FDA's Response to External Peer Review - Model Review[1] on FDA's "Draft Report for Peer Review: Risk-Ranking Model for Product Tracing as Required by Section 204 of FSMA (September 2015)"

August 2020

[1] "Summary Report: External Peer Review of FDA's Draft Risk-Ranking Model for Product Tracing as Required by Section 204 of FSMA – Model Review." Versar, Inc., March 18, 2016.

I. INTRODUCTION

As part of FDA's ongoing effort to provide the Agency with a risk-based decision support tool to assist the Agency in the process of designating high-risk foods, the Agency developed a Risk-Ranking Model for Product Tracing (RRM-PT).[2] The FDA Food Safety Modernization Act (FSMA) section 204 (21 U.S. Code § 2223) requires the Agency to designate high-risk foods for which additional recordkeeping requirements are necessary to protect public health. The Agency is required to determine which foods may warrant additional record-keeping by the food industry, so that individual items of those commodities may be more easily and quickly traced in the food chain, should they be implicated in an outbreak of foodborne illness or other events.

To address the statutory requirements of 21 U.S. Code § 2223, FDA developed a draft risk-ranking model for product tracing, which includes seven criteria related to public health risk. The RRM-PT took into account the characteristics of both foods and hazards, i.e., food-hazard pairs. It is adaptive to a variety of hazards (microbial and chemical hazards), is flexible enough to consider different food or categories of food and provides a means for considering the seven public health criteria and linking the criteria for risk ranking. The semi-quantitative risk ranking model scores food-hazard pairs according to data and seven criteria: (1) Frequency of Outbreaks and Occurrence of Illnesses, (2) Severity of Illness, (3) Likelihood of Contamination, (4) Growth Potential, with Consideration of Shelf Life, (5) Manufacturing Process Contamination and Industry-wide Intervention, (6) Consumption, and (7) Cost of Illness. These criteria are consistent with the requirements in section 204(d)(2)(A) (21 U.S. Code § 2223(d)(2)(A)).

The RRM-PT is a data-driven science-based decision support tool to assist the Agency in the process of designating a Food Traceability List as required by FSMA Section 204. The FDA Center for Food Safety and Applied Nutrition (CFSAN) developed the RRM-PT in consultation with an FDA Project Advisory Group (PAG) consisting of members from CFSAN, the FDA's Office of Foods and Veterinary Medicine, Center for Veterinary Medicine, and Office of Regulatory Affairs, and the Centers for Disease Control and Prevention. Contracts with the Institute of Food Technologists and RTI International provided technical assistance that included several expert elicitations from external expert panels.

Two separate peer-review panels of independent external experts reviewed a draft model and the underpinning data used to generate risk scores with the model, respectively. This report describes the comments from the peer review of the model. The peer review experts were asked to answer 12 charge questions, and to evaluate and provide written comments on the draft RRM-PT model, methodology report, and user's guide for a model user interface. The peer review focused on the conceptual framework of the model and associated equations and the underpinning relational

[2] Following this peer review, FDA changed the name of the model in 2019 to Risk-Ranking Model for Food Tracing.

database, as well as selected food-hazard combinations to evaluate the accuracy of scoring and assess the usability of the draft model interactive interface.

Peer reviewers[3] and affiliations:

Panos G. Georgopoulos, Ph.D.
Environmental and Occupational Health Sciences Institute (EOHSI)
Piscataway, NJ 08854

Igor Linkov, Ph.D.
Environmental Laboratory
U.S. Army Corps of Engineers
Boston, MA 02130

Thomas Ross, Ph.D.
School of Land and Food – Tasmanian Institute of Agriculture
University of Tasmania
Private Bag 54, Hobart, TASMANIA 7001 AUSTRALIA

Moez Sanaa, DVM, MSc., PhD
French Agency for Food, Environmental and Occupational Health & Safety
Risk Assessment Department
94701 MAISONS-ALFORT Cedex
France

Nga A. Tran, Dr.P.H., M.P.H., CIH
Exponent, Inc.
Washington, DC 20036

[3] Note: All five reviewers received the charge and model documentation and initiated their independent peer reviews. Four of the reviewers submitted comments while one reviewer was not able to submit comments due to unexpected family emergency.

II. CHARGE QUESTIONS

The draft Risk-Ranking Model for Product Tracing 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 model, in the context of the overall risk-ranking approach, criteria and results. Note: a separate panel is reviewing the underpinning data.

1. The draft risk-ranking model uses the FSMA statutory factors to define seven criteria for scoring food-hazard pairs.
    a. Are the seven criteria used in the draft model appropriate for a multicriteria decision approach? If not, please explain what changes might be considered and why.
    b. Within the bounds of the FSMA-mandated factors, are there additional criteria beyond the seven criteria that should be considered? If so, please describe these additional criteria that might be considered and why.

2. Are the scoring definitions for all criteria appropriate?
    a. Are the definitions appropriately defined for the various types of hazards considered (i.e., microbial hazards, chemical hazards (including chronic exposure) and undeclared allergens)? If not, please describe changes that might be considered and why.
    b. Is the value function 0-1-3-9 and scoring matrix appropriate for the intended purpose to inform the designation of high-risk foods? If not, please describe changes that might be considered and why.

3. Is the algorithm that combines criteria scores and weights into an overall score appropriate? If not, please provide suggestions on what improvements should be considered.

4. Considering the five different criteria weighting schemes described in Sections 6.2 and 6.3, are any one of these schemes (equal and non-equal weighting) not appropriate to consider for the intended purpose? For example, what weighting scheme is most useful? What weighting scheme should be avoided? Please make any additional recommendations on weighting schemes that might be considered for the proposed criteria and criteria indicators. Please explain the rationale behind your suggestions.

5. Considering the various scenarios (described in Section 7.3 and Appendix N) to aggregate food-hazard pairs in order to identify the foods which should be identified as high-risk vs. not

high-risk, which option(s) are more appropriate to consider and why? Are there additional aggregation method(s) that might be considered? Please explain.

6. Given the underpinning 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?
   a. What are the pros and cons in establishing a threshold considering all three types of hazards in the model, vs. drawing a line separately for microbial, chemical and undeclared allergens in foods?

7. Are the seven criteria and scoring definitions implemented appropriately in the SAS codes (Appendix P) and the Access Model?
   a. Does the scoring logic described in Section 4 Figures 4-1 to 4-7) appropriately represent the scoring definitions described for each of the criteria in Section 3? If not, please describe what changes need to be made to correct it.
   b. Are the scoring logic and order of preference accurately implemented in the SAS codes? (Please select 2-3 out of the 7 criteria for this evaluation). If not, please specify what changes need to be made.
   c. Are equations 1 through 4 and data weighting factors accurately implemented in the model (either the SAS codes or the Access Model)?

8. In the Access Model, is the underpinning relational database including lookup tables and algorithm appropriately designed and implemented? If not, please explain what changes should be considered.

9. Is the user interface of the Access Model sufficiently described for the user to understand each component of the model, e.g., foods, hazards, ranking criteria, results, and cited references?

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

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

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

III. SUMMARY OF PEER REVIEWERS COMMENTS AND FDA RESPONSE

General Impressions. Reviewers provided a number of general comments including: the document offers a robust approach for the purpose of risk ranking; the model is a major improvement over traditional dashboards or individual risk indicators and is straightforward, easy to understand and logical; use of multi-criteria models in risk prioritization are particularly suitable when it is impractical to build and populate a full causal risk model and when there is variability in the precision and accuracy of data across all foods and hazards; the scoring and underlying data supporting scoring for each of the seven criteria have been thoroughly examined by the Agency in the development of the model. Two reviewers highlighted the inclusion of both microbial and chemical hazards in the model, indicating significant progress has been made to adequately capture chemical risk in the model while some degree of unbalance emphasis on microbial hazards remains and that this initiative that includes both chemical and microbial hazards will be successful and recognized by other food safety agencies, respectively. Two reviewers shared concerns (elaborated in more detail under the relevant charge question) with criterion scoring and aggregation methods/algorithms used, suggesting alternatives including a different MCDA method and a more complex/quantitative risk assessment approach using tools such as FDA-iRISK.

FDA response.
We thank the reviewers for their thorough review and their general comments in this section regarding the appropriateness of the systematic approach and risk ranking methodology used in the model. We have considered concerns and issues raised by the reviewers and revised our approach and method as detailed below.

Question 1. There was general consensus that the seven criteria were appropriate. There were a number of specific comments/suggestions with regard to the definitions of criteria and relation to the statutory factors in FSMA 204(d)(2)(A) including a suggestion to separate Criteria 1 into two separate criteria (one for the number of cases and a different criterion for the frequency of outbreaks), a suggestion to enhance the definition of the consumption criteria to include consumption frequency, a suggestion to include dose-response relationships and level of contamination, a suggestion to change economic impact to the average cost per illness, and comments on the potential for some of the criteria to be correlated/confounded. There was also a suggestion that we consider weighing criterion 5 more heavily and a comment on the difference in scoring definitions for different types of hazards.

FDA response.
We appreciate the detailed comments from the reviewers. The scoring definitions for criteria 1, 4, 6 and 7 have been revised. We did not separate Criterion 1, which considers both the number

of cases and the frequency of outbreaks, into two separate criteria because we believe the matrix approach used in Criterion 1 better captures the relative public health impact of outbreaks. If we used separate criteria for numbers of cases and frequency, then one-off outbreaks would be too heavily weighted – an outbreak happened once but there is no additional indication in the outbreak data that it is a recurring problem - and frequent outbreaks that impact few people would be too heavily weighted – outbreaks occur relatively frequently but few people are impacted. The matrix approach appropriately combines these.

The risk ranking model is not a risk assessment and thus does not include a dose-response relationship nor does it rank based on per serving basis. The rank is based on a risk score which is obtained when combining the scores for the seven criteria in the model. Level of contamination was not included in the definition of Criterion 3 because these data are generally not available for most food-hazard pairs. The seven criteria for the RRM-PT encompass the range of factors suggested by the reviewer. We chose not to include the likelihood of a disease-causing dose as a criterion because an explicit consideration of dose-response relationship is more feasible in a quantitative approach; the RRM-PT is a semi-quantitative model. Furthermore, we believed dose is considered indirectly in the RRM-PT, in Criteria 1, 2 and 4 because the number of reported cases in an outbreak (Criterion 1) and the severity of illness (Criterion 2) may reflect the dose in the contaminated food, and the amount of growth (Criterion 4) may contribute to likelihood of illness.

Regarding the concern that criteria are correlated or confounded. We acknowledge there can be correlation between epidemiological data such as number of cases and risk assessment factors such as prevalence of hazards in foods; FSMA 204(d)(2)(A) requires inclusion of both. We were mindful of maintaining mutual independence among the criteria to the extent possible. We added additional text to the Methods document to clarify.

A sensitivity analysis considering all seven criteria was designed to explore different weighting schemes for the criteria in the model, including a scenario with a higher weight for criterion 5.

Regarding the comment on criterion 6 (consumption), we did not include consumption frequency because of the lack of data on consumption frequency. An additional comment later in the review suggested including “amount consumed”. We included amount consumed in the revised model.

For two of the criteria (Criteria 1 and 2), we proposed parallel but different scoring definitions for different types of hazard (microbial, chemical, undeclared allergen) or human health effects (acute, chronic) because directly comparable data across these are not available. We acknowledge that these differences in definition introduce some uncertainty into scoring when multiple types of hazards are considered.

Question 2. Reviewers provided several specific comments on the criteria scoring definitions and value functions, including suggesting FDA provide more detail on scoring chemical hazards and expanding the number of toxicology experts consulted in the future. One reviewer raised a concern that because chemicals do not grow in foods, the model may be preferentially scoring microbial hazards higher and suggested we combine elements of Criteria 3 and 4 into a single criterion. One reviewer recommended criterion 6 include the amount consumed. Regarding the value function 0-1-3-9, two reviewers pointed out potential concerns in scoring when varying size bins/ranges used across criteria. One reviewer noted use of scoring bins rather than a continuous scale can lead to some anomalies and asked that we provide additional justification as to why we didn't use a fully quantitative risk assessment approach. One reviewer was concerned that a majority of the highest ranked food-hazard pairs involved allergens.

FDA response.
We thank the reviewers for their suggestions to further improve the criteria scoring definitions. We have added more detail on scoring chemical hazards in the Methods document.

We do not agree with combining and redefining criteria 3 and 4 into one criterion; the potential for growth is an additional risk factor associated with microbes that is not present for chemicals. We agree that this difference should be noted when interpreting results of the model when both microbial and chemical hazards in foods are considered.

Criterion 6 was modified to include amount consumed in the scoring matrix.

We modified criteria scoring bins across the criteria to generally correspond to changes in order of magnitude. Through contracts, we also engaged external expert panels that evaluated and commented on the scoring scale. The external expert panels strongly supported the use of the 0, 1, 3, 9 value function. The experts were also asked to comment on the use of a 1, 3, 5 scoring scale compared to the 1, 3, 9 scoring scale in light of some public comments received on the risk ranking model and its approach. The SMEs strongly recommended that the scale should remain 1, 3, 9, because food safety risks are not expected to scale linearly, and this scale is based on a logarithmic scale, not a linear one.

We recognize that using scoring bins can cause some anomalies in ranking (particularly for values at the edge of the bins) but given the wide range in precision of the data available across all foods and hazards, including both quantitative and qualitative data, we believe it is the best compromise. We added text to the Methods document to explain our reasoning. We further enhanced the discussion of our modeling choice to use a semi-quantitative approach for the risk ranking model in the Methods document, including the fact that quantitative risk assessment cannot capture all the required FSMA 204 (d)(2)(A) factors, sufficient precise quantitative farm-to-table risk assessment data are not available for the full range of foods and associated hazards

identified, and it would not be feasible to complete such a task in a timely manner, despite the availability of computer software such as FDA-iRISK.

Due to the severity of illness according to expert opinion associated with consumption of foods with undeclared allergens by susceptible individuals, food-hazard pairs involving allergens tend to score higher in the model. Following this peer review, we conducted additional sensitivity analysis and examined an alternative scenario for Criterion 2 (severity of illnesses), where for undeclared allergens the scoring was based on illness severity measured by the loss of quality-adjusted life days (QALDs) reported by Minor *et al.* (2015). The alternative Criterion 2 scenario resulted in a shift to a lower risk score for some food-allergen pairs. We will further consider the reviewers' suggestion if undeclared allergens are included in the finalized rule.

Question 3. One reviewer stated that the overall process was reasonable, utilizing a structured knowledge based. Several suggestions were provided for alternative methods to combine criteria scores, including an outranking model, a multiplicative model, a weighted average of the seven criteria, and separate rankings for observed illnesses and potential illnesses.

FDA response. We considered the suggested alternatives to the ordinal scoring. Several multicriteria decision analysis (MCDA) algorithms used in food safety can result in some differences in ranking results (Garre et al., 2019). However, a critical review of methods for risk ranking of food related hazards based on risks for human health concluded there is no single best method (Van der Fels-Klerx et al., 2018). The authors recommended methods be selected on the basis of risk manager/assessor requirements, data availability and the characteristics of the method (Van der Fels-Klerx et al., 2018). In a study published by Duret et al. (2019), two MCDA methodologies (*i.e.* AHA and ELECTRE III) were evaluated to rank intervention options. Results showed that while the ranking of different actions resulted in slightly different scores, the ranking results between the methods remained similar (Duret et al., 2019).

We acknowledge that most food safety risk-affecting factors are not expected to scale linearly; rather, they are usually multiplicative in nature operating on logarithmic scales. We revised the scoring definitions for C1, C3 and C7 to approximately an order-of-magnitude difference. For example, the C1 scoring definitions represent a difference in order-of-magnitude, signifying large differences in public health impact while accommodating data with differing precision. When summing criteria scores to determine a risk score for a food-hazard pair, this scoring strategy is reflective of a risk model that operates on a logarithmic scale.

Question 4. There was general a consensus that selecting weighting schemes is challenging. One reviewer suggested a weighting scheme that considered differences in microbial, chemical

and allergen potential scoring (i.e., C4). Another reviewer suggested equal weighing for C1 (observed illnesses with a weight of 30) and C3 to C6 (potential illnesses with a weight of 7.5 each). Sensitivity analysis was suggested to explore the impact of different weighting schemes.

FDA response. There was no general consensus among the reviewers for an alternative weighting.

The different weighting schemes for sensitivity analysis considered available methodologies and were identified using an expert elicitation process. Four options were selected for non-equal weighting schemes as described in the revised report, i.e., the methods document (U.S. FDA 2019). These options give emphasis to different aspects of the risk-ranking model. While there are small shifts in some of the risk scores, high ranking foods generally remain higher-ranking and lower-ranking foods generally remain low ranking. Therefore, in the absence of a consensus of alternatives, FDA chose to continue with equal weighting of the criteria.

Question 5. The reviewers generally believed that the options presented were not appropriate. Reviewers recommended several alternative approaches.

FDA response. We appreciate the peer review comments on the various proposed approaches and the challenges noted for each option. Among the alternatives suggested, we selected the aggregation method that involves exponential transformation, summing and log transformation taking into account the risk scores, as the best scientific choice (see equation below).

$$\boldsymbol{aggrRS_{i\_}C = log_{10}(\sum_{j=1}^{n_i} 10^{RS_{i,j}})}$$

Where,

aggrRSi_C = Aggregated risk score associated with i^{th} commodity

$RS_{i,j}$ = Risk score associated with i^{th} food and j^{th} hazard

n_i = Number of hazards associated with i^{th} commodity

As the reviewer noted, when a food has risks attributable to multiple hazards (multiple food-hazard pairs for the same food), the overall risk is the sum of the risks, not the product of the individual risks. More specifically, adding risk scores that are based on logarithmic scales is logically equivalent to multiplying the risks from each hazard, rather than adding them together to achieve an overall risk estimate.

The equation was applied similarly to calculate an aggregated risk score for a commodity category. The risk scores associated with all food-hazard pairs in the category (which includes not only multiple hazards but also multiple commodities) were used to determine the aggregated

score for the commodity category. Aggregate risk score for a commodity category (aggrRS_CC), with risk score for each food-hazard pair given by RS_1, RS_2 … RS_n for *n* food-hazard pairs is:

$$\boldsymbol{aggrRS\_CC = 10 * (log_{10} (10^{RS_1\_C/10} + 10^{RS_2\_C/10} + 10^{RS_3\_C/10} + \ldots + 10^{RS_n\_C/10}))}$$

In this more logical approach, the aggregated risk score scales more naturally, and will be less affected by differences in the total number of hazards ascribed to the various food-hazard pairs, or by the total number of food-hazard pairs attributed to various commodities.

Question 6: The reviewers provided several suggestions for considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods, including: 1) use different thresholds for different types of hazards (microbial, chemical, undeclared allergens), including (i) one separated at the level of hazard type, and (ii) one collectively for the food; 2) take into consideration a range of factors that influence foodborne risk, including contamination, normal use, the existence of reliable critical control points, the severity of the symptoms, the frequency of consumption of the food, whether the hazard accumulates in the body of the consumer or whether each exposure is a discrete event; 3) discuss what to do in areas where there is lack of confidence of the data and results are highly uncertain; 4) use quantitative risk assessment to calibrate the model and set an Acceptable Limit of Protection (ALOP) across all foods (and hazards) and 5) include consideration of susceptible populations. One reviewer reiterated his/her concern that food-undeclared allergen pairs tend to score high in the model.

FDA Response. We appreciate the reviewers' suggestions for considerations to take into account when identifying high risk vs. not high risk food-hazard pairs or foods. Consistent with a risk analysis framework, the (risk management) process for this policy decision is functionally separate from the (risk assessment) model development, implementation, and risk ranking results. The model has been designed flexibly so different options can be explored. We also note that the seven criteria for the RRM-PT encompass the range of factors suggested by one of the reviewers.

We appreciate the reviewers' comments regarding the challenges in developing a transparent and objective risk evaluation system given the variety and complexity of the foods (and food processing steps), consumers with different susceptibility, and hazards (and how they respond to treatments) involved. We developed the RRM-PT to be transparent and objective, while keeping it as simple as possible. The reviewer also pointed out that more data than currently available are needed. We recognize data gaps exist and have utilized several expert elicitations as well as consultations with FDA subject matter experts to fill data gaps and documenting the use of different data sources in the model.

We recognized that for undeclared allergens, expert opinions were used for scoring several criteria and, for Criterion 2, the score based on expert opinion was 9 (high) for all the pairs involving the "big eight" undeclared allergens. We reviewed Minor *et al.* (2015, Risk Anal., 35:1135-1139) and examined an alternative scenario in which the scoring definitions for Criterion 2 (severity of illnesses) for microbial hazards and undeclared allergens were based on illness severity measured by the loss of quality-adjusted life days (QALDs) reported by Minor *et al.* The alternative Criterion 2 scenario resulted in a shift to a lower risk score for some of food-hazard pairs, particularly food-undeclared allergen pairs and pairs involving *Brucella* spp., *Cryptosporidium parvum* and other spp., *Salmonella* spp., *Shigella* spp., and *Y. enterocolitica, V. parahaemolyticus* and *V. cholerae.*

Question 7. The reviewers who had expertise in SAS commented that the codes were correct. One reviewer identified some differences in scores when comparing the Excel in Access files.

FDA response. We thank the reviewers for the confirmation of the accuracy of the SAS codes. Regarding the differences observed between the two files, we have reviewed and resolved the issue. On a separate note, in response to comments from a separate peer-review on the underpinning data, we have added more food-hazard pairs to the model, as suggested.

Question 7a. Overall, the reviewers stated that the scoring logic in Section 4 appropriately represents the scoring definitions in Section 2. One reviewer wanted additional information provided for some of the scoring definitions, including use of expert judgement in cases where quantitative data are not available.

FDA response. We thank the reviewers for the overall confirmation. We have revised the text in the Methods document to clarify scoring and improve transparency. On a separate note, we have revised the definition for Criterion 4 in response to other comments from the peer reviewers; Criterion 4 is now scored based on the estimated amount of growth of bacterial pathogens, when present, given the food characteristics and customary shelf life (one descriptor); this estimate is based on growth studies in the peer-reviewed scientific literature and predictive microbiology databases (e.g., ComBase).

Question 7b. Overall the reviewers who reviewed SAS codes (not all the reviewers had expertise in SAS codes) or selected examples stated that the scoring logic appeared to be correctly implemented in the SAS codes, including additions, aggregated scores, and weighting. One reviewer found that the Uncertainty score description in the Methods document did not describe inclusion of weighting factors in its calculation, but that the scores appeared to do so. The reviewer also suggested it was debatable whether inclusion of weighting factors in the Uncertainty score is logically justified. The reviewer also stated that there was little value in

including both a confidence score and uncertainty score as model outputs, as they effectively measure the same properties of the risk-ranking scores.

FDA response. We thank the reviewers for the overall confirmation, observations, and suggestions. We agree with the reviewer who suggested that there is little value in including both the confidence score and uncertainty score as model outputs; the model now only evaluates a confidence score. Further, we have determined that it is not appropriate to use weighting factors in the determination of a confidence score for data quality. Criteria weights reflect judgements on the importance of each criterion to the overall risk score, which are not related to data quality. Calculating a confidence score using the correct equation (not including a weighting factor) is now implemented.

Question 7c. Overall the reviewer who responded to this question stated that equations 1 and 2 (for criterion 3) appeared to correctly reflect the data weighting description in the draft report. The reviewer could not find the code for equations 3 and 4.

FDA response. We thank the reviewer for the confirmation on equations 1 and 2. Equations 3 and 4 were coded in the Access model provided for peer review. We have since combined and verified all equations in one set of modeling script.

Question 8. Overall, the reviewer who responded to this question stated that the underpinning relational database, look-up tables, and algorithms appear to be appropriately designed and implemented.

FDA response. We thank the reviewer for the confirmation.

Question 9. Overall, the reviewers who reviewed the Access model user interface and guide stated that these are sufficiently described to understand each component of the model, given an understanding of the structure and logic of the model provided in the Methods document. Reviewers noted some technical difficulties with input/output and pointed out some edits or typographical errors that when fixed would improve clarity. One reviewer found some missing or incomplete reference information when using the interface to identify source data for criterion scores.

FDA response. We thank the reviewers for confirmation that the Access model user interface was sufficiently described for an experienced user to understand each component of the model and we are glad that the information provided assisted with their review of model functions. We have revised the model interface and database to address the issues mentioned and to make it

easier to add new food-hazard pairs and add/update data and references for each of the criteria, as needed. We have also added sections to the Methods document and developed an interactive results table to make it easier for stakeholders to review food-hazard pairs considered, criterion and risk scores for each, and overall risk scores for each food.

Question 10. The reviewers provided several suggestions and considerations for how often the model should be updated, including: 1) the frequency of updates of the model should be determined by the underpinning data upon which the scores for the seven criteria are developed (e.g., CDC surveillance data release, TDS monitoring data release, NHANES data release, etc.); the reviewer suggested a 2- to 4-year period for update may be reasonable; 2) the frequency of updating the model should consider the robustness of the ranking; if the uncertainty score is high and the rank is not robust that means the score of those food-hazard combinations needs to be updated as soon as possible; 3) the frequency of updating the model should consider the rate of change of food processes and product formulation; the reviewer suggested criterion values be reviewed every 3 to 5 years with the recommendation that when 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.

FDA response. We thank the reviewers for the suggestions. Overall, the reviewers suggested a frequency of updating the model every 2-5 years, and suggested some case-by-case updating of food-hazard pairs on a more frequent basis may be advisable (e.g., where confidence score is low or radically different processes or products are introduced). FDA will take these suggestions into consideration in model updating efforts.

Question 11. Overall, the reviewers found the report clear, stating that the information provided in the draft report and associated appendices were well laid out and easy to follow with adequate details in most areas; the report was rather comprehensive and appears to be clearly written; it has an excellent and logical structure; the report was mostly clear and reflective of what the current model intends to do; overall the report did a very good job of explaining what has been done and the elements of the model and provided the needed information to understand the risk ranking process. The reviewers also made specific comments on many minor presentation errors, identified some places where additional details were needed, and places where clarification was needed (e.g., use of jargon and idioms that might not have unambiguous meaning to all readers).

FDA response. We thank the reviewers for the positive feedback on the report structure and clarity. We corrected minor errors and typos as pointed out by the reviewers. In revising the report, we also added text to improve transparency and clarity.

Question 12. The reviewers provided some additional suggestions/comments, including a suggestion to enhance the sensitivity analysis with additional scenarios and elaboration on a suggestion regarding aggregation.

FDA response. We thank the reviewers for the additional comments. The sensitivity analysis provided includes varying criteria weights which provides insights into the impact of each criterion on the results. The RRM-PT by mandate must consider all six factors required in FSMA 204 (21 U.S. Code § 2223), i.e., all seven criteria in the RRM-PT are necessary for RRM-PT model.

With regard to the elaboration on the reviewers suggestion for aggregation (Question 3), we have implemented the approach suggested by the peer reviewer, where the aggregation equation involves exponential transformation, summing and log transformation taking into account the risk scores for all food-hazard pairs under the food (see details above in response to Question 3).

Sections III, IV, and V. Specific Observations

The reviewers provided numerous specific comments, page by page, on language clarifications, errors/inconsistences, and typos, and provided suggested edits/changes.

FDA response. We thank the reviewers for their careful review, specific comments, and suggested edits. We have considered each in our revision of these documents.

REFERENCES

Duret, S., Hoang, H. M., Derens-Bertheau, E., Delahaye, A., Laguerre, O., & Guillier, L. (2019). Combining quantitative risk assessment of human health, food waste, and energy consumption: The next step in the development of the food cold chain? *Risk Analysis, 39*(4), 906-925.

Garre, A., Boué, G., Fernández, P. S., Membré, J. M., & Egea, J. A. (2019). Evaluation of multicriteria decision analysis algorithms in food safety: A case study on emerging zoonoses prioritization. *Risk Analysis*, https://doi.org/10.1111/risa.13391.

Minor, T., Lasher, A., Klontz, K., Brown, B., Nardinelli, C., & Zorn, D. (2015). The per case and total annual costs of foodborne illness in the United States. *Risk Analysis, 35*(6), 1125-1139.

Van der Fels-Klerx, H., Van Asselt, E., Raley, M., Poulsen, M., Korsgaard, H., Bredsdorff, L., Nauta, M., D'agostino, M., Coles, D., Marvin, H. J. P., & Frewer, L. J. (2018). Critical review of methods for risk ranking of food-related hazards, based on risks for human health. *Critical Reviews in Food Science and Nutrition, 58*(2), 178-193.