the criterion value selected, the other pairs presumably relying on 'expert opinion'.

Of the Criterion 4 values that required definition, i.e., for which estimation of growth potential is required, 629 product:hazard pairs involved bacterial pathogens. To support those estimates, 597 discrete sources of information are cited, with between 1 and 8, and an average of ~3, sources per product:hazard pair, for a total of 1990 citations across all 629 product:hazard pairs. Notably, of the primary references provided, 450 of the product:hazard pairs relied on 'Expert opinion (IFT/RTI expert elicitation)' as one, or the main, basis of the estimated growth potential. Of the 629 product:hazard pairs with potential for microbial growth, 83 had only one supporting reference for the growth potential "decision", and for all but five of those 83 the source of information was one or other source of 'expert opinion'. The most cited sources are shown in Table 1, below.

**Table 1. Sources of Information for Classification of Growth Potential in Food-hazard Pairs (Criterion 4)**

Ref. No.	Times cited	Details
10	450	Expert opinion (IFT/RTI expert elicitation)
434	165	NSW (2008) Potentially hazardous foods: Foods that require temperature control for safety. NSW/FA/CP016/0810
6035	51	FDA expert opinion, June 2015
6041	46	FDA SME August 2014
454	27	AEM 39(5):943-949 (1980) Cameron <i>et al.</i>
435	26	Betts <i>et al.</i> (2006) Scientific Review of the Microbiological Risks Associated with Reductions in Fat and Added Sugar in Foods
2181	19	FDA (2001) Evaluation and Definition of Potentially Hazardous Foods - Chapter 3. Factors that Influence Microbial Growth
2187	19	FDA. Quantitative Assessment of Relative Risk to Public Health from Foodborne <i>Listeria monocytogenes</i> Among Selected Categories of Ready-to-Eat Foods. Appendix 8: Growth of <i>Listeria monocytogenes</i> in Foods. <a href="http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf">http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf</a>

It is noted that the above analysis (Table 1) doesn't seem to accord closely with the text at Section 4.2.4 in the document, e.g., the apparently strong reliance on NSW Food Authority's (NSWFA) 'Potentially Hazardous Foods' document (P.52, Lines 3 to 13) in which NSWFA isn't mentioned. Also, it's clear that there is a strong reliance on three sets of expert opinions. Thus, it seems that many of the 'growth potential' classifications were based on expert opinion but, without scrutinizing each published reference (not feasible within the time commitment), it is not possible to comment on whether the additional references cited support the expert opinion. However, in this reviewer's experience some of the growth potential scores appear 'unrepresentative' of expected growth potential in some foods. (This topic is discussed again later in this report).

Given the above discussion, it is difficult to provide expert comment on whether the data available on growth potential support the level of granularity of food categories but the large number of references cited suggest that it *should* be. More importantly (and also elaborated later in this report, *see response to '3b'*) growth potential could have been estimated using one or more of the on-line predictive microbiology databases, e.g., ComBase ([www.combase.cc](http://www.combase.cc)), which includes tens of thousands of observations on microbial growth rates in foods, the USDA

Pathogen Modelling Program (<http://pmp.errc.ars.usda.gov/PMPOnline.aspx>), or the French SymPrevious database ([www.symprevious.net/](http://www.symprevious.net/)), etc.

*c. Are the method and data used to identify the food-hazard pairs adequate?*

As I understand it, the food-hazard pairs were predominantly based on food outbreak records from USA (i.e., the food-hazard pair has been demonstrated to actually cause illness), but also other surveys and reports that identified the presence of hazards *whether or not there was any evidence of adverse health outcomes*. This includes recall data and various surveys. The authors of the report/model indicate that they also consulted other databases (e.g., EFSA) to identify further relevant product:hazard pairs.

Upon examination of Appendix D, (outbreak data), it is apparent that there are ~411 discrete product:hazard pairs based on USA outbreak data. The model, however, considers 1286 food-hazard pairs and it is not clear how the additional 875 product:pathogen pairs were derived. Greater transparency (documentation) around the process of identifying product:hazard pairs seems required, not the least because some of the food-hazard pairs are somewhat ‘surprising’ selections.

By way of comparison, it could be argued that a holistic approach be taken and that for all 258 primary and secondary commodities, all 95 hazards should be considered, leading to some ~24,500 potential food-hazard pairs. Clearly this approach was not taken, and neither do I think it would be necessary. However, it accentuates that the model developers considered many product:hazard pairs to be irrelevant/trivial and eliminated them from further consideration but do not explain how (with the exception of the 411 product:hazard pairs specifically associated with outbreaks) only 1286 product:hazard pairs of the ~24,500 potential pairs were selected for inclusion in the ranking process. This question become more important because of the issue around how to rank product with multiple hazards. A product that has more hazards associated with in the list of product:hazard pair, irrespective of how trivial the actual risk, will have a higher risk ranking because any product:hazard pair in the list will have a risk score greater than 1. This inherent bias needs to be considered in the combination of scores, and will be discussed in greater detail in the complementary report on the model.

Given that the potential list (i.e., of all possible combinations) was ‘culled’, it seemed relevant to examine whether there were any inconsistencies in the database, e.g., similar products but with different hazards considered to be associated with them. Scanning through the lists in Appendix C it was apparent that there are such inconsistencies. While my search through the list was not at all exhaustive, some examples are presented below:

- i. Differences in hazard considered relevant to ice or to bottled water. Both products should involve potable water so its not clear why a range of protozoan parasites are considered relevant hazards in ice, but not in bottled water. If the argument is that bottled water receives a more severe treatment, then why would *Salmonella* be included as a hazard in bottled water and not a hazard in ice? Alternatively, parasites are more likely to be eliminated by freezing than are bacterial pathogens, i.e., ice might be expected to represent less chance of exposure to viable protozoan parasites. Also, the list of chemical hazards considered relevant to bottled water apparently does not apply to ice. This seems illogical.
- ii. A long list of vegetative bacterial pathogens are considered as hazards in Grade A, pasteurized, fluid white milk. The aim of pasteurization is to eliminate bacterial pathogens,

something which pasteurization does overwhelmingly well and has done so (as the records indicate) for about the last 100 years. Similarly, while *L. monocytogenes* is considered a relevant hazard in Grade A, pasteurized, fluid white milk it apparently is not considered relevant in flavored milk. Based on the composition and preparation of these products, they represent the same potential for contamination and potential growth, so this difference in hazard lists for such similar products requires explanation.

iii. Inclusion of *Cyclospora cayatensis* as a hazard in Gravies (Dry and liquid). This seems to require explanation because it is not listed as a hazard in other products of similar composition and processing.

iv. *Listeria monocytogenes* in dried egg. I'm not aware that *Listeria* spp. survive desiccation unusually well, nor that dried egg would be used in such a way as to allow *L. monocytogenes* to grow to high enough levels to have a significant likelihood of causing human illness, even in the severely immunocompromised. Some explanation/justification for this pair would be helpful.

v. Norovirus as a hazard in game meats. This seems to require explanation particularly as game meats will usually be cooked before eating. Norovirus on game meats could only arise from external contamination and so would be on the outside of the cuts of meats and would be expected to be easily eliminated during cooking.

vi. Dried pasta. Dried pasta is normally boiled for 5 – 15 minutes before consumption and would be expected to eliminate non-spore forming pathogens such as *Salmonella*, or STECs or *S. aureus*. Is the hazard that *S. aureus* might have grown and produced toxins during manufacture of dried pasta? The inclusion of these bacterial hazards in dried pasta requires some explanation.

vii. *Cyclospora cayatensis* as a credible hazard in pasta salads (673). If it is a hazard in pasta salads, why are other food-borne protozoan parasites such as *Cryptosporidium* or *Giardia*, etc. not included?

viii. STECs in RTE deli salads. Why are STECs not considered to be hazards in RTE deli salads (690-691) when they are considered to be a hazard in other fresh cut vegetable products?

ix. *Salmonella* in hot smoked finfish (992) also seems a very unlikely scenario (hot smoking should eliminate them) and this product:hazard pair doesn't appear in the outbreaks list (Appendix D).

x. *Listeria monocytogenes* risk from popcorn (1114) seems negligible, given that popping of corn would eliminate it and that pre-popped corn won't support its growth. Furthermore, this product:hazard pair doesn't appear in Appendix D. Some explanation/justification seems to be required.

While many hazards have been detected in foods, it doesn't always mean that there was any significant risk to consumers, and many reports have been published on detection of hazards in foods without any corresponding evaluation or even consideration of risk. In other words, some survey data are not risk-related and their relevance as indicators of risk and to this risk-ranking tool needs to be more closely scrutinized.

Similarly, in the outbreak data, the source of contamination does not always seem to have been clearly articulated. It is noted that the report states that the outbreak data were included only if food, rather than food handlers, were considered as the source of the contamination. This aspect might also be further elaborated in the main text.

For the scoring to work properly, and particularly in connection with Criterion 2, Criterion 3 (and the data used to support) cannot simply be about prevalence (i.e., simple detection of the hazard in the product) but would have to relate to detection of the hazards at levels approaching that required to produce symptoms in a significant proportion (e.g., >1%?) of those exposed so that the scoring of Criterion 2, about disease severity, is meaningful. This consideration also impinges on Question 8, and is discussed further in the response to that Question.

A more robust search of outbreak data from other nations, particularly nations with analogous life-style and food cultures, would probably yield further valuable information. For example, there have been numerous outbreaks of Hepatitis A in Europe (and more recently Australia) linked to frozen berry fruits but, while strawberries can be contaminated with HepA (and it is included in the list) many other frozen berries can be as well but are not included in the list (or was Strawberries:Hep A included as a representative?) Also, it is noted that the EFSA outbreak data were also explored and used to build the list of product:hazard pairs but the only reference for HepA and strawberries is the somewhat obscure “FDA (2015) Orange Book Database, Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations. <http://www.accessdata.fda.gov/scripts/cder/ob/>”, and in which I could find no reference to Hepatitis A in strawberries! Conversely, Maunula *et al.* (2014, *Int. J. Food Microbiol.*, 167(2)177-185) does provide a good introduction to enteric virus problems in berry fruits in Europe and is listed in the references as REF 3016.

- d. If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.***

As noted earlier, there is likely no perfect method for identification of product:hazard pairs and the approach presented seems reasonable, but can still be improved. Alternative approaches might involve a “Failure Mode Effects Analysis”, starting with each of the 49 FDA-regulated commodity groups and asking the question: “is it possible for any foods in any category to cause human food-borne illness and, if so, how could that occur?” This would naturally lead to identification of sub-divisions of the main categories into categories of product that were more likely (whether due to natural hazard occurrence in the food, differences in processing, potential for microbial growth, etc.) to be sources of human illness. Any such approach would also have to include assignment of (relative) probabilities to the potential fault modes as part of the means of identifying the most relevant product:hazard pairs.

Additive scoring schemes, even if based on exponential scales, cannot capture the full complexity of factors that contribute to food-borne risk. As a simple example, a listericidal process in a hermetically sealed product effectively eliminates the risk from that hazard in that product, irrespective of what occurred before or what will happen later to that product, as long as package integrity is preserved. Equally, proper cooking can completely eliminate many hazards. The current scoring scheme does not enable a “reset to zero” for cidal processes, as could be achieved with a multiplicative scoring scheme that includes ‘zero’ values for some of the inputs.

Importantly, (and as noted above) as no simple scoring scheme is likely to correctly rank all product:hazard pairs and will of necessity be a compromise, it will be important to achieve consensus among stakeholders that they accept the risk scoring decision tool as being the best, or

at least as good as any other approach. Allowing stakeholder comment and input and requests for clarification of the basis of the scoring scheme and its outputs, and researching and documenting those answers, will be critically important in achieving that consensus and also lead to refinement and improvement of the food categories and identification of important factors that discriminate relative risk by improving the logical robustness and broad applicability of the model.

**2. Please evaluate the scoring definitions for the seven criteria:**

- a. Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.**

The hazards are appropriately described for microbial, chemical and allergen hazards. Some of the chemical hazards listed are of microbial origin (algal and fungal toxins), but are usually already present in the food (prior to processing) and would not increase in level during processing, distribution, etc. Histamine levels can increase as microbial metabolism continues, and this difference is appropriately identified in the approach developed, and dealt with by considering histamine as a microbial hazard.

- b. For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).**

I presume that the intent of this question is to ask whether NRTE foods would be expected to receive a heat treatment prior to consumption, i.e., that treatments received by NRTE foods would greatly reduce the risk to consumers if the hazard were present. If so, the risk reduction would be orders of magnitude in an NRTE for infective microbiological hazards (see also comments under Question 1d), compared to RTE foods, and that should be explicitly 'factored in' in the scoring.

Allergens and chemical hazards would be unaffected by cooking, but so too would histamine, currently listed as a microbiological hazard, as would the risk from *S. aureus*, or *B. cereus*, etc., because both produce heat-stable toxins as they grow in foods, i.e., cooking for these hazards does not lessen the consumer risk. As such, the adjustment to the scoring scheme that would be required to accommodate this issue is rather more complex than a simple weighting for RTE *cf.* NRTE could accurately reflect. It's likely that a new "Criterion" might be necessary to include the importance of cooking as a terminal risk reduction step. Such a Criterion might need to include negative values in the scoring scheme (i.e., indicating that the risk is greatly reduced in some cooked foods), and would relate principally to microbial hazards for which the mode of illness is gastrointestinal infection or those few cases (e.g., *C. botulinum*) of microbial growth producing a heat labile toxin, or other chemical hazards that are heat labile.

**3. The draft model integrates a substantial amount of public health and commodity-specific data about food-hazard pairs, as well as information regarding manufacturing and processing of different foods.**

**a. Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.**

The generic data sources (i.e., rather than specific documents or research publications, which are also appropriate) are appropriate sources for the data required. In general, where ideal data can not be identified and surrogate data are used, the reasons for those choices are identified and the process is transparent (i.e., well-documented). In some cases greater specificity/clarity about the data could be presented. For example, in the identification of pathogen:product pairs, I had to read closely to determine whether any pairs identified (apart from the 411 derived from outbreak data) were selected simply because some-one/organization had done a screen for those hazards in a range of products and found the hazard, i.e., whether there was any reason to believe that those product:hazard pairs had, or were ever likely to, cause human foodborne illness. Also, the severity data used for Criterion 2 could be handled in the same way as suggested by Minor *et al.* (2015), as was used in Criterion 7, (i.e., the *mean severity* of illness was calculated based on the weighted mean of i) the estimated unreported cases, ii) cases seeking medical attention, iii) deaths, etc.). This is discussed further below.

**b. Are there data sources not yet used but that should be considered? If so, please provide specific examples of data source(s) for each criterion, and explain why the additional data sources might enhance the criteria scoring.**

As noted in response to Question 1b, through analysis of the data sources used (e.g., assessing the product:pathogen pair evidence) it seemed that a wider range of credible sources was used than was identified in the text of the main report. Also, as noted in response to Question 1b, predictive microbiology databases (ComBase, Pathogen modelling program, others) probably should have been used to assess pathogen growth potential in foods, whether by matching foods to those in the databases, or by deriving composition data (e.g., pH,  $a_w$ , organic acids levels, etc.), for foods of interest and using that to generate growth rate predictions, which can in turn be used to estimate potential pathogen growth within the typical packaging (vacuum packed, aerobic, enriched CO<sub>2</sub>, modified-atmosphere packed, and storage environment (temperature, gaseous atmosphere), etc.

If required, further information on consumption of specific products could be obtained by inference from supermarket sales figures and the market share that that business controls, or for major producer's production figures in conjunction with information on their market share.

**4. Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.**

In some cases it is relevant to discount older studies, such as when it is known that industry/regulators have taken measures to reduce the incidence of a specific foodborne illness by new methods and interventions, i.e., that might be expected to have resulted in a sustained lower prevalence of specific hazards. This would be true for *Listeria monocytogenes* in many

RTE foods in USA, where incidence of illness and prevalence in relevant products has probably declined by 30 – 50% in the last 10 – 15 years due to greater awareness and risk management by regulators and industry. Equally, there have been strong regulatory programs and industry actions in USA to reduce rates of STEC illness, and *Salmonella*, from red meats, or produce, etc.

Similarly, as chemical detection methodologies have improved (i.e., their detection sensitivity has increased) the apparent prevalence (i.e., ‘detections’, regardless of contamination level) of chemical hazards in foods has probably increased even though there has been no actual change to consumer risk.

As another example, the advent of PCR and even newer pathogen detection technologies based on specific chemical marker (e.g., DNA, or RNA, infra-red spectroscopy) has enable much greater throughput of samples, but not always with confirmation that positive signals came from live (and infective) organisms, *cf.* molecular residues from non-viable (‘dead’) cells. As such, caution must be exercised when collating older data to generate information relevant to the current situation, and weighting of different data sources, whether temporally or by geographic regions, may be necessary and useful. However, in the absence of specific evidence that older literature on product:hazard pairs is no longer relevant, more justification is required for weighting of different subsets of the data. Importantly, the weighting applied needs to be determined on a hazard-by-hazard basis: for reasons discussed above, appropriate weightings are likely to differ for different hazards.

- 5. Consumption data from the NHANES What We Eat in America database were used and, for certain commodities (e.g., ice) where no data or limited data were available, expert opinion was used in scoring Criterion 6. Is the timeframe used (2011-2012 or 2009-2010 survey cycle) appropriate?**
- a. Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.**

The NHANES data is reported in the document to enable consumption of all 385 commodity groups to be estimated. The report states that because Day 1 and Day 2 estimates of consumption were not always equal, the highest daily rate estimated was used, rather than the mean rate. While the differences are probably trivial in most cases, the text of the document states that: *“It should be mentioned that in addition to calculating consumption rates based on the sum of a commodity, the use of the mean and maximum rates were also explored. After reviewing these methods, using the sum of consumption rates was chosen here as the most appropriate”*, but without clear explanation of *why* it was considered the most appropriate approach. Given the size of the samples, the average seems a more reliable and logical way of synthesizing the available data to make the most representative assessment. While the difference between the mean and maximum of the two day’s values would be expected to be small in most cases (given the size of the samples), it is not possible to independently verify that expectation from the data presented in the C6 results spreadsheet. Again, justification/explanation of the use of the maximum (or sum?) and more evaluation of the effects of that decision are required for transparency in the scoring tool and its documentation.



6. **Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?**
- a. **Are there any underlying data not appropriately used?**
  - b. **Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.**

A number of product:hazard pairs were selected for detailed evaluation and comment. These were:

- i. *Bacillus cereus* in pasteurized milk
- ii. *Listeria monocytogenes* in processed (cooked) meats
- iii. *Clostridium botulinum* in canned fruits and vegetables
- iv. *Listeria monocytogenes* in soft surface-ripened cheeses
- v. *Listeria monocytogenes* in cold smoked fish
- vi. STEC *E. coli* in bagged leafy green salads
- vii. Hepatitis A in Frozen Berry Fruits (472)

The evaluations are described below.

- i. *B. cereus* in pasteurized milk

*B. cereus* is a spore-forming bacterium, and pasteurization cannot eliminate it from milk. Growth and toxin-production to high levels is possible during the shelf life of pasteurized milk. While outbreaks due to *B. cereus* in milk are infrequent they have been reported, and numerous studies have demonstrated potential for growth and toxin production in milk within the normal shelf life (see e.g., Christiansson *et al.* (1989), *Appl. Env. Microbiol.*, 55(10): 2595-2600; Te Giffel *et al.*, (1996), *Int. J. Food Microbiol.*, 34(3):307-18; Notermans *et al.*, (1997), *Food Microbiol.*, 14:143–151).

**This product:hazard pair was not identified in Criterion 1 and is not considered in the model or report. Its absence requires some explanation.**

- ii. *Listeria monocytogenes* in processed (cooked) meats

*L. monocytogenes* in processed meats was identified as the main cause of Listeriosis (>50% of cases) in USA by the USDA/FDA/CDC (2004) risk ranking of RTE foods associated with Listeriosis.

**This product:hazard pair was not identified in Criterion 1**, and is not considered in the model or report. I assume that the reason is because processed meats are not within the USFDA jurisdiction, but it might be useful to state that overtly (earlier) in the report.

- iii. *Clostridium botulinum* in canned fruits and vegetables  
*Criterion 1*

It is well documented that canned fruits and vegetables (low acid) have been causes of botulism. In USA this has often been attributed to home bottling. The Criterion 1 results spreadsheet, however, cites CSPI (2011; REF 1338) and The Orange Book (REF 1339) as the basis for inclusion of this product:hazard pair. I could find nothing about *C. botulinum* in REF 1339, and the CSPI outbreak database links cited are no longer fully functional. From the 2007 outbreak list, however, there was one reference to botulism involving a hot chili, sauce, and another

involving canned meat, seafood, or pet food. **Neither of these references seem appropriate to support the criterion value assigned.**

*Criterion 2*

The ranking of *C. botulinum* as '9' is appropriate, and accords with the severity estimates of Minor *et al.* (2015).

*Criterion 3*

A single reference to a French study is used to estimate prevalence/likelihood of contamination. Importantly, the prevalence reported in that study was in raw product (ingredients), not finished (heat treated) product (*see* Methods and Materials S2.1, p. 264, of Sevenier *et al.*). Because of the potential confounding of responses to Criterion 3 with responses to Criterion 6, it is stated in the report (p.56, L22/23) that contamination at Criterion 3 relates to contamination in the finished product. **Accordingly, the prevalence reported in REF 211 (IJFM 155(3):263-8 (2012) Sevenier *et al.*) is not relevant or appropriate to support the prevalence estimates assigned for Criterion 3 for this product:hazard pair.**

*Criterion 4*

Growth potential is assessed as 1 (i.e., "low") citing REF 10 (expert opinion) and others (e.g., NSWFA, Betts *et al.*). These seem to be credible and authoritative references. However, the inclusion of REFS 453, 454, and 455 seems spurious. REF 453 (Severnier *et al.*) doesn't seem to consider growth in finished product, REF 454 is about pH and heat resistance in thermal processing (not growth) and REF 455 is a survey of pH and water activity in acidified bottled vegetables. The category being considered is LOW ACID products, not acidified product. **Accordingly, the interpretation of the data cited seems to require reconsideration.**

*Criterion 5*

The values presented seem credible for retorted product, but are based on expert opinion. However, the long history of safety of canned foods supports this evaluation.

*Criterion 6*

The consumption criterion score (9) is appropriate and accords with the estimates from NHANES.

*Criterion 7*

Economic impact per case was estimated by Minor *et al.* (2015) which ranked the per case cost of botulism as the 3<sup>rd</sup> highest of the foodborne hazards they considered. This probably arises from the high fatality rate and complete life support required, potentially for months or years, for survivors of botulism intoxication. **Importantly, in the C7 spreadsheet the cost per botulism illness that is reported to be based on the Minor *et al.* (2015) paper is misquoted. The value should be \$1,514,289 but in the C7 spreadsheet is given as \$15142, i.e., a factor of 100 too low.**

iv. *Listeria monocytogenes* in soft surface-ripened cheeses

*Criterion 1*

The outbreak evidence for listeriosis from soft surface-ripened cheeses in USA is 3 outbreaks involving 44 cases. This rate seems surprisingly low in comparison to other nations, but may arise from a relatively low consumption of these types of cheese in USA. However, the NHANES data suggest that ~10% of the population consume such cheeses every day. It might

have been expected that more recalls had occurred, as recalls for this product:hazard combination are relatively common in other Western nations. There are numerous outbreaks recorded internationally, including a 2013/2014 outbreak in Australia that lead to >20 cases and three deaths including one miscarriage. Nonetheless, based on the USA outbreak data and the scoring criteria, the value applied is correct.

#### Criterion 2

The ranking of listeriosis as a '9' for severity of illness is appropriate for susceptible populations (which potentially comprise 15-20% of the population) and which, on average, are ~100x more likely than a typical (non-pregnant) healthy adult to develop a systemic infection.

Many studies and reviews report that 20 to 30% of systemic listeriosis cases result in death of the patient. The severity ranking also accords with Minor *et al.* (2015). However, the 'death rate' cited in the C2 Results spreadsheet (15.9%) seems a factor of 2 too low. On investigation and re-reading the Scallan *et al.* paper (Scallan *et al.* (2011). *Emerg Infect Dis.* 17(1): 16-22) which was cited in support of 15.9% - it seems that that value is a misinterpretation. As stated in Scallan *et al.* (see footnote to Table 3, p.12) "These rates [i.e., presented in the table] were doubled to adjust for underdiagnosis before being applied to estimate the total number of hospitalizations and deaths".

#### Criterion 3

The prevalence score seems appropriate based on this reviewer's experience and reading of relevant literature. However, there appear to be some errors among the references cited. Specifically, REF 3004 has the wrong title in the MASTER REFERENCES spreadsheet. It should be "*Incidence of Salmonella, Listeria monocytogenes.... in two types of Mexican soft cheese*". Given the title, and that there is a separate product category for Mexican soft cheese, **the inclusion of this reference to calculate the prevalence seems inappropriate. Also, I could not locate REF 6006 from the information given in the MASTER REFS spreadsheet. Also, details for REFS 6008, 6009, 6011, and 6012 are incomplete.**

#### Criterion 4

Growth potential/shelf life values also accord with expectations. However, there are far better references available to support the conclusions made here, including numerous papers that model the rate of *L. monocytogenes* in various products, or summarize observations. Predictive microbiology databases would also have been more reliable. In particular, the predictive model of Mejlholm and Dalgaard and colleagues (Mejlholm *et al.*, 2010, *Int J. Food Microbiol.*, 141:137-150; Mejlholm and Dalgaard, 2009. *J. Food Protect.*, 72, 2132–2143) has been demonstrated to provide very reliable predictions of the growth rate and growth limits of *L. monocytogenes* in a wide variety of products, including cheeses. **The references cited are not authoritative sources for the evaluation and scoring of this criterion. REFS 4107 and 4108 are incompletely described in the MASTER REFS spreadsheet, but look like websites. REF 4116 ("Bishop *et al.* Storage Temperatures Necessary to Maintain Cheese Safety") is also incomplete and equally does not seem to be authoritative. REF 4149 is not included in the MASTER REFS spreadsheet.**

#### Criterion 5

The score given seems reasonable, and average prevalence rates of a few percent to 10% in finished product might be expected despite that pasteurization of milk is effective in eliminating *L. monocytogenes* and reflecting the potential for post-processing contamination from processing

environments. **However, again the references provided are limited to ‘expert opinion’, when there is a large published literature, including many reviews and book chapters, on *L. monocytogenes* in cheeses, including soft cheese.** Elliot Ryser’s chapter “*Incidence and Behavior of Listeria monocytogenes in Cheese and Other Fermented Dairy Products*” in Ryser and Marth’s “*Listeria, Listeriosis, and Food Safety, 3<sup>rd</sup> Edn*” (CRC Press, 2007) would have been a useful place to start.

*Criterion 6*

**The consumption criterion score (‘3’) based on the NHANES data seems incorrect. The highest daily consumption rate was 10.7%, which should have been scored ‘9’.**

*Criterion 7*

Economic impact per case was estimated by Minor *et al.* (2015) who ranked the per case cost of listeriosis as the 4<sup>th</sup> highest of the foodborne hazards they considered. Accordingly, **the score given in the report (‘3’) seems too low compared to the evaluations in Minor *et al.* (2015) and should be ‘9’, in the same group as botulism, *V. vulnificus*, *Cronobacter*, etc. Importantly, in the C7 spreadsheet the cost per illness that is reported to be based on Minor *et al.* (2015) is misquoted. The value should be \$1,456,676 but in the C7 spreadsheet is given as \$14566, i.e., a factor of 100 too low.**

v. *Listeria monocytogenes* in cold smoked fish

*Criterion 1*

There is a large literature that shows that *Listeria monocytogenes* is commonly detected in cold smoked fish products and clearly has the potential to grow, although growth is limited in most vacuum-packed products by the presence of benign lactic acid bacteria that, through their growth, exert the Jameson Effect. Again, as a first step, the report authors might have consulted Ryser and Marth’s “*Listeria, Listeriosis, and Food Safety, 3<sup>rd</sup> Edn*” (CRC Press, 2007) which has a chapter on ‘Incidence and Behavior of *Listeria monocytogenes* in Fish and Seafood’ with US seafood HACCP expert Mel Eklund as the senior author. An even more authoritative series of papers on hazards and risks in smoked seafoods was prepared by IFT involving a group of international experts, and commissioned by USFDA. The entire series of articles including Lone Gram’s paper “Potential Hazards in Cold-Smoked Fish: *Listeria monocytogenes*” was published in *Journal of Food Science* in 2001. It also appears in full on FDA web-sites (<http://www.fda.gov/Food/FoodScienceResearch/SafePracticesforFoodProcesses/ucm092227.htm>).

*L. monocytogenes* also has a very high ID<sub>50</sub> and studies with pregnant animals that are suitable experimental models for humans have confirmed earlier inferential studies (FAO/WHO, 2004) that the ID<sub>50</sub>, even for highly susceptible people is millions, or billions, of cells. As such, while the hazard is recognized, there are very few reports of outbreaks from cold-smoked fish. Nonetheless, USA has rejected numerous shipments of European cold smoked fish after detection of *L. monocytogenes*. Similarly, studies in USA by Wiedemann and colleagues indicate that *L. monocytogenes* is also common in product and plants producing cold smoked fish in USA. The absence of recall data then, is somewhat surprising. **Thus, the C1 score of ‘0’ in this case seems quite anomalous. No references or explanations are given to support this ‘unexpected’ score.** However, it is noted that cold smoked fish is relatively infrequently consumed by USA citizens.

**What is REF 9999 meant to indicate?**

### *Criterion 2*

Many studies and reviews report that 20 to 30% of systemic listeriosis cases result in death of the patient. The severity ranking also accords with Minor *et al.* (2015). However, the 'death rate' cited in the C2 Results spreadsheet (15.9%) seems a factor of 2 too low. On investigation and re-reading the Scallan *et al.* paper (Scallan *et al.* (2011). *Emerg Infect Dis.* 17(1): 16-22) which was cited in support of 15.9% - it seems that that value is a mis-interpretation. As stated in Scallan *et al.* (see footnote to Table 3, p.12) "These rates [i.e., presented in the table] were doubled to adjust for underdiagnosis before being applied to estimate the total number of hospitalizations and deaths".

### *Criterion 3*

The prevalence score of '9' is based on a single publication, from Sweden in 2011. While the prevalence in that report is not unusual (and a weighted mean prevalence rate of ~18% was estimated from numerous studies prior to 2000 by WHO/FAO (2004; "Risk assessment of *Listeria monocytogenes* in ready-to-eat foods", Microbiological risk Assessment Series, 5) there many more recent studies that should be considered. Also, REF 211 reported that for cold smoked fish produced in Sweden the prevalence rate was ~8%. A quick literature search found several other recent studies, e.g., Leong *et al.* (2015) found a prevalence of 2.5% (mostly environmental some food samples) in smoked salmon processing in Ireland. Gonzalez *et al.* (2014) found 4.7% positives at retail in smoked salmon in Spain. Given this, it is suggested that a more thorough search of recent literature be undertaken to develop a more robust estimate of prevalence.

### *Criterion 4*

The scores given to the components of this criterion, and the overall evaluation and score, seem appropriate based on my experience and reading. However, some of the references cited do not seem very authoritative and better references might be found. Also, FDA 2013 is the IFT expert panel report commented on under Criterion 1 for this product:hazard pair and it might be more informative to refer to the original report, given that the data in the report relate to pre-2000, not pre-2013.

### *Criterion 5*

Given the limited effectiveness of control measures and, particularly, the lack of any listericidal treatment during the cold smoking process, the scores given for contamination prevalence in raw product (or potential for re-contamination or cross-contamination) seem appropriate as does the overall score.

### *Criterion 6*

The consumption criterion score (0) is appropriate and accords with the estimates from NHANES.

### *Criterion 7*

Economic impact per case was estimated by Minor *et al.* (2015) who ranked the per-case cost of listeriosis as the 4<sup>th</sup> highest of the foodborne hazards they considered. Accordingly, **the score given in the report ('3') seems to low and should be '9', in the same group as botulism, *V. vulnificus*, *Cronobacter*, etc. Importantly, in the C7 spreadsheet the cost per illness that is**

reported to be based on the Minor *et al.* (2015) paper is misquoted. The value should be \$1,456,676 but in the C7 spreadsheet is given as \$14566, i.e., a factor of 100 too low.

vi. STEC *E. coli* in bagged leafy green salads

*Criterion 1*

The scoring '1' seems appropriate given the relatively small number of cases associated with outbreaks due to this product:hazard pair.

*Criterion 2*

**In this case the score seems too high, e.g., in comparison with the evaluations and rankings of Minor *et al.* (2015).** Minor *et al.* (2015) developed a comprehensive assessment of mean severity, i.e., taking into account the proportion of cases that were unreported, those that sought/required medical intervention, and those that proceeded to serious illness requiring medical attention and, potentially, long term sequelae. In their assessment three main groups can be discerned based on Mean QALD lost. Six hazards (*Cronobacter*, *Clostridium botulinum*, *Listeria monocytogenes*, *Vibrio vulnificus*, *Trichinella spiralis*, Ciguatoxin) were clearly associated with *much* more severe illness, and had QALD loss scores greater than 10. The least severe group of hazards had, on average, QALD scores less than 1 and included *S. aureus*, *B. cereus*, 'other' *E. coli*, *Vibrio cholera*, *V. parahaemolyticus*, other (non-*vulnificus*) *Vibrio* spp., Astrovirus, Sapovirus, Scombrototoxin (= histamine), and "Food Allergic Reaction". All other hazards, including STECs fell in the range 1 to 10. **Based on this evaluation, the hazards that scored QALD>10, should score '9' in Criterion 1, those that had a QALD score <1, should score '1' for Criterion 1, and all others should score '3'.** Following this suggestion, *E. coli* O157:H7 and other STECs should score '3' in Criterion 1, not '9'.

*Criterion 3*

The score given for Criterion 3 (Likelihood of Contamination) for this product:hazard pair is 1. However, the data sources used to develop this estimate seem questionable. REF 316 (Sagoo *et al.*), reporting on prevalence in the United Kingdom, found no *E. coli* O157 in nearly 4000 samples of salad vegetables (the proportion that are 'bagged leafy greens' could not be easily discerned). REF 307 is the USDA's microbiological data program, and the bibliographical details are incomplete so I can't comment on the data being referred to. REF 326 relates to a Swiss study of product at retail which reported a 'low occurrence of contamination with ...Shiga-toxin producing *E. coli*'. While the body of evidence generally supports the score of '1', there could be more discussion of the limits of the data cited.

*Criterion 4*

The score ('1') for this criterion appears appropriate and to be supported by credible references, although firstly relying on 'expert opinion' (REF 10).

*Criterion 5*

The scores for this criterion, composed of both contamination probability and intervention efficacy, are based on 'expert opinion' (REF 10). Contamination is here scored '3', while Criterion 3 was scored '1'. This seems inconsistent. The intervention efficacy is scored '3', which is possibly debatable. Interventions in the field and prior to processing are probably most important but affect probability of contamination, not efficacy of processing interventions. Processing interventions in common use in the industry might achieve a 1 to 2 log reduction, which is not a strong intervention (by comparison 5 log reductions have been required of other

industries). While the overall result of Contamination scored '1' and 'Intervention Efficacy' scored '1' also results in an overall Criterion 5 score of '3', it is suggested that relevant studies be consulted. For example, the studies of Danyluk, Niemira, Perez Rodriguez, and many others provide quantification of intervention efficacy of currently used technologies for disinfection of leafy greens.

#### *Criterion 6*

The consumption criterion score (9) is probably appropriate but is based on expert opinion, not the NHANES database. The reasons for relying on expert opinion should be explained. Moreover, in deciding on a '9' were the experts thinking only of consumption of bagged leafy greens, or all sources of leafy greens? This should be clarified in the document.

#### *Criterion 7*

The economic impact is consistent with the evaluation of Minor *et al.* (2015) and the expected numbers of cases from Scallan *et al.* (2011).

#### vii. Hepatitis A in Frozen Berry Fruits (472)

##### *Criterion 1*

The score given ('1') seems somewhat low and may relate only to the situation in USA. As noted above in Europe and, more recently, in Australia there have been numerous outbreaks of Hepatitis A from frozen berry fruits and the situation in Europe should have been known to the report authors because they have cited a relevant review paper (Maunula *et al.* 2013, REF 3016). **As such, it is suggested that this score be reevaluated.**

##### *Criterion 2*

The score given is a '9' which is inconsistent with Minor *et al.* (2015), who score Hepatitis A as a more moderate hazard, akin to Salmonellosis, yersiniosis, or a range of parasitic infections. **It is suggested that the score given be reevaluated and reasons for the selection of '9' be documented.**

##### *Criterion 3*

For this criterion a score of '1' is given but with no supporting data or references at all. **More explanation is needed to maintain the transparency of the process.** The available evidence suggests that the frequency and level of contamination is very low but that, because Hepatitis A has a very low ID<sub>50</sub>, outbreaks can still occur when contamination levels are below practical limits of detection.

##### *Criterion 4*

Appropriately the C4 criterion score for this product:hazard pair, is '0', indicating that no growth is possible.

##### *Criterion 5*

The scores given for this criterion are '3' both for contamination and for intervention. I'm unaware of any highly effective interventions against Hepatitis A on fresh fruit, other than chemical sanitizers, and they're not very effective. Equally, freezing berries wouldn't be expected to cause significant reduction in Hepatitis A levels. The score of '3' for contamination seems inconsistent with the score given for Criterion 3. **It is suggested that the score given be reevaluated and reasons for the selection of the scores of '3' be documented.**

*Criterion 6*

The NHANES data indicates that much less than 1% of the population consume frozen berries on any given day. As such, and using the scoring criteria outlined in the model and report, a score of '1' for this criterion is correct.

*Criterion 7*

The economic impact is consistent with the evaluation of the cost per case of Hepatitis A of Minor *et al.* (2015) and the expected numbers of cases from Scallan *et al.* (2011).

- 7. *The draft model relies on expert elicitation to fill data gaps, in particular for the scoring of Criterion 5 and for the scoring of criteria for chemical hazards and undeclared allergens. Is the expert elicitation process (which involves external panels and FDA subject matter experts) adequate to address data gaps?***
- a. *Is the scoring logic used to combine data and expert opinions and the order of preference in section 4 (e.g., Figure 4.3) appropriate? If not, please explain what changes might be considered and why.***

The range of experts used seems appropriate and sufficient to generate credible responses to the missing data, particularly when combined with a public comment process to enable further input and opinion. Sometimes experts will need to have relevant, though imperfect, data presented to them to assist them to make an informed decision. In this regard, the integration of the expert's background knowledge and perspectives might be brought to bear earlier in the process by asking the experts to evaluate the significance of prevalence rates from RFR, or recall data, as part of the expert elicitation, rather than requesting the experts to make estimates when no other data has been found upon which to make a reasoned estimate.

- 8. *Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.***

It is difficult to respond to this question unequivocally, because the answer depends on whether the aim of the FSMA requirement to designate high-risk foods is intended to apply to the average risk to the entire population, or the risk to susceptible populations. Logically, the most objective approach is to establish the relative risk for the whole population exposed to the foods, particularly since none of the foods considered in the model seem to be specifically intended for an 'at risk' population. A potential problem with this approach, however, is that it effectively disenfranchises susceptible populations because those sub-populations will experience the bulk of the disease burden for specific hazards. Furthermore, the 'treatment' of susceptible populations in the model seems inconsistent between hazards, based upon reading Sections 2.2.2 and 4.2.2. In some cases, mortality and hospitalization rates are used to evaluate severity, but that ignores that those rates are related to the laboratory confirmed cases, not the total estimated number of cases. This potentially introduces a bias because the same denominator is not used for each hazard. While it might be argued that hospitalization rates based on laboratory confirmed cases do effectively reflect average disease severity, it is an indirect metric and better, more direct metrics are available, e.g., in Minor *et al.* (2015) discussed further below.



Furthermore, while it is stated that susceptible populations are ‘considered’ in the evaluation/scoring for Criterion 2 it is not clearly explained *how* susceptible populations are considered nor the relative size of the susceptible population, which is needed to establish the correct relative risk. Presumably this ‘issue’ is the genesis of Charge Question 8. Also, as noted earlier there is inconsistency with the relative severity estimates of Minor *et al.*, 2015, and the inconsistency for the ‘severity’ scores of allergens in the model *cf.* Minor *et al.*, 2015 is discussed in more detail in the response to Question 12).

To elaborate, disease severity would be expected to be greater in susceptible populations, or disease much more probable in susceptible individuals for a given level of exposure. If the disease severity score is based on the responses of susceptible populations, then all other Criteria would have to be evaluated with respect to susceptible populations as well, because the risk from a particular hazard pair would be overestimated by assuming that the entire population is ‘susceptible’ to that hazard to the same extent as the specifically susceptible population. Using *Listeria monocytogenes* as an example, the ‘susceptible’ population is, on average, about 100 times more susceptible than healthy non-pregnant adults but represents only 15–20% of the population. Additionally, the consequences of infection in the ‘susceptible’ population are much greater than in healthy non-pregnant adults. Thus, assuming all people are equally susceptible to listeriosis would overestimate the relative risk by 500-fold, or more. Also, even among the ‘susceptible population (i.e., very young, old, immunocompromised, pregnant), susceptibility to listeriosis varies over two orders of magnitude (*see e.g., Goulet et al., 2012, Clinical Infectious Diseases, 54:652-660*) for different pre-disposing conditions. Similarly, while 2-4% of the population suffer allergic responses to specific food-borne allergens, not all have severe responses leading to anaphylactic shock, but this does not seem to have been considered in the model in the scoring of the severity of exposure to allergens.

These relative susceptibilities to hazards and relative severity of consequences for susceptible populations will not necessarily be the same for all hazards. Accordingly, the lack of a clear statement of whether the severity rating is averaged across the entire population, or derived only for the susceptible population.

The suggestion for a ‘correction’ to be included to account for this by scaling down the % consumption should not be contemplated because it is not transparent unless it is explicitly made clear why that ‘adjustment’ is being made and the basis of the reduction in consumption assumed. Furthermore, I imagine that it would be difficult to implement because the adjustment would have to be accommodated with the incremental scoring scheme (0, 1, 3, 9) which would distort the true proportions and would be difficult if it were not applied to every product:hazard pair.

As suggested previously in this review, a more consistent and scientifically defensible approach would be to base the severity score on the average disease severity across all consumers, based on a weighted average of the number of cases not requiring medical intervention, those that do require intervention and those that lead to death or permanent disability together with a rating of the disease severity for each of those categories, e.g., as was presented by Minor *et al.* (2015) in terms of Quality Adjusted Life Days (QALD), in addition to their estimates of economic cost of various diseases. Given that the economic estimates of disease severity from Minor *et al.* (2015)

were adopted, it seems reasonable to adopt their QALD estimates as the basis of scoring for Criterion 2.

**9. Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?**

The intent of this question seems to be to seek advice on establishment of criteria for determining the “cut-off” in the risk ranking that delineates foods that will require additional record keeping and those that will not. This is a risk management decision and would seem to be the responsibility of the risk manager(s) who commissioned this risk ranking tool and exercise. There is no unequivocal, objective basis for deciding the “cut off”. Rather, that will require risk communication to understand the expectations of relevant stakeholders to determine an acceptable level of protection of public health compared to the costs associated with that. The goal of the risk ranking tool and exercise is to: “rapidly and effectively . . . prevent or mitigate a foodborne illness outbreak . . . and to address credible threats of serious adverse health outcomes” and the instructions given to the modelers/risk assessors were that the designation of high-risk foods shall depend on: “the known safety risks .. including the history and severity of foodborne illness outbreaks” and “the likelihood that a particular food has a high potential for microbiological or chemical contamination . . . or would support microbial growth due to the nature of the foods or the processes used”.

These instructions correctly identify considerations that need to be taken into account when identifying high-risk foods but, to elaborate a little:

Risk is the combination of the probability of an adverse event, and the severity of that event. In terms of foodborne illness, the severity is a combination of the severity of illnesses experienced and the number of people affected. Severity is explicitly considered in the model although, as discussed above, it might be characterized in other, more defensible, ways. The number of illnesses would be expected to be correlated with the number of people exposed to a particular product:hazard pair and, in the model, this is represented through the consumption data and criterion.

The probability of an ‘adverse event’ occurring is correlated with the inherent likelihood of contamination of the product at a level able to cause illness, or the potential for the contamination and subsequent growth of microbial hazards to a level able to cause illness. This also is represented in the model but, as a marker of such product/process/hazard combinations, it might be beneficial to consider:

- i) whether there is one, or more, credible hazards associated with the product;
- ii) whether there are clearly identifiable Critical Control Points necessary to control those hazards to enable the safe production/processing of the product and
- iii) whether those CCPs can be, and have been demonstrated to be, reliably controlled and monitored.

If the answer to the first question is ‘no’, the product should probably not be considered a HRF. If the answer to the first question is ‘yes’ and the answer to the second question is ‘no’, the product should probably be considered a HRF. If the answer to the both the first and second questions is ‘yes’ and the answer to the third question is ‘no’, the product should probably be considered a HRF. If the answer to all three questions is ‘yes’, the product is possibly not a HRF.

Further considerations may relate to the feasibility of identifying a potential public health risk and instituting a recall, e.g., less weight might be given to products with very short shelf lives (e.g., less than 10 days) because of the feasibility of “rapidly and effectively” preventing or mitigating a foodborne illness outbreak by recalling the product. Similarly, more weight might be given to food-hazard pairs that cause acute illness because these are also more likely to lead to overt outbreaks of illness that will be mitigated rapidly by a recall.

[Clarification: If the aim of the rule is to establish additional record keeping so as to be able to identify and recall foods associated with an outbreak, foods with a short shelf life after they are sold at retail, may not be easily identified and recalled before they are consumed. If so, there's no point keeping the additional record.

Note, though, that I'm not saying it is futile to maintain records for those products, but that it ought to be thought about, i.e., whether the additional records would enable rapid and effective prevention/mitigation of further cases of illness. If not, another strategy would be preferable.]

Finally, even though the primary objective of the HRF record-keeping initiative seems to relate to protection of public health, the economic costs of illness outbreaks might also affect decisions about cost vs. benefit of the HRF initiative, or other risk management actions. The economic costs of an outbreak are not, however, limited to medical costs (*see* Scharf, 2012 cited in the report). Thus, for an effective cost-benefit analysis a broader analysis of costs than has been considered in the current model, including, e.g., loss of consumer confidence leading to reduced sales for an industry/product, loss of national productivity due to days off work, etc. should also be considered.

***10. How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?***

I am not sufficiently expert to be able to offer a reliable response to this question but would suggest that the rate of change of food processes and product formulation might mean that the model and criterion values would need to be reviewed every 3 to 5 years. However, if it becomes evident that radically different processes or products are introduced, or products are sourced from new/different suppliers, it would be prudent to evaluate *before introduction of those products* whether those changes introduce a different level of public health risk.

***11. Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.***

Many comments were made above about the transparency of the process and the model, identification and documentation of relevant and authoritative sources, and limitations in the data and the model. Other comments will be made in the separate report on the model and document itself.

***12. Do you have any additional comments? Please share them in your review.***

During the evaluation of selected product:hazard pairings it was noted that for Criterion 1, of the 72 scores of '9', 52 (70%) were allocated to allergens but all were based entirely on expert

opinion and without any evidence nominated. The other 20 were for microbial hazards. *Prima facie*, this seems incongruous, and is perhaps a consequence of different experts being used to assess the severity of different categories of hazards. Importantly, the Minor *et al.* (2015) evaluation of hazard severity gave much lower ratings to allergens than were derived by experts in this risk ranking process. The scoring for allergens should be reviewed to ensure that all experts (not just allergen experts) believe that scoring is consistent across all hazards, and is also consistent with other published opinions/reports based on critical analysis of the available data. As discussed in relation to Criterion 2 for STEC *E. coli* in bagged leafy green salads, there seems to be an inconsistency between the hazard severity ratings developed in this model/risk ranking tool and the severity estimates developed by Minor *et al.* (2015). This aspect of the model/risk ranking tool should be re-evaluated to ensure that the relative severity measures in the model are scientifically defensible.

The requirement to identify “iii) the point in the manufacturing process of the food where contamination is most likely to occur” does not really seem to have been addressed in the model presented, nor in the evaluation of criterion scores. This requirement possibly relates to another perceived difficulty in the model, i.e., that it does not explicitly recognize the importance of sequences of events in the evaluation of relative risk, which was commented on above in relation to limitations of additive scoring schemes for microbial food safety (*see* response to Question 1d).

It is noted that “Section 204(d)(1) of the FDA Food Safety Modernization Act (FSMA), requires that the Secretary establish additional record-keeping requirements for high-risk foods in order to rapidly and effectively identify recipients of a food to prevent or mitigate a foodborne illness outbreak and to address credible threats of serious adverse health consequences or death to humans or animals.” Given that mandate, what is the intent of the questions about ‘economic impact’, i.e., how does economic impact affect the decision about high-risk foods requiring additional record-keeping?

Also, in Appendix C (page C-8, last paragraph) of Appendix K to the Sept 2015 draft report, it is stated that the estimates of Scharff (2012), in conjunction with Scallan *et al.* (2011) were used to assess economic impact and noted that the Scharff estimates included “financial losses due to medical expenses, lost productivity and lost utility”. In the September 2015 draft report, however, it states (p. 65) that: “The cost-per-case estimates for the current analysis were largely drawn from the most recent study (Minor *et al.* 2015). Cost-per-case estimates from Scharff (2011 [*sic*]) were used for hazards that did not have estimates from Minor *et al.* (2015).” (Note that on P. 65, Lines 1- 8, the Scharff paper is mis-cited several times, i.e., the publication year is 2012, not 2011). However, **having checked through the lists of hazards in both Scharff (2012) and Minor *et al.* (2015), I see no hazard considered in Scharff (2012) that is not also considered in Minor *et al.* (2015). Accordingly, the product:hazard pair estimates for which the Scharff (2012) estimates were used should be clearly articulated or the text of the report appropriately modified.**

Further, it should probably be clarified in the document that, to estimate the economic impact, ‘non-public health impacts such as potential industry costs and loss of market costs are not be [*sic*] included in this criterion’ (Draft report, p.24, line 7, 8) and that the earlier use of the Scharff (2012) estimates was abandoned because lost productivity costs were not to be included (if that is, in fact, the explanation).

**III. SPECIFIC OBSERVATIONS ON DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Page	Paragraph/Line	Comment
		Specific observations were presented above in the responses to the Charge Questions. Suggested corrections will be presented in the complementary report on the Model.

**IV. SPECIFIC OBSERVATIONS ON APPENDICES TO THE DRAFT REPORT FOR PEER REVIEW: RISK RANKING MODEL FOR PRODUCT TRACING AS REQUIRED BY SECTION 204 OF FSMA (RRM-PT Draft Report) WITHIN THE CONTEXT OF THE SUPPORTING DATA.**

Appendix	Page/Row	Paragraph/Line/Column	Comment
List of Appendices	P4, Line 77		“Master” is misspelt
All Appendices and throughout Entire document			Inconsistent use of RPM-PT cf. HRF as abbreviation for model. It is noted that Appendix M1 explains that that name/acronym for the model changed in August 2015, yet nearly all the document (dated September 2015), including appendices, refers to the ‘HRF’. The main report includes a ‘header’ on each page that describes the report as the RRM-PT report. This needs to be made consistent throughout, or an explanation provided very early in the document.
Various			Several appendices (esp. App. D, App. N) are previously submitted reports. While those documents also contain various grammatical errors, I have not recorded them on the assumption that they cannot be changed in this report unless the original documents are also changed and re-issued.
Throughout			Inconsistent page numbering conventions. Some appendices (e.g., C, ) include the Appendix ‘letter’, others (e./g., J, M1) just restart the page numbering from 1
C	All pages		Pages numbers suggest that it is Appendix D, not ‘C’
J	2	Line 10	Refers to Appendix J2, but there was no Appendix J2 provided to this reviewer
K	C-1	8 <sup>th</sup> line from bottom	Insert ‘the’ before IFT
K	C-11	Table 11	Why is ‘livestock feed’ included in this Table?
K	all		There is nothing in Appendix K to indicate that it is an appendix to the main report. It has its own ‘Appendix Letter – Page Number’ numbering, that makes it confusing when all the Appendices are combined as supplements to the main report. Needs some ‘tidying up’

#### **IV. PEER REVIEWER COMMENT TABLE**

<b>I. General Impressions</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	<p>This report describes an impressive effort to rank foodborne diseases and foods on the basis of an extensive list of criteria. The report is very clear, well written and structured, and the authors undertook a notable task of data gathering and analysis. Overall, the seven criteria chosen to score food-hazard pairs are adequate and comprehensive.</p> <p>The purpose of the project is to identify high-risk foods to “inform proposed rule-making”, and “establish recordkeeping requirements for the designated high-risk foods to prevent and mitigate outbreaks”. In this context, the model focuses largely on foods at the end of the food chain. Still, I have concerns about the general reliance of the model on outbreak data (specifically for microbial hazards). Outbreak events are random and can have multiple causes, from contaminated primary foods to some type of failure in the production or distribution process (e.g., environmental contamination, disrespect of optimal time-temperature). Even though these factors are potentially accounted for in the model, I believe that prevention of outbreaks (and of foodborne disease in general) requires the control and prevention of contamination of primary ingredients as well. Because random failures in the chain are difficult to predict and prevent, reduction of hazards at a point in the food chain that is closer to the original production of the food may actually be the most effective way to prevent many types of events. In this context, I have concerns about the use outbreak-related data for criteria 1 and 7. Outbreak cases correspond only to a fraction of the total cases of infection by most pathogens, and this proportion varies from pathogen to pathogen (e.g., 6% for <i>Salmonella</i> in the U.S. in 2013; 7% for <i>Listeria m.</i>; 15% for STEC (CDC data, see link below)). On the other hand, these data were also used for C7 (economic impact), but here outbreak cases were corrected for underreporting based on Scallan <i>et al.</i> (2011). However, Scallan’s estimates were derived for correcting sporadic cases for underreporting, which are less likely to be reported than outbreak cases. Furthermore, the relative contribution of different foods for outbreak cases is not necessarily representative of the causes of all cases (i.e., of sporadic cases also), because some foods are more prone to cause large-scale outbreaks than others. This can lead to an overrating of the economic impact of disease by some foods. Lastly, and as mentioned in the discussion of the report, not all hazards cause outbreaks (or at least not frequently).</p> <p>It is apparent that the authors have focused on outbreak data because the essential piece in the model is the food-hazard pair, but I believe that an alternative approach would be to start by ranking foodborne hazards on the basis of their disease burden in the population (e.g., by applying health metrics), and then estimating the proportion of this burden that is caused by different foods (using source attribution, which can also use outbreak data).</p>	
Reviewer #2	The FDA has done an admirable job in building a science-based model for risk ranking and in developing a report that makes it accessible to a reader with no previous experience in risk ranking or modeling. The	

<b>I. General Impressions</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>targeting of biological, chemical, and undeclared allergen hazards is appropriate and need not be expanded to include other potential hazards (e.g., physical hazards that are addressed in GMPs). The fifty Primary Commodities proposed appear to encompass all foods regulated by FDA, but because some high-risk and low-risk foods are combined in the Secondary Commodity groupings, some of these secondary groupings may require further refinement in the next iteration of the model. (For the current model, &gt;1200 food pairs is enough. One has to stop somewhere.) The seven criteria incorporate the six factors required by FSMA. The scoring and the underlying data appear to be accurate for each of the criteria. However, because low risk ranking the finfish:methyl mercury pair provided in the report is inconsistent with previous strong FDA advisories on methyl mercury in fish, the finfish:methyl mercury pair and its underlying data should be further scrutinized. This might be done in the broader context of discussing the issues associated with sensitive populations and the scoring of chronic versus acute illnesses. By the end of the report, readers have a clear understanding of the risk ranking process and the procedures which implement it. However, the beginning of the document is not user friendly; the incorporation of the relatively few editorial suggestions given in Section III of this review would help address this issue. A glossary would also be very helpful. The conclusions of each document follow from the underlying data. In summary, FDA has done an impressive job of tackling the very complex task of developing a semi-quantitative method which will allow foods to be regulated according to their risk as required by FSMA.</p>	
Reviewer #3	<p>The aim of the document to determine the risk of food-hazard pairs is clear and as such it is a useful and valuable approach which could benefit public health authorities as well as other stake-holders of the food chain. Although in certain cases and for certain food-hazards pairs, the need for an assessment may be questioned and whether the fact that undeclared allergens appear systematically in any type of food is justified and also, generally speaking, the outcome provides this and as such the goal is achieved.</p> <p>The document is not very clear in terms of language and structure – contains a lot of repeats between different sections/chapters and even within single paragraphs. While section 2 seems designed to provide the background and details to all criteria used in the model, often it is fragmentary and part of the information/explanations are provided much later in the document, e.g., certain calculations (equations are provided in section 2, others in section 4, in text or figures), aspects such as confidence, uncertainty appear in section 4, discussions on certain foods in section 4, while a clear understanding would be needed earlier on, e.g., cascade from primary, secondary and tertiary foods which makes it difficult to understand why certain foods are classified in the tertiary list (e.g., chicken sandwich) but other sandwich types are only kept at secondary level. Terminology is also not consistent or new terms are introduced later in the document. Clear definitions would be useful for example for terms such as cases, illness, outbreaks, sporadic cases – generally speaking a section for definitions and acronyms would be helpful.</p>	



<b>I. General Impressions</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>There are quite a number of inconsistencies between Appendices – e.g., numbering of hazards in annex B (in principle the master) is not always reflected in other appendices.</p> <p>The rationale for assigning certain “chemicals” as “biological” due to growth leading to their formation is not applied consistently – mycotoxins are typically associated with growth and the step of the food chain is also a key factor in their formation but they are handled as “inert” chemicals.</p> <p>It is also not completely clear what is included in the assessment, for example of C5 – which steps of the food chain are exactly considered is unclear as the text is very vague.</p> <p>Weighing is not always understandable – C3 is about contamination of a food – i.e., an event occurring before consumption and is therefore not linked. Nevertheless in the model consideration is given to importance of the food in the diet with a weighing factor. However, for the determination of C6, the importance of the food in the diet is again used for the scoring, which is the appropriate way to consider the impact a contamination will have on consumers.</p> <p>Section 4 – it is not clear why in any case the whole process is run for the different criteria. While from the text it seems that if information is available to determine a score, then the process stops, whether additional information exists or not. In the case of the figures and examples – it seems however that the whole process needs to be run, e.g., even if available quantitative data are sufficient to score “0”, scoring process continues. In other words – processes do not seem to follow a Y/N decision tree, with an end at the level where the score should be considered as substantiated.</p>	
Reviewer #4	<p>The ultimate aim of this model/tool appears to be to identify foods under USFDA oversight that represent a ‘high’ risk of causing food-borne illness and which, consequently, may warrant extra record keeping, e.g., to facilitate identification of outbreaks and hasten recalls of implicated products to minimize public health risk. As the model authors indicate in the report, they do not attempt to identify the ‘cut-off’ in the risk ranking that would identify foods that <i>do</i> require the additional record keeping.</p> <p>There are potentially two main approaches to identifying and ranking the relative risk of foodborne illness from specific foods. One is to rely on the available epidemiological evidence for foodborne illness to identify “problem” foods and hazards. There is a large amount of relevant data, however, not all foodborne diseases are reported, or even required to be reported. As such, this empirical approach alone is not sufficiently robust to meet the objective.</p>	

<b>I. General Impressions</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>A second approach is based on synthesis of relevant knowledge about factors, and their interactions, that influence risks of foodborne illness. To do so requires, however, requires characterization of the kind and magnitude of those factors for each product, process, intended consumer population, intended end-use, etc. Additionally, it would be ideal to be able to identify all hazards that are, or might be, associated with that product. Using some form of formalized risk assessment, these data could be used to estimate relative risks from different foods, processes, hazards, etc. This approach, being more fundamentally based on an <i>understanding</i> of risk-affecting factors than the empirical epidemiological data, might also be applied to estimate risks from foods that represent new formulations and processes, and for which there are currently no epidemiological data or, conversely, a log record of safety. This approach, however, could be used proactively to <i>predict</i> relative risk, rather than a reactive approach based on epidemiological data.</p> <p>There are, however, a large number of variables that influence food safety risk, even on a relative scale. Assessing all of these factors for each food-hazard pair creates a significant challenge to the achievement of the ambition of risk ranking. Many of these variables have potential to profoundly affect the risk from a single category of product, e.g., the addition, or not, of preservative; the time and temperature of storage of the product; whether the product is consistently manufactured hygienically and under GMP; the intended consumer group and whether some groups are predisposed to specific hazards that may be present in the product. In many cases data needed to estimate risk will not be available and the data can change over time, e.g., as new technologies and products are introduced to the market, when processes or suppliers change, and if products or ingredients begin to be sourced from regions outside of the USFDA's regulatory oversight.</p> <p>Thus, such (inferential) ranking schemes are likely to involve compromises and assumptions that potentially lead to over- or under-estimation of the relative risks from specific products, under specific processes, for specific populations. The challenge is to produce a decision tool that is simple enough to be workable but that is scientifically/logically robust enough to achieve the correct relative risk for most products and circumstances, but also to be able to recognize when the answer is inappropriate and to recognize that special considerations, not explicitly considered in the model/decision tool, might sometimes have to be applied to achieve a credible relative risk ranking. (Recognition and accommodation of such 'anomalies' should also help to improve the logic of the model and its robustness).</p> <p>The two approaches described above are not mutually exclusive and each has its own strengths and weaknesses. <b>However, it seems reasonable that, if a risk model were correctly formulated, it should produce predictions consistent with the observed epidemiological data (at least where such data are</b></p>	

<b>I. General Impressions</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p><b>available). If so, it would give confidence that the predictions of the model for other products that have not (yet) been linked with food-borne illness outbreaks would also be reliable.</b></p> <p>Upon reading the report and working through the model and its output what seemed to be missing were ‘reality checks’ to test whether the model, and the data it is based on, produce relative risks estimates that are close to, or at least consistent with, what is observed from the available epidemiological data. However, it is noted that the development of the model has received oversight, comment and insight from industry and scientific experts, and that there has been public consultation enabling stakeholders to share their knowledge, insights and concerns about the relative risks from relevant product:hazard pairs. If after those opportunities the stakeholders agree that the model is sufficiently credible for its intended use, there may be no need for further ‘reality checks’, i.e., if consensus/acceptance has been achieved.</p> <p>In looking closely at some of the data it also seems there are some inconsistencies in the approaches taken and decisions made, and some potentially flawed logic in the scoring scheme, particularly because some of the criteria are confounded, e.g., as suggested above, the epidemiological data would be expected to be a reflection of other risk-influencing factors (Criteria 3 – 6), so that effectively a “double-counting” of some factors is occurring in the logic of the model. It is also noted, however, that the model is based on an EFSA food risk model which used a similar approach of including both epidemiological data and data on products and processes to characterize risk.</p> <p>Nonetheless for reason discussed above, it seems that it would be a valuable experiment to remove the epidemiological data from the scoring, and use it instead to assess the reliability of the relative risk estimates of the risk assessment model based on Criteria 2 – 7 only, for example.</p> <p>Many minor typographical and grammatical errors were also evident, which are described in the accompanying reports (Sections III and IV on the model and report).</p> <p>Principle 6 of the Codex Alimentarius Guidelines<sup>2</sup> for the conduct of microbial food safety risk assessment, of which the report under review is an example, states that: “Any constraints that impact on the Risk Assessment such as cost, resources or time, should be identified and their possible consequences described”. Given some of the potential shortcomings described it may be appropriate to include a relevant statement concerning potential constraints in the report itself.</p> <p>Detailed comments against the Charge Questions are presented below and elaborate on comments made above.</p>	

<sup>2</sup> Codex Alimentarius Commission (1999). Principles and Guidelines for the Conduct of Microbiological Risk Assessment CAC/GL-30 (1999)

## II. Response to Charge Questions

**CHARGE QUESTION 1:** *In order to apply the FSMA factors it is necessary to first take into account both the characteristics of foods and known or reasonably foreseeable hazards, i.e., food-hazard pairs. The food categorization scheme involves a list of 1,286 food-hazard pairs (candidates) and 335 foods (secondary commodities) linked to approximately 50 food categories (primary commodities).*

**CHARGE QUESTION 1.a:** *Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?*

REVIEWER	COMMENT	RESPONSE
Reviewer #1	<p>The authors justify the choice of the categorization scheme well. The described arguments for the use of RFR-adapted primary categories [a) is in line with preventive control measures by FDA, b) considers processing, (c) no need for re-categorization] are reasonable, and the definition of secondary commodities under these primary ones facilitates the aggregation of information at different levels. The list is representative of (and more detailed than) the FDA's RFR commodities definitions.</p> <p>I have one comment: some secondary commodities seem too unspecific (in the sense that they correspond to multi-ingredient foods that can potentially include almost any type of single food), and their link with hazards (to generate food-hazard pairs) somewhat arbitrary. For example, "sandwiches" can include nearly any type of food commodity (in addition to bread), which means that many of these foods can be contaminated with a wide range of hazards (which is reflected in the number of food-hazard pairs with sandwiches: 15). However, information on this food-hazard pair will not give any information on the contaminated original ingredient or point of contamination in the processing chain. Another example is "RTE dinners", which can be constituted by an even wider variety of food commodities. The fact that these are paired with hazards which have been associated with outbreaks or have been recalled due to mislabeling (undeclared allergens) at some point in time, does not ensure that all potential hazards associated with these type of foods are listed, or that the relative occurrence of different hazards in these foods is well-established.</p> <p>That said, I would suggest that all secondary commodities have some kind of description of the basic food (ingredient) that can be linked to the hazard in the pair. In my view, this would allow for generating food-hazard pairs that are more useful for regulation purposes.</p>	
Reviewer #2	It appears that all FDA-regulated food products can be placed in this classification scheme.	
Reviewer #3	At the level of primary commodities it is helpful to categorize broad food categories allowing to fine tune at secondary and then at tertiary level.	
Reviewer #4	The list of foods seems appropriate having been based on commodity groups that fall within USFDA jurisdiction. It might be useful however to accentuate foods that do not fall under USFDA regulatory	

<b>CHARGE QUESTION 1.a: Is the food classification scheme appropriate and adequate to identify a comprehensive list of foods representative of FDA-regulated products?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>oversight, e.g., the reasons for not considering red meats and chicken products (risks of STECs, <i>Salmonella</i> and <i>Campylobacter</i>) as high-risk food:product pairs. The further refinement of the main categories into secondary categories is also appropriate because the 49 primary commodities groups are very broad and encompass products that would be expected to represent vastly different risk due to differences in product formulation, processing, packaging, shelf-life, intended end-use, that affect potential for pathogen growth in particular, but also likelihood of contamination.</p> <p>(<i>n.b.</i>, Table A-1 still lists pet foods/feed even though the text says they were excluded from consideration).</p>	

<b>CHARGE QUESTION 1.b: Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	Both primary and secondary commodities appear appropriate, and data to populate the model at the secondary-commodity level seems to be available in most cases. However, because of the point in the food chain that most foods are classified (i.e., prepared foods), contamination data for some seem to be lacking. Following the same example as above, contamination data from sandwiches seem to be sparse for most hazards. Categorizing foods on the basis of their ingredients (like foods that constitute a sandwich) may allow for collecting contamination data.	
Reviewer #2	<p>As noted in the public comments, many “risky” and “nonrisky” food products are aggregated into the same food-hazard pair. This is problematic and should be addressed in the next iteration of the model.</p> <p>There is one example of aggregating that is of particular concern; treating finfish as a single group for the secondary commodity, secondary food-hazard combination “finfish:methyl mercury.” A secondary commodity, “small fish” would not present a methyl mercury hazard in this category, but a secondary commodity “large fish”, e.g., swordfish, tile fish, king mackerel and shark), (i.e., a “large fish”:methyl mercury pair) present such a large risk that the FDA/EPA Joint Advisory (2004, 2014) states that pregnant women, nursing mothers, and young children “<i>should not eat</i>” swordfish, tile fish, king mackerel and shark. There should be intra-agency agreement regarding the risk of consuming large fish.</p>	
Reviewer #3	A clear scheme or diagram on the three levels would be helpful to understand – the challenge is to determine on how far a food-hazard pair is valid applicable for the higher (or lower) level – for example what is valid for a particular cheese might be valid for others, but not for cream; what is valid for infant formula is not valid for milk powder due to the different sensitivity of consumers; what is valid for a	

<b>CHARGE QUESTION 1.b: <i>Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>particular food is valid for others with a very similar ecology and manufacturing conditions, although not involved so far in outbreaks.</p> <p>The food classification does not always appear appropriate and the difference between a secondary and a tertiary commodity is not always understandable.</p> <p>Example: Secondary – Acidified vegetables and beans: these are in my opinion two different levels. If beans are not considered vegetables, then a secondary commodity would be “pulses”, encompassing more than beans. Aside of acidified vegetables and beans there is also a category canned vegetables and fruits – seems that they are overlaps making it difficult to either identify a food hazard pair or to “forget” some in case a food is considered by individuals in one of the three possible options.</p> <p>Another example – shelf-stable milk as secondary commodity to LACF (assuming this means low acid canned food) considering that there are other types of shelf-stable milk such as condensed milk (usually also canned) which is however part of “Dairy” as first commodity.</p> <p>Frozen vegetables as a secondary commodity, frozen vegetables (beans, potatoes) as well – in my opinion beans and potatoes are subsets and sometimes this gives the impression that the assignment to a category of commodity is driven by the hazard (in this example different hazards for the three).</p> <p>Salads – certain of the categories seem to describe products at a tertiary level – what is the difference between a mixed RTE salad and specific ones such as taco salad, fish salad...why not chicken salad, meat salad, tomato salad. Avocados are listed under dressings – why not salads? Most salads are “mixed” – and if there is a specificity then this would fit better at tertiary level. As an analogical example – sandwich is dealt with as secondary commodity, apparently irrespective of the type of sandwich.</p> <p>Grouping completely different forms at secondary level, e.g., dry and liquid gravies does not seem to allow for an appropriate assessment of certain of the criteria such as 4 and 5. Hence, it would probably be more appropriate to differentiate between dry and wet (low and high water activity products), also to be able to take into consideration elements pertaining to the rating of individual criteria, in this case “supporting growth or not”. By having a single commodity – which one should be used to answer these questions.</p>	
Reviewer #4	As noted above, risk assessments are often a compromise between i) the desire to include all relevant details and differences to be able to account for all important variability, and ii) the availability of relevant	

<b>CHARGE QUESTION 1.b: <i>Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>data to support those distinctions in the risk assessment. In risk assessment terminology this might be considered the paradox of disaggregation.</p> <p>The food groups selected seem to be “fine” enough to enable identification and differentiation of relevant classes of product, particularly as they relate to potential for pathogen growth, although details of packaging and storage conditions (as well as shelf life) might also be relevant for correct characterization of potential for growth. The other consideration, however, is the availability of relevant data to support assessment of relative risk of each of those sub-divisions.</p> <p>From the text of the report it seems that the consumption data from NHANES did support the level of ‘granularity’ adopted in the model. However, the data to assess relative outbreak incidence and hazard prevalence seem less complete. From Appendix 4b (Prevalence data) only ~220 of the 1286 nominated product:hazard pairs appear to have published (refereed or otherwise) data to support the prevalence estimates needed. Similarly, for outbreak data, only ~411 of the 1286 product:hazard pairs have data to support the criterion value selected, the other pairs presumably relying on ‘expert opinion’.</p> <p>Of the Criterion 4 values that required definition, i.e., for which estimation of growth potential is required, 629 product:hazard pairs involved bacterial pathogens. To support those estimates, 597 discrete sources of information are cited, with between 1 and 8, and an average of ~3, sources per product:hazard pair, for a total of 1990 citations across all 629 product:hazard pairs. Notably, of the primary references provided, 450 of the product:hazard pairs relied on ‘Expert opinion (IFT/RTI expert elicitation)’ as one, or the main, basis of the estimated growth potential. Of the 629 product:hazard pairs with potential for microbial growth, 83 had only one supporting reference for the growth potential “decision”, and for all but five of those 83 the source of information was one or other source of ‘expert opinion’. The most cited sources are shown in Table 1, below.</p>	

**CHARGE QUESTION 1.b: Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?**

REVIEWER	COMMENT	RESPONSE																											
	<p><b>Table 1. Sources of Information for Classification of Growth Potential in Food-hazard Pairs (Criterion 4)</b></p> <table border="1"> <thead> <tr> <th data-bbox="382 342 569 415">Ref. No.</th> <th data-bbox="569 342 695 415">Times cited</th> <th data-bbox="695 342 1661 415">Details</th> </tr> </thead> <tbody> <tr> <td data-bbox="382 415 569 456">10</td> <td data-bbox="569 415 695 456">450</td> <td data-bbox="695 415 1661 456">Expert opinion (IFT/RTI expert elicitation)</td> </tr> <tr> <td data-bbox="382 456 569 529">434</td> <td data-bbox="569 456 695 529">165</td> <td data-bbox="695 456 1661 529">NSW (2008) Potentially hazardous foods: Foods that require temperature control for safety. NSW/FA/CP016/0810</td> </tr> <tr> <td data-bbox="382 529 569 570">6035</td> <td data-bbox="569 529 695 570">51</td> <td data-bbox="695 529 1661 570">FDA expert opinion, June 2015</td> </tr> <tr> <td data-bbox="382 570 569 610">6041</td> <td data-bbox="569 570 695 610">46</td> <td data-bbox="695 570 1661 610">FDA SME August 2014</td> </tr> <tr> <td data-bbox="382 610 569 651">454</td> <td data-bbox="569 610 695 651">27</td> <td data-bbox="695 610 1661 651">AEM 39(5):943-949 (1980) Cameron <i>et al.</i></td> </tr> <tr> <td data-bbox="382 651 569 724">435</td> <td data-bbox="569 651 695 724">26</td> <td data-bbox="695 651 1661 724">Betts <i>et al.</i> (2006) Scientific Review of the Microbiological Risks Associated with Reductions in Fat and Added Sugar in Foods</td> </tr> <tr> <td data-bbox="382 724 569 797">2181</td> <td data-bbox="569 724 695 797">19</td> <td data-bbox="695 724 1661 797">FDA (2001) Evaluation and Definition of Potentially Hazardous Foods - Chapter 3. Factors that Influence Microbial Growth</td> </tr> <tr> <td data-bbox="382 797 569 911">2187</td> <td data-bbox="569 797 695 911">19</td> <td data-bbox="695 797 1661 911">FDA. Quantitative Assessment of Relative Risk to Public Health from Foodborne <i>Listeria monocytogenes</i> Among Selected Categories of Ready-to-Eat Foods. Appendix 8: Growth of <i>Listeria monocytogenes</i> in Foods. <a href="http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf">http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf</a></td> </tr> </tbody> </table> <p>It is noted that the above analysis (Table 1) doesn't seem to accord closely with the text at Section 4.2.4 in the document, e.g., the apparently strong reliance on NSW Food Authority's (NSWFA) 'Potentially Hazardous Foods' document (P.52, Lines 3 to 13) in which NSWFA isn't mentioned. Also, its clear that there is a strong reliance on three sets of expert opinions. Thus, it seems that many of the 'growth potential' classifications were based on expert opinion but, without scrutinizing each published reference (not feasible within the time commitment), it is not possible to comment on whether the additional references cited support the expert opinion. However, in this reviewer's experience some of the growth potential scores appear 'unrepresentative' of expected growth potential in some foods. (This topic is discussed again later in this report).</p> <p>Given the above discussion, it is difficult to provide expert comment on whether the data available on growth potential support the level of granularity of food categories but the large number of references cited suggest that it <i>should</i> be. More importantly (and also elaborated later in this report, <i>see response to '3b'</i>) growth potential could have been estimated using one or more of the on-line predictive microbiology</p>	Ref. No.	Times cited	Details	10	450	Expert opinion (IFT/RTI expert elicitation)	434	165	NSW (2008) Potentially hazardous foods: Foods that require temperature control for safety. NSW/FA/CP016/0810	6035	51	FDA expert opinion, June 2015	6041	46	FDA SME August 2014	454	27	AEM 39(5):943-949 (1980) Cameron <i>et al.</i>	435	26	Betts <i>et al.</i> (2006) Scientific Review of the Microbiological Risks Associated with Reductions in Fat and Added Sugar in Foods	2181	19	FDA (2001) Evaluation and Definition of Potentially Hazardous Foods - Chapter 3. Factors that Influence Microbial Growth	2187	19	FDA. Quantitative Assessment of Relative Risk to Public Health from Foodborne <i>Listeria monocytogenes</i> Among Selected Categories of Ready-to-Eat Foods. Appendix 8: Growth of <i>Listeria monocytogenes</i> in Foods. <a href="http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf">http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf</a>	
Ref. No.	Times cited	Details																											
10	450	Expert opinion (IFT/RTI expert elicitation)																											
434	165	NSW (2008) Potentially hazardous foods: Foods that require temperature control for safety. NSW/FA/CP016/0810																											
6035	51	FDA expert opinion, June 2015																											
6041	46	FDA SME August 2014																											
454	27	AEM 39(5):943-949 (1980) Cameron <i>et al.</i>																											
435	26	Betts <i>et al.</i> (2006) Scientific Review of the Microbiological Risks Associated with Reductions in Fat and Added Sugar in Foods																											
2181	19	FDA (2001) Evaluation and Definition of Potentially Hazardous Foods - Chapter 3. Factors that Influence Microbial Growth																											
2187	19	FDA. Quantitative Assessment of Relative Risk to Public Health from Foodborne <i>Listeria monocytogenes</i> Among Selected Categories of Ready-to-Eat Foods. Appendix 8: Growth of <i>Listeria monocytogenes</i> in Foods. <a href="http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf">http://www.fda.gov/downloads/Food/FoodScienceResearch/UCM197330.pdf</a>																											



<b>CHARGE QUESTION 1.b: Is the granularity of the food classification appropriate and supportable by available data (particularly the secondary commodities)?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	databases, e.g., ComBase ( <a href="http://www.combase.cc">www.combase.cc</a> ), which includes tens of thousands of observations on microbial growth rates in foods, the USDA Pathogen Modelling Program ( <a href="http://pmp.errc.ars.usda.gov/PMPOne.aspx">http://pmp.errc.ars.usda.gov/PMPOne.aspx</a> ), or the French SymPrevious database ( <a href="http://www.symprevious.net/">www.symprevious.net/</a> ), etc.	

<b>CHARGE QUESTION 1.c: Are the method and data used to identify the food-hazard pairs adequate?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	After establishing the categorization scheme, the authors listed all hazards that historically have been associated with each food. As stated in the report, this food list has been associated with known or reasonably foreseeable hazards. Even though this methodology is reasonable and potentially allows for the inclusion of most combinations of foods-hazards, I have one concern: it does not allow for the flexibility of allowing considering new, unseen events. Acknowledging that food contamination and consequently foodborne outbreaks are to a wide extent random and a consequence of so many different factors (e.g., environmental contamination, cross-contamination, carrier food-handlers), I think it would be important to identify food-hazard pairs not only on the basis of outbreaks and recalls, but also on other types of epidemiological evidence on the potential sources of foodborne hazards (see “general impressions”).	
Reviewer #2	Yes, the method and data build from the ground up, starting with the primary data from the literature or other sources. This provides a sound scientific basis for all of the food-hazard pairs.	
Reviewer #3	<p>To some extent it seems driven by publications which do not use the same classification of foods or do not precisely describe the food. Example – <i>Cronobacter</i> spp. (different from <i>Cronobacter sakazakii</i> as per annex B is listed for a few commodities – with the exception of infant formula, none of them has ever been involved in an outbreak. If the purpose is to list all commodities for which isolation of <i>Cronobacter</i> spp. has been described, then the list of commodities is far from being complete.</p> <p>In the report mention is made on the inclusion of “potential” hazards – from my understanding, the possibility to assign a hazard to a commodity based on expert knowledge. One element could be the microbial ecology, manufacturing processes, behavior of microorganisms – while this option seems to exist – it does not seem to have been applied systematically: for example 1249 identifies <i>L. monocytogenes</i> as microbial hazard, 1251 <i>Salmonella</i>. The products are probably very similar by nature, process, etc. – why should there be a difference in type of contamination ...even if not published? <i>C. botulinum</i> and ETEC are hazards for Tofu but not for Tofu products – does not seem consistent in terms of ecology and with the assessment of the process contamination (Tofu as raw material for tofu products).</p>	

<b>CHARGE QUESTION 1.c: Are the method and data used to identify the food-hazard pairs adequate?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>The same question could be raised for other categories – e.g., seafood with completely different microbial hazards for octopus, squids, although quite similar in terms of origin, processing etc., <i>C. botulinum</i> and histamine in smoked finfish but not in dried.</p>	
Reviewer #4	<p>As I understand it, the food-hazard pairs were predominantly based on food outbreak records from USA (i.e., the food-hazard pair has been demonstrated to actually cause illness), but also other surveys and reports that identified the presence of hazards <i>whether or not there was any evidence of adverse health outcomes</i>. This includes recall data and various surveys. The authors of the report/model indicate that they also consulted other databases (e.g., EFSA) to identify further relevant product:hazard pairs.</p> <p>Upon examination of Appendix D, (outbreak data), it is apparent that there are ~411 discrete product:hazard pairs based on USA outbreak data. The model, however, considers 1286 food-hazard pairs and it is not clear how the additional 875 product:pathogen pairs were derived. Greater transparency (documentation) around the process of identifying product:hazard pairs seems required, not the least because some of the food-hazard pairs are somewhat ‘surprising’ selections.</p> <p>By way of comparison, it could be argued that a holistic approach be taken and that for all 258 primary and secondary commodities, all 95 hazards should be considered, leading to some ~24,500 potential food-hazard pairs. Clearly this approach was not taken, and neither do I think it would be necessary. However, it accentuates that the model developers considered many product:hazard pairs to be irrelevant/trivial and eliminated them from further consideration but do not explain how (with the exception of the 411 product:hazard pairs specifically associated with outbreaks) only 1286 product:hazard pairs of the ~24,500 potential pairs were selected for inclusion in the ranking process. This question become more important because of the issue around how to rank product with multiple hazards. A product that has more hazards associated with in the list of product:hazard pair, irrespective of how trivial the actual risk, will have a higher risk ranking because any product:hazard pair in the list will have a risk score greater than 1. This inherent bias needs to be considered in the combination of scores, and will be discussed in greater detail in the complementary report on the model.</p> <p>Given that the potential list (i.e., of all possible combinations) was ‘culled’, it seemed relevant to examine whether there were any inconsistencies in the database, e.g., similar products but with different hazards considered to be associated with them. Scanning through the lists in Appendix C it was apparent that there are such inconsistencies. While my search through the list was not at all exhaustive, some examples are presented below:</p> <ol style="list-style-type: none"> <li>i. <u>Differences in hazard considered relevant to ice or to bottled water.</u> Both products should involve potable water so its not clear why a range of protozoan parasites are considered</li> </ol>	

CHARGE QUESTION 1.c: <i>Are the method and data used to identify the food-hazard pairs adequate?</i>		
REVIEWER	COMMENT	RESPONSE
	<p>relevant hazards in ice, but not in bottled water. If the argument is that bottled water receives a more severe treatment, then why would <i>Salmonella</i> be included as a hazard in bottled water and not a hazard in ice? Alternatively, parasites are more likely to be eliminated by freezing than are bacterial pathogens, i.e., ice might be expected to represent less chance of exposure to viable protozoan parasites. Also, the list of chemical hazards considered relevant to bottled water apparently does not apply to ice. This seems illogical.</p> <p>ii. <u>A long list of vegetative bacterial pathogens are considered as hazards in Grade A, pasteurized, fluid white milk.</u> The aim of pasteurization is to eliminate bacterial pathogens, something which pasteurization does overwhelmingly well and has done so (as the records indicate) for about the last 100 years. Similarly, while <i>L. monocytogenes</i> is considered a relevant hazard in Grade A, pasteurized, fluid white milk it apparently is not considered relevant in flavored milk. Based on the composition and preparation of these products, they represent the same potential for contamination and potential growth, so this difference in hazard lists for such similar products requires explanation.</p> <p>iii. <u>Inclusion of <i>Cyclospora cayatensis</i> as a hazard in Gravies (Dry and liquid).</u> This seems to require explanation because it is not listed as a hazard in other products of similar composition and processing.</p> <p>iv. <u><i>Listeria monocytogenes</i> in dried egg.</u> I'm not aware that <i>Listeria</i> spp. survive desiccation unusually well, nor that dried egg would be used in such a way as to allow <i>L. monocytogenes</i> to grow to high enough levels to have a significant likelihood of causing human illness, even in the severely immunocompromised. Some explanation/justification for this pair would be helpful.</p> <p>v. <u>Norovirus as a hazard in game meats.</u> This seems to require explanation particularly as game meats will usually be cooked before eating. Norovirus on game meats could only arise from external contamination and so would be on the outside of the cuts of meats and would be expected to be easily eliminated during cooking.</p> <p>vi. <u>Dried pasta.</u> Dried pasta is normally boiled for 5 – 15 minutes before consumption and would be expected to eliminate non-spore forming pathogens such as <i>Salmonella</i>, or STECs or <i>S. aureus</i>. Is the hazard that <i>S. aureus</i> might have grown and produced toxins during manufacture of dried pasta? The inclusion of these bacterial hazards in dried pasta requires some explanation.</p> <p>vii. <u><i>Cyclospora cayatensis</i> as a credible hazard in pasta salads (673).</u> If it is a hazard in pasta salads, why are other food-borne protozoan parasites such as <i>Cryptosporidium</i> or <i>Giardia</i>, etc. not included?</p>	

<b>CHARGE QUESTION 1.c: Are the method and data used to identify the food-hazard pairs adequate?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>viii. <u>STECs in RTE deli salads</u>. Why are STECs not considered to be hazards in RTE deli salads (690-691) when they are considered to be a hazard in other fresh cut vegetable products?</p> <p>ix. <u>Salmonella in hot smoked finfish</u> (992) also seems a very unlikely scenario (hot smoking should eliminate them) and this product:hazard pair doesn't appear in the outbreaks list (Appendix D).</p> <p>x. <u>Listeria monocytogenes risk from popcorn</u> (1114) seems negligible, given that popping of corn would eliminate it and that pre-popped corn won't support its growth. Furthermore, this product:hazard pair doesn't appear in Appendix D. Some explanation/justification seems to be required.</p> <p>While many hazards have been detected in foods, it doesn't always mean that there was any significant risk to consumers, and many reports have been published on detection of hazards in foods without any corresponding evaluation or even consideration of risk. In other words, some survey data are not risk-related and their relevance as indicators of risk and to this risk-ranking tool needs to be more closely scrutinized.</p> <p>Similarly, in the outbreak data, the source of contamination does not always seem to have been clearly articulated. It is noted that the report states that the outbreak data were included only if food, rather than food handlers, were considered as the source of the contamination. This aspect might also be further elaborated in the main text.</p> <p>For the scoring to work properly, and particularly in connection with Criterion 2, Criterion 3 (and the data used to support) cannot simply be about prevalence (i.e., simple detection of the hazard in the product) but would have to relate to detection of the hazards at levels approaching that required to produce symptoms in a significant proportion (e.g., &gt;1%?) of those exposed so that the scoring of Criterion 2, about disease severity, is meaningful. This consideration also impinges on Question 8, and is discussed further in the response to that Question.</p> <p>A more robust search of outbreak data from other nations, particularly nations with analogous life-style and food cultures, would probably yield further valuable information. For example, there have been numerous outbreaks of Hepatitis A in Europe (and more recently Australia) linked to frozen berry fruits but, while strawberries can be contaminated with HepA (and it is included in the list) many other frozen berries can be as well but are not included in the list (or was Strawberries:Hep A included as a representative?) Also, it is noted that the EFSA outbreak data were also explored and used to build the list of product:hazard pairs but the only reference for HepA and strawberries is the somewhat obscure "FDA (2015) Orange Book Database, Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations. <a href="http://www.accessdata.fda.gov/scripts/cder/ob/">http://www.accessdata.fda.gov/scripts/cder/ob/</a>", and in which I could find no</p>	

<b>CHARGE QUESTION 1.c: Are the method and data used to identify the food-hazard pairs adequate?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	reference to Hepatitis A in strawberries! Conversely, Maunula <i>et al.</i> (2014, <i>Int. J. Food Microbiol.</i> , 167(2)177-185) does provide a good introduction to enteric virus problems in berry fruits in Europe and is listed in the references as REF 3016.	

<b>CHARGE QUESTION 1.d: If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	The authors explain in the report that they have investigated alternative categorization schemes (like CDC's scheme by Painter <i>et al.</i> ), and justify well why they have chosen to adopt the here presented one. Another option would be some kind of hybrid, which could allow having the advantages of the currently used scheme (i.e., in line with preventive control measures by FDA, and considering processing), but also the inclusion of evidence other than outbreak data in the risk ranking model. More specifically, including "meat and poultry" primary and secondary commodities would allow for listing a wide variety of microbial hazards (which are very common causes of foodborne disease, such as <i>Campylobacter</i> -chicken), as well as chemical hazards (such as PAHs and meats (grilled, fried, etc.)). Such an approach could also allow for classifying ingredients of unspecific composite foods (like sandwiches and RTE foods).	
Reviewer #2	See comment 1b.	
Reviewer #3	Not sure there is a "magic" way to establish for the classification other than establishing rules on what needs to be considered which includes as well a cross-check between categories – for example if a hazard is identified in a commodity, then it must be assessed whether another commodity including this has a similar profile – see Tofu example.	
Reviewer #4	As noted earlier, there is likely no perfect method for identification of product:hazard pairs and the approach presented seems reasonable, but can still be improved. Alternative approaches might involve a "Failure Mode Effects Analysis", starting with each of the 49 FDA-regulated commodity groups and asking the question: "is it possible for any foods in any category to cause human food-borne illness and, if so, how could that occur?" This would naturally lead to identification of sub-divisions of the main categories into categories of product that were more likely (whether due to natural hazard occurrence in the food, differences in processing, potential for microbial growth, etc.) to be sources of human illness. Any such approach would also have to include assignment of (relative) probabilities to the potential fault modes as part of the means of identifying the most relevant product:hazard pairs. Additive scoring schemes, even if based on exponential scales, cannot capture the full complexity of factors that contribute to food-borne risk. As a simple example, a listericidal process in a hermetically sealed product effectively eliminates the risk from that hazard in that product, irrespective of what	

**CHARGE QUESTION 1.d: *If not, please describe another practical alternative to this food classification scheme that might be considered. Please describe what changes, if any, might be considered and provide examples of additional food-hazard pairs (if any) that might be considered in the future.***

REVIEWER	COMMENT	RESPONSE
	<p>occurred before or what will happen later to that product, as long as package integrity is preserved. Equally, proper cooking can completely eliminate many hazards. The current scoring scheme does not enable a “reset to zero” for cidal processes, as could be achieved with a multiplicative scoring scheme that includes ‘zero’ values for some of the inputs.</p> <p>Importantly, (and as noted above) as no simple scoring scheme is likely to correctly rank all product:hazard pairs and will of necessity be a compromise, it will be important to achieve consensus among stakeholders that they accept the risk scoring decision tool as being the best, or at least as good as any other approach. Allowing stakeholder comment and input and requests for clarification of the basis of the scoring scheme and its outputs, and researching and documenting those answers, will be critically important in achieving that consensus and also lead to refinement and improvement of the food categories and identification of important factors that discriminate relative risk by improving the logical robustness and broad applicability of the model.</p>	

**CHARGE QUESTION 2: *Please evaluate the scoring definitions for the seven criteria:***

**CHARGE QUESTION 2.a: *Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.***

REVIEWER	COMMENT	RESPONSE
Reviewer #1	<p>The seven criteria chosen to score food-risk pairs are adequate and comprehensive. In general, the scoring criteria are well defined. When explaining data sources for the criteria, it is extremely helpful that the same 3 (at least) examples (1 micro, 1 chemical, 1 allergen) are followed.</p> <p>I have some specific comments on the description of some of the criteria (in addition to the responses given below):</p> <p>Criterion 1:</p> <ul style="list-style-type: none"> <li>It is stated that this criterion is applicable to microbial and chemical hazards, and that the epidemiological link is based on the frequency and number of outbreak-related cases. However, it is also mentioned that chemicals and undeclared allergens have not been involved in outbreaks, so evidence that these hazards cause disease, and how much disease they cause in the population, is based solely on experts. It is of general agreement within the scientific community that the</li> </ul>	

**CHARGE QUESTION 2.a: Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.**

REVIEWER	COMMENT	RESPONSE
	<p>epidemiological evidence for the casual effect of exposure to most chemicals and disease is frequently inconsistent, and the estimation of the incidence of disease that can actually be attributed to these chemicals complicated. Still, epidemiological and animal studies assessing these relationships exist, and an assessment of all available evidence seems crucial to make inferences on the relative occurrence of foodborne disease due to exposure to chemicals.</p> <ul style="list-style-type: none"> <li>• It is described that “both the frequency of reported outbreaks and the occurrence of illnesses” (page 14 of the main report) (i.e., the number of reported outbreaks AND sporadic cases per year) are used in scoring. However, it is mentioned in the following page (page 15, line 10) that only outbreak data were used for this criterion (i.e., data on reported sporadic cases or estimates of total illnesses corrected for underreporting – burden of illness estimates) were not included. Aside from my concerns about the exclusion of sporadic disease in this criterion, the initial statement should be corrected.</li> </ul> <p>Criterion 2:</p> <ul style="list-style-type: none"> <li>• I have concerns about the indicators used to describe the severity of disease by the different hazards. Specifically: <ul style="list-style-type: none"> <li>○ For hazards that cause acute disease and for which quantitative data were available, only hospitalization and mortality rates are used. However, some foodborne diseases may lead to severe sequelae that do not necessarily reflect on a high mortality or hospitalization rate. It is described that these were not considered, but they can be relevant to describe severity. For example, Irritable Bowel Disease and Guillian-Barré Syndrome (potential sequelae of <i>Salmonella</i> and <i>Campylobacter</i> infection, respectively) can be very severe in the sense that they have large negative influence on the quality of life of the individual, but do not necessarily reflect in high rates of hospitalization (and certainly not mortality rates). One way of taking these differences into account and to be able to compare across hazards would be to use the disability weights (DWs) generated by the Global Burden of Disease (GBD) study as a basis for the indicator of severity. DWs reflect the impact of a health condition in terms of health-related quality of life estimated based on preferences obtained from a panel of judges. Preferences are defined as quantitative expressions or valuations for certain health states, which reflect the relative desirability of the health states. (See Salomont <i>et al.</i>, 2013). DWs for a wide range of conditions (which I expect include the vast majority of adverse health effects of foodborne hazards) have been estimated, and using DWs would allow for being much more comparative across hazards.</li> </ul> </li> </ul>	

<b>CHARGE QUESTION 2.a: Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<ul style="list-style-type: none"> <li>○ I feel that the semi-quantitative scale used for Criterion 2 does not adequately characterize the severity of different hazards. As described, the severity score for <i>Salmonella</i> and <i>Listeria</i> is the same (9), and – somewhat surprisingly - would be the same for adverse effects caused by exposure to some chemicals, such as cancer. However, the severity of disease by these hazards is very different. The authors argue that this will be compensated by Criterion 7, but the example of <i>Salmonella</i> and <i>Listeria</i> illustrates that this is not necessarily the case – a 3 times higher cost of disease does not necessarily compensate for the difference in severity. It may however balance if (true) incidence is taken into account (which is the highest for <i>Salmonella</i>), but it seems to me that this would require using corrected incidence estimates (BoI) in Criterion 1.</li> </ul>	
Reviewer #2	<p>C1. The scoring is appropriate for biological, undeclared allergens, and chemical hazards.</p> <p>C2. Table 2-1 is unclear. Are hospitalization rates and hazards being scored separately, or do hospitalization rates follow from the severity of the hazard, in which case they refer to the same thing? That is, severe hazards result in high hospitalization rates. It's difficult to imagine a severe hazard with a low hospitalization rate or a moderate hazard with a high hospitalization rate. Thus, I think the scoring double counts for the same characteristic.</p> <p>C3. The scoring is appropriate if the percentiles refer to % of samples.</p> <p>C4. The scoring is only for “the temperature at which the food is <i>intended</i> to be held and stored.” The potential for growth under abuse conditions should also be considered. The growth potential for e.g., potatoes salad would be high under abuse conditions, whereas foods with <math>a_w &lt; 0.85</math>, <math>pH &lt; 4.6</math> would have low growth potential. Use of three factors would, of course, require presentation as a cube rather than grid. This may be beyond the scope of the program.</p> <p>C5. The scoring is appropriate.</p> <p>C6a. The scoring for consumption is appropriate.</p> <p>C6b. This may be off-topic but should be noted somewhere: increased consumption does not equate to increased risk. Consider “Food A” which has 50 cases and 100,000 units of consumption versus “Food B” which has 5 cases and 1,000 units of consumption. Under the present scheme, based on the number of cases without normalization for consumption, Food A would appear to have greater risk, when in fact</p>	



<b>CHARGE QUESTION 2.a: Are the definitions appropriately defined for microbial hazard, chemical hazard, and undeclared allergens? If not, please describe changes that might be considered and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>Food B is ten-fold more risky. Because of this, the score for C6 should be weighted, perhaps as 7 rather than 10.</p> <p>C7. The scoring is appropriate, as is a weighting of 5.</p>	
Reviewer #3	<p>The rationale for biological toxins is not completely clear and consistent – on one side some microbiological toxins are missing, such as staphylococcal enterotoxins which can be present in the absence of <i>S. aureus</i> (die-off), on the other side the element “ability of growth” if used to classify metabolites such as histamine. For mycotoxins however, where elements such as ability to grow, the step at which contamination takes place, and shelf-life are important to the formation of mycotoxins, it is considered as “inert” as would be arsenic or lead.</p> <p>In terms of allergens, 8 are considered and it is understood that “undeclared allergens” is considered a “worst case” that covers all of them, i.e., without taking into account whether it is likely to occur or not (e.g., fish in dairy products...). It is however not clear why hazard sub-types 91–95 have been defined, what about eggs, soybean, tree nuts, for which no sub-types exist. What is the rationale for this difference?</p>	
Reviewer #4	<p>The hazards are appropriately described for microbial, chemical and allergen hazards. Some of the chemical hazards listed are of microbial origin (algal and fungal toxins), but are usually already present in the food (prior to processing) and would not increase in level during processing, distribution, etc. Histamine levels can increase as microbial metabolism continues, and this difference is appropriately identified in the approach developed, and dealt with by considering histamine as a microbial hazard.</p>	

<b>CHARGE QUESTION 2.b: For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	<p>If Criterion 5 is defined for different food-hazard pairs taking into account if a food is RTE or not – i.e., if the indicator “steps taken to reduce contamination” for pairs with RTE foods are classified as “weak”, and thus Criterion 5’s score would be minimum 3 for that pair, I think it is reasonable to have the same definition for RTE and non-RTE foods in Criterion 3.</p>	
Reviewer #2	<p>It is generally accepted that RTE foods present a greater risk than NRTE foods. This should be taken into account in the model. The suggested C3 weight of 15 for RTE food and 10 for NRTE foods is appropriate.</p>	
Reviewer #3	<p>The likelihood of contamination is, in my opinion, not dependent on whether it’s an RTE or NRTE food but is the outcome of food supply chain and the control measures applied during the manufacture up to</p>	

**CHARGE QUESTION 2.b: For Criterion 3 (C3), the same definition is used for ready-to-eat (RTE) and not-ready-to-eat (NRTE) foods. Should a different criterion weight to C3 be applied to the food risk score for an RTE vs NRTE food-hazard pair? If so, please specify the weighing scheme that might be considered (e.g., C3 weight of 15 for RTE food vs. weight of 10 for NRTE food).**

REVIEWER	COMMENT	RESPONSE
	<p>consumption. Hence the data gathered reflect the percent contamination rate and should already reflect these differences. Criteria 4 and in particular 5 are also linked to the differentiation of the two categories. Different weighing would also require to take into consideration the likelihood of misuse – i.e., how likely is it that a NRTE food is not cooked before consumption – should it then be the weighed as a RTE food?</p>	
Reviewer #4	<p>I presume that the intent of this question is to ask whether NRTE foods would be expected to receive a heat treatment prior to consumption, i.e., that treatments received by NRTE foods would greatly reduce the risk to consumers if the hazard were present. If so, the risk reduction would be orders of magnitude in an NRTE for infective microbiological hazards (see also comments under Question 1d), compared to RTE foods, and that should be explicitly ‘factored in’ in the scoring.</p> <p>Allergens and chemical hazards would be unaffected by cooking, but so too would histamine, currently listed as a microbiological hazard, as would the risk from <i>S. aureus</i>, or <i>B. cereus</i>, etc., because both produce heat-stable toxins as they grow in foods, i.e., cooking for these hazards does not lessen the consumer risk. As such, the adjustment to the scoring scheme that would be required to accommodate this issue is rather more complex than a simple weighting for RTE <i>cf.</i> NRTE could accurately reflect. It’s likely that a new “Criterion” might be necessary to include the importance of cooking as a terminal risk reduction step. Such a Criterion might need to include negative values in the scoring scheme (i.e., indicating that the risk is greatly reduced in some cooked foods), and would relate principally to microbial hazards for which the mode of illness is gastrointestinal infection or those few cases (e.g., <i>C. botulinum</i>) of microbial growth producing a heat labile toxin, or other chemical hazards that are heat labile.</p>	

**CHARGE QUESTION 3: The draft model integrates a substantial amount of public health and commodity-specific data about food-hazard pairs, as well as information regarding manufacturing and processing of different foods.**

**CHARGE QUESTION 3.a: Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.**

REVIEWER	COMMENT	RESPONSE
Reviewer #1	<ul style="list-style-type: none"> <li>• Criterion 1: The report uses official data for outbreak data (CDC and FDA), which are the appropriate data sources. Still (as described earlier), I have concerns about the use of outbreaks (only) as an indicator of the occurrence of disease in the population (i.e., for the epi link). On one hand, outbreaks caused by foodborne pathogens are recognized as a small fraction of cases by</li> </ul>	

<b>CHARGE QUESTION 3.a: Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>foodborne pathogens (this proportion varies, and it's particularly small for some pathogens, like <i>Campylobacter</i> (&lt;1%); for <i>Salmonella</i> was 6% in 2013: <a href="http://www.cdc.gov/foodnet/reports/annual-reports-2013.html">http://www.cdc.gov/foodnet/reports/annual-reports-2013.html</a>). Even though foodborne illnesses by microbial hazards are known to be highly underdiagnosed and underreported, several studies have made efforts to estimate the “true” incidence of a wide variety of foodborne diseases. In the U.S., Scallan and co-authors have estimated the burden of illness of 31 pathogens. This study’s estimated multiplication factors are used in this report in Criterion 7 (see comments below), and I would think that it would make sense to use them in this criterion as well. In other words, I believe the overall incidence of disease (sporadic + outbreak cases) should be used as an indicator of the epidemiological link. On the other hand, and as well described in the report, chemicals and undeclared allergens seldom or never cause outbreaks, and therefore this type of evidence cannot be used as an indicator for disease. In addition, because many chemical hazards cause chronic disease, which is difficult to link with exposure, incidence data are lacking, and I believe experts would also have difficulties comparing disease by different hazards in the absence of incidence estimates. This task is particularly difficult because many of these chemical hazards are (considered potentially) carcinogenic, but the probability of a given individual developing disease is determined by many other factors. I recognized that these are difficult to derive, and there are few studies available so far, but believe that it is an important piece of information and a research gap that would be valuable to address, particularly in a risk ranking effort like this one. I am aware of ongoing research efforts that would be worth including in the analysis. Also worth mentioning is that the WHO’s initiative to estimate the global burden of foodborne diseases (FERG, <a href="http://www.who.int/foodsafety/areas_work/foodborne-diseases/ferg/en/">http://www.who.int/foodsafety/areas_work/foodborne-diseases/ferg/en/</a>) also looked into the burden of some chemical hazards (at the regional level), and results will be publicly available latest in December 2015.</p> <ul style="list-style-type: none"> <li>• Criterion 2: To describe severity of different diseases, hospitalization and mortality rates from Scallan <i>et al.</i> (for microbial hazards), and expert opinions (for chemicals and undeclared allergens) were used. It seems that the data source is appropriate for these rates, but the fact that Scallan did not consider potential sequelae of foodborne diseases hampers a complete picture of the severity of the different pathogens, and thus comparability of diseases. As examples, in addition to acute diarrhea, which is common to many of the considered hazards, <i>Salmonella</i> infections can lead to reactive arthritis, irritable bowel disease, irritable bowel syndrome, <i>Campylobacter</i> to these and to Guillian-Barré Syndrome, VTEC to hemolytic uremic syndrome, etc. Even though these diseases are rare and may also result in hospitalization and even death (in some cases), most will probably not be linked with the foodborne infection, because they occur</li> </ul>	

<b>CHARGE QUESTION 3.a: Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>relatively long after exposure and acute disease (HUS by VTEC and abortion by <i>Listeria</i> are examples of exceptions). Methodologies that allow for the inclusion of sequelae and their severity are available (health metrics like disability adjusted life years - DALYs), and would allow for a better comparability between diseases. In addition, this measure of severity (DW) would allow for a better comparability with disease by chemical hazards. As examples, the DW for cancer as estimated by the GBD 2010 is around 0.5, for diarrhea from 0.061 (for mild cases, &gt;80% for most pathogens) to 0.281 (for severe, ~3%), and for HUS 0.21 (Salomon <i>et al.</i>, 2013). I believe that the fact that the final score for Criterion 2 (which is 9, the maximum, for ~50% of considered hazards) is for example the same for <i>Salmonella</i> spp. and <i>Salmonella Typhi</i>, and the same for <i>Yersinia</i> and acrylamide (which can cause cancer) illustrates the limitations of the current data for comparing between diseases.</p> <ul style="list-style-type: none"> <li>• Criterion 7: I have some concerns about the approach and data used for this criterion. Like in Criterion 1, the authors use outbreak cases, but here these data are corrected for underreporting using Scallan's multipliers. However, these multipliers were derived to correct all illnesses captured by different laboratory surveillance systems, i.e., including sporadic illnesses. As mentioned before, sporadic cases are, for many pathogens, the majority of all reported cases. It is recognized that outbreak-related cases are more likely to be reported, and thus applying an overall underdiagnosis/underreporting factor to these is not necessarily appropriate. Even though this could be irrelevant in terms of disease burden in general – because the estimated burden of a disease would still be lower if outbreak cases are used as a basis for estimations, as opposed to all reported cases –, it may make a difference when the risk ranking model has as a starting point the food-hazard pair. In other words, as I understand it, in this criterion the model is correcting the (outbreak) cases caused by a food-hazard pair using the UR factors derived for that disease, which could lead to an overestimation of the relative contribution of that specific source for disease with that pathogen. The limitations of using outbreak cases for source attribution (which I think also apply here) have been widely discussed (e.g., Pires <i>et al.</i>, 2009; Hald <i>et al.</i>, 2004, Batz <i>et al.</i>, 2004, etc.).</li> </ul> <p>I also have trouble understanding why UR factors of 1 were used for pathogens for which multipliers have not been estimated. For example, even though Scallan did not publish a multiplier for norovirus (because a top down approach was used, as opposed to a bottom-up/pyramid approach), that study did estimate burden of illness for norovirus (i.e., unreported cases), and the gap between reported cases and estimated burden was extremely large. This study estimated a total of ~21 million cases (which corresponds to 5,461,731 foodborne cases) in the study year (U.S. Census population data from 2006 was used, just for reference). This can be compared with the number of outbreak-related reported cases used in this project.</p>	

<b>CHARGE QUESTION 3.a: Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>Data from 2009-2010 shows 69,145 norovirus outbreak-related reported cases (<a href="http://wwwnc.cdc.gov/eid/article/19/8/13-0482_article">http://wwwnc.cdc.gov/eid/article/19/8/13-0482_article</a>). (From appendix D1 I can see 2,701 for 2006?).</p> <p>In addition, the model used surrogates for the cost of illness (CoI) for some pathogens; it would be helpful to understand why specific surrogate data were chosen. For example, the cost for norovirus illness was used for parasites. Were these defined on the basis of similarities between symptoms and duration of disease?</p> <p>It was also not clear how the experts provided input for chemicals and allergens. For e.g., some chemicals' health effect is cancer, and the costs of disease have been estimated. The problem with these hazards is estimating the incidence of disease – for which there are no data and few studies. Was there some type of scientific basis for the experts' input?</p>	
Reviewer #2	The data sources are comprehensive and appropriate.	
Reviewer #3	<p>In principle all the sources of information are appropriate – the only question is whether, taking the amount of documents, publications, etc., everything has been taken into account. From a few examples it seems that there could be a bias – e.g., <i>Cronobacter sakazakii</i> which has been isolated from numerous commodities but only involved in outbreaks with infant formula. In terms of criteria, there is probably much more information available than what is indicated – one example is ICMSF Book 8 which discussed the relative role and impact of contamination along the food chain. This does not seem to have been taken into consideration – compared to other ICMSF books. At least it is not mentioned in chapter 2.2.5 or 4.2.5.</p> <p>It seems that only U.S. data have been used for outbreaks (appendix D1) while other references are listed in appendix E. Some product and product categories have an extensive listing of data – while for others, relevant outbreaks (not in the U.S.) do not seem to have been taken into consideration. But it is very difficult to judge from appendix E what has been included (sometimes no title) and what not as the list is valid for all criteria, includes reports based or including themselves on reviews and compilations What can be commented is that for certain publications or books, probably much more references are used.</p>	
Reviewer #4	The generic data sources (i.e., rather than specific documents or research publications, which are also appropriate) are appropriate sources for the data required. In general, where ideal data can not be identified and surrogate data are used, the reasons for those choices are identified and the process is transparent (i.e., well-documented). In some cases greater specificity/clarity about the data could be presented. For example, in the identification of pathogen:product pairs, I had to read closely to determine whether any pairs identified (apart from the 411 derived from outbreak data) were selected simply because some-one/organization had done a screen for those hazards in a range of products and found the hazard,	

<b>CHARGE QUESTION 3.a: Are any of the data sources used not appropriate for any of the seven criteria? If so, please explain which data source(s) should not be used and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	i.e., whether there was any reason to believe that those product:hazard pairs had, or were ever likely to, cause human foodborne illness. Also, the severity data used for Criterion 2 could be handled in the same way as suggested by Minor <i>et al.</i> (2015), as was used in Criterion 7, (i.e., the <u>mean severity</u> of illness was calculated based on the weighted mean of i) the estimated unreported cases, ii) cases seeking medical attention, iii) deaths, etc.). This is discussed further below.	

<b>CHARGE QUESTION 3b: Are there data sources not yet used but that should be considered? If so, please provide specific examples of data source(s) for each criterion, and explain why the additional data sources might enhance the criteria scoring.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	Please see detailed answers above.	
Reviewer #2	The underlying data used for the FDA/EPA advisory on methyl mercury in fish would be relevant in determining the risk score for the finfish:methyl mercury pair.	
Reviewer #3	See comments on C5.	
Reviewer #4	<p>As noted in response to Question 1b, through analysis of the data sources used (e.g., assessing the product:pathogen pair evidence) it seemed that a wider range of credible sources was used than was identified in the text of the main report. Also, as noted in response to Question 1b, predictive microbiology databases (ComBase, Pathogen modelling program, others) probably should have been used to assess pathogen growth potential in foods, whether by matching foods to those in the databases, or by deriving composition data (e.g., pH, a<sub>w</sub>, organic acids levels, etc.), for foods of interest and using that to generate growth rate predictions, which can in turn be used to estimate potential pathogen growth within the typical packaging (vacuum packed, aerobic, enriched CO<sub>2</sub>, modified-atmosphere packed, and storage environment (temperature, gaseous atmosphere), etc.</p> <p>If required, further information on consumption of specific products could be obtained by inference from supermarket sales figures and the market share that that business controls, or for major producer's production figures in conjunction with information on their market share.</p>	

<b>CHARGE QUESTION 4: Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	The weighting factors seem appropriate.	

**CHARGE QUESTION 4: Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.**

REVIEWER	COMMENT	RESPONSE
Reviewer #2	<p>If “age” of the data is being used as a proxy for “quality” of data, weighting is not appropriate since the best science available may not be the most recent data. If, however, the weighting is meant to reflect the current situation in a progressive fashion (rather than retrospective) in keeping with the intent of the model, the weighting is appropriate.</p>	
Reviewer #3	<p>One can argue that outbreak is outbreak, whether it occurred today or in the past. The weighing is based on an assumption that control measures were not as effective in the past which is probably questionable for quite a number of food commodities – e.g., control measures for numerous food commodities are much older than 1998 – e.g., heat processes such as pasteurization, sterilization, hygiene control measures related to environmental contamination have been described, or included in guidance documents, e.g., for chocolate or dairy products before 1998.</p> <p>However if this is accepted as assumption – then the question is whether or how far this is not taken care in the determination of C5.</p>	
Reviewer #4	<p>In some cases it is relevant to discount older studies, such as when it is known that industry/regulators have taken measures to reduce the incidence of a specific foodborne illness by new methods and interventions, i.e., that might be expected to have resulted in a sustained lower prevalence of specific hazards. This would be true for <i>Listeria monocytogenes</i> in many RTE foods in USA, where incidence of illness and prevalence in relevant products has probably declined by 30 – 50% in the last 10 – 15 years due to greater awareness and risk management by regulators and industry. Equally, there have been strong regulatory programs and industry actions in USA to reduce rates of STEC illness, and <i>Salmonella</i>, from red meats, or produce, etc.</p> <p>Similarly, as chemical detection methodologies have improved (i.e., their detection sensitivity has increased) the apparent prevalence (i.e., ‘detections’, regardless of contamination level) of chemical hazards in foods has probably increased even though there has been no actual change to consumer risk.</p> <p>As another example, the advent of PCR and even newer pathogen detection technologies based on specific chemical marker (e.g., DNA, or RNA, infra-red spectroscopy) has enable much greater throughput of samples, but not always with confirmation that positive signals came from live (and infective) organisms, <i>cf.</i> molecular residues from non-viable (‘dead’) cells. As such, caution must be exercised when collating older data to generate information relevant to the current situation, and weighting of different data sources, whether temporally or by geographic regions, may be necessary and useful. However, in the absence of specific evidence that older literature on product:hazard pairs is no longer relevant, more justification is</p>	

**CHARGE QUESTION 4:** *Data weighting is used for Criteria 1 and 3, where more weight is given to more recent data and studies more relevant to the U. S. food supply. Are there any of the weighting factors not appropriate? If so, please explain what other weighting factors should be considered.*

REVIEWER	COMMENT	RESPONSE
	required for weighting of different subsets of the data. Importantly, the weighting applied needs to be determined on a hazard-by-hazard basis: for reasons discussed above, appropriate weightings are likely to differ for different hazards.	

**CHARGE QUESTION 5:** *Consumption data from the NHANES What We Eat in America database were used and, for certain commodities (e.g., ice) where no data or limited data were available, expert opinion was used in scoring Criterion 6. Is the timeframe used (2011-2012 or 2009-2010 survey cycle) appropriate?*

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Refer to 5a.	
Reviewer #2	Yes, the time frame is appropriate.	
Reviewer #3	Yes.	
Reviewer #4	Refer to 5a.	

**CHARGE QUESTION 5a:** *Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.*

REVIEWER	COMMENT	RESPONSE
Reviewer #1	The timeframe used seems appropriate, and, because the purpose is to estimate the proportion for the population that is exposed (i.e., consumes) each food, I agree with using the rate from the day with highest consumption.	
Reviewer #2	The use of highest consumption is appropriate since it will produce the most conservative model.	
Reviewer #3	Seems Ok – could possibly (if not already done) be compared with the approach applied in FAO/WHO risk assessments to ensure consistency on the way it's done.	
Reviewer #4	The NHANES data is reported in the document to enable consumption of all 385 commodity groups to be estimated. The report states that because Day 1 and Day 2 estimates of consumption were not always equal, the highest daily rate estimated was used, rather than the mean rate. While the differences are probably trivial in most cases, the text of the document states that: <i>“It should be mentioned that in addition to calculating consumption rates based on the sum of a commodity, the use of the mean and maximum rates were also explored. After reviewing these methods, using the sum of consumption rates was chosen here as the most appropriate”</i> , but without clear explanation of <i>why</i> it was considered the most appropriate approach. Given the size of the samples, the average seems a more reliable and logical	



**CHARGE QUESTION 5a: *Is it appropriate to use the highest consumption rate from day 1 or day 2 in the scoring? If not, please explain what changes should be considered and what other data sources for consumptions might be considered to improve consumption estimates.***

REVIEWER	COMMENT	RESPONSE
	<p>way of synthesizing the available data to make the most representative assessment. While the difference between the mean and maximum of the two day's values would be expected to be small in most cases (given the size of the samples), it is not possible to independently verify that expectation from the data presented in the C6 results spreadsheet. Again, justification/explanation of the use of the maximum (or sum?) and more evaluation of the effects of that decision are required for transparency in the scoring tool and its documentation.</p>	

**CHARGE QUESTION 6: *Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?***

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Refer to 6b	
Reviewer #2	<p>All food:botulism pairs were evaluated and compared, seeking to understand both the scoring of each pair and why the scoring of some pairs differed for some of criteria in a given food:risk pair. The data for each pair were consistent and the differences among them transparent.</p> <p>The finfish:methyl mercury pair was also evaluated in greater depth, and for chemical hazards in general, the scoring was found to be more subjective. There are several issues that make the finfish:methyl mercury pair difficult to evaluate: 1. There is high risk for the sensitive population (pregnant women, nursing mothers, and young children), but low risk for the general population. 2. High-risk fish are aggregated with low-risk fish into a single category of "finfish." 3. There is a huge data gap. There is compelling evidence that consumption of methyl mercury can lead to deficits in neurological and cognitive development. The PTWI has been determined to be 1.6 µg/kg body weight and a guideline for methyl mercury levels in fish set at 0.1-1.0 mg/kg (ref 47). However, the distribution of methyl mercury concentrations in fish is unknown, so there cannot be a direct connection between fish consumption and the PTWI.</p> <p>n.b. The FDA/EPA Joint Advisory on Methyl Mercury in Fish (2004 and 2014) states that pregnant women, nursing mothers, and young children "<i>should not eat</i>" swordfish, tile fish, king mackerel and shark. This implies a large risk even though the risk ranking score is low. There should be intra-agency agreement regarding the risk of consuming large fish. (A large volume of supporting data was provided to the FDA</p>	

<b>CHARGE QUESTION 6: Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>Food Advisory Committee in the 2004 advisory but only one (Choy, 2009) was cited in the 2014 document).</p> <p>Although I was furnished with the documents which were cited as references for the underlying data, it would not be possible for an end-user to determine the “correctness” of the scoring if the references were not readily available. In many cases it is impossible to access the underlying data because the relevant documents are not available online (i.e., the paper had to be purchased if one was not a subscriber). This issue could be overcome if there were hot-links from the score to the reference and from the reference to the actual document. Some of references are incomplete, e.g., a reader may not know that JFP refers to the Journal of Food Protection. At the very least, the full citation, including the title of the document, should be provided in the reference list.</p> <p>C1 is scored as a “1” based on refs 4, 42, 52, and 60, which as noted below, are inappropriate. Based on my knowledge and experience, a C1 score of 9 would be appropriate because there is compelling evidence that consumption of methyl mercury causes adverse health effects in the U.S. There is wide consensus that consumption of methyl mercury can lead to deficits in neurological and cognitive development in sensitive populations.</p> <p>C5 is scored as a “3” based on ref 6041, input from subject matter expert. This reference is difficult to evaluate, but C5 could be rated as a 9 because there is a high incidence of contamination and weak control. (Granted that the distribution of methyl mercury in fish is unknown and if available, data for the distribution in large fish could justify a score of 3).</p> <p>The differences between risk scores generated by the program (190) and the alternate scoring proposed here (250) could be enough to move the finfish:methyl mercury pair from a low-risk to a high-risk food. Whatever the final risk score is, it should be consistent with the risk presented in the FDA/EPA joint advisory.</p>	
Reviewer #3	<p>Scones – why are microbial pathogens included? As indicated somewhere else <i>Cronobacter</i> spp. are ubiquitous, and finding it in any type of commodity is not too difficult – relevance is however very limited.</p> <p>Beverage bases – relevance of <i>C. botulinum</i> is questionable, no reference on why it’s included (even if rated 0).</p> <p>Coffee – virus – without more details on what type of coffee, it is not very useful – liquid coffee, roasted coffee?? Same hold true for <i>Salmonella</i> – “0” but nevertheless included and no reference.</p>	

<b>CHARGE QUESTION 6: Please select primary commodities and hazards within your expertise and review the scoring of a few food-hazard pairs in the criteria files. Are any of the references used not appropriate?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>Tea – <i>Salmonella</i>, there have been 2-3 outbreaks in Europe (herbal tea) which do not seem to have been considered</p> <p>Chocolate – <i>Salmonella</i>, several outbreaks (e.g., Germany) and recalls (UK, Germany). On the other hand never heard about STEC.</p> <p>Confections – <i>Cronobacter</i> spp., see earlier comments – not relevant.</p> <p>Dried milk – <i>Cronobacter</i> spp., irrelevant as only infants sensitive. On the other hand <i>Salmonella</i> outbreak occurred in France 2005.</p> <p>Frozen fruits (berries) – Several outbreaks – at least one with several hundreds of cases – in Europe.</p> <p>Dry instant breakfast – <i>Cronobacter</i> spp. Same comment as above.</p> <p>Infant formula – <i>Cronobacter</i> spp. Strange to see a C1 = 0 considering that WHO/FAO classifies this a high-risk food and that a number of outbreaks have involved more than one baby. Same comment for <i>Salmonella</i> – there have been several outbreaks since 1998 but they do not seem to have been taken into consideration.</p>	
Reviewer #4	Refer to 6b	

<b>CHARGE QUESTION 6.a: Are there any underlying data not appropriately used?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	Refer to 6b	
Reviewer #2	<p>Re: the finfish:methyl mercury pair:</p> <p>Ref 4 gives data for outbreaks, but not the chronic exposure data which are required for the scoring.</p> <p>Ref 47 sets the PTWI as 1.6 µg/kg body weight, but does not relate this to consumption of fish. The report notes that if there were a known distribution of methyl mercury for the aggregated “finfish,” it would be dominated by species that do not have a high concentration of methyl mercury.</p> <p>Ref 52 provides information similar to ref 47, but focuses on the neurodevelopmental effects. It does not provide the consumption or incidence data needed for the scoring.</p> <p>Ref 60 grapples with the issue of risk v. benefits and sensitive populations. It concludes that for finfish consumption the benefit of polyunsaturated fats outweighs the risk from methyl mercury in women of childbearing age. However, this report aggregates high-risk and low-risk species of fish. This reference notes that FDA advises the sensitive population not to eat the four species of fish that are high in mercury.</p>	

<b>CHARGE QUESTION 6.a: Are there any underlying data not appropriately used?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	At the end of the day, the scoring of both C1 and C5 of the finfish:methyl mercury pair might come down to expert opinions. If this is the case, the opinions which gave rise to the FDA/EPA joint advisory should be given considerable weight.	
Reviewer #3	[Reviewer did not comment.]	
Reviewer #4	Refer to 6b	

<b>CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	<p>My expertise is within microbial hazards, and so I will mainly review in detail pairs with some pathogens. <i>Salmonella</i> appears in 4 of the first 20 pairs in the ranking (in this order):</p> <ul style="list-style-type: none"> <li>- Tomatoes-<i>Salmonella</i></li> <li>- Sandwiches-<i>Salmonella</i></li> <li>- Shell eggs-<i>Salmonella</i></li> <li>- Fresh salsa-<i>Salmonella</i></li> </ul> <p>Assessing the epidemiological evidence for all pairs (Criterion 1), it is evident that the number of outbreaks and outbreak-related illnesses varies substantially (from 492 (weighted) outbreak cases from fresh salsa to 2,678 from eggs); still, they all fit into scoring 9 according to the scoring bins defined, with the exception of salsa, which had &lt;13 outbreaks reported in the time period). Data sources seem appropriate.</p> <p>The criteria that drive these different rankings are mainly prevalence and economic impact. For sandwiches, contamination data were very poor. The score was based on one single study (from Greece). Was this assumed to be representative of “sandwiches” to which the U.S. population is exposed? I also did not understand the data – e.g., total samples appear in decimals. (Also see my comments on the use of composite food categories in the ranking).</p> <p>For fresh salsa, I did not understand why the score for Criterion 1 is 3, while for Criterion 7 is 9. The inverse happens for sandwiches.</p> <p>I was also interested in how <i>Campylobacter</i> pairs scored in the model. It is evident that all food-hazard pairs with this pathogen scored very low, and I believe there are two important reasons for this: a) <i>Campylobacter</i> seldom causes outbreaks, and this is recognized worldwide, and b) no meat and poultry products were considered in the model, and these are the most important sources of <i>Campylobacter</i> (which has been proven by a wide variety of source attribution studies). However, <i>Campylobacter</i> is the most</p>	

<b>CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>reported foodborne illness in the U.S., and estimated to be the fourth most imported foodborne pathogen in terms of total (corrected) incidence (Scallan <i>et al.</i>). I would think that foods that typically cause <i>Campylobacter</i> infections would be worth labeling as “high risk”, along with the others included in this report.</p> <p>I could see that fresh salsa appears highlighted at some points in the results table (appendix L), and was interested in knowing why. I noted it myself for the same reasons I have pointed for sandwiches – the fact that salsa is a composite food, potentially constituted by a wide variety of ingredients (commodities) makes me question the utility of including it in such a risk ranking exercise.</p> <p>Looking into a chemical hazard-pair in more detail, I can see that arsenic in rice is the first on the ranking. It is widely accepted that the link between exposure to arsenic and potential onset of cancer is one of the better established scientifically. Still, I wondered about the score for Criterion 3 (which is 9). I could not find the data used to support this score. Was the prevalence above the accepted level &gt;1%? Or have there been recalls? This sounds very high, especially for the U.S. (i.e., maybe it would be more likely in high producers and highly exposed populations).</p>	
Reviewer #2	See note above.	
Reviewer #3	[Reviewer did not comment.]	
Reviewer #4	<p>A number of product:hazard pairs were selected for detailed evaluation and comment. These were:</p> <ol style="list-style-type: none"> <li>i. <i>Bacillus cereus</i> in pasteurized milk</li> <li>ii. <i>Listeria monocytogenes</i> in processed (cooked) meats</li> <li>iii. <i>Clostridium botulinum</i> in canned fruits and vegetables</li> <li>iv. <i>Listeria monocytogenes</i> in soft surface-ripened cheeses</li> <li>v. <i>Listeria monocytogenes</i> in cold smoked fish</li> <li>vi. STEC <i>E. coli</i> in bagged leafy green salads</li> <li>vii. Hepatitis A in Frozen Berry Fruits (472)</li> </ol> <p>The evaluations are described below.</p> <ol style="list-style-type: none"> <li>i. <u><i>B. cereus</i> in pasteurized milk</u>  <i>B. cereus</i> is a spore-forming bacterium, and pasteurization cannot eliminate it from milk. Growth and toxin-production to high levels is possible during the shelf life of pasteurized milk. While outbreaks due to <i>B. cereus</i> in milk are infrequent they have been reported, and numerous studies have demonstrated potential for growth and toxin production in milk within the normal shelf life (see e.g., Christiansson <i>et al.</i> (1989),</li> </ol>	

<b>CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p><i>Appl. Env. Microbiol.</i>, 55(10): 2595-2600; Te Giffel <i>et al.</i>, (1996), <i>Int. J. Food Microbiol.</i>, 34(3):307-18; Notermans <i>et al.</i>, (1997), <i>Food Microbiol.</i>, 14:143–151).</p> <p><b>This product:hazard pair was not identified in Criterion 1 and is not considered in the model or report. Its absence requires some explanation.</b></p> <p>ii. <u><i>Listeria monocytogenes</i> in processed (cooked) meats</u>  <i>L. monocytogenes</i> in processed meats was identified as the main cause of Listeriosis (&gt;50% of cases) in USA by the USDA/FDA/CDC (2004) risk ranking of RTE foods associated with Listeriosis.  <b>This product:hazard pair was not identified in Criterion 1</b>, and is not considered in the model or report. I assume that the reason is because processed meats are not within the USFDA jurisdiction, but it might be useful to state that overtly (earlier) in the report.</p> <p>iii. <u><i>Clostridium botulinum</i> in canned fruits and vegetables</u>  <i>Criterion 1</i>                      It is well documented that canned fruits and vegetables (low acid) have been causes of botulism. In USA this has often been attributed to home bottling. The Criterion 1 results spreadsheet, however, cites CSPI (2011; REF 1338) and The Orange Book (REF 1339) as the basis for inclusion of this product:hazard pair. I could find nothing about <i>C. botulinum</i> in REF 1339, and the CSPI outbreak database links cited are no longer fully functional. From the 2007 outbreak list, however, there was one reference to botulism involving a hot chili, sauce, and another involving canned meat, seafood, or pet food. <b>Neither of these references seem appropriate to support the criterion value assigned.</b></p> <p><i>Criterion 2</i>                      The ranking of <i>C. botulinum</i> as ‘9’ is appropriate, and accords with the severity estimates of Minor <i>et al.</i> (2015).</p> <p><i>Criterion 3</i>                      A single reference to a French study is used to estimate prevalence/likelihood of contamination. Importantly, the prevalence reported in that study was in <u>raw product</u> (ingredients), not finished (heat treated) product (<i>see</i> Methods and Materials S2.1, p. 264, of Sevenier <i>et al.</i>). Because of the potential confounding of responses to Criterion 3 with responses to Criterion 6, it is stated in the report (p.56, L22/23) that contamination at Criterion 3 relates to contamination <u>in the finished product</u>. <b>Accordingly, the prevalence reported in REF 211 (IJFM 155(3):263-8 (2012) Sevenier <i>et al.</i>) is not relevant or appropriate to support the prevalence estimates assigned for Criterion 3 for this product:hazard pair.</b></p>	

**CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.**

REVIEWER	COMMENT	RESPONSE
	<p><i>Criterion 4</i> Growth potential is assessed as 1 (i.e., “low”) citing REF 10 (expert opinion) and others (e.g., NSWFA, Betts <i>et al.</i>). These seem to be credible and authoritative references. However, the inclusion of REFS 453, 454, and 455 seems spurious. REF 453 (Severn <i>et al.</i>) doesn’t seem to consider growth in finished product, REF 454 is about pH and heat resistance in thermal processing (not growth) and REF 455 is a survey of pH and water activity in acidified bottled vegetables. The category being considered is LOW ACID products, not acidified product. <b>Accordingly, the interpretation of the data cited seems to require reconsideration.</b></p> <p><i>Criterion 5</i> The values presented seem credible for retorted product, but are based on expert opinion. However, the long history of safety of canned foods supports this evaluation.</p> <p><i>Criterion 6</i> The consumption criterion score (9) is appropriate and accords with the estimates from NHANES.</p> <p><i>Criterion 7</i> Economic impact per case was estimated by Minor <i>et al.</i> (2015) which ranked the per case cost of botulism as the 3<sup>rd</sup> highest of the foodborne hazards they considered. This probably arises from the high fatality rate and complete life support required, potentially for months or years, for survivors of botulism intoxication. <b>Importantly, in the C7 spreadsheet the cost per botulism illness that is reported to be based on the Minor <i>et al.</i> (2015) paper is misquoted. The value should be \$1,514,289 but in the C7 spreadsheet is given as \$15142, i.e., a factor of 100 too low.</b></p> <p>iv. <u><i>Listeria monocytogenes</i> in soft surface-ripened cheeses</u> <i>Criterion 1</i> The outbreak evidence for listeriosis from soft surface-ripened cheeses in USA is 3 outbreaks involving 44 cases. This rate seems surprisingly low in comparison to other nations, but may arise from a relatively low consumption of these types of cheese in USA. However, the NHANES data suggest that ~10% of the population consume such cheeses every day. It might have been expected that more recalls had occurred, as recalls for this product:hazard combination are relatively common in other Western nations. There are numerous outbreaks recorded internationally, including a 2013/2014 outbreak in Australia that lead to &gt;20 cases and three deaths including one miscarriage. Nonetheless, based on the USA outbreak data and the scoring criteria, the value applied is correct.</p>	

**CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.**

REVIEWER	COMMENT	RESPONSE
	<p><i>Criterion 2</i>                      The ranking of listeriosis as a ‘9’ for severity of illness is appropriate for <u>susceptible populations</u> (which potentially comprise 15-20% of the population) and which, on average, are ~100x more likely than a typical (non-pregnant) healthy adult to develop a systemic infection.                      Many studies and reviews report that 20 to 30% of systemic listeriosis cases result in death of the patient. The severity ranking also accords with Minor <i>et al.</i> (2015). However, the ‘death rate’ cited in the C2 Results spreadsheet (15.9%) seems a factor of 2 too low. On investigation and re-reading the Scallan <i>et al.</i> paper (Scallan <i>et al.</i> (2011). <i>Emerg Infect Dis.</i> 17(1): 16-22) which was cited in support of 15.9% - it seems that that value is a mis-interpretation. As stated in Scallan <i>et al.</i> (see footnote to Table 3, p.12) “These rates [i.e., presented in the table] were doubled to adjust for underdiagnosis before being applied to estimate the total number of hospitalizations and deaths”.</p> <p><i>Criterion 3</i>                      The prevalence score seems appropriate based on this reviewer’s experience and reading of relevant literature. However, there appear to be some errors among the references cited. Specifically, REF 3004 has the wrong title in the MASTER REFERENCES spreadsheet. It should be “<i>Incidence of Salmonella, Listeria monocytogenes.... in two types of Mexican soft cheese</i>”. Given the title, and that there is a separate product category for Mexican soft cheese, the <b>inclusion of this reference to calculate the prevalence seems inappropriate. Also, I could not locate REF 6006 from the information given in the MASTER REFS spreadsheet. Also, details for REFS 6008, 6009, 6011, and 6012 are incomplete.</b></p> <p><i>Criterion 4</i>                      Growth potential/shelf life values also accord with expectations. However, there are far better references available to support the conclusions made here, including numerous papers that model the rate of <i>L. monocytogenes</i> in various products, or summarize observations. Predictive microbiology databases would also have been more reliable. In particular, the predictive model of Mejlholm and Dalgaard and colleagues (Mejlholm <i>et al.</i>, 2010, <i>Int J. Food Microbiol.</i>, 141:137-150; Mejlholm and Dalgaard, 2009. <i>J. Food Protect.</i>, 72, 2132–2143) has been demonstrated to provide very reliable predictions of the growth rate and growth limits of <i>L. monocytogenes</i> in a wide variety of products, including cheeses. <b>The references cited are not authoritative sources for the evaluation and scoring of this criterion. REFS 4107 and 4108 are incompletely described in the MASTER REFS spreadsheet, but look like websites. REF 4116 (“Bishop <i>et al.</i> Storage Temperatures Necessary to Maintain Cheese Safety”) is also incomplete and equally does not seem to be authoritative. REF 4149 is not included in the MASTER REFS spreadsheet.</b></p>	



<b>CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p><i>Criterion 5</i> The score given seems reasonable, and average prevalence rates of a few percent to 10% in finished product might be expected despite that pasteurization of milk is effective in eliminating <i>L. monocytogenes</i> and reflecting the potential for post-processing contamination from processing environments. <b>However, again the references provided are limited to ‘expert opinion’, when there is a large published literature, including many reviews and book chapters, on <i>L. monocytogenes</i> in cheeses, including soft cheese.</b> Elliot Ryser’s chapter “<i>Incidence and Behavior of Listeria monocytogenes in Cheese and Other Fermented Dairy Products</i>” in Ryser and Marth’s “<i>Listeria, Listeriosis, and Food Safety, 3<sup>rd</sup> Edn</i>” (CRC Press, 2007) would have been a useful place to start.</p> <p><i>Criterion 6</i> <b>The consumption criterion score (‘3’) based on the NHANES data seems incorrect. The highest daily consumption rate was 10.7%, which should have been scored ‘9’.</b></p> <p><i>Criterion 7</i> Economic impact per case was estimated by Minor <i>et al.</i> (2015) who ranked the per case cost of listeriosis as the 4<sup>th</sup> highest of the foodborne hazards they considered. Accordingly, <b>the score given in the report (‘3’) seems too low compared to the evaluations in Minor <i>et al.</i> (2015) and should be ‘9’, in the same group as botulism, <i>V. vulnificus</i>, <i>Cronobacter</i>, etc. Importantly, in the C7 spreadsheet the cost per illness that is reported to be based on Minor <i>et al.</i> (2015) is misquoted. The value should be \$1,456,676 but in the C7 spreadsheet is given as \$14566, i.e., a factor of 100 too low.</b></p> <p>v. <u><i>Listeria monocytogenes</i> in cold smoked fish</u></p> <p><i>Criterion 1</i> There is a large literature that shows that <i>Listeria monocytogenes</i> is commonly detected in cold smoked fish products and clearly has the potential to grow, although growth is limited in most vacuum-packed products by the presence of benign lactic acid bacteria that, through their growth, exert the Jameson Effect. Again, as a first step, the report authors might have consulted Ryser and Marth’s “<i>Listeria, Listeriosis, and Food Safety, 3<sup>rd</sup> Edn</i>” (CRC Press, 2007) which has a chapter on ‘Incidence and Behavior of <i>Listeria monocytogenes</i> in Fish and Seafood’ with US seafood HACCP expert Mel Eklund as the senior author. An even more authoritative series of papers on hazards and risks in smoked seafoods was prepared by IFT involving a group of international experts, and commissioned by USFDA. The entire series of articles including Lone Gram’s paper “Potential Hazards in Cold-Smoked Fish: <i>Listeria monocytogenes</i>” was</p>	

<b>CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>published in <i>Journal of Food Science</i> in 2001. It also appears in full on FDA web-sites (<a href="http://www.fda.gov/Food/FoodScienceResearch/SafePracticesforFoodProcesses/ucm092227.htm">http://www.fda.gov/Food/FoodScienceResearch/SafePracticesforFoodProcesses/ucm092227.htm</a>).</p> <p><i>L. monocytogenes</i> also has a very high ID<sub>50</sub> and studies with pregnant animals that are suitable experimental models for humans have confirmed earlier inferential studies (FAO/WHO, 2004) that the ID<sub>50</sub>, even for highly susceptible people is millions, or billions, of cells. As such, while the hazard is recognized, there are very few reports of outbreaks from cold-smoked fish. Nonetheless, USA has rejected numerous shipments of European cold smoked fish after detection of <i>L. monocytogenes</i>. Similarly, studies in USA by Wiedemann and colleagues indicate that <i>L. monocytogenes</i> is also common in product and plants producing cold smoked fish in USA. The absence of recall data then, is somewhat surprising. <b>Thus, the C1 score of '0' in this case seems quite anomalous. No references or explanations are given to support this 'unexpected' score.</b> However, it is noted that cold smoked fish is relatively infrequently consumed by USA citizens.</p> <p><b>What is REF 9999 meant to indicate?</b></p> <p><i>Criterion 2</i>            Many studies and reviews report that 20 to 30% of systemic listeriosis cases result in death of the patient. The severity ranking also accords with Minor <i>et al.</i> (2015). However, the 'death rate' cited in the C2 Results spreadsheet (15.9%) seems a factor of 2 too low. On investigation and re-reading the Scallan <i>et al.</i> paper (Scallan <i>et al.</i> (2011). <i>Emerg Infect Dis.</i> 17(1): 16-22) which was cited in support of 15.9% - it seems that that value is a mis-interpretation. As stated in Scallan <i>et al.</i> (see footnote to Table 3, p.12) "These rates [i.e., presented in the table] were doubled to adjust for underdiagnosis before being applied to estimate the total number of hospitalizations and deaths".</p> <p><i>Criterion 3</i>            The prevalence score of '9' is based on a single publication, from Sweden in 2011. While the prevalence in that report is not unusual (and a weighted mean prevalence rate of ~18% was estimated from numerous studies prior to 2000 by WHO/FAO (2004; "<i>Risk assessment of Listeria monocytogenes in ready-to-eat foods</i>", Microbiological risk Assessment Series, 5) there many more recent studies that should be considered. Also, REF 211 reported that for cold smoked fish produced <u>in Sweden</u> the prevalence rate was ~8%. A quick literature search found several other recent studies, e.g., Leong <i>et al.</i> (2015) found a prevalence of 2.5% (mostly environmental some food samples) in smoked salmon processing in Ireland. Gonzalez <i>et al.</i> (2014) found 4.7% positives at retail in smoked salmon in Spain. Given this, it is suggested</p>	

<b>CHARGE QUESTION 6b: <i>Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>that a more thorough search of recent literature be undertaken to develop a more robust estimate of prevalence.</p> <p><i>Criterion 4</i> The scores given to the components of this criterion, and the overall evaluation and score, seem appropriate based on my experience and reading. However, some of the references cited do not seem very authoritative and better references might be found. Also, FDA 2013 is the IFT expert panel report commented on under Criterion 1 for this product:hazard pair and it might be more informative to refer to the original report, given that the data in the report relate to pre-2000, not pre-2013.</p> <p><i>Criterion 5</i> Given the limited effectiveness of control measures and, particularly, the lack of any listericidal treatment during the cold smoking process, the scores given for contamination prevalence in raw product (or potential for re-contamination or cross-contamination) seem appropriate as does the overall score.</p> <p><i>Criterion 6</i> The consumption criterion score (0) is appropriate and accords with the estimates from NHANES.</p> <p><i>Criterion 7</i> Economic impact per case was estimated by Minor <i>et al.</i> (2015) who ranked the per-case cost of listeriosis as the 4<sup>th</sup> highest of the foodborne hazards they considered. Accordingly, <b>the score given in the report ('3') seems to low and should be '9', in the same group as botulism, <i>V. vulnificus</i>, <i>Cronobacter</i>, etc. Importantly, in the C7 spreadsheet the cost per illness that is reported to be based on the Minor <i>et al.</i> (2015) paper is misquoted. The value should be \$1,456,676 but in the C7 spreadsheet is given as \$14566, i.e., a factor of 100 too low.</b></p> <p>vi. <u>STEC <i>E. coli</i> in bagged leafy green salads</u></p> <p><i>Criterion 1</i> The scoring '1' seems appropriate given the relatively small number of cases associated with outbreaks due to this product:hazard pair.</p> <p><i>Criterion 2</i> <b>In this case the score seems too high, e.g., in comparison with the evaluations and rankings of Minor <i>et al.</i> (2015).</b> Minor <i>et al.</i> (2015) developed a comprehensive assessment of mean severity, i.e., taking into account the proportion of cases that were unreported, those that sought/required medical intervention, and</p>	

**CHARGE QUESTION 6b: Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.**

REVIEWER	COMMENT	RESPONSE
	<p>those that proceeded to serious illness requiring medical attention and, potentially, long term sequelae. In their assessment three main groups can be discerned based on Mean QALD lost. Six hazards (<i>Cronobacter</i>, <i>Clostridium botulinum</i>, <i>Listeria monocytogenes</i>, <i>Vibrio vulnificus</i>, <i>Trichinella spiralis</i>, Ciguatoxin) were clearly associated with <i>much</i> more severe illness, and had QALD loss scores greater than 10. The least severe group of hazards had, on average, QALD scores less than 1 and included <i>S. aureus</i>, <i>B. cereus</i>, ‘other’ <i>E. coli</i>, <i>Vibrio cholera</i>, <i>V. parahaemolyticus</i>, other (non-<i>vulnificus</i>) <i>Vibrio</i> spp., Astrovirus, Sapovirus, Scombrototoxin (= histamine), and “Food Allergic Reaction”. All other hazards, including STECs fell in the range 1 to 10. <b>Based on this evaluation, the hazards that scored QALD&gt;10, should score ‘9’ in Criterion 1, those that had a QALD score &lt;1, should score ‘1’ for Criterion 1, and all others should score ‘3’.</b> Following this suggestion, <i>E. coli</i> O157:H7 and other STECs should score ‘3’ in Criterion 1, not ‘9’.</p> <p><i>Criterion 3</i> The score given for Criterion 3 (Likelihood of Contamination) for this product:hazard pair is 1. However, the data sources used to develop this estimate seem questionable. REF 316 (Sagoo <i>et al.</i>), reporting on prevalence in the United Kingdom, found no <i>E. coli</i> O157 in nearly 4000 samples of salad vegetables (the proportion that are ‘bagged leafy greens’ could not be easily discerned). REF 307 is the USDA’s microbiological data program, and the bibliographical details are incomplete so I can’t comment on the data being referred to. REF 326 relates to a Swiss study of product at retail which reported a ‘low occurrence of contamination with ...Shiga-toxin producing <i>E. coli</i>’. While the body of evidence generally supports the score of ‘1’, there could be more discussion of the limits of the data cited.</p> <p><i>Criterion 4</i> The score (‘1’) for this criterion appears appropriate and to be supported by credible references, although firstly relying on ‘expert opinion’ (REF 10).</p> <p><i>Criterion 5</i> The scores for this criterion , composed of both contamination probability and intervention efficacy, are based on ‘expert opinion’ (REF 10). Contamination is here scored ‘3’, while Criterion 3 was scored ‘1’. This seems inconsistent. The intervention efficacy is scored ‘3’, which is possibly debatable. Interventions in the field and prior to processing are probably most important but affect probability of contamination, not efficacy of processing interventions. Processing interventions in common use in the industry might achieve a 1 to 2 log reduction, which is not a strong intervention (by comparison 5 log reductions have been required of other industries). While the overall result of Contamination scored ‘1’ and ‘Intervention</p>	

<b>CHARGE QUESTION 6b: <i>Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>Efficacy' scored '1' also results in an overall Criterion 5 score of '3', it is suggested that relevant studies be consulted. For example, the studies of Danyluk, Niemira, Perez Rodriguez, and many others provide quantification of intervention efficacy of currently used technologies for disinfection of leafy greens.</p> <p><i>Criterion 6</i> The consumption criterion score (9) is probably appropriate but is based on expert opinion, not the NHANES database. The reasons for relying on expert opinion should be explained. Moreover, in deciding on a '9' were the experts thinking only of consumption of bagged leafy greens, or all sources of leafy greens? This should be clarified in the document.</p> <p><i>Criterion 7</i> The economic impact is consistent with the evaluation of Minor <i>et al.</i> (2015) and the expected numbers of cases from Scallan <i>et al.</i> (2011).</p> <p>vii. <u>Hepatitis A in Frozen Berry Fruits (472)</u></p> <p><i>Criterion 1</i> The score given ('1') seems somewhat low and may relate only to the situation in USA. As noted above in Europe and, more recently, in Australia there have been numerous outbreaks of Hepatitis A from frozen berry fruits and the situation in Europe should have been known to the report authors because they have cited a relevant review paper (Maunula <i>et al.</i> 2013, REF 3016). <b>As such, it is suggested that this score be reevaluated.</b></p> <p><i>Criterion 2</i> The score given is a '9' which is inconsistent with Minor <i>et al.</i> (2015), who score Hepatitis A as a more moderate hazard, akin to Salmonellosis, yersiniosis, or a range of parasitic infections. <b>It is suggested that the score given be reevaluated and reasons for the selection of '9' be documented.</b></p> <p><i>Criterion 3</i> For this criterion a score of '1' is given but with no supporting data or references at all. <b>More explanation is needed to maintain the transparency of the process.</b> The available evidence suggests that the frequency and level of contamination is very low but that, because Hepatitis A has a very low ID<sub>50</sub>, outbreaks can still occur when contamination levels are below practical limits of detection.</p> <p><i>Criterion 4</i></p>	

<b>CHARGE QUESTION 6b: <i>Is any score assigned to a criterion not appropriate? If so, please provide suggestions on how the process of data collection and documentation might be refined.</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>Appropriately the C4 criterion score for this product:hazard pair, is '0', indicating that no growth is possible.</p> <p><i>Criterion 5</i> The scores given for this criterion are '3' both for contamination and for intervention. I'm unaware of any highly effective interventions against Hepatitis A on fresh fruit, other than chemical sanitizers, and they're not very effective. Equally, freezing berries wouldn't be expected to cause significant reduction in Hepatitis A levels. The score of '3' for contamination seems inconsistent with the score given for Criterion 3. <b>It is suggested that the score given be reevaluated and reasons for the selection of the scores of '3' be documented.</b></p> <p><i>Criterion 6</i> The NHANES data indicates that much less than 1% of the population consume frozen berries on any given day. As such, and using the scoring criteria outlined in the model and report, a score of '1' for this criterion is correct.</p> <p><i>Criterion 7</i> The economic impact is consistent with the evaluation of the cost per case of Hepatitis A of Minor <i>et al.</i> (2015) and the expected numbers of cases from Scallan <i>et al.</i> (2011).</p>	

<b>CHARGE QUESTION 7: <i>The draft model relies on expert elicitation to fill data gaps, in particular for the scoring of Criterion 5 and for the scoring of criteria for chemical hazards and undeclared allergens. Is the expert elicitation process (which involves external panels and FDA subject matter experts) adequate to address data gaps?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	Refer to 7a	
Reviewer #2	Yes, the process is adequate, but it is unclear how heavily FDA subject matter experts (versus expert panels) are used and the number of subject experts involved for each determination. If, for example, the scoring reflects the opinion of a single subject matter expert, that scoring might be open to question.	
Reviewer #3	If experts are chosen correctly, then this is certainly an appropriate procedure. What would need however to be captured is the rationale for certain decisions, rating. In some of the examples (chapter 4) it is often only mentioned that this was done based on expert elicitation – the rationale is however not given, also not when a calculated score is changed.	
Reviewer #4	Refer to 7a	

<b>CHARGE QUESTION 7.a: Is the scoring logic used to combine data and expert opinions and the order of preference in section 4 (e.g., Figure 4.3) appropriate? If not, please explain what changes might be considered and why.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	<p>Expert elicitations are a widely used and well accepted method to address knowledge gaps and produce evidence when basic data are lacking. Within food safety, they have been successfully used for instance with source attribution, for which they are often the only method available to answer certain public health questions. In this project, I feel they were the best approach to address gaps, and the panels seemed to be varied in the sense that they included experts with experience in different fields within food safety. They were however small (maximum 4 experts in each panel), and I understood that they were not provided with an answer sheet, but rather discussed questions and reached agreement. I wondered if, in this process, they did not influence each other? Also, in structured EEs the experts' knowledge is often validated (through initial validating questions) and eventually weighted, and the answers are analyzed to derive uncertainty around the final estimates. I understand that this would not be possible with such panels, but these limitations should be discussed.</p> <p>The scoring method used to combine available data and expert input seems reasonable.</p>	
Reviewer #2	The hierarchy is appropriate.	
Reviewer #3	<p>What does not look very logical is that the process from top to the bottom in the different processes illustrated in figures chapter 4 is not stopped as soon as there is a quantitative “zero” with a high confidence – e.g., no outbreaks ever registered, no recalls, no contamination. Otherwise the process seems quite heavy – e.g., to assess C4, C5 for foods which have never been implicated in outbreaks, e.g., scones and <i>Cronobacter</i> spp. It does somehow not fit with other provisions, for example the fact that a food is not or only marginally consumed.</p>	
Reviewer #4	<p>The range of experts used seems appropriate and sufficient to generate credible responses to the missing data, particularly when combined with a public comment process to enable further input and opinion. Sometimes experts will need to have relevant, though imperfect, data presented to them to assist them to make an informed decision. In this regard, the integration of the expert's background knowledge and perspectives might be brought to bear earlier in the process by asking the experts to evaluate the significance of prevalence rates from RFR, or recall data, as part of the expert elicitation, rather than requesting the experts to make estimates when no other data has been found upon which to make a reasoned estimate.</p>	

<b>CHARGE QUESTION 8: Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	I think the approach taken is appropriate, and would not agree with scaling down consumption: even if the foods would only be of high risk for certain subgroups in the population, there would be no way of protecting these populations if foods of this risk were not considered in the same way as foods/hazards that can cause disease to the wider population.	
Reviewer #2	Scaling down would present a slippery slope. If allergens are scaled, shouldn't the sensitive populations for <i>Listeria monocytogenes</i> , and methyl mercury also be scaled? Would it be possible to give two separate risk scores, one for the general population and one for susceptible populations, perhaps with an explanatory note for each case?	
Reviewer #3	<p>It seems inappropriate to rate a pathogen for a whole population but which has a specific “target”, e.g., <i>Cronobacter</i> only causes severe disease in infants below 6 months – which will not be consuming some of the foods which have been associated to, such as chocolate, scones, foods for adults consuming this type of food severity is “0”...and they would probably also not be consuming formula. In such a case the score 9 should only be used for that particular population.</p> <p>In terms of undeclared allergens – one could approach it in a similar way and consider only the fraction of the population which is susceptible. This population is however probably less well defined than in the above example. Considering that for undeclared allergens the worst case scenario is taken, i.e., (a) just “presence” and not a threshold below which it would no longer represent a severe hazard, and (b) significant percentage of mortality in case of presence, the risk is 9 and weighing in terms of consumer population is not really going to contribute to a change in the rating.</p> <p>If a weighing is to be considered, then this would, in my opinion, be better done with respect to the likelihood of finding an undeclared allergen above a certain threshold in a certain food – i.e., high likelihood to find a dairy ingredient or residues in a culinary preparation or nuts in confectionery, than finding shellfish in milk. The fact that no differentiation is made between allergens, excludes however this option. A possibility of weighing is C5 – which however would imply a very good knowledge on processing lines and processes (One could even argue on why for certain commodities “undeclared allergens” are not mentioned – as in annex C; for example butter and buttermilk, while it is included for cream and cheese, both used to manufacture these. A further example is nuts, where undeclared allergens is found for some but not for others or wrongly undeclared allergens other than nuts in the case of hummus which is (at least in the rest of the world) manufactured with chick peas and not nuts and contains other ingredients.</p>	



**CHARGE QUESTION 8: Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.**

REVIEWER	COMMENT	RESPONSE
Reviewer #4	<p>It is difficult to respond to this question unequivocally, because the answer depends on whether the aim of the FSMA requirement to designate high-risk foods is intended to apply to the average risk to the entire population, or the risk to susceptible populations. Logically, the most objective approach is to establish the relative risk for the whole population exposed to the foods, particularly since none of the foods considered in the model seem to be specifically intended for an ‘at risk’ population. A potential problem with this approach, however, is that it effectively disenfranchises susceptible populations because those sub-populations will experience the bulk of the disease burden for specific hazards. Furthermore, the ‘treatment’ of susceptible populations in the model seems inconsistent between hazards, based upon reading Sections 2.2.2 and 4.2.2. In some cases, mortality and hospitalization rates are used to evaluate severity, but that ignores that those rates are related to the laboratory confirmed cases, not the total estimated number of cases. This potentially introduces a bias because the same denominator is not used for each hazard. While it might be argued that hospitalization rates based on laboratory confirmed cases do effectively reflect average disease severity, it is an indirect metric and better, more direct metrics are available, e.g., in Minor <i>et al.</i> (2015) discussed further below.</p> <p>Furthermore, while it is stated that susceptible populations are ‘considered’ in the evaluation/scoring for Criterion 2 it is not clearly explained <i>how</i> susceptible populations are considered nor the relative size of the susceptible population, which is needed to establish the correct relative risk. Presumably this ‘issue’ is the genesis of Charge Question 8. Also, as noted earlier there is inconsistency with the relative severity estimates of Minor <i>et al.</i>, 2015, and the inconsistency for the ‘severity’ scores of allergens in the model <i>cf.</i> Minor <i>et al.</i>, 2015 is discussed in more detail in the response to Question 12).</p> <p>To elaborate, disease severity would be expected to be greater in susceptible populations, or disease much more probable in susceptible individuals for a given level of exposure. If the disease severity score is based on the responses of susceptible populations, then all other Criteria would have to be evaluated with respect to susceptible populations as well, because the risk from a particular hazard pair would be overestimated by assuming that the entire population is ‘susceptible’ to that hazard to the same extent as the specifically susceptible population. Using <i>Listeria monocytogenes</i> as an example, the ‘susceptible’ population is, on average, about 100 times more susceptible than healthy non-pregnant adults but represents only 15-20% of the population. Additionally, the consequences of infection in the ‘susceptible’ population are much greater than in healthy non-pregnant adults. Thus, assuming all people are equally susceptible to listeriosis would overestimate the relative risk by 500-fold, or more. Also, even among the ‘susceptible population (i.e., very young, old, immunocompromised, pregnant), susceptibility to listeriosis varies over two orders of</p>	

**CHARGE QUESTION 8: Susceptible populations are considered as part of the definitions in Criterion 2. Are there other ways to account for susceptible populations more specifically in the model? For example, would it be appropriate to scale down the % consumption (Criterion 6) for undeclared food allergens-pairs to reflect the proportion of allergic consumers? Please explain.**

REVIEWER	COMMENT	RESPONSE
	<p>magnitude (see e.g., Goulet <i>et al.</i>, 2012, <i>Clinical Infectious Diseases</i>, <b>54</b>:652-660) for different pre-disposing conditions. Similarly, while 2-4% of the population suffer allergic responses to specific food-borne allergens, not all have severe responses leading to anaphylactic shock, but this does not seem to have been considered in the model in the scoring of the severity of exposure to allergens.</p> <p>These relative susceptibilities to hazards and relative severity of consequences for susceptible populations will not necessarily be the same for all hazards. Accordingly, the lack of a clear statement of whether the severity rating is averaged across the entire population, or derived only for the susceptible population.</p> <p>The suggestion for a ‘correction’ to be included to account for this by scaling down the % consumption should not be contemplated because it is not transparent unless it is explicitly made clear why that ‘adjustment’ is being made and the basis of the reduction in consumption assumed. Furthermore, I imagine that it would be difficult to implement because the adjustment would have to be accommodated with the incremental scoring scheme (0, 1, 3, 9) which would distort the true proportions and would be difficult if it were not applied to every product:hazard pair.</p> <p>As suggested previously in this review, a more consistent and scientifically defensible approach would be to base the severity score on the average disease severity across all consumers, based on a weighted average of the number of cases not requiring medical intervention, those that do require intervention and those that lead to death or permanent disability together with a rating of the disease severity for each of those categories, e.g., as was presented by Minor <i>et al.</i> (2015) in terms of Quality Adjusted Life Days (QALD), in addition to their estimates of economic cost of various diseases. Given that the economic estimates of disease severity from Minor <i>et al.</i> (2015) were adopted, it seems reasonable to adopt their QALD estimates as the basis of scoring for Criterion 2.</p>	

**CHARGE QUESTION 9: Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?**

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I am not sure I understand this question. I believe that risk scores should be compared, taking into account that the epidemiological link that constitutes the basis of the identification of the food-hazard pairs, and which contributes to the overall scores as well, is informed by outbreak data, which may not allow for the flexibility of considering new events, or for measuring the relative contribution of specific foods for disease	

<b>CHARGE QUESTION 9: Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	by a given hazard if this pair has not caused (many) outbreaks before but may have caused sporadic illnesses (e.g., <i>Campylobacter</i> pairs).	
Reviewer #2	<p>If this question is asking where the line between high-risk and non-high-risk foods should be drawn, it's subjective and akin to scoring the cutoff number for a student's average % grade on an exam with specific letter grades. One might default to the traditional A = <math>\geq 90\%</math>, B = 80 -89%, C = 70-79% etc. But this may be inappropriate for a given class, and if the grade distribution curve is smooth, even these cut-offs would be arbitrary (i.e., is there really a difference between a student who gets an 89.9 and one who gets a 90?).</p> <p>That having been said, there has to be some sound scientific basis for making the cut-off. One might consider making some % of all food-hazard pairs the cut-off (e.g., defining the top quadrille of risk scores as the cut-off). How this can be done remains to be determined, but clearly a cut-off score that places 90% of the food as high risk has too low a cut-off and one that places 90% of foods as low risk wouldn't be very useful either.</p> <p>Perhaps the cut-off could be approached statistically by looking at the risk score distribution and finding the 95% (or 90%, or whatever) confidence interval (i.e., the risk ranking score that is significantly different from the population) and basing the cut-off on that.</p>	
Reviewer #3	To have a cross-check to see how far it makes sense and whether there is a bias due for example to the structure primary, secondary, and tertiary commodity, impact of a single report, contamination not related directly with the food as could happen in a home (see comments under IV for annex C). What would also be important is to check for consistency – it seems illogical to have hazards associated with certain commodities but not with others which contain the first as an ingredient – or to make distinctions such as mycotoxins (overall) for apple and apple products and specifically patulin for apple juice concentrate; acrylamide in peanuts but no longer in peanut butter – unless there is a clear rationale. A further example is nuts, where undeclared allergens are found for some but not for others or wrongly undeclared allergens other than nuts in the case of hummus which is (at least in the rest of the world) manufactured with chick peas and not nuts and contains other ingredients.	
Reviewer #4	The intent of this question seems to be to seek advice on establishment of criteria for determining the “cut-off” in the risk ranking that delineates foods that will require additional record keeping and those that will not. This is a risk management decision and would seem to be the responsibility of the risk manager(s) who commissioned this risk ranking tool and exercise. There is no unequivocal, objective basis for deciding the “cut off”. Rather, that will require risk communication to understand the expectations of relevant stakeholders to determine an acceptable level of protection of public health compared to the costs associated with that.	

<b>CHARGE QUESTION 9: <i>Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>The goal of the risk ranking tool and exercise is to: “rapidly and effectively .... prevent or mitigate a foodborne illness outbreak....and to address credible threats of serious adverse health outcomes” and the instructions given to the modelers/risk assessors were that the designation of high-risk foods shall depend on: “the known safety risks .. including the history and severity of foodborne illness outbreaks” and “the likelihood that a particular food has a high potential for microbiological or chemical contamination...or would support microbial growth due to the nature of the foods or the processes used”.</p> <p>These instructions correctly identify considerations that need to be taken into account when identifying high-risk foods but, to elaborate a little:</p> <p>Risk is the combination of the probability of an adverse event, and the severity of that event. In terms of foodborne illness, the severity is a combination of the severity of illnesses experienced and the number of people affected. Severity is explicitly considered in the model although, as discussed above, it might be characterized in other, more defensible, ways. The number of illnesses would be expected to be correlated with the number of people exposed to a particular product:hazard pair and, in the model, this is represented through the consumption data and criterion.</p> <p>The probability of an ‘adverse event’ occurring is correlated with the inherent likelihood of contamination of the product at a level able to cause illness, or the potential for the contamination and subsequent growth of microbial hazards to a level able to cause illness. This also is represented in the model but, as a marker of such product/process/hazard combinations, it might be beneficial to consider:</p> <ul style="list-style-type: none"> <li>i) whether there is one, or more, credible hazards associated with the product;</li> <li>ii) whether there are clearly identifiable Critical Control Points necessary to control those hazards to enable the safe production/processing of the product and</li> <li>iii) whether those CCPs can be, and have been demonstrated to be, reliably controlled and monitored.</li> </ul> <p>If the answer to the first question is ‘no’, the product should probably not be considered a HRF.                      If the answer to the first question is ‘yes’ and the answer to the second question is ‘no’, the product should probably be considered a HRF. If the answer to the both the first and second questions is ‘yes’ and the answer to the third question is ‘no’, the product should probably be considered a HRF. If the answer to all three questions is ‘yes’, the product is possibly not a HRF.</p> <p>Further considerations may relate to the feasibility of identifying a potential public health risk and instituting a recall, e.g., less weight might be given to products with very short shelf lives (e.g., less than 10 days) because of the feasibility of “rapidly and effectively” preventing or mitigating a foodborne illness</p>	

<b>CHARGE QUESTION 9: <i>Given the underlying data supporting the scoring, what are the considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>outbreak by recalling the product. Similarly, more weight might be given to food-hazard pairs that cause acute illness because these are also more likely to lead to overt outbreaks of illness that will be mitigated rapidly by a recall.</p> <p>[Clarification: If the aim of the rule is to establish additional record keeping so as to be able to identify and recall foods associated with an outbreak, foods with a short shelf life after they are sold at retail, may not be easily identified and recalled before they are consumed. If so, there's no point keeping the additional record.</p> <p>Note, though, that I'm not saying it is futile to maintain records for those products, but that it ought to be thought about, i.e., whether the additional records would enable rapid and effective prevention/mitigation of further cases of illness. If not, another strategy would be preferable.]</p> <p>Finally, even though the primary objective of the HRF record-keeping initiative seems to relate to protection of public health, the economic costs of illness outbreaks might also affect decisions about cost vs. benefit of the HRF initiative, or other risk management actions. The economic costs of an outbreak are not, however, limited to medical costs (<i>see</i> Scharf, 2012 cited in the report). Thus, for an effective cost-benefit analysis a broader analysis of costs than has been considered in the current model, including, e.g., loss of consumer confidence leading to reduced sales for an industry/product, loss of national productivity due to days off work, etc. should also be considered.</p>	

<b>CHARGE QUESTION 10: <i>How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	<p>The model uses an impressive amount of data, which were retrieved from numerous data sources and collected over more than ten years. I could see that more recent data from e.g., surveillance programs (namely outbreak data, used for criteria 1 and 7) are publicly available, and this could now be included. However, for most other criteria new data would probably come slowly (e.g., scientific articles), and thus updating the model would make sense only every few years.</p> <p>That said, recent efforts to estimate the burden of foodborne diseases, particularly caused by chemical hazards, would be worth evaluating and including.</p>	
Reviewer #2	<p>Once the model is in use, areas for improvement and new data will become apparent, so it would be useful to update the model after three years of usage.</p>	

<b>CHARGE QUESTION 10: <i>How often should the model be updated, considering the data sources and data currently available and data that might become available in the future?</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #3	It is difficult to give an answer but this should probably be based on changes of criteria which could evolve such as C1. C6 may be the most critical as in today's world eating habits may change rapidly to follow new social trends and globalization (see comment on C6, page 19 in the specific observations).	
Reviewer #4	I am not sufficiently expert to be able to offer a reliable response to this question but would suggest that the rate of change of food processes and product formulation might mean that the model and criterion values would need to be reviewed every 3 to 5 years. However, if it becomes evident that radically different processes or products are introduced, or products are sourced from new/different suppliers, it would be prudent to evaluate <i>before introduction of those products</i> whether those changes introduce a different level of public health risk.	

<b>CHARGE QUESTION 11: <i>Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	As mentioned in the general comments, the report is very clear, well written and well structured. The level of detail is adequate in the sense that it allows for understanding the approach and decisions, and refers to very detailed, complete and transparent appendices. I would perhaps suggest having a chapter on the general approach on uncertainty earlier on the report. Even though the current description is clear, I did wonder about it while I was reading the earlier chapters.	
Reviewer #2	6.1 The risk ranking approach is clear. Figure 6-1 is especially useful.	
Reviewer #3	<p>Overall the purpose of the report is apparent after repeated reading. However, aside of more technical question as the previous charges, the report would, as discussed above as well in the following section III, greatly benefit from serious editorial review, to streamline the structure, avoiding numerous repetitions, grouping of related text currently spread across different sections/chapters, introduction of new terms without explaining them (e.g., score used in relation with criteria and then later for uncertainty/confidence – is it really a score?), one equation in section 2, one in section 4 and another one in one of the figures + others somewhere else. The figures in chapter are very difficult to read and not always aligned with text (or contain new elements such as equations not explained before).</p> <p>Simple elements such as sub-sections on the individual hazards would help better understanding of “what is what” without having to read several times the text, list of definitions, list of acronyms, etc., changes in terminology. Expert conclusions not always understandable – e.g., in the examples...”has been decided or changed because of expert ...” does not help in transparency and understanding of the rationale.</p>	

<b>CHARGE QUESTION 11: <i>Is the draft report clear in its description of the risk ranking approach, criteria, data and results, and model limitations? If not, please identify aspects that are unclear or could be more transparent.</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	It is also quite cumbersome to find associated information across the numerous annexes and excel-spread sheets.	
Reviewer #4	Many comments were made above about the transparency of the process and the model, identification and documentation of relevant and authoritative sources, and limitations in the data and the model. Other comments will be made in the separate report on the model and document itself.	

<b>CHARGE QUESTION 12: <i>Do you have any additional comments? Please share them in your review.</i></b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
Reviewer #1	No additional comments, except for the editorial ones in section III. Many of these are very important for clarity in the presentation of the report.	
Reviewer #2	[Reviewer did not comment.]	
Reviewer #3	[Reviewer did not comment.]	
Reviewer #4	<p>During the evaluation of selected product:hazard pairings it was noted that for Criterion 1, of the 72 scores of '9', 52 (70%) were allocated to allergens but all were based entirely on expert opinion and without any evidence nominated. The other 20 were for microbial hazards. <i>Prima facie</i>, this seems incongruous, and is perhaps a consequence of different experts being used to assess the severity of different categories of hazards. Importantly, the Minor <i>et al.</i> (2015) evaluation of hazard severity gave <u>much</u> lower ratings to allergens than were derived by experts in this risk ranking process. The scoring for allergens should be reviewed to ensure that all experts (not just allergen experts) believe that scoring is consistent across all hazards, and is also consistent with other published opinions/reports based on critical analysis of the available data.</p> <p>As discussed in relation to Criterion 2 for STEC <i>E. coli</i> in bagged leafy green salads, there seems to be an inconsistency between the hazard severity ratings developed in this model/risk ranking tool and the severity estimates developed by Minor <i>et al.</i> (2015). This aspect of the model/risk ranking tool should be re-evaluated to ensure that the relative severity measures in the model are scientifically defensible.</p> <p>The requirement to identify “iii) the point in the manufacturing process of the food where contamination is most likely to occur” does not really seem to have been addressed in the model presented, nor in the evaluation of criterion scores. This requirement possibly relates to another perceived difficulty in the model, i.e., that it does not explicitly recognize the importance of sequences of events in the evaluation of relative risk, which was commented on above in relation to limitations of additive scoring schemes for microbial food safety (<i>see</i> response to Question 1d).</p>	

<b>CHARGE QUESTION 12: Do you have any additional comments? Please share them in your review.</b>		
<b>REVIEWER</b>	<b>COMMENT</b>	<b>RESPONSE</b>
	<p>It is noted that “Section 204(d)(1) of the FDA Food Safety Modernization Act (FSMA), requires that the Secretary establish additional record-keeping requirements for high-risk foods in order to rapidly and effectively identify recipients of a food to prevent or mitigate a foodborne illness outbreak and to address credible threats of serious adverse health consequences or death to humans or animals.” Given that mandate, what is the intent of the questions about ‘economic impact’, i.e., how does economic impact affect the decision about high-risk foods requiring additional record-keeping?</p> <p>Also, in Appendix C (page C-8, last paragraph) of Appendix K to the Sept 2015 draft report, it is stated that the estimates of Scharff (2012), in conjunction with Scallan <i>et al.</i> (2011) were used to assess economic impact and noted that the Scharff estimates included “financial losses due to medical expenses, lost productivity and lost utility”. In the September 2015 draft report, however, it states (p. 65) that: “The cost-per-case estimates for the current analysis were largely drawn from the most recent study (Minor <i>et al.</i> 2015). Cost-per-case estimates from Scharff (2011 [<i>sic</i>]) were used for hazards that did not have estimates from Minor <i>et al.</i> (2015).” (Note that on P. 65, Lines 1- 8, the Scharff paper is mis-cited several times, <i>i.e.</i>, the publication year is 2012, not 2011). However, <b>having checked through the lists of hazards in both Scharff (2012) and Minor <i>et al.</i> (2015), I see no hazard considered in Scharff (2012) that is not also considered in Minor <i>et al.</i> (2015). Accordingly, the product:hazard pair estimates for which the Scharff (2012) estimates were used should be clearly articulated or the text of the report appropriately modified.</b></p> <p>Further, it should probably be clarified in the document that, to estimate the economic impact, ‘non-public health impacts such as potential industry costs and loss of market costs are not be [<i>sic</i>] included in this criterion’ (Draft report, p.24, line 7, 8) and that the earlier use of the Scharff (2012) estimates was abandoned because lost productivity costs were not to be included (if that is, in fact, the explanation).</p>	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
Reviewer #1			[Reviewer did not comment.]	
Reviewer #2	General		In general, all of the scoring grids should be presented in the same format. For example, scores are given within the grid in Figure 2-2, but as column labels in Table 2-1. Placing the scores consistently in the grid would have greater transparency and make them easier to use.	



<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	General		A glossary of technical terms used (e.g., criteria, factors, bin, risk, hazard, acute, chronic, foodborne disease, granularity, etc.) would greatly improve the report's utility for the nonexpert.	
	12	33 and others	"Granularity" can be considered jargon. It might be helpful to define it here as, e.g., "level of detail." The definition could go in the glossary.	
	8	20-29	How the 7 criteria relate to the 6 factors was unclear in the first reading. Figure 2.6 is a great explanation. Could it be moved up to follow the initial paragraphs on criteria and factors? The figure on Slide 33 from the webinar was very instructive in putting everything together and should be included in the report, perhaps after Figure 2.1.	
	C1	Figure 2.2	The figure would be clearer if the column headings were above rather than below the columns. As labeled, the difference between "frequency of outbreaks" "and occurrence of illness" is unclear. (Is this per year or over the whole period?) Does the latter refer to the number of cases since 1998 or the number of cases per outbreak?	
	Table 2.1		The rows should be labeled, e.g., "hospitalization data," "severity data." % relative to what should be defined. Presumably it's % of cases. It is unclear whether score is based on "hospitalization data," and/or, "severity data."	
	Table 2.3		Same concerns as Table 2.1.	
	40		This refers to two methods of scoring. It was unclear what these are.	
	73	10-15	It would be useful to give a sample calculation here.	
	73	27	"PAG" is not included in list of abbreviations.	
	75	25	This paragraph is very difficult to follow, especially without having the model in hand. For example, line 35 refers to Risk Scores, Food Rank, and FRRS. It took several readings to understand that the Risk Score gives rise to the Food Rank, which is the FRRS. FRRS is not included in list of abbreviations.	
	85	1	This sentence needs a verb. Should it read "Appendix L <i>includes</i> the food risk score and ..."?	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	General		Biological hazards, chemical hazards, and undeclared hazards are not defined until page 27-28 of the report. They should be defined at the beginning of the report or in the glossary.	
Reviewer #3	8	25/26 (iii)	The term processing step is probably more appropriate than point of manufacturing process	
	8	27 (iv)	Normally one refers to control or preventive measures rather than “steps taken”, in addition “step” is used in (iii) with a different meaning.	
	8	27 (iv)	The reference to “manufacturing process” does not seem to be aligned with further sections dealing with the subject – e.g., 2.2.5 talking about “the entire food supply chain” which encompasses more than the manufacturing process.	
	9	1	What are “ <u>the</u> statutory factors”?	
	9	14	What is RTI – acronyms should be spelled out, at least the first time it’s used. Valid throughout the document. A table summarizing them would be useful.	
	9	34	What is meant by “additional analysis” – information, data?	
	10	5	Since animal food/feed is not covered – should this category not be deleted from Appendix A nos. 3 and 4, or at least a comment provided?	
	11	19	“...and adapted it to account...” to which model does this refer?	
	12	11	See section Expert Elicitations below...where below?	
	12	37	An annex with secondary commodities would be helpful	
	13	14	Assume you are referring to factors listed on page 8 – cross-reference would be helpful	
	13	22	Section 2.2.5 speaks of “manufacturing process” as one of the elements of the whole supply chain – should Criterion 5 not also reflect this?	
	13	26	Different for undeclared allergens which are not mentioned here?	
	14	11	A numerical value <b>from 0 to 9</b> – this is not really correct as only agreed upon values are used, 0, 1, 3 and 9 and not any value between these two limits.	
14	16	Since this section covers the three hazards – it would certainly help the reader to have a further sub-division: 2.2.1.1, 2.2.1.2. and 2.2.1.3 for the individual hazards.		

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	14	23	What about an anaphylactic shock caused by undeclared allergens such as peanuts- is this not part of the acute effects?	
	15	2 / 3	Included or include? In the rest of the document undeclared allergens are handled separately (text, comments, examples), why then saying they are included in chemical hazards. Should the “hazards” not be described/defined up front as they are generally valid and not just for 2.2.1	
	15	Figure	Should read low <b>tens</b>	
	15	9	“food hazard pair” – is this valid for all three hazards or just micro (see my comment on a further page)	
	15	12	“Sporadic cases” – how is this defined?	
	15	26	What does “including the outbreak itself” mean?	
	15	27	Why is “timing of outbreak data” defined by (C1)?	
	15	27	Is C1 definition correct – not rather scoring?	
	15	6, 11-13	Are marine biotoxins such as algal toxins not chemical hazards – why not include them after row 24 which refers broadly to “chemical hazards”? See comment page 14/Line 16 on sub-division – this would help avoiding confusions.	
	16	8 / 9	Fig 2.2 “Have the potential to be involved – what does this mean? Unclear since, as understood, Figure 2-2 shows “real” outbreaks. How could potential be classified as low, medium or high?	
	16	15	Why e.g., undeclared allergens and chemicals – does this imply also microbiological hazards which are the only ones also included in the model?	
	16	26	Reference is made in this part on “U.S. food consumption” – does this mean this aspect is only valid for chemical hazards, not for other hazards? Does this also apply to the data used – only U.S. data for chemical contaminants? Is this not an element which is anyways considered for C6?	
	17	5 – 10	It is ICMSF (2002) – since the chemical hazards are mycotoxins and not any other “traditional” chemical hazard (e.g., arsenic used in later examples, it would be good to specify to avoid confusion).	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	17	15	Also here it is confusing as here the chemical hazards referred to are not of microbial origin (which include mycotoxin).	
	17	Tables 2-1	Title –hazard – all of them? 1 <sup>st</sup> column/3 <sup>rd</sup> row Should probably read “not a hazard” to be consistent with the rest of the row. 2 <sup>nd</sup> column/3 <sup>rd</sup> row – with little or <b>no</b> medical 3 <sup>rd</sup> column/3 <sup>rd</sup> row – is “not life threatening” aligned with 0.5% mortality?	
	18	Table 2-2	1 <sup>st</sup> column/3 <sup>rd</sup> row should probably read “not a hazard” to be consistent with the rest of the row. Since the descriptions of the hazards are identical to those in Table 2-1 – why making a difference? The text itself is more descriptive of acute symptoms short duration, “self-limiting” effects, little medical intervention, life threatening – does not sound like effects of chronic exposure leading at some stage to effects. What is IQ reduction considering that sequelae are infrequent?	
	19	5	Is food supply system – the food supply chain as described under 2.2.5? Suggest to use same terminology throughout report.	
	19	21	Unclear why the geographical origin is used to weight as this chapter deals with the likelihood of presence in a finished product (irrespective of where it’s produced) and not with the exposure of the consumer. The fact that a food is not consumed in the U.S. or only in limited quantities is considered under 2.2.6. What if the consumption of a certain food is increasing in the U.S. or completely new – does it mean you will have to change C3 and C6? Why then not consider outbreaks related to those foods which have happen outside of the U.S. differently than those within the U.S.?	
	19	36	Is a scoring bin a definition?	
	19/20	35 – 40/1 – 8	Suggest to move discussion around what is a RTE/NRTE to a separate section –e.g., along with discussion on food categories. It is not specific to C3 and it also detracts from the content of the section – likelihood of contamination.	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	20	Table 2-3	<p>1<sup>st</sup> column/1<sup>st</sup> row – why not just “0” since n (positive samples) = 0 or no reported occurrence? “Known” sounds “qualitative.”</p> <p>1<sup>st</sup> column/2<sup>nd</sup> row – what are “indicators” – defined or explained somewhere before?</p> <p>2<sup>nd</sup> column/1<sup>st</sup> row – is <math>\leq 0.1\%</math> which includes 0 correct?</p> <p>3<sup>rd</sup> column/1<sup>st</sup> row – should probably read <math>&gt;0.1</math>.</p> <p>3<sup>rd</sup> column/1<sup>st</sup> row – should probably read <math>&gt;1</math>.</p> <p>How are the two rows connected is it one <b>and</b> the other? For example what if 0.1% occurrence but no recalls?</p>	
	21	1 – 6	Why is the same parameter “includes conflicting studies...” valid to assign to both Moderate and Low?	
	21	21	C5 refers to manufacturing process, while the text below explains that this is only one of the elements. C5 is, according to this text the entire food supply chain – from the farm to the consumer (as per line 27). Believe this needs clarification.	
	21	24	All three hazards included here- would be good to specify here to avoid confusion.	
	21/22	28 / 1-6	Would certainly be clearer if sentences on C5 are grouped and not mixed with C3. C3 is in principle the outcome of C5.	
	22	16	<p>According to the introductory part of 2.2.5, manufacturing is just one element – wording should be consistent as it is also a possibility to establish the contribution of individual steps in the food chain, e.g., before and after a kill step.</p> <p>As a consequence what is assessed in terms of contamination probability – the occurrence in finished products, of raw materials, intermediate product or processing environments leading to contamination of product? Believe it would be important to understand the rationale to understand Figure 2-4, as to some extent it seems that it is very much focused on finished product and hence not very different to C3.</p>	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	22	18/19	Moderate: Sporadic; Low: Infrequent detection of contamination – what is the difference? What means “contamination introduced post manufacturing” – post-kill step, during the distribution, during preparation by consumer? Why should this type of contamination be of a low probability?	
	22	25/26	Control measure is probably a more usual term in food processing than “step”. Would Effective not be more accurate than “strong”? Is the assessment and rating done for a whole industry or for individual branches or factories?	
	22	34	Seems that discussion around manufacturing should be moved up front at the beginning of this section for a more logical structure.	
	23	1	What are activities? Unclear why the physical location is not taken into account as it is an essential part of the hygiene control measures – e.g., related to post-process (kill) step contamination.	
	23	Figure 2-4	What is the contamination potential – same as probability used before? Scoring not completely clear – if control measures are effective, why is the contamination potential rated 3 for “high”, which would rather correspond to “not completely effective”, the same for the combination weak/low – in the case of post-kill step contamination even low levels of, for example <i>Salmonella</i> , will invariably lead over time to a high probability of contamination.	
	23	5	What does this mean – negative results for all elements contributing to contamination, e.g., raw materials, processing environment etc.? It would seem more appropriate to have the first column scored 1, the second scored 3 and the last scored 9. ICMSF Volumes 6 and 8 provide estimates on the relative impact of different steps in the food chain in terms of contamination – does not seem to have been taken into consideration.	
	24	10/11	What is distinct from C1 and C2?	
	24	15	“Where appropriate....” – according to Figure 4-7 it’s done in any case...	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	24	16/17	For other criteria details (e.g., values) are provided on factors to weight scores as well as calculations (Equation) – why not in this section and only in Figure 4-7?	
	24	Table 2-5	How are the qualitative ratings established, lower or higher than what?	
	25	8	Should the rationale to decide on “reasonably foreseeable hazards” not be explained somewhere? E.g., same ecology, same behavior, same sources....	
	28	3	The rationale for dealing with histamine is clear – why was not the same rationale used for mycotoxins whose formation is also dependent on growth of molds and therefore elements considered for bacterial growth equally valid?	
	31	5	Should probably read scored instead of defined	
	31	14	Term attribute introduced here – along with indicator, underlying indicator (line 19), data indicator in Table 4-2. None of them seems to have been defined and explained in the introductory part of the document.	
	32	Table 4-2	<p>Title – what is the difference between ranking criteria and model criteria as per 2.2? Is this table not a summary of elements used to determine individual criteria 1 to 7?</p> <p>5<sup>th</sup> column/1<sup>st</sup> and 3<sup>rd</sup> row: would probably be useful to qualify the chemical hazard (acute, chronic) which are handled differently.</p> <p>4<sup>th</sup> column/4<sup>th</sup> – 6<sup>th</sup> row – also here difference acute/chronic +Likelihood of contamination” – “Average number/year – seems to be the number of recalls or reports per year – hence an absolute number and not an average.</p> <p>C4 – Is it not time rather than days as for certain products shelf life can be months or even years?</p>	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	35	Figure 4-1	First step – why not just referring to Table 4-1, rather citing a few elements of info (not complete – rationale)? Second box/right – 2. Is only if several data sets are available. Third box/right – why introducing this table here and not page 15 where the subject is dealt with in details? Fourth box/right – discrepancies between table and text as text page 15/16 do not speak about “experts” just “0” or does not refer to “the occurrence of illnesses includes outbreak-associated cases <b>only</b> ” – Figure 4.1 does not seem to be aligned.	
	37	2	Why e.g., in brackets? This gives the impression there are others. However considering that only few hazards are dealt with, would it not be clearer just to cite the ones for which the statement is valid? Otherwise there is ambiguity.	
	37	11	Why not introducing and explaining the concept of confidence and uncertainty scores (is it really a score not a level?), e.g., in chapter 2?	
	37	35	Why not using the example Cantaloupe- <i>Salmonella</i> as for the other criteria?	
	39	28	Estimates of estimates – or should it read CDC reports?	
	40	5	Any particular reason to write – including the definitions for microbiological and chemical hazards?	
	41	Figure 4-2	First boxes right – examples of references of the boxes point to chemical/allergen hazards for the “Quantitative” one and to micro for the “Qualitative”, while text on page 40/line 4 – 17 says something different. 2 <sup>nd</sup> box/left – should it not read “the most instead of “more”?”	
	42	24	Wrong reference – ICMSF (2002) does not provide any information on arsenic.	
	43	6	Same comment – Book 7 does not address allergens.	



<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	44	Figure 4-3	2 <sup>nd</sup> box /right – why referring to ICMSF 2002 considering that Books 6 and 8 are much better references for information on contamination? 4 <sup>th</sup> boxes/right – should probably read “study weight” 6 <sup>th</sup> box/right – Equation shows $n_p$ text below n. Why does this equation appear here and not earlier in the text as well $n_p$ as in the case of Equation 1?	
	45	Figure 4-3	Several indexes not explained 1 <sup>st</sup> box/ right For score = should it not read 0%, and then $>0 - <1\%$ ? For score 3 it should probably read $>0.1$ as the 0.1% is already included in score 2. Are the qualitative qualifiers needed since quantitative % are given? 2 <sup>nd</sup> box/left – if weighted prevalence is 0, what is the purpose for moving it to the next step? Does it mean a calculated and weighted can be changed further down? 3 <sup>rd</sup> box/right – should score 1 not read 1 – 5, considering that score 0 is “0” and $<5$ is then undefined (could also be 0...)? 5 <sup>th</sup> box/left – would eLEXNET not be used in the previous step by an expert anyways?	
	46	27	Why is this equation (probably not 2, considering the one on weighted positive samples) not introduced and explained earlier – chapter 2.	
	48	11	Unclear – what if they have a quantitative “0”? Should then the scoring not stop at this level.	
	48	15	Unclear – according to Figure 4-3 it’s a number of reports per year and not an average (not sure anyway to understand how one can do an average of reports per year).	
	48	34	Not sure to understand why eLEXNET is needed if recall $>0$	
	48	2	According to Figure 4-3 expert opinion comes before eLEXNET. Unclear why a process is continued for scores determined as score “0”	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	48/49	Example C3.1	It is said that if no “quantitative”, then RFR – in this example there is however a quantitative prevalence, hence a score of 3 based on the 0.246% - what is the aim- to continue with the process with other elements? Seems that whether data are available or not the process goes on. ... In terms of example, it would certainly be more illustrative to have one where certain elements are missing and would really require to go to a next step.	
	49	34/35	Seems to be inverted – according to Figure 4.3 RFR comes before recall – but also here there is a weighted positive rate, so why pursuing the process? Seems that is principle of the two options – response available = “stop at this level” and response not available = “continue to next level” is not applied.	
	51	9	Does seem as if Figure 4-3 is not followed – in principle with a positive rate of > 1% then the score should be 9 (as indicated in line 27 of previous page). Does C3 = 3, now mean that scoring is changed afterwards....or the rationale is not aligned with the example? Clarification would be needed.	
	51	Example C3.3	Same comment – according to Figure 4-3 RFR is considered first: C = 1. What is the purpose of looking at recalls which provides a C = 3... which is then ignored? Or is the purpose to determine confidence/uncertainty? If this is the case, then this should be explained better in the text.	
	52	13	Growth potential – has a reference rate been established to assess growth potential, e.g., expressed as minimum log increase to differentiate between foods supporting or not growth? This would also provide experts a common reference and avoid different interpretations on what “potential for growth” means.	
	52	16	Microbial or bacterial pathogen – two different things and putting one in brackets does not help a lot.	
	52	17	Also undeclared allergens do not grow	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	52	17 – 21	Unclear – the shelf-life will have an impact on the levels reached if growth is supported. The term the “extent” seems not appropriate – also because in the bracket there is a need to explain further and then again in another sentence (line 23). To be logical the determination on whether a food can support growth comes first and then the shelf-life which will determine the levels. In this text the shelf-life seems to be considered first...at least according to the sequence of the discussion and numerous repetitions of the same. These two indicators – which one, as the text before is quite general?	
	52	22	Unclear what is meant by taking into account the point where contamination takes place. How does this relate to potential for growth and shelf-life and how is this taken into consideration?	
	52	30	What about molds which can grow at lower water activities? The potential for growth and associated potential mycotoxins formation does not seem to have been considered in this model as mycotoxins are handled as “pure chemicals” unlike histamine, enterotoxins.	
	52	34	Unclear – microbial hazards is wider than bacterial hazards, e.g., molds. Is a pathogen not a hazard per definition?	
	53	Figure 4-4	Why assigning a shelf-life for foods which do not support growth? 4 <sup>th</sup> box/right – is it here necessary to indicate days which are only an example for a specific commodity? Why not providing more guidance (introductory part) on how to proceed with other commodities? The table does not include foods not supporting growth, while they seem to have to undergo the whole process.	
	54	Example C4.1	Is it whole cantaloupe, fresh cut? Why assigning confidence/uncertainty levels for both elements – not as in Figure 4-3? What means moderate – not exceeding a certain level?	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	55	Examples C4.2 & C.4.3	Why considering shelf-life since there is anyways no potential for growth – or were other aspects of arsenic considered which would modify toxicity over time? Same comment applies for allergens – is it not a bit of a waste of time to look for shelf-life in this case. While the confidence /uncertainty is 9/1 for both elements, the overall is 1/9. What is the rationale for this change?	
	57	7	Book 8 ICMSF provides lot of information of contamination in the food chain – does not seem to have been used.	
	58	Figure 4-5	3 <sup>rd</sup> box/right – the wording seems very quantitative, while most is based on a qualitative assessment.	
	59	Example C5.1	The scoring will very much depend on the type of product – whole fruit, fresh-cut. Since there are several “1s” in the table – which one is it? Unclear how two scores of “1”, both on expert advice can then give a final score of “3”, also on expert advice (different ones?). Would be good to explain the rationale. Same for example C5.2 the final score of 1 is just a fact – what is the rationale for different expert judgements as this is what is expect in a section entitled “rationale behind scoring”?	
	60	Example C5.3	Not sure I understand why a high score of “9” has a very low confidence level – this seems quite contradictory. These two elements would certainly benefit from more explanations in the text as to their determination, role, etc.	
	62	Figure 4-6	4 <sup>th</sup> and 5 <sup>th</sup> box/right there is overlap for scores 0 and 1 in terms of %, since the 1 in 1-5% belongs also to $\leq 1$ of score 0. Should probably read $>1-5\%$ .	
	66	Figure 4-7	2 <sup>nd</sup> box/right – Equation is missing on page 24, the location where one would expect explanations on calculations.	
	67	3	Cases – illnesses, outbreaks, sporadic cases? Would be helpful to have definitions somewhere clarifying terminologies used throughout the document (and which are not always used consistently).	

<b>III. Specific Observations on Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the supporting data.</b>				
<b>REVIEWER</b>	<b>Page</b>	<b>Paragraph/ Line</b>	<b>Comment</b>	<b>RESPONSE</b>
	75	28	(See section below) – where below, there is no further section in this chapter – it would be clearer if cross-references within the report – see above or see below would be identified with the relevant reference (e.g., see 2.2.3).	
Reviewer #4			Specific observations were presented above in the responses to the Charge Questions. Suggested corrections will be presented in the complementary report on the Model.	

<b>IV. Specific Observations on Appendices to the Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the model itself</b>					
<b>REVIEWER</b>	<b>Appendix</b>	<b>Page/ Row</b>	<b>Paragraph/ Line/Column</b>	<b>Comment</b>	<b>RESPONSE</b>
Reviewer #1				[Reviewer did not comment.]	
Reviewer #2				None to report.	
Reviewer #3	A	1/no 1		LACF – abbreviation should be explained somewhere.	
	A	1/no 1		The definition” of baby is quite vague. Infant is normally defined as 0 – 12 months. Junior does also not seem to be an official definition – to what age does this correspond – 1 – 6 years?	
	A	1/no.2		What is N.E.C?	
	A	1/ nos. 3 & 4		See comments page 10 of main document	
	A	1/ no.7		What means “Waters” – flavored water? If the meaning is mineral or bottled water, then this would be quite a different category from soft-drinks.	
	A	1/ no. 9		Liquid coffee and tea or dry?	
	A	1/ no. 17		Does this include all types of products: raw, pasteurized, ESL, UHT?	
	A	1/ no. 25		Are these prepared dishes or also frozen vegetables	

IV. Specific Observations on Appendices to the Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the model itself					
REVIEWER	Appendix	Page/ Row	Paragraph/ Line/Column	Comment	RESPONSE
	A	1/ no. 28		Are simulated meats an example? If the case, then “e.g.,” would be appropriate as there are other example such as soya or bean flours.	
	A	2 /no. 37		Are refrigerated RTE salads an example of refrigerated RTE foods such as delis, cold cuts refrigerated pasta or prepared dishes etc., or are those included in NEC (whatever it means)? What is the rationale to single out salads?	
	A	2/ no. 45		What is “soup” – dry, refrigerated, canned? No secondary category found.	
	B	1	<i>Aeromonas</i>	To our knowledge it is <i>Aeromonas hydrophila</i> and should be specifically mentioned to be consistent with other pathogens such as <i>L. monocytogenes</i> , not just <i>Listeria</i> .	
	B	1	Hazards 4, 20, 21, 22, 23, 34	Hazards are listed but have not been considered further – e.g., absent in appendix D2.	
	B	1	Hazard 14	<i>Enterococcus faecalis</i> – absent in appendix D2, but ETEC listed as hazard 14.	
	B	1	Hazard 18	Norovirus – but in appendix D2, it’s <i>Listeria monocytogenes</i> .	
	B	1	Hazard 19	Parasite – but absent in appendix D2 and becomes Norovirus.  These are three examples but there are more where there is a discrepancy between Annex B and others. In principle B should be the master and the others aligned.	
	C	1	15	Wonder why <i>Cryptosporidium parvum</i> appears – have tried to find information in the numerous spreadsheets and annexes to understand– considering the type of product, preparation, etc. Believe it is this type of example which would benefit from a review to decide whether it is really beneficial for public health to perform the whole risk ranking – while beans are probably just a vehicle and not a systematic carrier of this parasite.	

<b>IV. Specific Observations on Appendices to the Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the model itself</b>					
<b>REVIEWER</b>	<b>Appendix</b>	<b>Page/ Row</b>	<b>Paragraph/ Line/Column</b>	<b>Comment</b>	<b>RESPONSE</b>
	C	2	23	Same problem with <i>Listeria</i> in canned foods – realistic or not, what is the cause? Poor hygiene in handling after opening but not a systematic problem.	
	C	5	112	Muffins and <i>norovirus</i> – is muffin really the issue and not, for example one of the ingredients such as berries? The assessment would certainly be more beneficial if a specific ingredient could be targeted – would also allow improvement – rather than “all muffins” irrespective of their recipe. If a specific ingredient is identified as high risk, this would allow to focus preventive measures to that one which may also be used in other products.	
	H	1		The text indicates that food/feed for animals is out of the scope – no. 1- 3 refer pet food. If this is due to the occurrence of human cases, then these are not the only outbreaks. Should be clarified in the text.	
	File 09b (I2)			File name does not match content – comments on Criteria rather than info on control measures.	
	L		Undeclared allergens	Rating, e.g., of C1 does not seem to be consistent – e.g., condiments is scored 1, while ingredients which are used to manufacture them such as flavorings, flour, spices are scored higher. Bakery products with scores of 1 up to 9 for C1 and C5 – basically manufactured with the same ingredients, frequently on the same lines and some being ingredients of others.	
	L		<i>Cronobacter</i>	C1 rated as “0” including for infant formula despite around 50 outbreaks and sporadic cases since 1998. The C2 score corresponds to the sensitivity of infants up 6 months, while at this age they will never consume chocolate. Does C6 correspond to the infant population which is consuming 100% of the products, C6 for coatings – not relevant for infants?	

<b>IV. Specific Observations on Appendices to the Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the model itself</b>					
<b>REVIEWER</b>	<b>Appendix</b>	<b>Page/ Row</b>	<b>Paragraph/ Line/Column</b>	<b>Comment</b>	<b>RESPONSE</b>
	L		Beverages	<i>C. botulinum</i> is mentioned as C2 = 9 and C4 = 0, hence no growth and therefore C2 is questionable as toxin formation depends on growth (only exception is infant botulism, but then it would be, if at all, a very defined category of beverages). The same is valid for some commodities rated as C4 = 0 – no growth = no toxin. The only exception is honey causing infant botulism.	
	L		<i>S. aureus</i>	Similar problem – microorganism as such is not causing any illness, only its toxin. This, however requires growth and in several cases there is a discrepancy between scores – e.g., scores of 1 or 3 for C1, C3 but C4 of 0 or 1 or C5 is wrongly scored.	
	L		<i>B. cereus</i>	Not sure why C3 was rated 0 and C4 as 1 – number of outbreaks are associated with cooked rice and growth is needed to permit toxin formation.	
Reviewer #4	List of Appendices	P4, Line 77		“Master” is misspelt	
	All Appendices and throughout Entire document			Inconsistent use of RPM-PT cf. HRF as abbreviation for model. It is noted that Appendix M1 explains that that name/acronym for the model changed in August 2015, yet nearly all the document (dated September 2015), including appendices, refers to the ‘HRF’. The main report includes a ‘header’ on each page that describes the report as the RRM-PT report. This needs to be made consistent throughout, or an explanation provided very early in the document.	
	Various			Several appendices (esp. App. D, App. N) are previously submitted reports. While those documents also contain various grammatical errors, I have not recorded them on the assumption that they cannot be changed in this report unless the original documents are also changed and re-issued.	



<b>IV. Specific Observations on Appendices to the Draft Report for Peer Review: Risk Ranking Model for Product Tracing as Required by Section 204 of FSMA (RRM-PT Draft Report) within the context of the model itself</b>					
<b>REVIEWER</b>	<b>Appendix</b>	<b>Page/ Row</b>	<b>Paragraph/ Line/Column</b>	<b>Comment</b>	<b>RESPONSE</b>
	Throughout			Inconsistent page numbering conventions. Some appendices (e.g., C, ) include the Appendix 'letter', others (e./g., J, M1) just restart the page numbering from 1	
	C	All pages		Pages numbers suggest that it is Appendix D, not 'C'	
	J	2	Line 10	Refers to Appendix J2, but there was no Appendix J2 provided to this reviewer	
	K	C-1	8 <sup>th</sup> line from bottom	Insert 'the' before IFT	
	K	C-11	Table 11	Why is 'livestock feed' included in this Table?	
	K	all		There is nothing in Appendix K to indicate that it is an appendix to the main report. It has its own 'Appendix Letter – Page Number' numbering, that makes it confusing when all the Appendices are combined as supplements to the main report. Needs some 'tidying up'	