Multicriteria-based Ranking Model for Risk Management of Animal Drug Residues in Milk and Milk Products

Peer Review Report:

External Peer Review Comments

and FDA Responses

Food and Drug Administration

U.S. Department of Health and Human Services

2015

Introduction

Versar, Inc. conducted an external peer review of the January 2014 version of the FDA's *Multicriteria-based Ranking Model for Risk Management of Animal Drug Residues in Milk and Milk Products*. This risk assessment serves as a decision-support tool to assist with re-evaluating which animal drug residues should be considered for inclusion in milk testing programs. The risk assessment also may be used to identify and prioritize research needs.

For this external peer review, Versar, Inc. under contract with FDA, selected five experts to evaluate and provide written comments on FDA's draft approach for the *Multicriteria-based Ranking Model for Risk Management of Animal Drug Residues in Milk and Milk Products*. The following describes the charge to the reviewers, as well as a brief introduction and background of the five independent external reviewers.

Reviewers:

Beth P. Briczinski, Ph.D.
National Milk Producers Federation
Arlington, VA

Igor Linkov, Ph.D.
Carnegie Mellon University (Adjunct Professor)
Pittsburgh, PA

Scott A. McEwen, DVM, DVSc
University of Guelph
Guelph, Ontario, Canada

Shirley Price, Ph.D.
University of Surrey
Guildford, Surrey, UK

Geoffrey W. Smith, DVM, Ph.D.
North Carolina State University
College of Veterinary Medicine
Raleigh, NC

CHARGE TO REVIEWERS

FDA is seeking your expert opinion and to comment on the following:

Charge Questions:

1. Is the draft technical approach, which has been described in the document, adequate and sufficient to answer the risk assessment charge questions (see "Purpose and Charge", page 8 of the document); and if not, what approach would be better to use with the data available to answer the questions?

2. Is the report clear in its description of the draft approach? If not, please identify areas that are unclear or could be more transparent.

3. Does the draft approach consider the relevant criteria needed to rank the public health risks associated with drug residues in milk and milk products? If not, please describe the additional criteria that should be included.

4. Does the draft approach consider all relevant sub-criteria, factors, or data sources needed to evaluate the likelihood and magnitude of drug administration to dairy cattle on U.S. dairy farms? If relevant sub-criteria, factors, or data sources are missing:

    a. Which are missing?
    b. Where can information about these sub-criteria, factors, or data sources relevant to the likelihood and magnitude of drug usage on the farm to treat dairy cattle be found and what is the expected impact of not incorporating these sub-criteria, factors, or data sources?

5. Does the draft approach consider all relevant sub-criteria, factors, or data sources needed to determine the likelihood of drug presence in bulk-tank milk? If relevant factors are missing:

    a. Which are missing?
    b. Where can information about these sub-criteria, factors, or data sources relevant to the probability of the drug entering the raw bulk-milk tank be found; and what is the expected impact of not incorporating these factors, sub-criteria, or data sources?

6. Does the draft approach consider all relevant sub-criteria, factors, or data sources needed to predict the impact of dairy processing on the relative concentration of drug residues in the final dairy product? If relevant sub-criteria, factors, or data sources are missing:

    a. Which are missing, and where can this information relevant to the dairy products and drug residues considered in the risk assessment be found; and
    b. What is the expected impact of not considering these sub-criteria, factors, or data sources?

7. Does the draft approach consider all relevant sub-criteria, factors, data sources and scoring standards needed to predict the impact of consumption of dairy products on drug residue intake? If relevant sub-criteria, factors or data sources are missing:

a. Which are missing, and where can this information relevant to the consumption of dairy products be found; and
b. What is the expected impact of not considering these sub-criteria, factors, or data sources?

8. Does the draft approach consider all relevant sub-criteria, factors or data sources needed to predict the public health effects on consumers? If relevant sub-criteria, factors or data sources are missing:

    a. Which are missing, and where can this information relevant to the consumption of dairy products be found; and
    b. What is the expected impact of not considering these sub-criteria, factors or data sources?

9. Is the proposed scoring for all criteria, sub-criteria, and factors appropriate? If not, what changes would you recommend and why?

10. Is the algorithm that combines criteria scores and weights into an overall score appropriate? If not, please provide suggested improvements.

11. What weighting is most useful for an accurate drug ranking, and what weighting should be avoided to prevent an inaccurate ranking?

12. How would you recommend weighting the proposed criteria, sub-criteria, and factors in this risk ranking? Please provide justification or rationale/reasoning for your assignments.

13. Is the drug list complete and inclusive of all drugs most likely to be used in dairy cattle in the U.S., and does the list include drugs that should not be on the list? If you answered yes to either of these questions, please describe how you would revise the list, and the justification for the revision.

14. Is the list of dairy products representative of all potential dairy products, or does the list include products that should not be on the list? If so, please describe how you would revise the list, and the justification for the revision.

15. Do you have any additional comments on other aspects of the document? Please share them in your review.

The FDA thanks each of the reviewers for their thoughtful and detailed review of the *Multicriteria-based Ranking Model for Risk Management of Animal Drug Residues in Milk and Milk Products*. We have considered all of the comments in our revision of the model and responded to each comment below. The revised risk assessment model has benefited significantly from this independent peer review.

I. GENERAL IMPRESSIONS

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Overall, I believe this is a solid risk assessment model designed to rank the public health risks associated with drug residues in milk and milk products. There are certainly some limitations to the model (for example, much of the data on drug use in dairy cattle are 20+ years old and we have very limited data on the incidence of drug residues in milk with the exception of β-lactam antibiotics) – however, I believe the FDA has done the best job they could with the data available and I think the conclusions of the model will be sound. In my opinion, we certainly should look long term towards changing the antibiotic residue testing program done in the dairy industry. As the report points out, all shipments of milk are currently tested for β-lactams and only randomly (*i.e.*, rarely) for other classes of antibiotics or other non-antibiotic drugs. In my work as a dairy veterinarian – it is clear that producers are aware that milk is frequently tested for β-lactams and purposely choose other classes of drugs on some occasions. A new system that will randomly test each load of milk for a different class of antibiotics would be more effective in my opinion (every load is tested for a class of antibiotics but not necessarily β-lactams). I think this risk assessment will provide the evidence for which classes propose the greatest risk – both in terms of their likelihood to show up in milk as well as their likelihood to cause significant health problems in humans. I believe the model clearly asks the correct questions (which drugs are most often used in dairy cattle, which drugs are most likely to cause residues in milk, what happens to these drug concentrations during the manufacture of different dairy products, and which drugs could have the most significant impact on human health if consumed). Given the data available to build the model – I believe the conclusions should be fairly accurate. I would like to see the results of running the 54 drugs of concern through the model to really comment on the "soundness of conclusions," however, I believe the proposed risk assessment provides accurate guidance to the industry as we consider ways to potentially change the current testing program.	We appreciate the comment. The full technical report will include a section to discuss limitations of the approach and available data and will identify data/information needs to further refine the model.
Reviewer #2	Overall, I believe the draft approach is very well done. It makes sense to use a risk-based approach to the monitoring of veterinary drug residues in milk, and the proposed approach should provide very valuable input into a monitoring program. For the most part, the approach to risk ranking is presented in a clear and transparent fashion that should facilitate understanding and discussion among stakeholders as well as building confidence in the process. In the absence of quantitative risk assessments on the full range of veterinary drugs, the proposed risk ranking approach, based on multi-criteria decision analysis, appears to be a good choice of method, given the available data and need to consider a wide range of	We appreciate the comment. In developing this model, we evaluated the potential correlations among the sub-criteria to ensure independence as appropriate for this modeling approach. The full technical report addresses this point.

REVIEWER	COMMENT	RESPONSE
	criteria. The most important criteria have been identified and incorporated into the model. There are a few areas, described in more detail below, where improvements could be made in clarity of presentation and explanation of rationale so that readers from a variety of backgrounds and perspectives can understand the approach. There is also a need to consider the implications of potential correlations among sub-criteria in the model, particularly pertaining to criteria A & B.	
Reviewer #3	The FDA has proposed a risk assessment model to develop a relative ranking of veterinary drugs in milk and dairy products based on the public health risk they present to consumers. The model considers five major criteria, all of which are relevant and appropriate: the likelihood of the drug to be administered to lactating dairy cattle and to be found in bulk tank milk, the impact of various processing steps on the concentration of the drug, the relative intake of a drug through consumption of dairy products, and the impact on human health. Overall, the explanation and rationale presented in the report was sufficient, with the exception of Criterion E, which seemed to be lacking in detail and clarity. While data gaps do exist and some data sources could and should be improved (specific comments for each criteria below), nonetheless, it is an impressive body of work. The model could easily be adapted as more information becomes available to further refine the relative ranking of the risks of the drugs over time. The final risk score will encompass the likelihood of drug residue intake through consumption of dairy products (Criteria A, B, C, and D) as well as the effect on human health (Criterion E). The key will be to determine how to weight the risk scores for each of these criteria, balancing the likelihood of residues being consumed with the severity of the consequence, in calculating that final score. The general approach is logical, is defensible based on available scientific data, and should allow FDA to meet their charge from NCIMS. The model will also help focus resources and future testing to best protect public health.	We appreciate the comment. In preparing the full technical document, we updated and enhanced the description of the model to improve clarity of descriptions of all criteria.
Reviewer #4	Drug residues in milk attract significant interest from multiple stakeholders. Multiple drugs are used in the milk supply chain, but drug use, frequency, and toxicity could vary significantly. Lacking clear decision guidance, risk-averse managers could spend significant resources testing for multiple drugs with limited risk potential, diverting scarce resources from higher-risk areas. The document proposes a solid general approach for prioritizing antimicrobial drugs used in the milk supply chain for testing. The approach incorporates a hierarchical multi-criteria scoring model to derive a risk score to inform decisions on priority testing for milk and milk	We appreciate the comment. In preparing the full technical document, we improved the description of the methodology and evaluated the structure of the model. We also provide justification for the selected MCDA approach.

REVIEWER	COMMENT	RESPONSE
	products for drug residues. The general idea is that drugs that score higher on risk should be a higher priority for testing. The use of multi-criteria models in risk prioritization can be suitable when it is impractical to create and populate a full causal risk model, so models instead act as proxies that are measurable and which are thought to be associated with risk. The proposed approach integrates significant volume of historic and measurable data with expert judgment. Overall, the methods used are reasonable and document is well structured and clear. There are several areas where the methodology and presentation could be improved. First, drug prioritization is part of an overall testing strategy and a clear link to the overall decision process should be provided. Second, more justification of the proposed MCDA approach should be given. Third, the structure of the model should be carefully evaluated. As discussed below, the proposed models may not be appropriate given the task at hand.	
Reviewer #5	Ranking any residues in bulk materials poses problems: concentration; parent compound vs. metabolite; mixtures; population at risk, etc. This dossier aims to address these issues using a matrix ranking scheme. The overall aims of the study were clear and are stated as: • To assess the public health risk to consumers from veterinary drug residues in milk and milk products and develop a ranking of the drugs, based on the risk to the consumer. • To assess whether the proposed ranking will support the development of future testing protocols to continue to ensure a safe milk supply. The Matrix Risk Ranking Scheme proposed used a multi criteria decision analysis (MCDA) involving four steps: **Step 1** Identification of the drugs/drug residues or metabolites to be evaluated. **Step 2** Identification of the milk and milk products to be evaluated. **Step 3** Definition of the model criteria, weighting, and scoring standards. **Step 4** Execution of the risk model to determine risk scores for each drug-commodity pair for relevant age groups. The rationale for the steps chosen was clear and provided a strong basis on which to evaluate the veterinary residues in milk from all sources identified and establish the potential risk to humans. A preliminary list of over 450 drugs was initially identified and then, using exclusion criteria, the number was reduced to 54 drugs. The MCDA used five criteria which were	We appreciate the comment. In preparing the full technical report, we improved the description of the model methodology to enhance clarity and transparency.

REVIEWER	COMMENT	RESPONSE
	selected based on risk management questions and available data: **Criterion A**: Likelihood and magnitude of use of the drug in dairy cattle **Criterion B**: Likelihood of the drug's presence in bulk-tank milk **Criterion C**: Impact of processing on drug residues present in raw milk **Criterion D**: Magnitude of consumption of dairy products **Criterion E**: Health effects from human exposure In my assessment, the matrix ranking proposed compared favourably with the one used by the Veterinary Residue Subcommittee in the UK. The ranking was a clear evaluation. However, the decision tree was complex and any rules set for evaluating risk in any one of the five criterion was complex and not as transparent to those who are not risk assessors. It would be useful for this to be reviewed and rules established to allow this ranking to be fully understood by a wider audience. However, the current scheme does provide a very strong foundation to assess drugs and their metabolites and their potential for risk in all the commodities provided. The ranking also considers the population at risk.	

II. RESPONSE TO CHARGE QUESTIONS

CHARGE QUESTION 1: Is the draft technical approach, which has been described in the document, adequate and sufficient to answer the risk assessment charge questions (see "Purpose and Charge", page 8 of the document); and if not, what approach would be better to use with the data available to answer the questions?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Yes – I believe the purpose and charge is both appropriate and clearly written. It is not really stated "WHY" the NCIMS would like to investigate changing the current milk residue testing program (although I feel fairly certain that some people think we are over testing for β-lactam antibiotics while not testing often enough for other drugs) – but perhaps this is beyond the need of this document. The purpose clearly lays out which are the important questions to ask during the process of building the risk assessment model.	In preparing the full technical report, additional background information was incorporated.
Reviewer #2	I believe that the proposed approach is a feasible and useful way of addressing the charge questions. It enables a ranking of veterinary drugs with respect to a variety of criteria of interest with respect to potential human exposure through milk and milk products. Although the document refers to the proposed approach as a "risk assessment" (line 2), I really wonder if this is the best term to use in this context. I suggest that "risk ranking" or something similar would be more appropriate. I think of a risk assessment as a structured approach (*e.g.*, hazard identification, hazard characterization, exposure assessment, risk characterization) to estimating the likelihood of adverse effects from exposure to chemical hazards, in which risk to human health is characterized in quantitative or qualitative terms.	The approach we used is consistent with approaches used by others for the purposes of ranking potential hazards (Anderson *et al.* 2011; FAO 2012; and Brookes 2014). The technical report provides a detailed justification of the selected approach. Regarding terminology, this multicriteria-based ranking model is a type of risk assessment and we have revised the title of the risk assessment to avoid any confusion as to its nature.

REVIEWER	COMMENT	RESPONSE
	In my view, the proposed approach does not really attempt to undertake full risk assessments of the veterinary drugs that could occur in bulk milk. Formal quantitative risk assessments, if they were already available or technically feasible to undertake, would I think be the best way to address the charge questions. However, I don't believe that they are available and I think there are formidable barriers (*e.g.*, many important data gaps) to performing risk assessments on the long list of veterinary drugs that could be found in milk. Having said that, I recognize that most of the criteria / sub-criteria provide information that is relevant to one or more steps in a formal risk assessment, but I wonder whether the outcome (final scores) of the proposed model could legitimately be considered to characterize risk. Perhaps some discussion on this is warranted in the text. In any case, I think that the proposed approach is the best available option. national and international policy documents (*e.g.*, EPA, NUREG, FAO, WHO, OIE),	
Reviewer #3	In general, the approach described in the report does seem to be appropriate to sufficiently address the risk assessment charge questions based on the request from NCIMS. For clarification of Question #2 (*Which of these drugs are likely to result in drug residues present in bulk-tank milk?*): is this question referring to *violative* drug residues, or drug residues at *any* concentration (which could, in theory, be any drug administered to a dairy cow during its lifetime)? The latter would be likely redundant to the drugs identified in Question #1 (*What drugs are most likely administered to lactating dairy cattle in the US and what is the magnitude of use?*).	We agree with the reviewer that the distinction between violative drug residues (*i.e.*, drug residue presence in concentrations above tolerance) and drug residues in concentrations below tolerance is important for public health as well as enforcement. We have differentiated between violative drug residues and non-violative drug residues in the model scoring. The criteria A and B are distinct from one another. We have added additional text in the technical report to make these distinctions clear.
Reviewer #4	The purpose statement includes two parts: (1) "assess the public health risk to consumers from veterinary drug residues in milk and milk products" and (2) develop a relative ranking of the drugs, based on the risk to the consumer." The approach incorporates a hierarchical multi-criteria scoring model to derive a risk score for drugs based on multiple criteria related to risk, but risk itself is not estimated. The risk score derived in this approach makes sense only as a comparative metric of risk associated with each drug, but can not be used as an absolute metric of risk. I do believe that the proposed approach is solid overall and given the data available, traditional risk assessment is not possible. With this said, I would recommend adding an explanation of how this prioritization will be used to design testing strategy. A clear link to the overall decision process should be provided. Depending on the type of testing decision, different portions of the entire model or the entire model itself might be most appropriate. The various risk scores can be inputs to simple "value of information" decision models that determine whether a particular test is likely to produce a useful result. For example, a prioritization of drugs to research might be different than the prioritization of drugs for which to test, and the prioritization for testing	The use of the prioritization to inform the overall testing strategy and decision process will be addressed when a sampling plan is developed. It is outside the scope of this assessment.

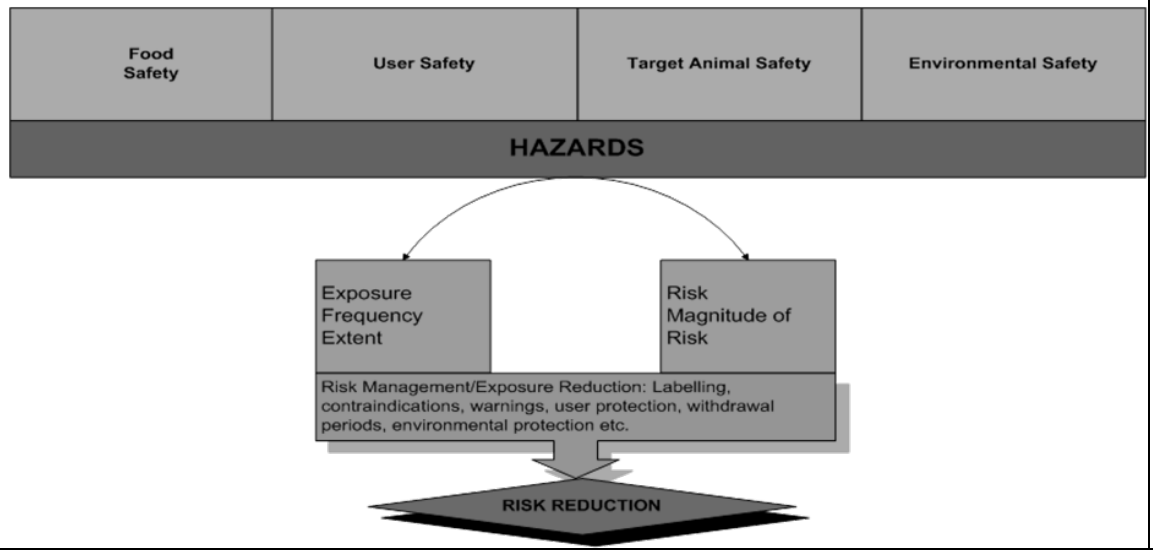
REVIEWER	COMMENT	RESPONSE
	might be different at one production stage vs. another.	
Reviewer #5	The document describes the objectives of the study and the background to the risk approach providing a clear rationale. As an introduction, the document describes how "the U.S. Food and Drug Administration (FDA) is developing a risk assessment to rank veterinary drugs based on human health risks posed by drug residues in milk and milk products. The purpose of the risk assessment is to assess the public health risk to consumers from veterinary drug residues in milk and milk products and develop a relative ranking of the drugs, based on the risk to the consumer. The risk assessment was initiated in response to a request from the National Conference on Interstate Milk Shipments (NCIMS), Appendix N Modification Committee. The NCIMS is a voluntary coalition of representatives from state and local regulatory agencies, which, together with the FDA, administers the national Grade "A" Milk Safety Program. The NCIMS Appendix N Modification Committee is re-evaluating the current drug residue testing requirements in the Pasteurized Milk Ordinance Appendix N. In December 2008, the Appendix N Modification Committee submitted a set of questions to the FDA and requested that FDA perform a risk assessment on potential risk associated with veterinary drug residues in the milk supply (NCIMS, 2008). This risk assessment will address the following questions, which were developed taking into consideration the original NCIMS request (NCIMS, 2008): I. What drugs are most likely to be administered to lactating dairy cattle in the U.S. and what is the magnitude of use? II. Which of these drugs are likely to result in drug residues present in bulk-tank milk? III. Of the drug residues found in bulk-tank milk, what is the fate of these residues during processing/manufacturing of various milk products (*i.e.*, in what milk products would these drug residues be found)? IV. Of the drugs used in dairy cattle and potentially present in bulk-tank milk, which have the potential for concentration in dairy products? V. What amount of drug residue is consumed in dairy products within each age group?" In my opinion, and having read through the dossier, the approach proposed in this paper for a risk assessment to rank veterinary drugs based on human health risks posed by drug residues in milk and milk products is adequate and sufficient to answer the risk assessment charge questions 1-15. The proposed approach considers hazard and reduced risk. There are areas that need to be clarified and these will be raised below.	No response is needed.

CHARGE QUESTION 2: Is the report clear in its description of the draft approach? If not, please identify areas that are unclear or could be more transparent.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Yes – The draft report is clear in both what the purpose of the risk assessment is and what questions it intends to ask in order to build the model. It talks about the type of model to be used, the criteria that will be considered, and how the total risk assessment score for each drug will be determined. It is not clear how each category will be weighted in the final model – however, perhaps this hasn't been determined yet.	The full technical report describes the weighting system.
Reviewer #2	Some detailed comments are provided at the end of this report and as marginal notes in the annotated text of the draft. Additional suggestions for greater clarity include the following: - In the interests of transparency, the master list of 450 drugs and the revised list of 140 drugs should be provided somewhere (Appendix? Online?), along with the reasons for exclusion from the final list.	The list of drugs and exclusion criteria are provided in the full technical report appendix as well as in the main text.
Reviewer #2	- I think it would be helpful to list the divisions or administrative groups within FDA that participated in developing the draft approach. Specifically, I wondered if the Center for Veterinary Medicine was involved, but didn't see reference to this in the document or during the teleconference. I think it would be helpful to show that there was cross-divisional communication and participation and that the relevant FDA resources contributed.	In conducting this risk assessment, we adapted the CODEX framework for risk analysis, including risk assessment, risk management and risk communication. The risk assessment team was multi-disciplinary with expertise from a variety of scientific disciplines and included subject matter experts from both CFSAN and CVM. The list of contributors is provided in the full technical report.
Reviewer #2	- Perhaps there should be mention that FDA has in their possession certain data that are not available for use in this risk ranking for confidentiality or legal reasons; this point was made during the teleconference. I assume this includes data provided by pharmaceutical companies in application dossiers (*e.g.*, pharmacokinetic / residue depletion studies), as well as pharmaceutical production or sales data. I wondered why some of this information was not used in the model and it was explained during the teleconference that only published data could be used.	The risk assessment report provides or cites all data and information used in the model, as per the transparency and reproducibility requirements of the Information Quality Act. Non-public confidential or proprietary data were not used.
Reviewer #2	- I think it would help to have some background on risk assessment of veterinary drug residues to help the reader understand why the MCDA approach was selected as the best among available options. For example, why was an additive model adopted?	The full technical report includes a literature review of published drug residue risk assessments and justification for the selected MCDA ranking approach. The technical report also justifies adoption of an additive linear model.
Reviewer #2	- If it is decided to call the proposed approach a "risk assessment" rather than a risk ranking (see point made above in response to question 1), I think it would be useful to discuss / acknowledge	We have revised the title of the risk assessment to avoid any confusion as to its

REVIEWER	COMMENT	RESPONSE
	how this approach compares in structure and outcome to the classical (*e.g.,* "Red Book") risk assessment approach.	nature. Further, the full technical report now contains a discussion of how this risk assessment compares to other types of risk assessment and why this structure determined to be the most appropriate for this assessment.
Reviewer #2	- Some drug preparations contain multiple drugs; was this factor a consideration in the risk ranking?	We did not include drugs containing multiple drugs to avoid double-counting of ingredients marketed as stand-alone and combination products. The full technical report addresses this.
Reviewer #2	- Perhaps point out that the report focuses exclusively on cow milk, and products made from sheep and goat milk are not considered	The full technical report includes clarification of the scope ("milk and milk products from cow's milk only.")
Reviewer #3	The description is reasonably clear.	No response needed.
Reviewer #3	For additional clarity, all assumptions that are made should be more clearly identified throughout the text (due to limited data, etc.).	We thank the reviewer for bringing this important issue to our attention. We have revised the text to identify and clarify our assumptions.
Reviewer #3	In the interests of transparency, additional information could be provided about "Step 1" (page 9), the identification of the drugs to be evaluated in this work.	The full technical report includes detailed information on how we identified drugs for evaluation in the text as well as in an appendix.
Reviewer #3	Also, Appendix I only shows information for 39 of the 54 drugs, rather than the full data set. The master list of drugs could be provided in an appendix, noting which drugs met which exclusion/inclusion criteria; additional information which served as the basis of the RA team expert opinion on these points would also be desirable.	The full technical report appendices include the master list of drugs, noting which drugs met exclusion/inclusion criteria. The text and appendices include additional information, which served as the basis of why the RA team excluded certain drugs.
Reviewer #3	For transparency, the "published literature" mentioned in B1 should be specifically cited. Without knowing what information has been considered and included, it is difficult to say what data might have been overlooked, what might not be appropriate to include, or what limitations such data sources would represent. Were there any specific criteria used to evaluate the evidence that was included in this sub-criterion (for example, published literature or test results based on the drug in the US milk supply rather than a non-domestic source, etc.)?	We revised criterion B1 in the full technical report using specific milk sampling data for transparency.
Reviewer #3	With respect to the list of dairy products, for clarity and transparency, it would be helpful if (instead of a list) the dairy products considered in this work were presented in a table, along with the "typical"	This information was added to the full technical report in the main text.

REVIEWER	COMMENT	RESPONSE
	composition that was assumed for this work (% water, % fat, % protein (casein/whey), etc.).	
Reviewer #3	Criterion E "Health effects from human exposure" seemed to be lacking in information or was not sufficiently clear in its description in the draft report.	We agree and have revised the description of this criterion [now criterion D (former criterion E)] in the full technical report.
Reviewer #4	The report is clearly written. It has an excellent and logical structure. Assumptions are clearly stated; Appendices are very useful in evaluating data and calculation algorithms.	No response needed.
Reviewer #4	More justification to the proposed MCDA approach should be given. In any model, there is a tradeoff between ease of use and precision and accuracy of the results. Specifically, a Bayesian Network model may be appropriate in similar situations, but possibly requires more data and more understanding of processes in the system. Nevertheless, a BN model may be appropriate for sub-components of the model and evaluation of individual criteria scores.	The full technical report (main text and appendix) includes a discussion of why we selected MCDA risk ranking as well as why we did not use a Bayesian Network model.
Reviewer #5	The approach provided is one that is equitable and transparent. The 5 phase procedure described in the dossier is as follows: • Phase I is to commission the risk assessment. In this phase, a risk management team provides the specific charge or questions to be answered and a risk assessment team of experts is formed. • Phase II involves efforts to collect and evaluate data. In this phase, a conceptual model is developed to identify the specific data needed. Targeted literature reviews are conducted and outreach to stakeholders (such as through issuance of a Federal Register Notice) are made. • Phase III involves development and validation of the model. Using information from Phase II, the model is developed and iteratively tested and validated. • The risk assessment is reviewed (both internal and external reviews) in Phase IV. In this phase, a risk assessment report is prepared, peer-reviewed and cleared by the Agency to issue in draft form for public comment. • A final report is developed that incorporates comments from the peer review and the public comments. • In Phase V, the final risk assessment is issued and posted on the FDA web site. • In assessing the document, Phases II and III are particularly relevant to the development of this model. The approach to developing this risk ranking model involves four steps: • Identification of the drugs/drug residues or metabolites to be evaluated. • Identification of the milk and milk products to be evaluated. • Definition of the model criteria, weighting, and scoring standards. • Execution of the risk model to determine risk scores for each drug-commodity pair for relevant age groups. The approach proposed is a very clear and a common approach to risk assessment allowing risk to be	We appreciate the comment and have developed a graphical overview of the approach which is similar in concept to the suggested model, to provide clarity about the procedure.

REVIEWER	COMMENT	RESPONSE
	based on evidence. A simple model of this approach is presented below: 	

CHARGE QUESTION 3: Does the draft approach consider the relevant criteria needed to rank the public health risks associated with drug residues in milk and milk products? If not, please describe the additional criteria that should be included.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Yes – I believe the criteria presented are correct. The key questions are what drugs are veterinarians using to treat dairy cattle and which of these drugs are most likely to show up in milk and other dairy products. The model also considers consumption of milk (and products) by consumers as well as the potential effects each of these drugs might have on public health. I cannot think of another major criterion that should be included in the model.	No response needed.
Reviewer #2	Yes, I think the relevant criteria are considered.	No response needed.
Reviewer #3	The draft approach considers five major criteria in ranking the public health risks associated with drug residues in milk/milk products. These five criteria seem to be reasonable; although additional data for each might be useful.	We appreciate the comment. The full technical report identifies additional data needs.
Reviewer #3	One significant limitation in the approach is that, by design, the model is considering the public health risks associated with drug residues in milk/milk products manufactured *from bulk-tank milk*. This assumption – that the prevalence of drug residues in bulk-tank milk is the same as in milk at the processor-level – is significant, and should not be overlooked.	As mentioned above, this model is semi-quantitative in nature and therefore does not consider drug residue concentrations. Therefore, this model is not able to distinguish between violative drug residue

REVIEWER	COMMENT	RESPONSE
	Without currently having sufficient data to estimate the prevalence of residues at the point of product manufacture (*i.e.*, very little finished product testing data or milk silo testing data), it would be difficult to add a criterion related to the pooling of milk and the concentration of drug residues.	concentrations and residue concentrations below tolerance, and the impact of co-mingling (on the farm and at the processor level) cannot be evaluated. While we agree with the reviewer that this is an important limitation of the model, it is a necessary result of our risk ranking approach. As discussed above, the issue of drug residue concentrations (on the farm and at the processor level) will have to be addressed at a subsequent step.
Reviewer #4	The approach is reasonable. Of course an ideal model (such as a full, vetted, detailed risk assessment model) would calculate frequency of tests to achieve risk reduction below threshold level based on solid toxicology models. In practice, with many drugs and limited data and analytic resources, such a model might not be possible to develop. At the other extreme, a simple model might just have a veterinary drug expert or panel of experts give a qualitative score to each drug and rank them accordingly. This proposed model lies in between, providing some knowledge structure – the data fields used, the algorithm for calculating risk scores, and the weighting parameters and scoring protocols used are intended to get more of the precision and transparency of the ideal model without creating onerous information requirements. To the extent that it does this in a way that approximates what an ideal model would do, the approach can lead to real practical improvement in risk prioritization and testing.	No response needed.
Reviewer #5	In the documentation provided, the procedure to reduce the chemicals for screening was described and had a clear rationale. In summary, the procedure was as follows: A preliminary list of over 450 drugs was developed using published information indicating any potential for administration of each drug to U.S. dairy cattle (FDA, 2013; USDA, 2008a and 2008b; Moore, 2010; USDA-FSIS, 2012; Wren, 2012; NMPF, 2011; Smith, 2005; Haskell, 2003). The master list of over 450 identified drugs was screened using the following exclusion criteria: • No evidence, or questionable information on drug administration to dairy cattle; • Information available indicating the drug will not enter the cow's milk; drug is hazardous for dairy cattle, and is therefore highly unlikely to be used in dairy cattle; or lack of availability that makes it highly unlikely the drug will be administered to dairy cattle in the U.S. After screening the drug list using the exclusion criteria, the revised drug list of 140 drugs was again screened using the following inclusion criteria:	The full technical report includes a discussion comparing our model to that of UK's as well as literature review of drug residue risk assessments and justification for selection of the MCDA ranking approach.

REVIEWER	COMMENT	RESPONSE
	• Drugs most likely to be administered to U.S. dairy cattle, and if likely to administered, also having the probability of entering the milk supply (based on the risk assessment team expert opinion). A final list of 54 drugs was screened and identified for the risk assessment (see Table 1). Different formulations of each drug were accounted for by including representative variations with different modes of administration. The list of 54 drugs was expanded to 104 formulations, which may result in different drug-residue levels in milk and milk products, for more in-depth evaluation in this risk assessment. The procedure to screen out certain drugs again was clearly defined within the dossier with clear guidance as to the exclusion criteria. The approach taken is a risk matrix approach which needs to consider a number of key elements. In the UK, a matrix ranking scheme is available, but there are often questions posed around hazard, potency and exposure. Careful consideration is required when assessing the types of hazard, exposure, route and the potential potency of the exposure and age range. In assessing a number of chemicals, their metabolites and class additivity, synergism and/or antagonicity should be considered; an area not fully considered in this Risk Assessment. Going forward this may need to be considered, but it is often the level of exposure that will define this. This is also the case when considering combinations of chemicals, which must also be considered. The success of the current scheme used in the UK for veterinary residues (VRC 2010 annual reports www.vmd.defra.gov.uk/vrc) is dependent on: • The assumptions, values and judgments inherent in the procedure. • The source, quality and quantity of the data provided. • Effective horizon scanning to monitor forthcoming trends and changes in veterinary medicine use. • Frequency of review and updating of the matrix ranking to encompass changes and new evidence. • The defined critical effect vs. the most sensitive effect if these differ. There are several advantages of using such a system:	

REVIEWER	COMMENT	RESPONSE
	• The objectives are clear and the process is relatively simple and transparent. • Diverse information and intelligence is used for making choices as to which substances to include in NSS surveillance plans and it is capable of dealing with a lack of information. • The use of Maximum Residue Levels (MRL) and/or Acceptable Daily Intake (ADI) to define potency as well as defining the nature of the hazard is considered to provide a reasonable characterization of hazard. • It is amenable to combining with other systems to develop a more holistic approach, e.g.: o environmental impact of veterinary medicines, o development of antimicrobial resistance, o assessing exposure to groups of substances, and o food security issues. The current proposed published Scheme is shown below:	

REVIEWER	COMMENT	RESPONSE

Criteria	Score						
	0	1	2	3	4	5	6
A: nature of hazard	No reported adverse effects	Reversible adverse pharmacological or microbiologi-cal effects	Reversible organ toxicity	Irritants. Evidence of allergic reactions in animals	Carcinogenic by mechanisms not relevant to humans, irreversible organ toxicity, reprotox, immunotox effects, non-genotoxic carcinogens	Includes irreversible neurotoxic effects, reproductive effects or evidence of mutagenicity	Evidence of carcinogenicity in humans. Carcinogenic by mechanisms relevant to humans
B: potency - based on Acceptable Daily Intake of substance (ADI)	> 10 µg/kg/bw/day	>0.1-10 µg/kg/bw/day	>0.001-0.1 µg/kg/bw/day	<0.001 µg/kg/bw/day			
C: % of whole population's diet that may come from treated animals	<2.5%	2.5 - 20%	20 - 50%	50 – 100%			
D: estimate of frequency of dosing/ percentage of animals treated	<2.5%	2.5 - 20%	20 - 50%	50 – 100%			
E: evidence for high exposure groups	Known that no high exposure groups	Unlikely to be high exposure groups	Likely to be high exposure groups	Known that high exposure groups or no data to be able to judge			
F: evidence for detectable residues		No evidence of detectable residues for a substance included in previous year's surveillance	Residues detected in previous year at concentrations below MRL/MRPL	Residues detected at MRL/MRPL or intelligence from other sources	Residues detected at ≥10x MRL/MRPL or no limit set or no previous testing		

Substance total score = (A+B) x (C+D+E) x F

CHARGE QUESTION 4: Does the draft approach consider all relevant sub-criteria, factors, or data sources needed to evaluate the likelihood and magnitude of drug administration to dairy cattle on U.S. dairy farms? If relevant sub-criteria, factors, or data sources are missing:

REVIEWER	COMMENT	RESPONSE
Reviewer #1	There are two primary problems with the data on frequency of drug use on dairy farms. One problem is that a good portion of the data is over 20 years old (Sundlof study) and there may be significant changes in which drugs are used now. Several new drugs have been approved for use in cattle over the past few years (*e.g.*, tulathromycin, tildipirosin, etc.) – however very few new drugs have been approved for use in dairy cattle. The 2nd problem is that both the USDA (NAHMS) and Sundlof studies rely on producer surveys to gather information on which drugs are used. Producers are aware of which drugs they are supposed to use (*i.e.*, which drugs are approved) versus which drugs are extra-label or illegal. Therefore, they certainly could omit the use of certain drugs (*e.g.*, fluoroquinolones) or under report the use of others (macrolides, florfenicol, etc.). However, I don't know of any other data sources that would provide more recent information.	We agree there are limitations to the available data on frequency of drug use on dairy farms. We commissioned an expert elicitation to assist with reducing the uncertainty. In the uncertainty section of the report, we compare model results obtained with and without the expert opinion data in the technical report.
Reviewer #1	I do think the prescription drug status and drug approval status sub-criteria are appropriate and I agree that over-the-counter drugs are slightly more likely to be used than prescription only drugs (particularly on small dairies).	No response required.
Reviewer #1	I would question how certain ranking scores are assigned to some of the drugs as listed in Table 9. For example – enrofloxacin would be AMDUCA prohibited for ELDU in dairy cattle – however, it is approved for use in dairy calves (for treatment of pneumonia). Therefore, it is legal for this drug to be on a dairy farm and experience says that makes it more likely to be used (illegally) when a producer has a really sick cow. I believe a drug like this or some of the sulfonamides (with the exception of sulfadimethoxine, all sulfas are banned in lactating cows but some are approved for treatment of diarrhea in calves) should get scores of at least 5 if not 7 (and not a score of 1). They are illegal in dairy cows – but are approved in calves or beef cattle – thus, they become "easier" for dairy producers to obtain and are more likely to be used as compared to chloramphenicol or nitrofurazone which are completely prohibited for use in all food animals and, in general, are not encountered on dairy farms. Remember that on many dairies (particularly the larger ones) you have treatment crews treating most of the cows. These crews are often Hispanic workers that have no understanding of what drugs are labeled for use in dairy and which ones are not. They generally have protocols they are supposed to follow – but when a cow does not respond to their first treatment – they often reach for something else. For example, ceftiofur is approved for the treatment of respiratory disease in adult dairy cattle, however experience may say that florfenicol (or potentially even enrofloxacin) works better. Drugs that are available to them (*i.e.*, approved for use in dairy calves or dairy cows) are much more likely to be used than those that aren't approved for use in cattle at all (*e.g.*, chloramphenicol).	The scoring standard for this sub-criterion has been revised to include drugs approved in cattle but not approved for lactating dairy cattle.

REVIEWER	COMMENT	RESPONSE
Reviewer #2	Yes, for the most part. I have a concern that there may be too much overlap between some of the sub-criteria. I think this could have the effect of adding more weight to certain factors given the additive nature of the model. See comments on this below.	We have carefully reviewed the structure of the risk assessment to avoid or minimize undue interactions among criteria. We tried to ascertain that the individual criteria and sub-criteria are value-independent to the maximum extent possible. The full technical report addresses this.
Reviewer #3	General comments on Criterion A: Regarding Criterion A, overall, this is an area where data seem to be lacking, or is at least somewhat limited, and many assumptions have been made to account for that lack of information.	We agreed there were limitations to the available data for criterion A. As a result, we commissioned an expert elicitation to assist with reducing the uncertainty.
Reviewer #4	This reviewer can not provide details on relevant factors and data sources since he has experience in risk assessment and decision analysis and does not have domain subject matter expertise. In general, models used to calculate scores at criteria level should be carefully reconsidered, as discussed in response to Question 10. For example, in this case there seem to be two main sub-criteria – magnitude and likelihood of use. Nevertheless, the sub-criteria listed are not structured accordingly. It looks like all of them are reflective of the likelihood of drug use and not magnitude. It may be an example of double counting and may require restructuring of the model.	We thank the reviewer for this comment. We have revised criterion A to address only likelihood of use.
Reviewer #5	In answer to this question, the same assumptions can be considered for Questions 4-7. In Questions 4 and 5 the assumptions are based on the MRL, Question 6 on the MRL and ADI, and Question 7 the ADI, but the following principle should be adapted across all areas. > The assumption made here is that a fair ranking-score of the likelihood of administration for each drug can be obtained from rough estimates, using the data from the surveys. Recognizing the limitations posed by the surveys, an attempt to reduce any impact from inaccurate estimates of drug usage is done by combining the scores from the data sets. A limitation in using these data for the purpose of the risk assessment is that the USDA survey did not collect data specifically on each of the 54 drugs to be considered in its evaluation. This can be a limiting step in the use of a matrix ranking model. In the dossier, a number of assumptions were made, and again if assumptions are made there needs to be a clear rationale, and where possible, experimental/epidemiological data to support these assumptions. The study considered the following: > Much of the drug administration information was reported at the drug class level, rather than for individual drugs. Also, the only data on anti-parasitic drug administration was for use in de-worming dairy cattle. To compensate for this,	We appreciate the comment. The full technical report includes a discussion of limitations of the available data and the modeling approach, and additional data needs.

REVIEWER	COMMENT	RESPONSE
	several assumptions will be made. We will assume that all anti-parasitic drugs are administered to dairy cattle as de-worming drugs. Also, we will assume that each drug within the reported drug class is utilized equally and any drugs in our evaluation not included in the drug classes in this report, will be evaluated with values reported in the "Other" (other drug classes) category in the 2007 NAHMS report (USDA,2008a; USDA 2008b). One limitation of this survey was the lack of specific data for drugs not included, which is a limitation to the work, such as newer drugs that did not exist at the time of this survey (*e.g.*, Tulathromicin, Enrofloxacin, Danofloxacin, Florfenicol, Meloxicam, Naproxen, Tilmicosin, Tildipirosin, Gamithromycin). For the purpose of this risk assessment, it will be assumed that these drugs had usage values equivalent to the reported usage for other drugs within the same drug group. It will also be assumed that all drug formulations for each drug have equivalent average-use scores, since more detailed information is not available. Another limitation is that the survey data from 1992 may not be reflective of today's dairy and veterinary management practices and disease-incidence patterns in U.S. dairy cattle. An assumption is made that by combining the scoring from this data set and the scoring from the USDA data set, we are reducing the impact of uncertainty in the data created by these limitations. A key limiting step in this and other Matrix Ranking systems is the limited data and the number of compounds reported at any one time. However, the robustness of the modelling proposed will ensure that there are rules to enable any drug residue to be put through the system. The set of "rules" used together with the documented assumptions will ensure that this system is scientifically sound and reproducible. One of the key questions that has arisen in reading the dossier is whether the critical hazard(s) driving the limits set for human health (*i.e.*, the hazard[s] occurring at the lowest dose levels) or for the most serious hazard the substance can cause regardless of dose considerations. This question arose in relation to substances with ADIs based on microbiological effects for which it could be suggested that higher ADIs were based solely on toxicological effects. It could alternatively be argued, however, that the score is based on the most sensitive effect, as not only is the ADI based on this effect, but so are the Maximum Residue Limits (MRLs) since they are usually derived from the lowest ADI. This is based on further evidence of detectable residues. The MRLs used should relate to the ADI used to score potency which should relate to the hazard score. If these scorings do not relate to each other, then the overall scorings will be inconsistent between chemicals, affecting the ranking.	

REVIEWER	COMMENT	RESPONSE
	The above illustrates there are other issues that must be considered when drugs are administered to food producing animals and the likelihood of any potential risk at the source and through the food chain.	

CHARGE QUESTION 4a: Which are missing?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	More current data on drug use in dairy cattle are missing – however, I believe this to be unavoidable.	The full technical report includes a discussion of limitations and data needs.
Reviewer #2	I wondered if it would be useful to include in one of the sub-criteria the route of administration of the drug. I suspect that intramammary infusions would be more likely to lead to residues in milk. During the teleconference it was mentioned that this may have been one of the considerations in screening drugs for the list to be evaluation with the model.	Although route of administration was not included as a separate sub-criterion, we did consider multiple routes of administration via the selection of the formulations included in this study. We selected drug formulations that were most likely to be used and to get into the milk (bulk-tank or bulk pickup tanker). The full technical report lists these drug formulations in the appendix. Each selected formulation was considered separately in criterion A and criterion B.
Reviewer #3	Assumptions from this data could significantly impact conclusions drawn from the outcome of the model – for example, because the USDA NAHMS data reported drug use *by class*, it was assumed that each drug within the reported drug class was utilized equally. However, this assumption is only reasonable if the regulatory status is the same for all the drugs within that class (for example, this would not be the case for sulfonamides).	We did not assume equal use of each drug in each class unless the drugs were actually used to treat the same conditions. In the model, overall use is cumulated over all conditions and considers marketing and regulatory status as well as inspection evidence of use on farms. The full technical report discusses this aspect.
Reviewer #3	It should be noted that the first charge question developed in response to the NCIMS request does specifically refer to *lactating* dairy cattle, not dairy cattle in general. Throughout the report, it is not clear if the sub-criteria and data for drug use are focused on dairy cows in general, or lactating dairy cows specifically (the language used throughout this section is inconsistent on this point). However, it is noted that the USDA data (A1.1) include drug use among lactating and non-lactating dairy cows, while the Sundlof data (A1.2) collected information specific to lactating animals. Sub-criterion A.3 seems to try to address the issue of using drugs in lactating dairy animals, but this is only with a slightly elevated ranking score (9 vs. 7); perhaps that gap should be widened to account for the limitations in the data.	We agree with the reviewer that the differentiation between lactating and non-lactating dairy cattle is important, albeit not completely clear-cut for the purposes of this risk ranking. The majority of drug residues are likely caused by administration of drugs to lactating dairy cattle. However, a small number of drug residues in the bulk tank milk may be due to drug administration to heifers or dry cows (*e.g.*, due to improperly administered dry-off medications) if the drug persists in the animal beyond the time at which the animal enters the (next) lactation. The full technical report has clarified how the data on lactating dairy cattle

REVIEWER	COMMENT	RESPONSE
		and administration to heifers (or dry cows) was included. The scores for drugs approved in lactating dairy cattle vs. drugs approved in cattle (not approved in lactating dairy cattle) was developed based on consultation with FDA veterinarians.
Reviewer #3	It is not clear what information is being included with sub-criterion A.4. The text describes the inspections that are required under the Grade "A" program, but I am not aware that those inspectors are collecting information about the drugs identified on the farm (as such, p 24-25, lines 388-400 should be deleted from the text). The requirements in the PMO for farm inspections are related to whether or not approved drugs are used, as well as if drugs are properly segregated for lactating/non-lactating animals.	A revised description of the data used as the basis of scoring for this sub-criterion is provided in the full technical report.
Reviewer #3	If the referenced on-farm inspections are related to compliance and enforcement activities that result from a tissue residue violation, then it would seem that an assumption is being made that the drugs stored on farms that have had a tissue residue violation are the same drugs that would be found on a typical dairy farm. This may be a limitation and an assumption of the available data, but it is not clear from the text 1) how this inspection information was obtained (*i.e.*, not likely through Grade 'A' farm inspections) and 2) if these data truly represent the population of dairy farms as a whole.	We have clarified how we obtained the inspection data and acknowledged that the population of dairy farms as a whole may not have been represented.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	No additional response required.

CHARGE QUESTION 4b: Where can information about these sub-criteria, factors, or data sources relevant to the likelihood and magnitude of drug usage on the farm to treat dairy cattle be found and what is the expected impact of not incorporating these sub-criteria, factors, or data sources?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I think the impact of both issues mentioned above on the final risk assessment model is minimal. While there certainly have been changes in the last 20 years – I believe the major drugs used in the dairy industry are similar to what they have been (penicillin, ceftiofur, oxytetracycline, sulfonamides). I believe some of the newer drugs (florfenicol, flunixin, and perhaps enrofloxacin) are used somewhat more commonly in dairy cows than they were 20 years ago – however at least these drugs are still included in the model and ultimately we should still be at least occasionally looking for them going forward. So although the data used in this portion of the model are old and may not have been reported 100% accurately by producers – I do believe these studies represent the most comprehensive datasets available right now and I believe the conclusions drawn by this portion of the assessment	We thank you for the comment.

REVIEWER	COMMENT	RESPONSE
	should be accurate.	
Reviewer #2	Data on the route of administration are available on drug product labels. Not incorporating this factor could reduce the ranking of drugs administered by intramammary infusion.	Although route of administration was not included as a separate sub-criterion, we did consider multiple routes of administration via the selection of the formulations included in this study. We selected drug formulations that were most likely to be used and to get into the milk (bulk-tank or bulk pickup tanker). The full technical report lists these drug formulations in the appendix. Each selected formulation was considered separately in criterion A and criterion B.
Reviewer #3	See above. There are definite limitations to the available data on magnitude of drug use in lactating dairy cattle. Additional research, survey work, or industry data in this area would be helpful. Where data are limited and assumptions are made, these should be more clearly identified in the text.	We agree. An expert elicitation was conducted to enhance the data; additional research in this area would provide more robust data. With regard to the magnitude of use, we removed it from the model. The full technical report includes a discussion of data limitations.
Reviewer #4	[No comments were provided by the reviewer.]	No response needed.
Reviewer #5	See above.	No additional response needed.

CHARGE QUESTION 5: Does the draft approach consider all relevant sub-criteria, factors, or data sources needed to determine the likelihood of drug presence in bulk-tank milk? If relevant factors are missing:

REVIEWER	COMMENT	RESPONSE
Reviewer #1	This is one criterion where I'm not sure I agree with the design. For many of the drugs on the list of 54 – there is not going to be much data available on whether the drug (or metabolite) enters the milk and how long it persists. In Table 12 – I have concerns about the "quantity of published evidence" that assigns a higher score to drugs that have greater than 5 published papers where it has been found in the milk supply or there is experimental evidence that the drug enters the milk. The available literature is strongly biased towards older drugs and those that are actually approved in dairy cattle (since the pharmaceutical companies have largely funded this research). I will give a couple of examples: 1) Tulathromycin – This is one of the newer macrolides that is not approved for use in dairy cattle and I don't know of a published paper that says residues can be found in milk. However, the reality is that this drug (along with the other new macrolides like gamithromycin and tildipirosin) is very fat soluble and has a very prolonged elimination in milk. The pharmaceutical companies may share data on this eventually. For	We agree and have re-defined this sub-criterion in the model.

REVIEWER	COMMENT	RESPONSE
	example, I know Zoetis has data that tulathromycin would be in milk for multiple months after a single administration (I was told none of these macrolides would ever be approved in dairy cattle because the milk withdrawal time would be far too long). 2) NSAIDs are on your list and have been very poorly studied in milk. Drugs like flunixin and meloxicam can be found in milk – but testing has not been done at all to date (perhaps very recently for flunixin) as part of our normal federal milk residue testing program because of the lack of rapid assays. There is also very little published literature on many of these drugs in terms of whether they enter the milk – but a recent publication indicates that flunixin residues can be found in comingled milk (Food Additives and Contaminants Part A 2013;30:1513-1516). Other newer drugs (*e.g.*, enrofloxacin, danofloxacin, etc.) are not going to have much published literature citing the ability of the drug to enter the milk because they are prohibited in dairy cattle and thus aren't studied. I think the current ranking score system for sub-criteria B1 will be bias towards older drugs (penicillin, ceftiofur, sulfas) that have been around for years. They have much more research done on them and historically have been tested for much more commonly during residue testing.	
Reviewer #2	I was a little puzzled by some of the explanation given for this criterion and sub-criteria (see detailed points below). I think it logical to include data from studies that tested bulk milk for drug residues, particularly monitoring studies using a wide range of tests. I suspect that available published studies of this type are included but it is hard to tell from the explanation provided. It would help if this information was available for review in at least summary (tabular) form with references. As indicated below, I really don't get the point of B2 "concern score."	Criterion B includes milk-testing data from NMDRD and FDA. The description of B2 has been clarified.
Reviewer #3	General comments on Criterion B: Sub-criterion B1 mentions evidence of the drug in cow's milk to include "published laboratory tests finding the drug" – does this refer to finding the drug at *any* level in milk, or only at levels *in excess* of the US safe/tolerance?	We have re-defined this sub-criterion in the model. It is now solely based on milk sampling data.
Reviewer #3	Sub-criterion B1 includes four forms of published evidence that were considered. Factor B1α refers to a positive milk sample test on the farm or "during processing." As subsequent parts of the risk ranking model take into account the effects of processing (and potential concentration of the drug residue during specific unit operations), positive results from processed products should not be included here; the data should be limited to raw milk either on-farm or *prior to* manufacture. (The bulleted list on p. 29, lines 469-470 does specify "raw cow's milk," but Table 12 includes milk samples from the farm or "during processing," which could imply that data on manufactured products are also included.)	We have re-defined this sub-criterion in the model. It is now solely based on milk sampling data.
Reviewer #3	Regarding sub-criterion B3: The rationale for the relationship between milk-discard time and identification of a drug residue in bulk tank milk was not clear (p. 31, lines 508-512). It	We revised this sub-criterion to include the milk discard time, which reflects the relative partitioning of drug

REVIEWER	COMMENT	RESPONSE
	would seem that the factor "a drug without an official milk-discard time" may be related to the regulatory status of the drug. Does sub-criterion B3 add discriminatory power to the model?	residue into milk, the actual clearance of the drug from the milk, and the hazard as reflected by the official tolerance.
Reviewer #4	This reviewer can not provide details on relevant factors and data sources since he has experience in risk assessment and decision analysis and does not have domain subject matter expertise. In general, models used to calculate scores at the criteria level should be carefully reconsidered, as discussed in response to Question 10. In general, is the likelihood of the drug's presence in bulk-tank milk dependant on usage on the farm assessed in Criterion A? The logic model and quantification of B1 and B2 seem to be arbitrary; assumption of score linearity and additivity requires justification.	A detailed justification of the linear additive model is provided in the full technical report. As discussed above, criteria A and B are measuring different and distinct factors.
Reviewer #5	In my opinion, I believe that a consistent approach has been taken, but again, clarity must be given as to the assumptions made in the points raised above.	The full technical report includes a discussion regarding assumptions and data limitations.

CHARGE QUESTION 5a: Which are missing?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	What is missing is some measure of how long drugs stay in milk. I believe there should be a rank (high number) given to drugs that reach very high concentrations in milk and persist for an extended period of time (*e.g.*, tilmicosin, florfenicol) versus drugs that are detected for a moderate period of time (*e.g.*, oxytetracycline) versus those drugs that don't reach very high concentrations in milk and are typically cleared very quickly (within days) such as flunixin or ceftiofur.	We agree and have revised sub-criterion B2 include the milk discard time, which reflects the relative partitioning of drug residue into milk, the actual clearance of the drug from the milk, and the hazard as reflected by the official tolerance.
Reviewer #1	I also think the current model puts too much emphasis on "number of publications" versus asking how likely are each of these drugs to enter and persist in the milk if they are used. We should consider the pharmacokinetics of drugs in dairy cattle – not just "yes" or "no" the drug can be found in milk.	We agree and have revised sub-criterion B1. In addition, we have incorporated aspects of pharmacokinetics in sub-criterion B2 as discussed in the response above.
Reviewer #2	None that I am aware of.	No response required**.**
Reviewer #3	Criterion B includes three sub-criteria: evidence of the drug in milk, a concern score, and a milk discard time. I agree that data from tissue residue violations should not be a factor in determining the likelihood of the drug presence in bulk-tank milk without data to justify a correlation and including it.	We have re-defined this criterion. We agree with the reviewer that tissue residue violations should not be included.

REVIEWER	COMMENT	RESPONSE
Reviewer #3	I agree that "evidence of drug in cow's milk" should be included as a sub-criterion in the model. However, it should be recognized that a limitation of including this is that the evidence will only exist for drugs which are being tested for (whether for ease of analysis, health concern, etc.). Drugs that are tested for more frequently are likely to be biased; there is likely to be a greater number of test results reporting the drug.	We agree. We have included sub-criterion B2 (likelihood and consequence of misuse) to include in the analysis drugs that may not have been sampled and detected.
Reviewer #3	In the Summary for Criterion B (p. 32, lines 525-529), it mentions a ranking score for each drug formulation on small, medium and large dairy farms. There was no information presented in Criterion B that would suggest data on drugs in bulk-tank milk were parsed by farm size (unless this is captured with the farm inspection data). I question whether or not such a distinction is warranted, but if the intent is to differentiate by farm size in some manner, this should be made more clear.	We carefully considered whether to make a distinction among farm types, with herd size representing one natural distinguishing factor. After analyzing the NAHMS data for small, medium, and large farms separately, we did not see any differences that warrant separating them out. So, small, medium, and large farms were not considered separately in the model.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	No additional response required.

CHARGE QUESTION 5b: Where can information about these sub-criteria, factors, or data sources relevant to the probability of the drug entering the raw bulk-milk tank be found; and what is the expected impact of not incorporating these factors, sub-criteria, or data sources?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I don't know whether data exists on the persistence of drug in milk for each of the 54 drugs listed – but there are data for many of them in the literature. The book "Tabulation of FARAD Comparative and Veterinary Pharmacokinetic Data" by Craigmill, Riviere and Webb is a good source for older data. Literature searches of each drug should turn up more recent papers.	We have updated the model to include data on milk discard time (which reflects the relative partitioning of drug residue into milk, the actual clearance of the drug from the milk, and the hazard as reflected by the official tolerance) and have included data from the cited reference. We thank you for the reference.
Reviewer #2	Not applicable.	No response required**.**
Reviewer #3	There is a significant amount of industry testing for non-beta-lactam drugs that is being performed that would be helpful in this area and may be obtained through a *Federal Register* notice.	We agree. The availability of the risk assessment and request for public comment will be published in the Federal Register, as is standard practice.
Reviewer #3	Was the NMDRD considered as a data source for sub-criterion B1? If so, what assumptions were made with respect to individual drugs (as the NMDRD reports information for drug families/classes)?	Yes, NMDRD is a data source for sub-criterion B1. The full technical report describes the assumptions made with respect to individual drugs.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	See response above.

CHARGE QUESTION 6: Does the draft approach consider all relevant sub-criteria, factors, or data sources needed to predict the impact of dairy processing on the relative concentration of drug residues in the final dairy product? If relevant sub-criteria, factors, or data sources are missing:

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I think so – I am not an expert in dairy processing or the manufacture of dairy products but I believe the model is looking at the correct factors (change in product composition and impact of heat degradation or water removal) to build this portion of the model. What is presented seems sound and I cannot think of anything that the model is missing.	No response required.
Reviewer #2	Yes, I think the relevant sub-criteria are included.	No response required.
Reviewer #3	General comments on Criterion C: The product-processing score takes into account the effect of heating and water removal. Is there sufficient data available to be able to make decisions for all 54 drugs in the model? Greater transparency about what is based on experimental/published data and what was inferred is needed in this section.	The full technical report provides the available data, information, and assumptions used to inform scoring including for example: heat inactivation and partitioning behavior.
Reviewer #3	As mentioned above, the approach does not consider the impact of commingling of bulk-tank milk at the processing level, which will also impact the relative concentration of drug residues in the final dairy product.	The risk assessment is a multicriteria-based ranking model to assist with re-evaluating which animal drug residues should be included in milk sampling programs. Commingling at the processing level would be a critical factor to consider if the objective was to estimate absolute values, as in the case of a quantitative risk assessment.
Reviewer #4	This reviewer can not provide details on relevant factors and data sources since he has experience in risk assessment and decision analysis and does not have domain subject matter expertise. In general, models used to calculate scores at criteria level should be carefully reconsidered, as discussed in response to Question 10. .	No response required**.** See Question 10**.**
Reviewer #4	The quantification of sub-criteria and model structure is unclear.	We have added more information to clarify the quantification of the sub-criteria and the model structure for criterion C. The full technical report provides additional details on scoring and calculation based on the model structure.
Reviewer #4	Product composition is listed twice as both main criteria and sub-criteria.	We thank the reviewer for bringing this to our attention. The full technical report describes the criterion and distinguishes between the criterion and subcriterion.
Reviewer #4	Scoring strategy is unclear and should be justified better.	The full technical report now clearly describes the scoring strategy.
Reviewer #5	In my opinion, I believe a consistent approach has been taken, but again the points raised in Question 4, above, are relevant to this question. All assumptions made must, where possible, be evidence based, or if this is not permissible, the use of the worst case scenario provided.	See response above Question #4.

CHARGE QUESTION 6a: Which are missing, and where can this information relevant to the dairy products and drug residues considered in the risk assessment be found.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	See above.	See response above.
Reviewer #2	Not applicable.	No response required.
Reviewer #3	There are definite limitations to the available data on the impact of unit operations and product processing. There were five processing steps that were described in the model (heating, culturing, cheese aging, freezing, condensing). However, it is not clear how these five processing steps (beyond heating or removal of water) were incorporated into the model. Aside from product selection (identifying products that involve these unit operations in their manufacture), how is the impact of each of these processing steps (*e.g.*, freezing) incorporated into the model? The product processing component assumes (p. 39, lines 651-652) that drug concentration decreases because of heating and increases through selective water removal. I recognize this is a simplification because of a lack of available scientific data. Drug concentration would also change as a result of filtration or removal of whey, and could also be impacted by acidification (either through culturing or direct addition), through salt addition, and through freezing. Additional research or industry data in this area would be helpful, although I am not sure it currently exists. This is a clearly identified research need.	We have revised the text of the report to better clarify the types of processing steps considered and evaluated in the model. We agree that additional data on the impact of processing on drug residue concentrations will enhance the accuracy of the model and that this is a research need.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	See response above.

CHARGE QUESTION 6b: What is the expected impact of not considering these sub-criteria, factors, or data sources?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	See above.	See response above.
Reviewer #2	Not applicable.	No response required.
Reviewer #3	See above. Also, additional data are needed on protein binding. Understanding how drugs partition between caseins and whey proteins will be necessary to determine the full impact of processing on drug residues present in milk. An assumption was made (p. 36, lines 598-599) that the concentration of water-soluble drugs only decreases moderately as fat content increases. This assumption is true as long as fat is the only component that is changing (for example, this would not be the case with cheese, where the casein/whey protein ratio is also changing). This is especially important for products that incorporate whey proteins as ingredients or in products (*e.g.*, cheese) where whey is	We agree. In order to include WPC to the products considered one would need data on the binding characteristics of the selected drugs to whey and casein proteins. These data are not currently available. FDA has commissioned a study to fill this data gap. We welcome additional data to inform the modeling of animal drug binding to casein and whey proteins.

REVIEWER	COMMENT	RESPONSE
	removed. The actual concentration of the drug in the final product could be significantly different than otherwise expected.	
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	See response above.

CHARGE QUESTION 7: Does the draft approach consider all relevant sub-criteria, factors, data sources and scoring standards needed to predict the impact of consumption of dairy products on drug residue intake? If relevant sub-criteria, factors or data sources are missing:

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I think this criterion is also designed correctly. I cannot think of any other factors that should be included here. The data used is very current and should be very accurate. This portion of the risk assessment appears to be very well done.	No response required.
Reviewer #2	Yes, I think the relevant sub-criteria are included.	No response required
Reviewer #3	General comments on Criterion D: Table 26 includes "raw milk" as a dairy food for consideration. Is this meant to be a part of the model? I am not aware of a good source of data that would allow for estimating the amount of raw milk that is consumed by various population subgroups (and would expect it to be much smaller than pasteurized product).	We are not considering the consumption of raw milk in this model. The term "raw milk" in table 26 was meant to describe the bulk-tank milk entering the processing plant. This term has been deleted to eliminate confusion about the intent of that table.
Reviewer #4	This reviewer can not provide details on relevant factors and data sources since he has experience in risk assessment and decision analysis and does not have domain subject matter expertise. In general, models used to calculate scores at criteria level should be carefully reconsidered, as discussed in response to Question 10.	See response to Question 10.
Reviewer #5	In my opinion, I believe a consistent approach has been taken, but again the points raised in Question 4, above, are relevant to this question. All assumptions made must, where possible, be evidence based, or if this is not permissible, the use of the worst case scenario provided.	See response to Question 4.

CHARGE QUESTION 7a: Which are missing, and where can this information relevant to the consumption of dairy products be found.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	See above.	See response above.
Reviewer #2	Not applicable.	No response required.
Reviewer #3	It may be worth considering the impact of gender on dairy consumption, using the NHANES data. Dairy product intake, and fluid milk consumption specifically, is significantly different between males and females, especially during the teen years. Perhaps the model could be run including	There were some gender differences in amounts consumed of certain milk and milk products in certain age groups; however, we did not

REVIEWER	COMMENT	RESPONSE
	gender as a factor in consumption to see if it affects the final conclusions.	incorporate such differences in our analysis because we evaluated lifetime average daily intakes of the 12 selected milk and milk products. This is described in the full technical report.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	See response above.

CHARGE QUESTION 7b: What is the expected impact of not considering these sub-criteria, factors, or data sources?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	See above.	See response above.
Reviewer #2	Not applicable.	No response required.
Reviewer #3	See above.	See response above.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	See response above.

CHARGE QUESTION 8: Does the draft approach consider all relevant sub-criteria, factors or data sources needed to predict the public health effects on consumers? If relevant sub-criteria, factors or data sources are missing:

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Yes – the hazard values used are appropriate (taken from ADIs where possible or toxicology data when ADIs are not available). The adjunct carcinogenicity score is also appropriate – however, one could question whether it is necessary. To truly be a carcinogen – the drug would likely need to be ingested at low concentrations over an extended period of time. Even if a residue were present in milk – it is unlikely that it would be present on a consistent basis. So human exposure would (in most cases) be limited to a few single exposures over an extended period of time. However, because of the seriousness devoted to carcinogens – I believe the inclusion of this score is still appropriate.	No response required**.**
Reviewer #2	This criterion essentially uses the ADI and whether or not the drug is carcinogenic. While these two drug attributes don't predict public health risk by themselves, they do provide information applicable to the hazard identification / hazard characterization steps of risk assessment.	No response required**.**
Reviewer #2	Regarding E1, I don't understand why the term "hazard-value" is introduced here when ADI would do. For drugs without an ADI, it can be stated that the ADI was extrapolated from other related drugs. Re E2 - I was under the impression that carcinogens, at least genotoxic ones, were not approved for use in food animals. If this is the case, does this sub-criterion add any information not already captured in criteria A or B (*e.g.*, ELDU or drugs prohibited for use in food animals)? I suppose there is the possibility that some carcinogenic drugs, such as furizolidone, will be used in dairy cows and there should probably be some recognition of that.	An ADI has not established for many of these drugs. Therefore, we used the term "hazard value" to describe the information we use to score each drug for this criterion. E2 has been deleted as a sub-criterion.

REVIEWER	COMMENT	RESPONSE
Reviewer #2	The toxicity, including carcinogenicity, of the veterinary drugs of interest should already have been classified by credible bodies (*e.g.*, JECFA) and publicly available, and should be considered for inclusion.	We agree and have considered data from other international bodies.
Reviewer #3	General comment on Criterion E: It was difficult to follow this section, and the details and rationale should be clarified. Additionally, it seems that some text is missing (p. 47, after line 791).	The full technical report provides additional details and the rationale of this section has been clarified.
Reviewer #4	This reviewer can not provide details on relevant factors and data sources since he has experience in risk assessment and decision analysis and does not have domain subject matter expertise. In general, models used to calculate scores at criteria level should be carefully reconsidered, as discussed in response to Question 10.	See response to Question 10.
Reviewer #5	The answer to Questions 6-8 should also ensure that the following criteria are considered at each step and the relevant score is used to determine the overall risk based on: • Exposure • Route • Class effect • Excipients present • Additivity, synergism, and antagonicity of the compounds • The effect noted is microbiological or toxicological • Age group exposed • If all the toxicological critical end points have been considered for human health (*i.e.*, from pre-conception [immunological status] to adult hood) • Should a greater uncertainty factor be included to counter the lack of toxicological information? • Should classes of compounds be considered to ensure no one class is missed? Again, to do this with little or no evidence may require additional factors in the Matrix ranking. • Polypharmacy (*i.e.*, lack of withdrawal periods, either intentional or accidental)	We agree with the reviewer that the factors are important and many of these factors have been incorporated into the multicriteria-based ranking model. Limitations in the available data limit which of these factors we can explicitly incorporate in the risk assessment.

CHARGE QUESTION 8a: Which are missing, and where can this information relevant to the consumption of dairy products be found; and

REVIEWER	COMMENT	RESPONSE
Reviewer #1	See above.	See response above.
Reviewer #2	I can't think of any other sub-criteria that would apply under this criterion.	No response required.
Reviewer #3	See above.	See response above.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.

REVIEWER	COMMENT	RESPONSE
Reviewer #5	See above.	See response above.

CHARGE QUESTION 8b: What is the expected impact of not considering these sub-criteria, factors or data sources?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	See above.	See response above.
Reviewer #2	Not applicable.	No response required.
Reviewer #3	See above.	See response above.
Reviewer #4	[No comments were provided by the reviewer.]	No response required.
Reviewer #5	See above.	See response above.

CHARGE QUESTION 9: Is the proposed scoring for all criteria, sub-criteria, and factors appropriate? If not, what changes would you recommend and why?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	If you look at the "Overview of Risk Ranking" on page 50 – it shows each of the five criteria that will have a score that is multiplied by a weight. What is not clear is what the weight multiplication factors will be. I believe the correct questions have been asked – however, it's not clear which you consider the most important in determining the total risk score for the drug (*i.e.*, which criterion is weighted higher than the others). In my opinion – health effects from human exposure has to be given considerable weight in this model. While we want to avoid all drug residues in milk if possible – we need to be particularly diligent about avoiding residues that would have a high likelihood of causing adverse effects in humans. I would weight criterion E fairly high followed by criteria A and B (placing emphasis on drugs that are frequently being used and those that are most likely to wind up in milk).	Thank you for this suggestion. We have used an expert panel to determine criterion weights.
Reviewer #2	Yes, I think the scoring of criteria and sub-criteria is reasonable. There is some subjectivity but I think that is understandable and reasonable given the extent of uncertainty that exists.	No response required**.**
Reviewer #3	Sub-criterion A.3 – drugs approved in lactating dairy cattle should be scored significantly higher than all other drugs (would recommend changing the scoring so that the first two categories are both a '1'). Sub-criteria (A2, A3, A5) seem to be overlapping (*i.e.*, status of prescription/approval/ELDU). Without seeing a data table comparing the 54 drugs included in this work, it may be reasonable to ask if each these three sub-criteria provide new information to the model and differentiate among the drugs in the model, or if they are non-discriminatory. While any number of sub-criteria and factors could be added to the model, if they do not differentiate among the drugs in a meaningful way, then their utility is limited. Perhaps the discriminatory power of all sub-criteria and factors could be determined.	We thank the reviewer for these comments. Sub-criteria, under criterion A, have been revised. The full technical report describes the sub-criteria and the data used in scoring. Expert elicitation was used to account for limitations in data, rather than widening gap in scoring in A3. We agree with the reviewer. Data to calculate the percentage of farm inspection was not

REVIEWER	COMMENT	RESPONSE
	Sub-criterion A.3 seems to try to address the issue of using drugs in lactating dairy animals, but this is only with a slightly elevated ranking score (9 vs. 7). Because there is information lacking about whether drugs were used in all dairy cows or only lactating dairy cows, perhaps this gap in scoring should be widened to compensate for the lack of distinction. Sub-criterion A.4: It is difficult to evaluate the scoring assigned to each of these categories (*i.e.*, "drugs identified in 11-40 inspections", etc.) without knowing the total number of inspections that comprised the data set. This should be clarified (*i.e.,* "> 80 out of 90 inspections" or expressed as a percentage). Expressing this data as a percentage will also allow the information to be easily updated as new inspection data become available over time. Sub-criterion B1 (Evidence of drug in cow's milk): Similar to A4, it is difficult to evaluate the ranking scores assigned based on evidence without knowing what the data are that were considered. For example, it is not clear what published data were included for consideration nor is it clear what a "test result" means (*e.g.,* Is the NMDRD included here? It represents a massive amount of testing data, but it is not clear how these data are evaluated for this sub-criterion. Is the source itself a "single test result"? Is each year a single test result? Or, for example, 777 positive beta-lactam tests were reported to the NMDRD in 2012 – would this represent 777 test results, or is the entire 2012 report a single test result?). Sub-criterion B1 evaluates published evidence and combines 4 factors (α,β,γ,δ) to calculate a ranking score. Of the four factors, only B1α represents (Criterion B) the "*likelihood* of the drug's presence in bulk-tank milk;" the others do not represent the *probability* of a drug occurring in milk, rather, they represent the *possibility* of the drug being identified. For that reason, I would re-distribute weights among these four factors such that α is given greater significance. Sub-criterion B2 represents a "Concern score," for which the ranking scores are assigned in direct contradiction to A3. Because the major criterion (B) is the "*likelihood* of the drug's presence" in milk, and not the *concern* for the drug, I would recommend omitting B2. Approval status was incorporated into Criterion A, and negative effects on health are considered in Criterion E.	available at the time. However, we have now incorporated a new data set that uses the percentage of farm inspections.
Reviewer #4	The proposed approach covers many factors which can be measured or for which data can be obtained. In general, the methodology translates actual measurements or expert estimates into scores on a 1-9 scale. In many cases this approach is justifiable, but overall it would be better to keep actual units in which individual sub-criteria are measured. As discussed below, it is important to check for criteria independence and discuss linearity. Nevertheless, it seems that in many cases, testing for drugs that receive the worst scores would indeed reduce risk more	We generally agree with the benefits of using quantitative data where appropriate. The justification for the linear additive model and discussion of factors such as criterion scales, scores, and potential interdependence among criteria are discussed in the revised report.

REVIEWER	COMMENT	RESPONSE
	than testing for drugs that receive better scores.	
Reviewer #5	With regard to the criteria used, the focus should be on the "critical" effects of each substance unless these occurred only at levels that were very unlikely to be found as residues in food. As a result, the following rules could be considered: • The hazard scored must relate to the potency score given. • For any substance, in most cases the scores for hazard (A) and potency (B) will be based on the sum of the toxicological effect (A) and related potency (B) which give the highest resulting score. Generally, this means that the score will be based on the highest score for hazard and this is likely to be the most critical effect of the substance. • In the first instance, the hazard score is based on the most sensitive no-observed-adverse-effect-level (NOAEL: often a body-weight effect). However, if a more critical adverse effect (*e.g.*, adverse reproductive effects or neurotoxic effect) occurs within 5-fold of the most sensitive NOAEL, the more critical effect will be used to derive the hazard score. • When the critical effects are only applicable at levels too high to be relevant (cut-off point nominally chosen as more than 5-fold of the most sensitive NOAEL) then the more sensitive effect will be used for arriving at a score. These rules may not cover all eventualities. Where scientific judgment is applied, the reasoning for any deviations from the application of these rules will be supplied.	We thank you for your comment. The hazard score assigned for each drug considers the potency and toxicological effect.

CHARGE QUESTION 10: Is the algorithm that combines criteria scores and weights into an overall score appropriate? If not, please provide suggested improvements.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	Yes – it's appropriate but we are not told what the weighting factors will be. However, the development of the overall risk ranking model taking scores from each criterion and forming a weighted score for each category that filters into the final risk score for each drug is fine.	No response required.
Reviewer #2	Yes.	No response required.
Reviewer #3	[No comments were provided by the reviewer.]	No response required.
Reviewer #4	At the top level, criteria A-E seem to approximate elements of a pathway/exposure model. The expected loss associated with such a pathway is the product of its elements. The relative magnitude of risk at each stage of the pathway corresponds to the log of a probability, that is, the risk stage i for drug j is proportional to 10^(xij-k). Here, k is an arbitrary constant associated with the scoring scale (1-9). A risk score that is the simple sum of the element scores thus corresponds to the logarithm of the expected loss; weights other than 1 imply that there are	Thank you for your comment.

REVIEWER	COMMENT	RESPONSE
	increasing or diminishing impacts of scale in terms of the associated factors, *e.g.*, dose-response. Equal weighting seems best. At lower levels in this approach, the integration of sub-criteria into a risk score for each top level criterion is not as simple. In some cases, it is not really directed toward the kind of multiplicative risk factors that are implicitly assumed at the top level. It seems the approach could be improved with further effort if further fidelity is required by – for these sub-criteria – considering what formal risk analytic models might actually look like and aligning the approximations to these models. In some cases, the sub-criteria sort of reflect probabilities, but not exactly (*e.g.*, number of instances reported in the literature). In some cases, the sub-criteria scores seem to represent conditioning variables (*e.g.*, approval status) that would be used in calculating conditional probabilities higher in the hierarchy, and in this case, simple addition might not be appropriate depending on what interactions apply. In theory, Bayes' nets might be used at the lower levels, but this would certainly complicate the use of the model. Simpler might be, defining variables at the bottom level to be as probabilistically independent as possible, and then considering the translation to the 1-9 scales so that the weighted sums of the sub-criteria are approximately proportional to the multiplicative risk factors they are producing at the top level. Again, with the use of an additive scoring model representing essentially multiplicative factors, weights can be thought of in terms of whether there are increasing or decreasing scale effects.	
Reviewer #5	This approach seems sensible and if the rules above could be incorporated then the Matrix Ranking will provide sufficient data to allow the public to be content that a spectrum of substances can be considered for risk using this table and the criteria set out.	No response required**.**

CHARGE QUESTION 11: What weighting is most useful for an accurate drug ranking, and what weighting should be avoided to prevent an inaccurate ranking?

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I'm not sure I have a great answer for this question. I don't have a strong background in epidemiology or the development of risk assessment models. Not sure I know what weighting system would be the most appropriate in building the overall drug risk score.	No response required**.**
Reviewer #2	This is a difficult question. I suggest running the model using equal weights, at least initially. This is probably a reasonable default in the absence of data to support different weights. Intuitively, I think criteria A and E are most important, but this is a highly subjective assessment that is difficult to justify. It might be possible to do a Delphi – type exercise with experts to arrive at a consensus on weighting but I am not sure. Perhaps the literature on MCDA provides guidance.	We appreciate the comment and we have run the model initially, as suggested. We obtained input on the weighting through an expert elicitation.
Reviewer #3	Determining the weighting of the various criteria is key. Regardless of how the weights are	We appreciate the comment. We obtained input

REVIEWER	COMMENT	RESPONSE
	assigned, the overall goal is to balance likelihood of risks with the severity of the consequences. One recommendation would be – in considering that four criteria (A, B, C, and D) all represent the *likelihood* of consumer exposure to the various drugs through consumption of dairy products, while a single criterion (E) represents the *risk* – to give a slightly higher weight to E.	on the weighting through an expert elicitation.
Reviewer #4	Decision analysis centers on the idea that by quantifying the preferences for each *criteria* rather than a specific drug, a more objective and systematic prioritization can be reached. By defining which criteria are most important and integrating these weights with scores representative of drug performance by each criteria, an integrated risk score can be quantified. There are different ways to elicit weights and to assign scores, some of which are tailored to specific MCDA methodologies. Three main methods of weighting include: ranking, pairwise comparison, and swing-weighting. Details of these methods are provided in multiple publications (*e.g.*, Linkov and Moberg, 2012). Swing-weighting and ranking are probably most appropriate for the problem at hand. Pairwise comparison is closely associated with AHP; AHP is being widely criticized in the field of decision analysis and its application in regulatory settings could be controversial.	We thank the reviewer for reiterating the importance of weighting. We have included a description of different weighting methods and a justification for the weighting methods used in this risk assessment in the revised report.
Reviewer #5	In reviewing this document together with published data from the VRC 2010, my view is that this approach is on a case by case basis, but that criteria are in place so that classes of drugs can be assessed and where there is any doubt additional uncertainty factors introduced. Again, this must be clear and equitable within the Ranking criteria**.**	The technical report now has a separate section that addresses the uncertainties in data and model structure

CHARGE QUESTION 12: How would you recommend weighting the proposed criteria, sub-criteria, and factors in this risk ranking? Please provide justification or rationale/reasoning for your assignments.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	As mentioned above – I think criterion E (health effects from human exposure) should be weighted the highest. I don't think too many of the drugs listed should pose extremely serious risks from consumption of residues – but if there are a few that receive high "health concern scores" (*e.g.*, phenylbutazone) – we need to make sure the industry is looking for them even if risk of use is low. The other two criteria I would weight above the others would be A and B. It is important to keep drugs that pose serious risks to humans out of the milk supply – but after that, we need to work to keep all drugs out of the milk supply. Therefore, focusing on which drugs are used most and which drugs are most likely to show up in milk is important. I am not suggesting criteria C and D are unimportant – but I would weight them lower.	We appreciate the comment. We obtained addition input on weighting through an expert elicitation.
Reviewer #2	See response to 11.	See response above.
Reviewer #3	See above.	See response above.
Reviewer #4	In my practice, we use individual interviews where experts are asked to provide their preferences towards the importance of the identified criteria and sub-criteria. This is done in the context of establishing the relative importance of each of the criteria and sub-criteria already	Thank you for this suggestion on soliciting preferences; we obtained additional input on weighting through an expert elicitation.

REVIEWER	COMMENT	RESPONSE
	identified in the evaluation framework. Prior to interviewing, a read-ahead packet is usually distributed to summarize the decision model and available information. Interviews are conducted by telephone. During the interviews, interviewers are asked to rank the criteria (or sub-criteria) from 1 to n, where n is the number of criteria or sub-criteria. Item #1 was given 100 points. We then ask, "If item #1 was given 100 points, how many points would you give to item #2 relative to item #1?" This is repeated for all criteria or sub-criteria. In addition, the experts are given the opportunity to add supplemental narrative responses along with their point allocations. The interview responses are then coalesced and summarized to show the distribution of priorities. The responses can be grouped and categorized in a number of ways depending on the classifications of the respondents.	
Reviewer #5	See the response to Questions 6-8.	See response above.

CHARGE QUESTION 13: Is the drug list complete and inclusive of all drugs most likely to be used in dairy cattle in the U.S., and does the list include drugs that should not be on the list? If you answered yes to either of these questions, please describe how you would revise the list, and the justification for the revision.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I do think the list is complete – as a dairy veterinarian for the past 15 years – I cannot think of any drug that I've seen used on a dairy (at least on multiple occasions) that is not included on this list. One exception would be corticosteroids (dexamethasone or isoflupredone acetate) – however, neither of these really presents a milk residue concern. There are a couple drugs that I don't believe are needed on the list. To begin with, thiabendazole has not been commercially available in the United States for at least 10 years. I don't believe it's marketed in the United States for any livestock species and I have not seen it used in cattle. While I was glad to see several NSAIDs included on this list – I don't think naproxen is necessary. I have never seen or heard of it being used in cattle. Amikacin and kanamycin are probably not needed – but if you had an assay that detected all aminoglycosides – they could certainly be included. However gentamicin, neomycin and streptomycin are much more important (more commonly used). I have not seen nitrofurazone or furazolidone used (with the exception of topical creams), however, due to the importance of these drugs and their potential health effects in humans, I believe they should stay on the list.	We thank the reviewer for this comment. We decided to include certain drugs even if they are no longer marketed in the U.S. (because the drug may still be available through imports, product remaining on farms, or from other sources), and drugs that may not commonly be used in dairy cattle (unless there was a clear and justifiable reason for exclusion). A more in-depth description of the process used to create the drug list for this risk assessment to the report to better highlight our rational for the inclusion or exclusion of drugs. For example, we decided not to include corticosteroids, as well as selected other drugs, such as hormones (see full technical report).
Reviewer #2	I don't believe the full list was made available, hence my suggestion above to do so. I comment on the final list below.	The full (initial) drug list, as well as a description of the rational used to arrive at the final drug list, is provided in the full technical report.
Reviewer #3	Is thiabendazole readily commercially available? If not, it may not be worth keeping it on the	See rational above.

REVIEWER	COMMENT	RESPONSE
	list. A more common drug for this purpose to include instead might be fenbendazole.	
Reviewer #4	This reviewer does not have knowledge in this domain.	No response required.
Reviewer #5	See the response to Questions 6-8.	See response above.

CHARGE QUESTION 14: Is the list of dairy products representative of all potential dairy products, or does the list include products that should not be on the list? If so, please describe how you would revise the list, and the justification for the revision.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	I think the list of dairy products is appropriate. I wonder if Greek yogurt should be considered differently from traditional yogurt. I'm not a dairy products expert but understand that significantly more milk goes into making Greek yogurt as compared to regular yogurt.	Greek yogurt consumption is included. However, Greek yogurt consumption has not, to date, been tracked separately from regular yogurt in NHANES, so we cannot separate consumption of Greek vs. non-Greek yogurt.
Reviewer #2	I think the list of products is reasonable.	No response required**.**
Reviewer #3	If NFDM is included on the list of dairy products, it might also be worth adding a whey-based powder (like a WPC). While not consumed directly as foods, both are used as ingredients in a variety of dairy products and other foods. To add WPC, additional data on protein binding would be needed. In terms of the specific products, the assumption that all dairy products are consumed in their standard (implied 'full-fat') forms is not accurate. Whole milk should be removed from the list, and reduced-fat milk added. Similarly, low-fat yogurt should be substituted for full-fat yogurt. Justification: Criterion C only considers products with standard fat content. While I recognize that, for feasibility, it may not be realistic to consider all possible fat levels, it may not be accurate to assume a full-fat content for all products when this doesn't represent the products actually being consumed. For example, recent (2012) consumption data for fluid milk indicate that reduced-fat milk consumption has exceeded that of whole milk since 2004 (currently 55 pounds per capita, versus 44.6). This difference is even more pronounced for yogurt, where 2012 retail sales of whole milk yogurt were 224 million pints, versus 1,561 and 1,435 million pints for low-fat and fat-free yogurt, respectively.	We agree. In order to include WPC to the products considered one would need data on the binding characteristics of the selected drugs to whey and casein proteins. These data are not currently available. FDA has commissioned a study to fill this data gap. The model considers consumption of all types of these products (*e.g.*, regular, reduced-fat, low-fat, and non-fat milk).
Reviewer #4	This reviewer does not have knowledge in this domain.	No response required.
Reviewer #5	See the response to Questions 6-8.	See response above.

CHARGE QUESTION 15: Do you have any additional comments on other aspects of the document? Please share them in your review.

REVIEWER	COMMENT	RESPONSE
Reviewer #1	In Appendix III., two questions are posed: 1) Is there a need to give more recognition in the health effects scoring system to antibiotics determined by FDA to be critically important in human medical therapy versus those determined by FDA to be very important? My answer would be no. I don't really think that aspect needs to be included in the risk assessment. The FDA controls which drugs they are concerned about being used in food animal medicine by regulating which drugs get approved for use and which drugs may be used in an extra-label manner. I think this risk assessment needs to focus on which drugs are actually being used (criterion A) and which drugs are most likely to show up in the milk supply (criterion B). I don't see a need for an additional sub-criteria that considers "importance to human medicine." 2) Should the health effects score be modified to capture the long-term unintended health consequences for antimicrobial resistance to the human population from chronic low level exposure to antibiotics in which the ADI is based on microbial rather than toxic effects? Again my answer would be no. Primarily because I just don't think we understand enough about the development of antimicrobial resistance over time due to exposure to low levels of antibiotics. I am more familiar with the data available in food animal species – but it would indicate that animals do have increased levels of resistance intestinal bacteria after exposure to antibiotics (for example E. coli O157:H7) – however, this susceptibility levels return to baseline levels after the antibiotic is discontinued. I think it is very unlikely that people would be exposed to antibiotics on a regular basis as a result of milk residues. It's much more likely to be a short-term exposure from one contaminated gallon of milk, etc. Therefore, I don't see the development of resistance bacteria in the intestine to be a major problem associated with consumption of drug residues through milk. Even if it might be possible – I don't believe our understanding of this subject is strong enough to justify this as a sub-criterion in the model. My final comment would be that I do think the current drug residue testing program for milk and milk products should be re-evaluated. In my opinion, a new system that would increase the frequency of testing for classes of drugs other than β-lactams would be helpful. Obviously the industry can't afford to test every load of milk for each of the 54 drugs – but if you considered β-lactams, lincosamindes, macrolides, sulfonamides, tetracyclines, and fluoroquinolones as the 6 primary classes of drugs used in dairy cattle - you could test each load of milk for one of these classes of drugs at random. The results of this risk assessment should determine which need to be tested for most often (for example 60% of the tests or more could still be aimed at β-lactams, but this would still significantly increase testing for other classes). Then drugs like	We appreciate the comments.

REVIEWER	COMMENT	RESPONSE
	NSAIDs (flunixin) or antiparasitics could be tested for randomly in addition to a class of antibiotics. In my opinion, this would be a smarter system and would keep producers from trying to "predict" what their milk is being tested for.	
Reviewer #2	No.	No response required.
Reviewer #3	[No additional comments provided by the reviewer.]	No response required.
Reviewer #4	*The following books and papers could be useful in strengthening methodological approach:* 1. Linkov, I., Moberg, E. (2012). *Multi-Criteria Decision Analysis: Environmental Applications and Case Studies.* CRC Press. 2. Belton, V.; Stewart, T. J., *Multiple criteria decision analysis an integrated approach*. Kluwer Academic: Boston Mass., 2002. 3. Mitchell J, Pabon N, Collier ZA, Egeghy PP, Cohen-Hubal E, Linkov I, Vallero DA. (2013). A Decision Analytic Approach to Exposure-Based Chemical Prioritization. *PLoS ONE* 8: e70911. 4. Linkov, I., Tkachuk, A., Canis, L., Mohan, M., Keisler, J. (2012) Risk Informed Decision Framework for Integrated Evaluation of Countermeasures against CBRN Threats. *Journal of Homeland Security and Emergency Management*. 9: 1547-7355. 5. Linkov, I., Coles, J.B., Welle, P., Bates, M., Keisle, J. (2011). Anthrax Cleanup Decision: Statistical Confidence or Confident Response. *Environmental Science and Technology* 45:9471-2.	The full technical report cites two of the references (the first two listed).
Reviewer #5	[No comments were provided by the reviewer.]	No response required**.**

III. SPECIFIC OBSERVATIONS

REVIEWER	Page	Paragraph/Line	Comment or Question	RESPONSE
Reviewer #1			None provided.	
Reviewer #2	5	2	I suggest that that the term "risk assessment" may not be the most appropriate. Other terms to consider include risk ranking model (as used on page 7, line 68), risk model, risk-based ranking, etc.	See response to previous comment. The full technical report uses the term "MCDA risk ranking." We have also added a discussion of how this risk ranking differs from other, more traditional risk assessments, such as those described by Codex.

REVIEWER	Page	Paragraph/Line	Comment or Question	RESPONSE
	5	4-6	Re: "The purpose of the risk assessment is to assess the public health risk to consumers from veterinary drug residues in milk and milk products and develop a relative ranking of the drugs, based on the risk to the consumer." – it appears to me that the main purpose is really the second aspect, to develop the ranking. The model and approach do not seem well fit to assess public health risk to consumers; the output of the model is a score with an unclear relationship to actual risk.	The full technical report has revised the purpose of this risk ranking report.
	8	72-84	It appears that this external review is being conducted before the model is tested and validated (with data from Europe etc. as discussed in the teleconference) – so phases 3 & 4 were reversed?	Risk assessment is an iterative process. This external peer review included evaluation of the multicriteria-based ranking model approach, the modeling structure (including criteria, sub-criteria, factors, and data used to inform these), and aggregation strategies. It also included a review of the animal drugs selected and dairy products included.
	9-10	104-122	I think in the interests of transparency that the master list of 450 drugs and the revised list of 140 drugs should be provided somewhere (Appendix? Online?), along with the reasons for exclusion. The inclusion criteria described on lines 112-155 would seem to overlap with some of the criteria (*e.g.*, parts of criteria A & B) later used in the MCDA? Also, the use of expert opinion here is not well described – what factors did they consider? Why not apply the risk ranking model to the 140 drugs rather than the 54?	The full technical report provides the full master drug list and the reasons some drugs were excluded from inclusion in the multicriteria-based ranking. We agree and revised criteria A and B to eliminate overlap. The full technical report now describes the use of expert opinion.
	10	Table 1	There are some drugs that I expected to see because of reported use in dairy cattle, but presumably were excluded for one or more reasons (trimethoprim, polymyxin B, bacitracin). Others may wonder about some drugs – this is a reason to post the full list and if possible the reason(s) for exclusion.	We agree and now have provided the full list of drugs and reasons for exclusion in the full technical report.

REVIEWER	Page	Paragraph/Line	Comment or Question	RESPONSE
	12	162-163	I think it would be helpful to briefly state why MCDA was selected over alternative approaches; the most obvious to me would be a quantitative classical risk assessment model as mentioned above, in which drugs could be ranked based on calculated risk estimates. I assume that the latter approach is less desirable because of the numerous data gaps that are likely to exist. I think it would also be useful to briefly describe why the additive model approach was selected.	The full technical report describes why we selected MCDA risk ranking over other approaches and why we selected the additive model.
	15	227-228	Should mention that the USDA survey was completed by producers and the Sundlof survey by vets, so this is likely to lead to some differences in estimated frequency of use (see note in margin of draft report) that is likely to be balanced out by using both surveys.	This information has been added to the full technical report.
	23	351-353	I'm not convinced that this factor adds information not already captured by the survey data.	We disagree. The marketing status provides a measure of drug availability. An expert elicitation evaluating the sub-criteria also found the marketing status to be a valuable sub-criterion.
	26	411-432	It appears to me that A5 addresses information already captured in A3?	We have re-defined the sub-criterion to address this concern.
	29	463-477	B1 seems not to be well thought out or described. For example, it is not clear whether "published evidence" refers to residue monitoring, pharmacokinetic studies, other types of studies, or all of the above. There are likely to be inherent biases that affect this criterion, such as screening tests used in monitoring programs. Table 12 – what is the rationale behind inclusion of the tolerance level? I don't see the point of δ – why double-count zeros from the first three?	We have re-defined the sub-criterion to address this concern.
	30	485-489	I don't understand this criterion as written – concern about what not already addressed in A3, A5?	We have re-defined the sub-criterion to address this concern.
	31	502	B3 is likely to be highly correlated with one or more of A criteria.	The actual discard time is now used in criterion B rather than just whether the drug had an official discard time.
	44	724-726	This question does not appear with those listed on lines 17-32.	Our risk assessment addresses the FDA risk management charge questions posed.
	44	728-730	I think there is a slight error here in regards to use of the term ADI. See note in margin of the annotated draft.	The full technical report describes use of the terms "ADI" vs. "hazard value."
	45	744	Please see marginal note regarding "hazard-value."	The full technical report describes use of the terms "ADI" vs. "hazard value."

REVIEWER	Page	Paragraph/Line	Comment or Question	RESPONSE
Reviewer #2			Additional Comments from Reviewer #2 were embedded in the draft risk assessment report. Responses are provided to comments other than those on formatting and typographical errors. These latter corrections were made.	Comments: Comments addressed previously. Comment: AMDUCA prohibition was not a screening factor for the selection of the drugs in this study. Comment: We re-defined the sub-criteria in Criterion A, so this is no longer a concern. Comment: We re-defined the sub-criteria in Criterion B, so this is no longer a concern. Comment: Our risk assessment addresses the FDA risk management charge questions posed. Comment: We clarified the text. Comment: We clarified the text. Comment: Thank you for the comment. The full technical report describes use of the terms "ADI" vs. "hazard value." Comment: Yes. There are four drugs for which a tolerance or tolerable level (safe concentration) could not be established.

REVIEWER	Page	Paragraph/Line	Comment or Question	RESPONSE
Reviewer #3			Additional Comments from Reviewer #3 were embedded in the draft risk assessment report. Responses are provided to comments other than those on formatting and typographical errors. These latter corrections were made.	The full technical report now addresses all the comments. Comment: Now incorporated NMDRD 2000-2013 in the model. Comment: Clarification provided in the report. Comment: Now referenced in the appendix (1.1). Comment: The full technical report now includes a detailed explanation of how drugs were excluded. Comment: The full technical report now includes a detailed explanation of how drugs were excluded. Comment: Table of dairy products is now included in the report. Comment: See previous responses on the use of the term lactating dairy cattle in criteria A and B. Comment: Text describing the discard time has been clarified. Comment: Table header changed to improve clarity. Comment: Composition of selected milk and dairy products has been added. Comment: See previous comment on suggestion for including whey protein powder. Comment: Revised text to remove redundant information in appendix.
			Throughout the document, the tense is inconsistent ("will be evaluated," "were selected," etc.) and should be re-written to *past* tense.	The full technical report is now written in past tense.
			Likewise, the text should also be consistent with respect to being written in *third* person throughout (*e.g.*, "We need data" should be re-written to "data are needed"…).	FDA promotes the use of active voice (first person view; plain language). When possible, we used active voice throughout the full technical report.
Reviewer #4			[No specific observations were provided by the reviewer.]	
Reviewer #5			[No specific observations were provided by the reviewer.]	