

REVISED DRAFT

PEER REVIEW SUMMARY REPORT

External Peer Review of the FDA/CFSAN Draft Report
*Fresh Produce Risk Ranking Tool Summary: Identification of Priority
Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment*

Prepared for:

Sherri Dennis, Ph.D.
FDA, CFSAN, RACT
5100 Paint Branch Pkwy
College Park, MD 20740

Prepared by:

Versar, Inc.
6850 Versar Center
Springfield, VA 22151

Contract No. HHSF223200730231G

Task Order 09

Peer Reviewers:

Linda J. Harris, Ph.D.
Mary S. Palumbo, Ph.D.
Jeffrey A. Soller, M.S.
Donald Schaffner, Ph.D.
Nga A. Tran, Dr.PH, MPH, CIH

October 20, 2009

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	CHARGE TO REVIEWERS	2
III.	INDIVIDUAL REVIEWER COMMENTS	3
	REVIEWER #1	4
	REVIEWER #2	8
	REVIEWER #3	16
	REVIEWER #4	22
	REVIEWER #5	28
IV.	PEER REVIEW COMMENTS BY CHARGE QUESTION.....	34

I. INTRODUCTION

The Produce Risk Ranking Tool is one of the efforts undertaken by FDA/CFSAN to further develop methodology to assist regulatory scientists in making resource decisions. Risk ranking, sometimes called hazard ranking or comparative risk assessment, is a technique that can be used to identify, and thereby prioritize, the most significant risks for a given situation. Through a contract with RTI International, FDA conducted a risk ranking of hazards associated with fresh produce. The purpose of the risk ranking was to identify priority commodity/ hazard pairs for further study and development of quantitative risk assessment. A total of 51 commodity/hazard pairs were identified and ranked according to risk criteria, using data available in the peer-reviewed literature and when necessary expert opinion.

Peer Reviewers:

Linda J. Harris, Ph.D.
University of California, Davis
Davis, CA 95616

Mary S. Palumbo, Ph.D.
California Department of Public Health
Sacramento, CA 95831

Donald Schaffner, Ph.D.
Center for Advanced Food Technology
Rutgers University
New Brunswick, NJ 08901

Jeffrey A. Soller, M.S.
Soller Environmental, LLC
Berkeley, CA 94703

Nga A. Tran, Dr.PH, MPH, CIH
Exponent
Washington, DC 20036

II. CHARGE TO REVIEWERS

Please provide written responses to the following questions:

Charge Questions:

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rationale and when appropriate, the sources/ data for the alternatives.

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

Question 6: Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.

III. INDIVIDUAL REVIEWER COMMENTS

REVIEWER #1

Peer Review Comments on FDA's Draft Document

Fresh Produce Risk Ranking Tool Summary: Identification of Priority Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment

Reviewer #1
October 5, 2009

I. GENERAL IMPRESSIONS

In my opinion, the Fresh Produce Risk Ranking Tool accomplishes its intended purpose of providing a means of prioritizing pathogen/commodity pairs based on past identified outbreaks, the severity of the resulting disease, and likelihood of exposure by the consuming public. While a strictly quantitative approach is not possible at this time with available data, the tool does provide a semi-quantitative means of ranking the risk associated with various pathogen/commodity pairs. This will allow ranking of pathogen/commodity pairs where fewer outbreaks have been detected, compared with the well-recognized combinations of leafy greens and *E. coli* O157:H7 (EHEC) as well as tomatoes and *Salmonella enterica*. The rankings obtained with these less frequent outbreaks are variable based on the weight assigned to criteria other than epidemiological link. There is no other means to assess the uncertainty in the rankings. The tool is easy to use once the weightings are decided. The tool is flexible and allows the user to exclude a criterion by assigning it a weight of zero, or to rank eight of the criteria in relation to the ninth. As the authors point out in Chapter 4, Synthesis Comments and Conclusions, the risk associated with each commodity/pathogen pair is relative only to other pairs and is not proportional to the true risk difference between pairs. It will be important to update the database frequently as new information becomes available.

II. RESPONSE TO CHARGE QUESTIONS

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

In my opinion the tool satisfies the purpose for which it was designed. It will be necessary to revise the database going forward in order for it to continue to serve this purpose. The quality of the data is important and not always obvious. For example, in reviewing historical data on outbreaks associated with EHEC and leafy greens I have observed that different sources will show different numbers. Looking more closely at the data, I found one period in the fall of 1999 where some counted 2 outbreaks and others 3 outbreaks. The difference depended on whether one counted a cluster in California as being separate from cases in the northwest or whether these were considered a single outbreak. The definition of an outbreak requires that there be two or more cases with a common exposure. The difficulty is in determining whether the cases did indeed have a common exposure. This is but one example of the difficulties in developing a risk

ranking tool such as this. The developers have accurately described it as a semi-quantitative tool.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

I believe the criteria are appropriate, and the user has the option of weighting the criteria differently. Maintenance of the database will be critical, as more data is accumulated in the Microbiological Data Program and in the literature on prevalence of contamination, infectious dose, and growth potential. The Growth Potential and Shelf-Life score is a simplification of a fairly complex criterion, in that cut produce has a shorter shelf-life but may have greater growth potential than uncut produce. It is common practice, for example, to chop tomatoes in a processing plant for distribution and sale to foodservice operations. The cut tomatoes will be held refrigerated, whereas the raw uncut tomato may be held at ambient temperature. Bagged salads contain leafy greens with cut surfaces, where pathogens have been shown to adhere and grow. Shredded carrots in the bagged salad may represent a different risk than whole carrots. As more data becomes available on growth potential of the pathogens on various commodities, it may be desirable to separate these into two criteria, but at the present time data does not exist for most commodities. Only leafy greens and EHEC are ranked as “strong” for growth potential.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

I do not know of another method of sensitivity analysis.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rationale and, when appropriate, the sources/data for the alternatives.

I believe the scoring categories are appropriate, and I understand that you have not attempted to match the scores to the relative risk under each criterion. But I wonder if it is possible to assign values other than 1, 2, 3, and 4 to the bins. In particular, I question the scores for consumption. According to the reported data (Table A-7 in Appendix A), the daily consumption rate for leafy greens is about 56 times higher than for herbs and 100 times higher than for green onions. The score for leafy greens is 4 times higher than for herbs and green onions. Could the scoring more

closely reflect the difference in consumption rates? Would it not better reflect the difference in risk? I realize that changing this scale would not affect the relative risk of the leafy greens/EHEC pair but it might affect the scores for other leafy green/pathogen pairs.

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

I think it would be extremely valuable to the user to know the uncertainty associated with rankings, but I do not have a method to accomplish this. The uncertainty is largely associated with the small number of outbreaks associated with some commodity groups as well as lack of data for other parameters. Produce items such as herbs that are consumed in small amounts may not be identified by the case patient when interviewed, and thus never enter the data stream. In addition, we know that foodborne illness is underreported both by the general population and by some local health jurisdictions.

Question 6: Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.

I think it is likely that listeriosis is an illness that may be associated with certain types of produce, and perhaps such outbreaks will be detected in the future. There was an outbreak associated with cole slaw in the 1980's, which is prior to the time period used for this data set. It was attributed to contamination of the cabbage in the field. *Listeria monocytogenes* has been found by surveillance sampling in commercial bagged salads, but no outbreaks have been detected from such contamination.

III. SPECIFIC OBSERVATIONS

[Specific Observations were not provided by this reviewer.]

REVIEWER #2

Peer Review Comments on FDA's Draft Document

Fresh Produce Risk Ranking Tool Summary: Identification of Priority Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment

Reviewer #2
September 28, 2009

I. GENERAL IMPRESSIONS

Overall, I believe that the Fresh Produce Risk Ranking Tool will be a valuable component in FDA's toolbox for identifying and prioritizing important pathogen / commodity pairs among the myriad possible combinations of pathogens and commodities that potentially could be of greatest public health concern. The approach that was employed in the development of this tool seems reasonable given the goals of the project, and the interface is elegant in its simplicity, clarity, and flexibility. Moreover, the information supplied documenting the development and technical basis for the tool is well written and clear.

I do, however, have a number of concerns with the tool in its current state and the companion documentation. In particular, I believe that there are issues that need to be addressed with 1) the technical merit of the "infectious dose" criterion, 2) the selection of and justification for default bin boundaries, and 3) the potential for misinterpretation of the results due to the unintuitive nature of the "weighted" semi-quantitative scoring scheme that is employed. I do believe that these issues can be addressed via enhanced discussions and technical justifications, coupled with a revision of the approach used to characterize the relative infectivity of the pathogens of concern. I believe addressing these concerns will enhance the transparency and technical merit of the tool and facilitate interpretation of the tool's output.

II. RESPONSE TO CHARGE QUESTIONS

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

I do believe that the tool (subject to the recommended changes) will satisfy the purpose for which it was designed and will be able to be used to identify high priority pathogen/commodity pairs.

There are a few issues that really need to be considered (as described in detail below) and resolved before the tool should be considered ready for use by federal agencies. First, it appears likely that the overall rankings from individual runs are highly sensitive to the specific bin boundaries that are selected for the run (for each criterion, scores range from 1-4, and default boundaries are provided). I am of the opinion that the default bin boundaries are very important because users of the tool may not feel like they have sufficient expertise to confidently override

the default values. Little justification is provided in the documentation for the default values selected for the bin boundaries. To enhance transparency, the justification for the choices that were made in this regard should be provided in the report.

Second, the technical basis for supporting the “infectious dose” criterion is inadequate. This criterion should be fundamentally based on peer-reviewed dose response relationships for a specific level of risk (for example, the number of organisms which cause infection to 10% of those exposed – ID₁₀).

Third, some discussion is needed to support the risk ranking algorithm that was selected for this tool. It is not clear to me that the summation of weighted binned data results in a metric that is easily, clearly, and deeply understood. Discussion on this topic would go a long way towards supporting the selection of one algorithm over others.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

The criteria seem generally reasonable. However, I found the discussion of risk dimensions to be confusing and distracting. In fact, it is the 9 criteria that the tool is based on, not the four dimensions. I suggest editing the text for clarity (specific suggestions provided below in Section III).

I found the use of a composite variable for growth potential and shelf life to be curious. Some additional justification is needed to explain why this approach was taken for this criterion and not others.

The approach employed for “infectious dose” needs a complete overhaul as detailed below. Briefly, peer reviewed dose response data should be the basis for this criterion and a specific and consistent level of infection (ID₁₀ for example) should be used for all pathogens. A source for dose response relationships is provided in Section III.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

Two types of sensitivity analyses were conducted: weighting of an individual criterion and randomized weighting of the nine criteria. The sensitivity analyses that were conducted by changing individual (and in some cases two) criteria are appropriate (however, I believe that the actual results could change substantially once the relative infectivity issue identified above is addressed). In terms of presentation, I found the results hard to follow and would prefer to see

the results for each of the sensitivity analyses presented by ranked results from that analysis (currently they are presented ranked by the baseline rank).

With respect to the sensitivity analyses conducted using randomized weighting, I am unconvinced that sufficient randomized samples were conducted (Section 3.3.2) to be representative of the solution space of interest (100 runs were conducted out of nearly 2 million possible combinations). Rather than enhancing this section, I would prefer to see an emphasis on the relative impact and/or synergy of weighting 2 or 3 criteria that are likely to be most important from a public health management perspective.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rational and, when appropriate, the sources/data for the alternatives.

I think that the use of four scoring categories is appropriate for each criterion. However, one of my principal concerns revolves around the selection of the default boundaries for each criterion category. The justification for each selection should be made explicit and in important cases, sensitivity analyses should be conducted to determine the relative sensitivity of the output to these selections. Without such explicit explanation it is difficult to assess the appropriateness of the choices that were made.

In several cases, it appears that the default values used for categories are not set such that maximum sensitivity is achieved (for example, in the epidemiological link the “very strong” category appears to be set too low to differentiate the pathogen / commodity pairs that caused the most number of outbreaks from others).

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

I am not convinced that this tool needs to be limited to pathogen/commodity pairs for which outbreaks have been reported. The epidemiological link criterion appears to account for this intersection, whereas each of the other criteria only requires data on either the pathogen of interest or the commodity (criterion 7: prevalence of contamination is the possible exception to this statement). It seems like this tool could be extended so that it could be used to identify potential future outbreaks of concern and thus the consideration of management alternatives to prevent them.

Section 4, paragraph 2 highlights the fundamental weakness of this work, which in my opinion raises questions about the conclusions. Specifically, “The categorization and binning approaches may seem somewhat arbitrary”. I agree with this statement. Moreover, I believe that the usability of the tool would be most strongly enhanced by addressing this issue. The potential benefits of this tool are great, but the technical underpinning must be sound and transparent, and I believe that addressing this issue is much more important than adding additional features.

Question 6: Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.

My additional comments are detailed below in Section III.

III. SPECIFIC OBSERVATIONS

Page	Line	Comment
ES-1	P2	The terms “semi-quantitative” and “dimension” should be defined. I found the term “risk dimension” to be confusing and unnecessary. I believe that the nine criteria speak for themselves without the need for “dimensions”. Alternatively, consider an explanation along on the lines of the following: “the pathogen commodity pairs were considered with respect to four general characteristics which were refined for specificity resulting in nine criteria”.
ES-1	Col 2	Infectious dose – see specific comments below, but this term should be deleted from the document and replaced, as it is antiquated and misleading.
ES-1	P4	Need some explanation as to why/how the overall risk score for a pathogen/commodity pair results in an outcome that can be meaningful for a risk manager.
ES-1	P5	For the reasons noted previously and below, I am not convinced of the “conclusions” for EHEC and leafy greens, and <i>Salmonella enterica</i> and tomatoes.
1	P4	Define what is meant by semi-quantitative.
2	P3	Need a stronger explanation as to why outbreaks based on multiple foods were excluded. The justification provided is not convincing.
3	P1	It is not clear that the pathogen/commodity pairs that were included are sufficient. Are we to believe that simply because an outbreak was not reported that it could not be important in the future? It seems that this tool could help to answer that question simply by including a score of 1 in the epidemiological link section for any pair for which no outbreaks were reported.
6	P2	Change “characterized by” to “refined into”.
6		The Tables in Appendix A would fit better into the report rather

		than in the Appendix. These data are critical for understanding each of the Criteria. I found myself constantly flipping back and forth between the Appendix and text. This change would ease interpretation of the report for the reader.
6	P4	Again, it is not clear to me that this needs to be limited to pairs that have been historically associated with outbreaks. If no outbreaks were reported, why not simply give them a score in this section of 1 and allow the tool to provide the answers as to whether there are cases in which such pairs could be quite important. If my assertion is incorrect, then a clear justification should be provided.
6	P4	<p>I list this comment here because it is the first time it arises, but it applies equally to all criteria. There needs to be discussion indicating the rationale that was used for selecting the “default” endpoints for each of the bins.</p> <p>In this case, I suggest that the values that are used for the “very strong” bin could be too low. The implications of this choice are that those pairs which scored “4” are not differentiated from each other. For example consider norovirus in mixed produce (112 outbreaks and 5390 cases). Given the current ranking, any pair with >5 outbreaks and >100 cases would have the same score and effectively be counted as equally important.</p> <p>Without some detailed explanation about how the decision was made about the bin endpoints, it is not possible to determine the appropriateness of the default value(s). I believe that in many cases, users will accept the default values as “correct”, thus, it is imperative that the justifications are transparent.</p>
7	P3	It is not clear why the multiplier for norovirus is higher than that for “mild and non-reportable illnesses”. Some more explanation about the difference is warranted.
8	P3	See subsequent discussion re infectious dose.
8	P4	Citations needed to support first sentence in this paragraph (Certain populations....).
8-9		Justification needed to support the sentence that spans pp 8-9.
9	P1	Much stronger citations from the peer reviewed literature are needed to support the first sentence on this page.
9	Table 8	Not clear why the ranking is established as it is for scores 2 and 3. Explanation needed.
9	S2.3.2	<p>The term infectious dose is antiquated and should be replaced throughout the document with something more meaningful. For example, ID₁₀ or ID₅₀ if you want to use a quantitative description or relative infectivity if you want to use a qualitative description.</p> <p>Most experts in the risk assessment field agree that 1 organism is sufficient to cause infection, thus the term infectious dose is</p>

		misleading. In fact, the text correctly summarizes this point in P3 on this page, (It is well recognized).
9	Table 9	These need to be based on peer reviewed dose response functions at a specific level of risk (ID ₁₀ or ID ₅₀ for example). Furthermore, I believe that the bin range for score 4 is likely too high to be meaningful in terms of differentiating pathogen levels of concern. See also comment on Table A-5.
12	P2	The differences between growth scores of 2 and 3 seem somewhat arbitrary. Better justification needed.
12	P3	A better explanation is needed why shelf-life is as important as the other criteria. This is not clear to a reader that is not intimately familiar with the field.
13	P1	Why was a composite variable created using these two pieces of data when no other composite variables were created for other criteria? It seems like the impact of this methodological choice is to reduce the importance of each of these two factors, was this intended? Justification for this approach and an explanation of the impacts of this choice are needed.
14	P3	Some justification for this approach in terms of precedent would be very useful. Essentially, you are taking mostly quantitative data, binning those data, weighting the binned scores, and summing those products. It is not clear to me that the resultant score (Rank) is an intuitive metric. This would be particularly true for conditions in which many attributes score close to a bin edge. I am sure that other risk ranking approaches have also been used by others for different applications. Some discussion would be helpful for understanding why this approach was taken, its advantages, and its drawbacks.
17	P1	This example seems to support my previous point about the critical nature of the bin edge selection. Table 18 summarizes the relative scores when epidemiological link is most important (weight 5), and all other criteria are equal (weight 2). Yet the pair with the strongest epidemiological link (most outbreaks and most cases) does not appear in the top 12 pairs. In some way, this should be addressed.
18	P2	I have a hard time believing the results that indicate that the relative risk ranking did not change when the “infectious dose variable” weight was changed to 5. Refer to my comments on Table A-5. The relative infectivity of the pathogens needs to be based on peer reviewed dose response modeling to the extent possible. The data on <i>Salmonella enterica</i> is not consistent with anything in the peer reviewed dose response literature that I am aware of. Furthermore, Noroviruses are more highly infectious than indicated (see Teunis et al 2008). I suspect that if the data in Table in A-5 are based on peer reviewed dose response information and bin edges are set appropriately, the results in

		Table 20 (and for that matter, perhaps other tables) may substantially change.
20	P1	(First sentence) This still does not explain why norovirus-mixed produce does not appear in the results.
21	P1	Using the described approach, there are nearly 2 million possible combinations (5 ⁹), please provide some explanation as to why a random sample of 100 of those, is sufficiently representative to draw inferences from.
22	P2	Suggest changing “emerging” to “unreported” in L3. Suggest paragraph break at “The categorization (food categories). Please explain what is meant by the last two sentences of this paragraph (The metric produced.....and/or uncertainty). To me, the fundamental question raised by P2 is how do you know that the bin boundaries are set to achieve maximum (or least sufficient) sensitivity? As indicated above, “infectious dose” and epidemiological link seem to be rather insensitive using the current values.
22	P3	The statement is made that most of the data used in the tool are from the peer reviewed literature. These linkages should be made much more explicitly in the text and the tables.
A-2	Table A-2	This comment applies to Tables A-2, 3, 4, 5, 6, 8, and 9. A column should be added to the far right of the table to hold the reference that supports the values shown. The primary reference should be shown in each case.
A-3	Tables A-4 &5	Much stronger technical basis is needed than that shown in Appendix B.
A-3	Table A-5	The data in this table are key and need a complete overhaul. This Table should be based on peer reviewed dose response information for a specific proportion of the population becoming infected. Refer to US EPA’s recent draft MRA Protocol document ¹ (Table 4) for a comprehensive summary of peer reviewed dose response relationships.
A-3 and A-5	Table A-6 &7	Please check number of significant digits. (4 sig. digits for % population consuming seems unreasonable).
B-1		This bulleted list seems to be missing the increased severity and duration observed in the immunocompromised populations due to exposure to <i>Cryptosporidium</i> .

1

[http://yosemite.epa.gov/sab/SABPRODUCT.NSF/7FAA3A556A92CF21852576160064DEC2/\\$File/Draft+MRA+Protocol+July+30+2009+for+DWC+Sept+21-22+2009+Meeting.pdf](http://yosemite.epa.gov/sab/SABPRODUCT.NSF/7FAA3A556A92CF21852576160064DEC2/$File/Draft+MRA+Protocol+July+30+2009+for+DWC+Sept+21-22+2009+Meeting.pdf)

REVIEWER #3

Peer Review Comments on FDA's Draft Document

Fresh Produce Risk Ranking Tool Summary: Identification of Priority Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment

Reviewer #3
October 11, 2009

I. GENERAL IMPRESSIONS

The Risk Ranking Tool (RRT) report provides a clear description of the model purpose, components, constructs and underlying data. Overall, the main strength of the model is that it is relatively simple and easy to understand and its results are easily replicated by users. The distinction that is made in the RRT between the data-driven component of the model (bins and scores for each model criterion) and the subjective component of the model (weights for each model criterion) provides the much needed transparency and separation between empirical evidence and value-laden expert judgment. As currently set up, if the value-judgment weights were set at 1 for all model criteria, the relative ranking of commodity (pathogen-produce) pairs are mostly based on objective/empirical evidence (see response to question 2 for more discussion here). As a simple and transparent tool, the RRT provides the relative ranking of the pathogen-produce commodities and allows for a rapid and systematic approach to prioritizing future work, i.e. follow-up robust quantitative risk assessment for commodity pairs that are ranked high. So long as the purpose remains that of prioritizing pathogen-produce commodities for more detailed risk assessment, the RRT is a reasonable approach to achieve this goal.

One of the major limitations of the RRT, and as acknowledged by the authors, is that it is not a predictive tool, and that it only relatively ranks known pathogen-produce commodities (as evidence by epidemiological database). As such, for the RRT to be useful in future efforts to identify high priority pathogen-commodity pairs, a periodic updated literature search to identify and incorporate new pathogen-commodity pairs and their associated information into the RRT will be necessary. In general, this RRT will remain useful as long as the list of the pathogen-produce commodities to be ranked and the underlying data for the 9 model criteria are periodically updated.

II. RESPONSE TO CHARGE QUESTIONS

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

The stated purpose for the RRT is to prioritize pathogen/commodity pairs for more robust risk modeling, i.e. highly ranked pathogen/commodity pairs would be subjected to further detailed quantitative risk assessment to predict their human health impact. The RRT as a simple risk

based ranking tool to provide a rapid, transparent and systematic approach to prioritize a large combination pathogen-produce commodities, is appropriate for the stated objective.

As currently developed, the RRT is limited to the relative ranking of the pathogen-produce commodities that have been identified in a previously conducted public data search (pre-2009?). More recent outbreaks involving pathogen-produce commodities are not included in the current tool. As such, for the RRT to be useful in future efforts to identify high priority pathogen-commodity pairs, a periodic updated literature search to identify and incorporate new pathogen-commodity pairs and their associated information into the RRT will be necessary. In general, this RRT will remain useful as long as the list of the pathogen-produce commodities to be ranked and the underlying data for the 9 model criteria are periodically updated.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

As described in the RRT report, there are 9 criteria in the model: 1) epidemiological link, 2) disease multiplier, 3) hospitalization rate, 4) death rate, 5) population susceptibility, 6) infectious dose, 7) prevalence of contamination, 8) consumption and 9) growth/potential shelf life. This reviewer agreed that these are the appropriate indicators for the hazard, exposure and public health risk potential associated with the pathogen-produce commodities. The cited literature/data sources associated with each criterion are appropriate. However, the reviewer also noted the following with respect to the current scoring of the 9 criteria:

- In general, the following broad aspects of the RRT can be described:
- The potential public health impact aspect has two dimensions: the epidemiological and health dimensions. These two dimensions include the epidemiological link, disease multiplier, hospitalization rate and death rate scores. Thus, the public health potential part of the model can score as much as 16 (4x4, assuming the unity weights for all 4 criteria)
- The hazard potential aspect has one dimension: the agent dimension, which includes the population susceptibility and infectious dose scores. Thus the maximum score that the hazard potential part of the model can get is 8 (2x4, assuming the unity weights for both criteria)
- The exposure potential aspect has one dimension, production and processing, with scores based on 3 criteria: prevalence of contamination, consumption and growth/potential shelf life scores. Thus, the maximum composite for the exposure potential is 12 (4x3, assuming the unity weights for all 3 criteria)

- Implicit in the overall imbalance in the total scores between the potential hazard, exposure and public health impact aspects of the model is the developer's judgment of the order of their importance. This implicit judgment if retained in the model should be explicitly noted. Since the RRT has a distinct subjective component in which additional weight (by expert judgment) can be incorporated to these criteria, this information needs to be made known so expert judgment to increase the score (using weights) for any of the criteria can be done appropriately and in context of the model emphasis on the public health impact scores. As it is, with the weights at unity, the ranking will be heavily influenced by the public health impact aspect of the model.

From NHANES, the amount (grams) per eating occasion (EO) can also be obtained. The current RRT only uses the % consumer data from the survey to establish the score for the consumption criterion. A more robust consumption score can be obtained based on the combination of % consumers and amount per EO. The reviewer also noted that the NHANES 03-04 data used for criterion 8 (consumption) is not the most current (NHANES 05-06 is available), and the RRT report incorrectly described the data from NHANES 03-04 as a three day dietary recall (it is only 2-day).

It is also noted that while the RRT report described 9 criteria, the Access tool (and user's manual), only has 8 criteria in the bins (score) portion of the model. The susceptibility criterion is not present in this portion of the tool. Further, while the RRT report described the score for the growth/potential shelf life criterion (Table 14 of the RRT report) as the combined growth score (Table 12 of RRT report) and shelf-life score (Table 13 of the RRT report), the Access tool (and user's manual) only shows the shelf-life bin/score in the bin/score portion of the RRT. The weight portion of the RRT has all 9 criteria. This discrepancy should be addressed by the developer.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

The sensitivity described in the report was based on changing the weights (value judgment) component of the model. While this is necessary to understand the influence of this component of the model, it is also important to perform a sensitivity analysis based on the bin/score component of the model. This can be done by setting the value judgment (weight component) at 1 (i.e. no impact on final score). There are two considerations in performing the sensitivity analysis for bin/score component of the RRT: 1) number of bins and the upper and lower limits for each bin, and 2) type of scores for each bin. In the current model, each of the 8 criteria (this reviewer only counted 8 criteria in this part of the tool – see comments above) has 4 bins and each bin is given an ordinal score (1, 2, 3 and 4). The ranking results can change if the upper and lower values of the 4 bins are modified, particularly when changes are made to the consumption criterion (by changing the limits of these consumption bins, this reviewer can influence the 2nd ranking of the commodity pairs currently in the system). It is also important to evaluate whether

more resolution in the ranking can be obtained with more than 4 bins for each criterion, for example 5 or 6 bins (see additional comments on this point under question 4). Lastly, the current model uses ordinal ranking. It may be worthwhile to consider another type of scoring, e.g. proportional (1, 10, 100, 1000, etc.) in the sensitivity analysis to see if improved resolution in the ranking and possible changes in the ranking can occur with an alternative scoring metric.

This reviewer recognizes that the possibilities of alternative number of bins and scores are limitless and one can over complicate the sensitivity analysis. However, the bin/score is objective (data driven) and central to the RRT and thus some minimum effort is needed to demonstrate that the current bin/score is the most optimal in providing the most stable and consistent relative ranking. Again, this sensitivity analysis should be done without the influence of the value judgment (weight =-1) component of the RRT.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rational and, when appropriate, the sources/data for the alternatives.

As indicated in the question, each criterion is unique in its possibility of values. Thus, while the current 4 bins/scores may be sufficient for one criterion, it may be inadequate to capture the granularity of the available data for another criterion. Significant loss of information in the minimalist approach to 4 bins is observed with the epidemiological link and consumption criteria. For the epi-link criterion, it appears that there is a need to differentiate the current “very strong” bin (>5 outbreaks and >100 cases) to capture situations where there are >10 outbreaks and >1000 cases. Similarly, additional bins may be needed to capture the range of data that are available with the consumption data. Further as noted above, consideration to incorporate g/EO with % consumers to derive the bin/score for this criterion is recommended to fully capture consumption/exposure potential. Clearly any modification to the no. of bins and scores will need to be made with an eye toward the discussed imbalance in the total scores for the 3 aspects of the model (hazard, exposure and public health impact potential) as discussed under question #2.

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

As discussed earlier, the limitation of the current RRT is the starting list of pathogen-produce pairs that the model ranks. This list needs to be updated periodically for this tool to remain useful. As far as capturing uncertainty associated with the model criteria, there is always a real risk for going beyond the simple relative ranking purpose of the tool when elaborate uncertainty model is introduced. To avoid doing so, perhaps an overall uncertainty weight could be

incorporated as the third part of the tool. However, as with the current “weight” component of the tool, this added component will be judgment based and the numerical value has to be such that it does not overwhelm the empirically-derived scores (bin component of the model).

Question 6: *Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.*

- As noted earlier, there is an inconsistency between the RRT report and the Access tool (and user guide). The report described 9 criteria, the Access tool (and user manual) only has 8 criteria in the bins (score) portion of the model. The susceptibility criterion is not present in this portion of the tool. Further, while the RRT report described the score for the growth/potential shelf life criterion (Table 14 of the RRT report) as the combined growth score (Table 12 of RRT report) and shelf-life score (Table 13 of the RRT report), the Access tool (and user’s manual) only shows the shelf-life bin/score in the bin/score portion of the RRT. The weight portion of the RRT has all 9 criteria. Perhaps there is a reason for the discrepancy. If this is the case, appropriate documentation should be incorporated.
- Access tool: While the values of all bins for all 8 model criteria can be changed, the lowest bin value cannot be changed by user. Is there a reason for this? In some instances, it may be justifiable to have lower value for the lowest bin, such as % hospitalization, % consumers, etc. and a form of sensitivity analysis around the values of the bins would be warranted.

III. SPECIFIC OBSERVATIONS

Page	Line	Comment
1	Last sentence in last paragraph	The following is stated: “Although not discussed in this report, the intention was that the higher risk combinations identified in this project would be used in subsequent quantitative microbial risk assessment efforts.” It is unclear to this reviewer how the higher risk combinations can be used in subsequent QMRA.
11	Section 2.4.2	Somewhere in this section, the fact that only fresh/raw produce consumption is included should be clearly noted.
6	Table 4	Total case column, the less than sign is incorrectly displayed as a period.
9	Table 9	Infectious dose column, the greater than sign is incorrectly displayed as a period.
13	Table 13	Shelf-life column, the greater than sign is incorrectly displayed as a period.
14	Algorithm equation	SusPopRank x SusPopWeight – don’t see this in user’s manual or tool.

REVIEWER #4

Peer Review Comments on FDA’s Draft Document
Fresh Produce Risk Ranking Tool Summary: Identification of Priority Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment

Reviewer #4
October 7, 2009

I. GENERAL IMPRESSIONS

[General Impressions were not provided by the reviewer.]

II. RESPONSE TO CHARGE QUESTIONS

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

First, let me say that I’m puzzled by the statement “In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development.” Is this tool supposed to complement those tools? Specifically, what other tools are available or under development? Does that mean that certain limitations are acceptable in this tool? If yes, what are those limitations and exactly how does the existence of other tools allow those limitations?

If the primary objectives for this risk ranking approach are “to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling,” I would say that the tool is moderately successful. I would caution however that there are two explicit biases that are built into the tool: the use of CDC data and the math used to combine the data. No matter what other choices the user makes, they cannot change the CDC data or the math used to combine the data, and these two biases may drive many of the “conclusions” arising out of the use of this tool.

Also, I should note that Microsoft Access is an interesting choice. I certainly understand reasons for this choice, but it presents some unique challenges: The user must run the Windows and they must have a copy of Microsoft Access.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

The list of criteria appears to be complete although perhaps somewhat biased. For example, by including hospitalization rate and death rate (since these two are often correlated), the severity of the disease gets double-counted.

As noted in the detailed comments below, the population susceptibility scale is somewhat odd, as it purports to be a numerical scale, but (as I understand it), diseases affecting the elderly more will score 2 and 3, but never 1 and 4, while diseases that affect children more will score 2 and 4 but never 1 and 3.

Data sources are cited in the body of the text when each criterion is mentioned, but the reader must consult the appendix to see agent-by-agent or food-by-food breakdown. A critical omission in these appendix tables are any citations what so ever. This is true for all the appendix tables, but let me use Table A-5 “Infectious Dose and Infectious Dose Categories by Pathogen” as an example. In the first line of this table it says “*Bacillus cereus* | 10^5 – 10^6 | 10^5 | High”, but the reader is left wondering what reference or references were used to deduce this information. This is a very critical omission, and must be fixed in the revised version of the document if the document is to have any use to the large risk modeling community at all.

It is difficult to recommend alternative data sources if the original data sources are not clear.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

The “simplistic form of sensitivity analysis” appears to be acceptable to show the inherent variability possible from the use of the tool; however, as noted below, more details are needed on exactly what was done to perform the Monte Carlo portion of the sensitivity analysis.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rational and, when appropriate, the sources/data for the alternatives.

The scoring categories are generally appropriate for each criterion; however, see my comments regarding the population susceptibility scale above and below.

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different

pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

One limitation of the model is the lack of an ability to edit the basic assumptions easily. For example, let’s say that I’ve done some experiments and I believe that Salmonella growth potential in tomatoes should be 4 (not 3). The only way to change this would be to edit the GROWTH_POTENTIAL table, which I can only do if I know the PATHOGEN_CATEGORY_ID and the COMMODITY_CATEGORY_ID. I could also make a form with a lookup tables to translate these into English, but this is likely beyond most users who are unfamiliar with Access.

I’m not sure that the absence of a means to characterize the uncertainty associated with the rankings for different pairings is a feature worth developing.

Question 6: Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.

See comments below.

III. SPECIFIC OBSERVATIONS

Note: If you want reviewers to respond with line numbers, it’s best to provide line numbers in the document.

Page	Line	Comment
Es-1		Use of CDC data biases the tool from the start. This is ok, but needed to be acknowledged.
Es-1		Document states that 51 pathogen-commodity pairs were IDed, but please also note here how many pathogens, how many commodities total.
Es-1		How the 4 dimensions were turned into the 9 criteria is not clear at this point, although it does become clear in the body of the text.
Es-1		The result showing leafy greens - EHEC consistently ranked first, followed by tomatoes–Salmonella could have been predicted without a model.
2		“We compiled 10 years of data... to take into account changes in food safety practices applied to fresh produce...” It’s unclear how collecting data that spans the introduction of GAPs “takes this into account”. It just merges the before and after data. Also, Guatemala no longer produces raspberries, etc.
3		Produce experts should evaluate lumping of items into general categories for suitability
6		Including “Health: Severity of disease” and “Agent: Pathogen characteristics that affect disease risk or severity” would appear to

		'double count' severity.
6		What does "dot 100" mean?
6		Unclear how Table 4 works. Providing several worked examples would help. Also is "Total cases" for all outbreaks? Clarify.
9		Table 8 contains some embedded value judgments, and there is no clear progression from category to category. Diseases affecting the elderly more will score 2 and 3, but never 1 and 4. Diseases that affect children more will score 2 and 4 but never 1 and 3. I suggest that this be changed to a 2 point scale: 1 - No one is more susceptible than others and 2 - No susceptible groups exist.
9		<p>"Some pathogens cause disease at very low doses, perhaps after consumption of as few as 1–10 infectious units." Please provide citations for all the organisms you wish to cite as examples.</p> <p>Virtually all microbial risk assessors would take issue with the term "infectious dose." Microbes have a dose-response function, with higher doses leading to higher probabilities of illness. If you must use a single value, please use ID50, which is the dose needed to cause illness in 50% of those exposed. The authors go on to clarify this in the second paragraph under 2.3.2, but my comments still stand.</p> <p>The authors should revise Table 9 to use ID50, and should provide citations for the dose response models.</p>
10		The Microbiological Data Program screened produce items have changed periodically. Please confirm, year by year, that cantaloupe, green onions, bagged lettuce, and tomatoes are in fact correct.
11		"dot" typo in front of "5 %" in Table 10.
11		"if was necessary" should be "it was necessary"
11		Table 11: A product consumed exactly 5% of the time would go in medium and high. Add greater than or less than as needed.
12		Calling a category "No evidence" when there are conflicting studies is not appropriate. I suggest "unclear" or "conflicting evidence".
12		It's unclear if "conflicting studies" exist, how to determine if something is a 2 or a 3.
12		Clarify if cut or whole produce is considered here. If it's cut produce, Salmonella will grow in cut tomatoes. If you are lumping cut and whole produce together here then the entire exercise is potentially flawed. Simply put, many pathogens don't grow on whole produce, but grow well on cut produce. This point must be made very clear.
13		Table 13, dot typo.
13		<p>"the numerical score for growth potential and the numerical score for shelf-life were summed to produce an ordinal number ranging from 2 to 8."</p> <p>Why were they summed? Multiplication would make more sense. Consider a pathogen with a growth potential of 4 in products with shelf</p>

		lives of 1 and 4 respectively. If the shelf life is 4 times as long, that pathogen has 4 times as long to grow. $4*1 = 4$, vs. $4*4 = 16$. In an additive model it's $4+1 = 5$ vs. $4+4 = 8$.
14		I'm not sure why "This scale is particularly appealing because it is more intuitive for expert elicitation." It's only intuitive to experts that think in 5-point scales. If a given expert thinks a particular factor is irrelevant, but another factor is 10 times more important than a third factor, this system won't help.
14		"Rank =..." variable should be given in English, not code.
16		"all criteria were set at the low (1) or high (5) values" not clear. Were all permutations and combinations evaluated? i.e. 5^9
16		"and the simulations repeated." How is this "simulation"?
16		"As perhaps expected, the leafy greens and E. coli O157:H7 (EHEC) pair had ... the largest difference between the lowest and highest scores." Why is this expected?
21		"The weights were randomly selected and the model rerun 100 times." Why 100 times? Why not 1000? Was this true Monte Carlo, or was Latin Hypercube used? Latin Hypercube would be preferred for such a small number of simulations. Was this done manually? With Access macros?
22		"In a majority of iterations, tomatoes–Salmonella enterica and leafy greens–Salmonella enterica ranked second and third, respectively. This bodes well for the use of the tool to prioritize future quantitative risk assessment efforts." I'm not sure this bodes anything. It may be more a reflection of the data and the algorithm that essentially stack the deck so that these pairs always come out on top.
A-2		Users guide: text in Constraint and Comment columns not clear.
Software		Weight tab: Names should match the report text exactly, i.e. not "contamination" but "Prevalence of contamination."
Software		No error trapping on weights, 0 and > 5 are allowed. This is ok, but it should be noted that entry of data outside the expected range is permitted.

REVIEWER #5

Peer Review Comments on FDA's Draft Document
Fresh Produce Risk Ranking Tool Summary: Identification of Priority Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment

Reviewer #5
October 16, 2009

I. GENERAL IMPRESSIONS

The Risk Ranking Tool was designed to assist in the prioritization of pathogen/produce pairs that warrant further risk modeling based on an overall impact to public health. The tool was easy to use, the presentation was clear and the conclusions were generally sound.

The tool considers four dimensions: 1) strength of the epidemiological association between the pathogen and the commodity; 2) severity of disease; 3) pathogen characteristics that affect disease risk or severity; 4) commodity characteristics. The nine criteria considered were 1) epidemiological link; 2) disease multiplier; 3) hospitalization rate; 4) death rate; 5) population susceptibility; 6) infectious dose; 7) prevalence of contamination; 8) consumption; 9) growth potential/shelf life. These dimensions of risk and their characterization seem reasonable for the intended purpose.

There are some limitations with the data sources that were which were discussed in the report. Although the types of data accessed in the preparation of the tool are listed and a cited reference list is included, a complete reference list was not provided. Therefore, it is difficult to assess if the appropriate references were used in all cases.

I think the following would be useful if it has not already been used (it was not cited):

Harris, L.J., J.N. Farber, L.R. Beuchat, M.E. Parish, T.V. Suslow, E.H. Garrett, F.F. Busta. 2003. Outbreaks associated with fresh produce: Incidence, growth and survival of pathogens in fresh and fresh-cut produce. *Comp. Rev. Food Sci. Food Safety*. 2S:78-141.

Some of the key database biases presented in the report are worth repeating. The epidemiology used to populate the tool is based on documented outbreaks. As pointed out in the report there is significant bias built in to CDC data and other published outbreak data. Outbreak investigations are not standardized across the country or even from outbreak to outbreak. Many outbreaks are not reported to the CDC. Details are often not available and the strength of the association to the "identified" produce item is not always known.

In order for an outbreak to be recognized sufficient illnesses must be recognized. Outbreak data is, by nature, biased to products consumed in greater per capita quantity. Outbreak data coupled with Criterion 8 (consumption) will bias the outcome to commodities consumed in greatest volume. I am not sure how the risk ranking would differ if a risk of illness/serving could be calculated. From a public health standpoint, per-capita risk rather than per-serving risk may be more important.

Some relatively well-documented produce outbreaks have resulted from cross contamination in the kitchen during final preparation. Some of these outbreaks are probably better characterized as “cross contamination” outbreaks and should probably be excluded or handled in an alternative fashion because the association to a specific produce item is largely by chance rather than a specific link. However, because citations for Table 3 are not provided, this is difficult to evaluate.

Additionally, there is no separation of whole or commodity-associated outbreaks and those associated with commercially-prepared fresh-cut product. This ties in to Criterion 9: growth potential/shelf life.

The ability of a pathogen to survive on or in a fruit or vegetable depends on the pH and water activity of the item (influenced by the presence/absence of peel), the humidity of the surrounding environment (including influence of package), temperature of storage, and time of storage. The ability of pathogens to grow in fresh-cut products is significantly enhanced but temperature control is more important and expected in distribution and retail. The shelf life of fresh cut products is also significantly different than for uncut product and reasons for shelf life termination are different. Self-life data included in Table 13 was for whole commodity rather than fresh-cut product. Although the citations are not provided, the growth potential score was most likely related to cut rather than intact product. This complicates use of growth potential/shelf life data. It is possible that the “conflicting” data cited for Growth score 2 and 3 are based on experimental design and type of product evaluated. Consumer or food service handling of commodities – essentially preparing fresh-cut product after purchase also complicates evaluation of this section.

Prevalence data for specific produce/pathogen are also relatively limited which influences Criterion 7. In many cases no data were available. This is not necessarily a problem with the current risk ranking tool but an acknowledged limitation. Even when survey data are available they often vary widely in sample size (both numbers of items and weight/volume), point of collection, and methods used. These factors may be significant. For example, the prevalence of *Salmonella* in the same lot of cantaloupe might be very different if surface swabs were compared to 25-g samples of excised rind or 100 g of melon flesh. These issues are a reality of the available published data – even the USDA MDP program, while reasonably consistent, has changed methodology over the years of the program.

The significant data collection that went into development of the risk ranking tool should be readily available to the users. Access to these data will be critical for Quantitative Microbial Risk Assessment and should be easily available.

II. RESPONSE TO CHARGE QUESTIONS

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on

whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

Comments are included in general narrative. The risk ranking tool satisfies the purpose for which it is designed. However, the ability to easily input new data especially on outbreaks and prevalence will be critical for long-term usefulness of the tool. These data directly influence Criterion 1 and 7. A tool was not really necessary for identification of lettuce/*E. coli* O157:H7 and tomatoes/*Salmonella* as important pathogen/produce pairs. It will be more useful for future pair identification but that will require continued data input.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

See general comments. The selection of dimensions of risk and criteria to characterize risk seem appropriate.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

I cannot think of another approach that could have been used for the sensitivity analysis.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rational and, when appropriate, the sources/data for the alternatives.

See general comments. Development of the tool and scoring of criterion is complicated by the number of produce items, and differences in their production, handling and processing. In many cases data are lacking or incomplete. Generalization was necessary and appropriate at this point. As more data become available, separation of commodity and fresh cut data may be appropriate.

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different

pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

See general comments. The statement “the role of sporadic disease is not necessarily considered to the fullest extent possible” could be shortened to “the role of sporadic disease is unknown”.

I am not sure a means to characterize uncertainty would be of value. A lot of uncertainty surrounds much of the data used.

The ability to update the database would be an important feature to using the tool more than one time. The purpose was to identify one or more pathogen/produce pairs to evaluate more thoroughly in a quantitative microbial risk assessment. *Salmonella*/tomatoes and *E. coli* O157:H7/lettuce and leafy greens are clear candidates based on the risk ranking tool (although the tool was probably not necessary to come to this conclusion). Once these risk assessments are underway it is not clear how the tool would be further used without updates.

Although the general data sources are mentioned they are not included and this significantly limits evaluation of the sources and interpretation of those data.

Some data inputs are unlikely to change in the short term: Disease multiplier, health dimension, hospitalization rate and death rate, population susceptibility. However, other information could change quickly as new data are generated and published. These include epidemiological link, prevalence, consumption, growth potential and shelf life.

Question 6: Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.

While it would significantly lengthen the report, inclusion of the publications used to develop the inputs of outbreak/pathogen combinations, growth potential, and prevalence would very useful to evaluating the tool.

The documented outbreaks were restricted to 1996 to 2006. Much has changed in the past decade with respect to both recognition of produce as an important vehicle of foodborne illness and implementation of risk reduction strategies pre- and post-harvest. In addition, a number of important produce outbreaks have occurred since 2006. It would be highly valuable to be able to enter new outbreak data in real time and to delete outbreaks that may eventually be considered irrelevant. As an example, the authors eliminated fresh juice outbreaks because of the introduction of mandatory pasteurization to the juice industry. Without this capability the usefulness of the tool will be limited and will become dated very quickly.

Data on sporadic illness associated with produce are non-existent and relative importance unknown. Should data become available incorporation would be useful.

Prevalence is another area where it would be important to be able to update the tool when new data became available. It is unclear if any attempt was made to standardize the available prevalence data. That said, it may be that there would be not impact overall given that the weighted average prevalence was given one of 4 scores.

III. SPECIFIC OBSERVATIONS

None.

IV. PEER REVIEW COMMENTS BY CHARGE QUESTION

Peer Review Comments on FDA’s Draft Document by Charge Question *Fresh Produce Risk Ranking Tool Summary: Identification of Priority Pathogen-Commodity Combinations for Quantitative Microbial Risk Assessment*

I. GENERAL IMPRESSIONS

Reviewer #1

In my opinion, the Fresh Produce Risk Ranking Tool accomplishes its intended purpose of providing a means of prioritizing pathogen/commodity pairs based on past identified outbreaks, the severity of the resulting disease, and likelihood of exposure by the consuming public. While a strictly quantitative approach is not possible at this time with available data, the tool does provide a semi-quantitative means of ranking the risk associated with various pathogen/commodity pairs. This will allow ranking of pathogen/commodity pairs where fewer outbreaks have been detected, compared with the well-recognized combinations of leafy greens and *E. coli* O157:H7 (EHEC) as well as tomatoes and *Salmonella enterica*. The rankings obtained with these less frequent outbreaks are variable based on the weight assigned to criteria other than epidemiological link. There is no other means to assess the uncertainty in the rankings. The tool is easy to use once the weightings are decided. The tool is flexible and allows the user to exclude a criterion by assigning it a weight of zero, or to rank eight of the criteria in relation to the ninth. As the authors point out in Chapter 4, Synthesis Comments and Conclusions, the risk associated with each commodity/pathogen pair is relative only to other pairs and is not proportional to the true risk difference between pairs. It will be important to update the database frequently as new information becomes available.

Reviewer #2

Overall, I believe that the Fresh Produce Risk Ranking Tool will be a valuable component in FDA’s toolbox for identifying and prioritizing important pathogen / commodity pairs among the myriad possible combinations of pathogens and commodities that potentially could be of greatest public health concern. The approach that was employed in the development of this tool seems reasonable given the goals of the project, and the interface is elegant in its simplicity, clarity, and flexibility. Moreover, the information supplied documenting the development and technical basis for the tool is well written and clear.

I do, however, have a number of concerns with the tool in its current state and the companion documentation. In particular, I believe that there are issues that need to be addressed with 1) the technical merit of the “infectious dose” criterion, 2) the selection of and justification for default bin boundaries, and 3) the potential for misinterpretation of the results due to the unintuitive nature of the “weighted” semi-quantitative scoring scheme that is employed. I do believe that these issues can be addressed via enhanced discussions and technical justifications, coupled with a revision of the approach used to characterize the relative infectivity of the pathogens of concern. I believe addressing these concerns will enhance the transparency and technical merit of the tool and facilitate interpretation of the tool’s output.

Reviewer #3

The Risk Ranking Tool (RRT) report provides a clear description of the model purpose, components, constructs and underlying data. Overall, the main strength of the model is that it is relatively simple and easy to understand and its results are easily replicated by users. The distinction that is made in the RRT between the data-driven component of the model (bins and scores for each model criterion) and the subjective component of the model (weights for each model criterion) provides the much needed transparency and separation between empirical evidence and value-laden expert judgment. As currently set up, if the value-judgment weights were set at 1 for all model criteria, the relative ranking of commodity (pathogen-produce) pairs are mostly based on objective/empirical evidence (see response to question 2 for more discussion here). As a simple and transparent tool, the RRT provides the relative ranking of the pathogen-produce commodities and allows for a rapid and systematic approach to prioritizing future work, i.e. follow-up robust quantitative risk assessment for commodity pairs that are ranked high. So long as the purpose remains that of prioritizing pathogen-produce commodities for more detailed risk assessment, the RRT is a reasonable approach to achieve this goal.

One of the major limitations of the RRT, and as acknowledged by the authors, is that it is not a predictive tool, and that it only relatively ranks known pathogen-produce commodities (as evidence by epidemiological database). As such, for the RRT to be useful in future efforts to identify high priority pathogen-commodity pairs, a periodic updated literature search to identify and incorporate new pathogen-commodity pairs and their associated information into the RRT will be necessary. In general, this RRT will remain useful as long as the list of the pathogen-produce commodities to be ranked and the underlying data for the 9 model criteria are periodically updated.

Reviewer #4

[General Impressions were not provided by the reviewer.]

Reviewer #5

The Risk Ranking Tool was designed to assist in the prioritization of pathogen/produce pairs that warrant further risk modeling based on an overall impact to public health. The tool was easy to use, the presentation was clear and the conclusions were generally sound.

The tool considers four dimensions: 1) strength of the epidemiological association between the pathogen and the commodity; 2) severity of disease; 3) pathogen characteristics that affect disease risk or severity; 4) commodity characteristics. The nine criteria considered were 1) epidemiological link; 2) disease multiplier; 3) hospitalization rate; 4) death rate; 5) population susceptibility; 6) infectious dose; 7) prevalence of contamination; 8) consumption; 9) growth potential/shelf life. These dimensions of risk and their characterization seem reasonable for the intended purpose.

There are some limitations with the data sources that were which were discussed in the report. Although the types of data accessed in the preparation of the tool are listed and a cited reference

list is included, a complete reference list was not provided. Therefore, it is difficult to assess if the appropriate references were used in all cases.

I think the following would be useful if it has not already been used (it was not cited):

Harris, L.J., J.N. Farber, L.R. Beuchat, M.E. Parish, T.V. Suslow, E.H. Garrett, F.F. Busta. 2003. Outbreaks associated with fresh produce: Incidence, growth and survival of pathogens in fresh and fresh-cut produce. *Comp. Rev. Food Sci. Food Safety*. 2S:78-141.

Some of the key database biases presented in the report are worth repeating. The epidemiology used to populate the tool is based on documented outbreaks. As pointed out in the report there is significant bias built in to CDC data and other published outbreak data. Outbreak investigations are not standardized across the country or even from outbreak to outbreak. Many outbreaks are not reported to the CDC. Details are often not available and the strength of the association to the “identified” produce item is not always known.

In order for an outbreak to be recognized sufficient illnesses must be recognized. Outbreak data is, by nature, biased to products consumed in greater per capita quantity. Outbreak data coupled with Criterion 8 (consumption) will bias the outcome to commodities consumed in greatest volume. I am not sure how the risk ranking would differ if a risk of illness/serving could be calculated. From a public health standpoint, per-capita risk rather than per-serving risk may be more important.

Some relatively well-documented produce outbreaks have resulted from cross contamination in the kitchen during final preparation. Some of these outbreaks are probably better characterized as “cross contamination” outbreaks and should probably be excluded or handled in an alternative fashion because the association to a specific produce item is largely by chance rather than a specific link. However, because citations for Table 3 are not provided, this is difficult to evaluate.

Additionally, there is no separation of whole or commodity-associated outbreaks and those associated with commercially-prepared fresh-cut product. This ties in to Criterion 9: growth potential/shelf life.

The ability of a pathogen to survive on or in a fruit or vegetable depends on the pH and water activity of the item (influenced by the presence/absence of peel), the humidity of the surrounding environment (including influence of package), temperature of storage, and time of storage. The ability of pathogens to grow in fresh-cut products is significantly enhanced but temperature control is more important and expected in distribution and retail. The shelf life of fresh cut products is also significantly different than for uncut product and reasons for shelf life termination are different. Self-life data included in Table 13 was for whole commodity rather than fresh-cut product. Although the citations are not provided, the growth potential score was most likely related to cut rather than intact product. This complicates use of growth potential/shelf life data. It is possible that the “conflicting” data cited for Growth score 2 and 3 are based on experimental design and type of product evaluated. Consumer or food service handling of commodities – essentially preparing fresh-cut product after purchase also complicates evaluation of this section.

Prevalence data for specific produce/pathogen are also relatively limited which influences Criterion 7. In many cases no data were available. This is not necessarily a problem with the current risk ranking tool but an acknowledged limitation. Even when survey data are available they often vary widely in sample size (both numbers of items and weight/volume), point of collection, and methods used. These factors may be significant. For example, the prevalence of *Salmonella* in the same lot of cantaloupe might be very different if surface swabs were compared to 25-g samples of excised rind or 100 g of melon flesh. These issues are a reality of the available published data – even the USDA MDP program, while reasonably consistent, has changed methodology over the years of the program.

The significant data collection that went into development of the risk ranking tool should be readily available to the users. Access to these data will be critical for Quantitative Microbial Risk Assessment and should be easily available.

II. RESPONSE TO CHARGE QUESTIONS

Question 1: The Risk Ranking tool was developed to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling. In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development. Considering the primary objectives for this risk ranking approach, please comment on whether the tool satisfies the purpose for which it was designed and whether it provides a useful tool for future efforts to identify high priority pathogen-commodity pairs.

Reviewer #1

In my opinion the tool satisfies the purpose for which it was designed. It will be necessary to revise the database going forward in order for it to continue to serve this purpose. The quality of the data is important and not always obvious. For example, in reviewing historical data on outbreaks associated with EHEC and leafy greens I have observed that different sources will show different numbers. Looking more closely at the data, I found one period in the fall of 1999 where some counted 2 outbreaks and others 3 outbreaks. The difference depended on whether one counted a cluster in California as being separate from cases in the northwest or whether these were considered a single outbreak. The definition of an outbreak requires that there be two or more cases with a common exposure. The difficulty is in determining whether the cases did indeed have a common exposure. This is but one example of the difficulties in developing a risk ranking tool such as this. The developers have accurately described it as a semi-quantitative tool.

Reviewer #2

I do believe that the tool (subject to the recommended changes) will satisfy the purpose for which it was designed and will be able to be used to identify high priority pathogen/commodity pairs.

There are a few issues that really need to be considered (as described in detail below) and resolved before the tool should be considered ready for use by federal agencies. First, it appears likely that the overall rankings from individual runs are highly sensitive to the specific bin boundaries that are selected for the run (for each criterion, scores range from 1-4, and default boundaries are provided). I am of the opinion that the default bin boundaries are very important because users of the tool may not feel like they have sufficient expertise to confidently override the default values. Little justification is provided in the documentation for the default values selected for the bin boundaries. To enhance transparency, the justification for the choices that were made in this regard should be provided in the report.

Second, the technical basis for supporting the “infectious dose” criterion is inadequate. This criterion should be fundamentally based on peer-reviewed dose response relationships for a specific level of risk (for example, the number of organisms which cause infection to 10% of those exposed – ID₁₀).

Third, some discussion is needed to support the risk ranking algorithm that was selected for this tool. It is not clear to me that the summation of weighted binned data results in a metric that is easily, clearly, and deeply understood. Discussion on this topic would go a long way towards supporting the selection of one algorithm over others.

Reviewer #3

The stated purpose for the RRT is to prioritize pathogen/commodity pairs for more robust risk modeling, i.e. highly ranked pathogen/commodity pairs would be subjected to further detailed quantitative risk assessment to predict their human health impact. The RRT as a simple risk based ranking tool to provide a rapid, transparent and systematic approach to prioritize a large combination pathogen-produce commodities, is appropriate for the stated objective.

As currently developed, the RRT is limited to the relative ranking of the pathogen-produce commodities that have been identified in a previously conducted public data search (pre-2009?). More recent outbreaks involving pathogen-produce commodities are not included in the current tool. As such, for the RRT to be useful in future efforts to identify high priority pathogen-commodity pairs, a periodic updated literature search to identify and incorporate new pathogen-commodity pairs and their associated information into the RRT will be necessary. In general, this RRT will remain useful as long as the list of the pathogen-produce commodities to be ranked and the underlying data for the 9 model criteria are periodically updated.

Reviewer #4

First, let me say that I’m puzzled by the statement “In developing this tool, it was recognized that other types of risk ranking tools were available and/or under development.” Is this tool supposed to complement those tools? Specifically, what other tools are available or under development? Does that mean that certain limitations are acceptable in this tool? If yes, what are those limitations and exactly how does the existence of other tools allow those limitations?

If the primary objectives for this risk ranking approach are “to provide a simple and transparent way to prioritize pathogen/commodity pairs for more robust risk modeling,” I would say that the tool is moderately successful. I would caution however that there are two explicit biases that are built into the tool: the use of CDC data and the math used to combine the data. No matter what other choices the user makes, they cannot change the CDC data or the math used to combine the data, and these two biases may drive many of the “conclusions” arising out of the use of this tool.

Also, I should note that Microsoft Access is an interesting choice. I certainly understand reasons for this choice, but it presents some unique challenges: The user must run the Windows and they must have a copy of Microsoft Access.

Reviewer #5

Comments are included in general narrative. The risk ranking tool satisfies the purpose for which it is designed. However, the ability to easily input new data especially on outbreaks and prevalence will be critical for long-term usefulness of the tool. These data directly influence Criterion 1 and 7. A tool was not really necessary for identification of lettuce/*E. coli* O157:H7 and tomatoes/*Salmonella* as important pathogen/produce pairs. It will be more useful for future pair identification but that will require continued data input.

Question 2: The tool considers four dimensions of risk and nine criteria to characterize the risk. The criteria were developed in consultation with a panel of food safety risk assessment experts. Is the list of criteria complete and balanced (i.e., not biased to one risk dimension)? If not, please identify any criteria that are unnecessary or additional criteria (and sources of data) that could be included in the future. Are appropriate data sources associated with each criterion? If not, please recommend alternative data sources.

Reviewer #1

I believe the criteria are appropriate, and the user has the option of weighting the criteria differently. Maintenance of the database will be critical, as more data is accumulated in the Microbiological Data Program and in the literature on prevalence of contamination, infectious dose, and growth potential. The Growth Potential and Shelf-Life score is a simplification of a fairly complex criterion, in that cut produce has a shorter shelf-life but may have greater growth potential than uncut produce. It is common practice, for example, to chop tomatoes in a processing plant for distribution and sale to foodservice operations. The cut tomatoes will be held refrigerated, whereas the raw uncut tomato may be held at ambient temperature. Bagged salads contain leafy greens with cut surfaces, where pathogens have been shown to adhere and grow. Shredded carrots in the bagged salad may represent a different risk than whole carrots. As more data becomes available on growth potential of the pathogens on various commodities, it may be desirable to separate these into two criteria, but at the present time data does not exist for most commodities. Only leafy greens and EHEC are ranked as “strong” for growth potential.

Reviewer #2

The criteria seem generally reasonable. However, I found the discussion of risk dimensions to be confusing and distracting. In fact, it is the 9 criteria that the tool is based on, not the four dimensions. I suggest editing the text for clarity (specific suggestions provided below in Section III).

I found the use of a composite variable for growth potential and shelf life to be curious. Some additional justification is needed to explain why this approach was taken for this criterion and not others.

The approach employed for “infectious dose” needs a complete overhaul as detailed below. Briefly, peer reviewed dose response data should be the basis for this criterion and a specific and consistent level of infection (ID₁₀ for example) should be used for all pathogens. A source for dose response relationships is provided in Section III.

Reviewer #3

As described in the RRT report, there are 9 criteria in the model: 1) epidemiological link, 2) disease multiplier, 3) hospitalization rate, 4) death rate, 5) population susceptibility, 6) infectious dose, 7) prevalence of contamination, 8) consumption and 9) growth/potential shelf life. This reviewer agreed that these are the appropriate indicators for the hazard, exposure and public health risk potential associated with the pathogen-produce commodities. The cited literature/data sources associated with each criterion are appropriate. However, the reviewer also noted the following with respect to the current scoring of the 9 criteria:

- In general, the following broad aspects of the RRT can be described:
- The potential public health impact aspect has two dimensions: the epidemiological and health dimensions. These two dimensions include the epidemiological link, disease multiplier, hospitalization rate and death rate scores. Thus, the public health potential part of the model can score as much as 16 (4x4, assuming the unity weights for all 4 criteria)
- The hazard potential aspect has one dimension: the agent dimension, which includes the population susceptibility and infectious dose scores. Thus the maximum score that the hazard potential part of the model can get is 8 (2x4, assuming the unity weights for both criteria)
- The exposure potential aspect has one dimension, production and processing, with scores based on 3 criteria: prevalence of contamination, consumption and growth/potential shelf life scores. Thus, the maximum composite for the exposure potential is 12 (4x3, assuming the unity weights for all 3 criteria)
- Implicit in the overall imbalance in the total scores between the potential hazard, exposure and public health impact aspects of the model is the developer’s judgment of the order of their importance. This implicit judgment if retained in the model should be explicitly noted. Since the RRT has a distinct subjective component in

which additional weight (by expert judgment) can be incorporated to these criteria, this information needs to be made known so expert judgment to increase the score (using weights) for any of the criteria can be done appropriately and in context of the model emphasis on the public health impact scores. As it is, with the weights at unity, the ranking will be heavily influenced by the public health impact aspect of the model.

From NHANES, the amount (grams) per eating occasion (EO) can also be obtained. The current RRT only uses the % consumer data from the survey to establish the score for the consumption criterion. A more robust consumption score can be obtained based on the combination of % consumers and amount per EO. The reviewer also noted that the NHANES 03-04 data used for criterion 8 (consumption) is not the most current (NHANES 05-06 is available), and the RRT report incorrectly described the data from NHANES 03-04 as a three day dietary recall (it is only 2-day).

It is also noted that while the RRT report described 9 criteria, the Access tool (and user's manual), only has 8 criteria in the bins (score) portion of the model. The susceptibility criterion is not present in this portion of the tool. Further, while the RRT report described the score for the growth/potential shelf life criterion (Table 14 of the RRT report) as the combined growth score (Table 12 of RRT report) and shelf-life score (Table 13 of the RRT report), the Access tool (and user's manual) only shows the shelf-life bin/score in the bin/score portion of the RRT. The weight portion of the RRT has all 9 criteria. This discrepancy should be addressed by the developer.

Reviewer #4

The list of criteria appears to be complete although perhaps somewhat biased. For example, by including hospitalization rate and death rate (since these two are often correlated), the severity of the disease gets double-counted.

As noted in the detailed comments below, the population susceptibility scale is somewhat odd, as it purports to be a numerical scale, but (as I understand it), diseases affecting the elderly more will score 2 and 3, but never 1 and 4, while diseases that affect children more will score 2 and 4 but never 1 and 3.

Data sources are cited in the body of the text when each criterion is mentioned, but the reader must consult the appendix to see agent-by-agent or food-by-food breakdown. A critical omission in these appendix tables are any citations what so ever. This is true for all the appendix tables, but let me use Table A-5 "Infectious Dose and Infectious Dose Categories by Pathogen" as an example. In the first line of this table it says "Bacillus cereus | 10^5 - 10^6 | 10^5 | High", but the reader is left wondering what reference or references were used to deduce this information. This is a very critical omission, and must be fixed in the revised version of the document if the document is to have any use to the large risk modeling community at all.

It is difficult to recommend alternative data sources if the original data sources are not clear.

Reviewer #5

See general comments. The selection of dimensions of risk and criteria to characterize risk seem appropriate.

Question 3: The report describes a simplistic form of sensitivity analysis to better understand the influence of different weighting schemes on the resulting risk ranking. Given the semi-quantitative approach of this tool, was an appropriate sensitivity analysis method chosen? If not, please suggest specific alternative methods (e.g., more robust methods) to better understand and characterize the sensitivity of the tool to weighing choices.

Reviewer #1

I do not know of another method of sensitivity analysis.

Reviewer #2

Two types of sensitivity analyses were conducted: weighting of an individual criterion and randomized weighting of the nine criteria. The sensitivity analyses that were conducted by changing individual (and in some cases two) criteria are appropriate (however, I believe that the actual results could change substantially once the relative infectivity issue identified above is addressed). In terms of presentation, I found the results hard to follow and would prefer to see the results for each of the sensitivity analyses presented by ranked results from that analysis (currently they are presented ranked by the baseline rank).

With respect to the sensitivity analyses conducted using randomized weighting, I am unconvinced that sufficient randomized samples were conducted (Section 3.3.2) to be representative of the solution space of interest (100 runs were conducted out of nearly 2 million possible combinations). Rather than enhancing this section, I would prefer to see an emphasis on the relative impact and/or synergy of weighting 2 or 3 criteria that are likely to be most important from a public health management perspective.

Reviewer #3

The sensitivity described in the report was based on changing the weights (value judgment) component of the model. While this is necessary to understand the influence of this component of the model, it is also important to perform a sensitivity analysis based on the bin/score component of the model. This can be done by setting the value judgment (weight component) at 1 (i.e. no impact on final score). There are two considerations in performing the sensitivity analysis for bin/score component of the RRT: 1) number of bins and the upper and lower limits for each bin, and 2) type of scores for each bin. In the current model, each of the 8 criteria (this reviewer only counted 8 criteria in this part of the tool – see comments above) has 4 bins and each bin is given an ordinal score (1, 2, 3 and 4). The ranking results can change if the upper and lower values of the 4 bins are modified, particularly when changes are made to the consumption criterion (by changing the limits of these consumption bins, this reviewer can influence the 2nd ranking of the commodity pairs currently in the system). It is also important to evaluate whether

more resolution in the ranking can be obtained with more than 4 bins for each criterion, for example 5 or 6 bins (see additional comments on this point under question 4). Lastly, the current model uses ordinal ranking. It may be worthwhile to consider another type of scoring, e.g. proportional (1, 10, 100, 1000, etc.) in the sensitivity analysis to see if improved resolution in the ranking and possible changes in the ranking can occur with an alternative scoring metric.

This reviewer recognizes that the possibilities of alternative number of bins and scores are limitless and one can over complicate the sensitivity analysis. However, the bin/score is objective (data driven) and central to the RRT and thus some minimum effort is needed to demonstrate that the current bin/score is the most optimal in providing the most stable and consistent relative ranking. Again, this sensitivity analysis should be done without the influence of the value judgment (weight =-1) component of the RRT.

Reviewer #4

The “simplistic form of sensitivity analysis” appears to be acceptable to show the inherent variability possible from the use of the tool; however, as noted below, more details are needed on exactly what was done to perform the Monte Carlo portion of the sensitivity analysis.

Reviewer #5

I cannot think of another approach that could have been used for the sensitivity analysis.

Question 4: For each criterion, a unique scoring system was developed to fit the data source(s) for that criterion using general categories from low (or unknown) to high. Because each criterion is unique, the category bins were designed to distinguish among different possibilities at a level of granularity appropriate to both the available data and intent to design a simple, easy-to-use ranking tool. Are the scoring categories appropriate for each criterion? If not, please suggest alternatives including the rational and, when appropriate, the sources/data for the alternatives.

Reviewer #1

I believe the scoring categories are appropriate, and I understand that you have not attempted to match the scores to the relative risk under each criterion. But I wonder if it is possible to assign values other than 1, 2, 3, and 4 to the bins. In particular, I question the scores for consumption. According to the reported data (Table A-7 in Appendix A), the daily consumption rate for leafy greens is about 56 times higher than for herbs and 100 times higher than for green onions. The score for leafy greens is 4 times higher than for herbs and green onions. Could the scoring more closely reflect the difference in consumption rates? Would it not better reflect the difference in risk? I realize that changing this scale would not affect the relative risk of the leafy greens/EHEC pair but it might affect the scores for other leafy green/pathogen pairs.

Reviewer #2

I think that the use of four scoring categories is appropriate for each criterion. However, one of my principal concerns revolves around the selection of the default boundaries for each criterion category. The justification for each selection should be made explicit and in important cases, sensitivity analyses should be conducted to determine the relative sensitivity of the output to these selections. Without such explicit explanation it is difficult to assess the appropriateness of the choices that were made.

In several cases, it appears that the default values used for categories are not set such that maximum sensitivity is achieved (for example, in the epidemiological link the “very strong” category appears to be set too low to differentiate the pathogen / commodity pairs that caused the most number of outbreaks from others).

Reviewer #3

As indicated in the question, each criterion is unique in its possibility of values. Thus, while the current 4 bins/scores may be sufficient for one criterion, it may be inadequate to capture the granularity of the available data for another criterion. Significant loss of information in the minimalist approach to 4 bins is observed with the epidemiological link and consumption criteria. For the epi-link criterion, it appears that there is a need to differentiate the current “very strong” bin (>5 outbreaks and >100 cases) to capture situations where there are >10 outbreaks and >1000 cases. Similarly, additional bins may be needed to capture the range of data that are available with the consumption data. Further as noted above, consideration to incorporate g/EO with % consumers to derive the bin/score for this criterion is recommended to fully capture consumption/exposure potential. Clearly any modification to the no. of bins and scores will need to be made with an eye toward the discussed imbalance in the total scores for the 3 aspects of the model (hazard, exposure and public health impact potential) as discussed under question #2.

Reviewer #4

The scoring categories are generally appropriate for each criterion; however, see my comments regarding the population susceptibility scale above and below.

Reviewer #5

See general comments. Development of the tool and scoring of criterion is complicated by the number of produce items, and differences in their production, handling and processing. In many cases data are lacking or incomplete. Generalization was necessary and appropriate at this point. As more data become available, separation of commodity and fresh cut data may be appropriate.

Question 5: Chapter 4 of the report identifies beneficial features and limitations of this risk ranking tool and approach. Are there additional features that could be added to the model that would improve its usability? For example, one limitation identified in the report is the absence of a means to characterize the uncertainty associated with the rankings for different pairings. Would developing such a feature be valuable given the objectives of the tool? If so, please recommend methods or approaches that should be considered.

Reviewer #1

I think it would be extremely valuable to the user to know the uncertainty associated with rankings, but I do not have a method to accomplish this. The uncertainty is largely associated with the small number of outbreaks associated with some commodity groups as well as lack of data for other parameters. Produce items such as herbs that are consumed in small amounts may not be identified by the case patient when interviewed, and thus never enter the data stream. In addition, we know that foodborne illness is underreported both by the general population and by some local health jurisdictions.

Reviewer #2

I am not convinced that this tool needs to be limited to pathogen/commodity pairs for which outbreaks have been reported. The epidemiological link criterion appears to account for this intersection, whereas each of the other criteria only requires data on either the pathogen of interest or the commodity (criterion 7: prevalence of contamination is the possible exception to this statement). It seems like this tool could be extended so that it could be used to identify potential future outbreaks of concern and thus the consideration of management alternatives to prevent them.

Section 4, paragraph 2 highlights the fundamental weakness of this work, which in my opinion raises questions about the conclusions. Specifically, “The categorization and binning approaches may seem somewhat arbitrary”. I agree with this statement. Moreover, I believe that the usability of the tool would be most strongly enhanced by addressing this issue. The potential benefits of this tool are great, but the technical underpinning must be sound and transparent, and I believe that addressing this issue is much more important than adding additional features.

Reviewer #3

As discussed earlier, the limitation of the current RRT is the starting list of pathogen-produce pairs that the model ranks. This list needs to be updated periodically for this tool to remain useful. As far as capturing uncertainty associated with the model criteria, there is always a real risk for going beyond the simple relative ranking purpose of the tool when elaborate uncertainty model is introduced. To avoid doing so, perhaps an overall uncertainty weight could be incorporated as the third part of the tool. However, as with the current “weight” component of the tool, this added component will be judgment based and the numerical value has to be such that it does not overwhelm the empirically-derived scores (bin component of the model).

Reviewer #4

One limitation of the model is the lack of an ability to edit the basic assumptions easily. For example, let’s say that I’ve done some experiments and I believe that Salmonella growth potential in tomatoes should be 4 (not 3). The only way to change this would be to edit the GROWTH_POTENTIAL table, which I can only do if I know the PATHOGEN_CATEGORY_ID and the COMMODITY_CATEGORY_ID. I could also make a

form with a lookup tables to translate these into English, but this is likely beyond most users who are unfamiliar with Access.

I'm not sure that the absence of a means to characterize the uncertainty associated with the rankings for different pairings is a feature worth developing.

Reviewer #5

See general comments. The statement “the role of sporadic disease is not necessarily considered to the fullest extent possible” could be shortened to “the role of sporadic disease is unknown”.

I am not sure a means to characterize uncertainty would be of value. A lot of uncertainty surrounds much of the data used.

The ability to update the database would be an important feature to using the tool more than one time. The purpose was to identify one or more pathogen/produce pairs to evaluate more thoroughly in a quantitative microbial risk assessment. *Salmonella*/tomatoes and *E. coli* O157:H7/lettuce and leafy greens are clear candidates based on the risk ranking tool (although the tool was probably not necessary to come to this conclusion). Once these risk assessments are underway it is not clear how the tool would be further used without updates.

Although the general data sources are mentioned they are not included and this significantly limits evaluation of the sources and interpretation of those data.

Some data inputs are unlikely to change in the short term: Disease multiplier, health dimension, hospitalization rate and death rate, population susceptibility. However, other information could change quickly as new data are generated and published. These include epidemiological link, prevalence, consumption, growth potential and shelf life.

Question 6: Are there any additional comments or significant issues not addressed in the above questions? If so, please note those issues along with any suggestions that you might have as to how the methodology, risk ranking tool, approach or documentation can be further improved upon.

Reviewer #1

I think it is likely that listeriosis is an illness that may be associated with certain types of produce, and perhaps such outbreaks will be detected in the future. There was an outbreak associated with cole slaw in the 1980's, which is prior to the time period used for this data set. It was attributed to contamination of the cabbage in the field. *Listeria monocytogenes* has been found by surveillance sampling in commercial bagged salads, but no outbreaks have been detected from such contamination.

Reviewer #2

My additional comments are detailed below in Section III.

Reviewer #3

- As noted earlier, there is an inconsistency between the RRT report and the Access tool (and user guide). The report described 9 criteria, the Access tool (and user manual) only has 8 criteria in the bins (score) portion of the model. The susceptibility criterion is not present in this portion of the tool. Further, while the RRT report described the score for the growth/potential shelf life criterion (Table 14 of the RRT report) as the combined growth score (Table 12 of RRT report) and shelf-life score (Table 13 of the RRT report), the Access tool (and user's manual) only shows the shelf-life bin/score in the bin/score portion of the RRT. The weight portion of the RRT has all 9 criteria. Perhaps there is a reason for the discrepancy. If this is the case, appropriate documentation should be incorporated.
- Access tool: While the values of all bins for all 8 model criteria can be changed, the lowest bin value cannot be changed by user. Is there a reason for this? In some instances, it may be justifiable to have lower value for the lowest bin, such as % hospitalization, % consumers, etc. and a form of sensitivity analysis around the values of the bins would be warranted.

Reviewer #4

See comments below. [In *Specific Comments.*]

Reviewer #5

While it would significantly lengthen the report, inclusion of the publications used to develop the inputs of outbreak/pathogen combinations, growth potential, and prevalence would very useful to evaluating the tool.

The documented outbreaks were restricted to 1996 to 2006. Much has changed in the past decade with respect to both recognition of produce as an important vehicle of foodborne illness and implementation of risk reduction strategies pre- and post-harvest. In addition, a number of important produce outbreaks have occurred since 2006. It would be highly valuable to be able to enter new outbreak data in real time and to delete outbreaks that may eventually be considered irrelevant. As an example, the authors eliminated fresh juice outbreaks because of the introduction of mandatory pasteurization to the juice industry. Without this capability the usefulness of the tool will be limited and will become dated very quickly.

Data on sporadic illness associated with produce are non-existent and relative importance unknown. Should data become available incorporation would be useful.

Prevalence is another area where it would be important to be able to update the tool when new data became available. It is unclear if any attempt was made to standardize the available prevalence data. That said, it may be that there would be not impact overall given that the weighted average prevalence was given one of 4 scores.