Final Report

External Letter Peer Review of FDA's

Post-market Assessment Prioritization Tool

October 27, 2025

Contract No. 75F40124A00002
75F40124F19001

Prepared for:
U.S. Food and Drug Administration
Human Foods Program
5001 Campus Dr
College Park, MD 20740

Peer Reviewers:
Barbara Engeli, MSc
Mary A. Fox, PhD, MPH
George Gray, PhD
Norbert E. Kaminski, PhD
Salomon Sand, PhD

Prepared by:

Versar Global Solutions
1025 Vermont Ave NW Suite 500
Washington, DC 20005

TABLE OF CONTENTS

I. INTRODUCTION

Versar Global Solutions (Versar), an independent Food and Drug Administration (FDA) contractor, coordinated an external letter peer review of the *Post-market Assessment Prioritization Tool* report. The peer review was conducted for FDA's Human Foods Program (HFP).

The mission of the US Food and Drug Administration (FDA) is to protect public health by ensuring safe food. In August 2024, the FDA published a discussion paper on the development of an enhanced systematic process for the post-market scientific (risk and safety) assessment of chemicals in food, including food additives; color additives; generally recognized as safe (GRAS) substances (including GRAS substances that have not been notified to FDA); food contact substances; and chemicals that are present as unintentional contaminants. The systematic post-market assessment of food chemicals consists of the following steps: signal detection, triage, prioritization, scoping, scientific assessment (safety, risk, and/or hazard), risk management review, and risk management action. The Post-market Assessment Prioritization Tool focuses on potential risk to public health (risk ranking) and also includes other decisional considerations, using a Multi-Criteria Decision Analysis (MCDA) method. Subject Matter Experts (SMEs) from a variety of disciplines within FDA's Human Foods Program (HFP) will use the prioritization tool to score a set of criteria evaluating candidate chemicals in food for priority for further review. From the individual criterial scores, an overall score is determined. In our MCDA method, the higher the total score, the higher the priority of that chemical for post-market assessment.

Versar contacted fifteen experts to fulfill FDA's requests for: 1-2 experts in multiple-criteria decision analysis (MCDA), risk assessment, decision analysis, decision-making in government; 2-4 experts in Toxicology including new-approach methods (NAMs) and conventional and exposure assessment of food chemicals including authorized substances or contaminants; and 1-2 experts in Epidemiology and public health, including other decisional criteria including public perception, international activities. As a result, Versar found five experts who were interested and available to participate and spanned the necessary expertise. For each interested and available peer reviewer, Versar conducted conflict of interest (COI) screening to ensure that the experts had no real or perceived COI.

Peer Reviewers:

Barbara Engeli, MSc
Ms. Engeli has twenty years of experience in chemical risk assessment/regulatory human toxicology and food safety at the Federal Food Safety and Veterinary Office and Federal Office of Public Health in Switzerland. Ms. Engeli has conducted chemical risk assessments on food and commodities, in particular contaminants, novel food, botanicals, food supplements, veterinary drug residues, cosmetics, and toys. Ms. Engeli holds a Master of Science in Biology from the University of Zurich.

Mary A. Fox, PhD, MPH
Dr. Fox is an Associate Practice Professor at the Johns Hopkins University School of Public Health in the Departments of Health Policy and Management and Environmental Health and Engineering. She develops and applies methods for cumulative impact and risk assessment to

reduce environmental health disparities, design and lead community and occupational health investigations, and policy and program evaluations. Dr. Fox holds a PhD from Johns Hopkins University Bloomberg School of Public Health, Environmental and Occupational Health Policy.

George M. Gray, PhD
Dr. Gray is a professor in the Department of Environmental and Occupational Health at George Washington University Milken Institute School of Public Health. He is Director for the Center for Risk Science and Public Health at GWU and Associated Dean for MPH Programs. Dr. Gray holds a PhD in Toxicology from the University of Rochester School of Medicine and Dentistry.

Norbert E. Kaminski, PhD
Dr. Kaminski is Endowed Chair for the Food and Consumer Product Ingredient Safety, Director for the Center for Research on Ingredient Safety, Director the Institute for Integrative Toxicology, and Professor of Pharmacology and Toxicology in the Cell and Molecular Biology Program at Michigan State University. His lab conducts research focused on elucidation of the molecular mechanisms for impairment of signal transduction cascades and gene expression during lymphocyte activation by drugs and chemicals. Dr. Kaminski holds a PhD in Toxicology and Physiology from North Carolina State University.

Salomon Sand, PhD
Dr. Sand is a Risk Assessor for the Department of Risk-Benefit Assessment at the Swedish Food Agency. He coordinates development of risk and benefit assessment methodology at the department across the Toxicology, Biohazards, and Nutrition units. Dr. Sand holds a Medical PhD in Toxicology and quantitative health risk assessment from the Karolinska Institute, Sweden.

II. CHARGE TO REVIEWERS

The focus of this review is the draft Post-market Assessment Prioritization Tool, which is a critical part of FDA's overall systematic post-market assessment process. Comments from the technical reviewers will be used to inform further development and refinement of the Post-market Assessment Prioritization Tool. The intention is for the reviewers to provide a technical review of the *Post-market Assessment Prioritization Tool* by answering the charge questions. The FDA posted the tool and questions for public comment on June 18, 2025; the docket closed August 18, 2025. We request the experts review these comments before answering the charge questions. A description of the tool and the charge questions were posted here (https://www.regulations.gov/document/FDA-2025-N-1733-0001).

It should be noted that since the tool was sent out for public comment in June 2025, FDA has continued to develop and refine the systematic post-market assessment process. As a result, several changes have been made to the process described in the section of the prioritization tool titled "Overview". An updated description of the post-market assessment process and how the prioritization tool would be applied within that process is described below.

The systematic post-market assessment of food chemicals consists of the following three phases and will be published in more detail later this year:

1. Review of Information (signal detection and triage; prioritization; and finalization of an annual workplan);
2. Scientific Assessment (safety, risk, and/or hazard); and
3. Risk Management

Phase 1, Review of Information, includes three steps: signal detection and triage; prioritization; and finalization of an annual workplan. The prioritization tool is FDA's proposed method of prioritizing chemicals to be added to the annual workplan. Food chemicals added to the yearly workplan move into Phase 2, of the systematic post-market assessment process, Scientific Assessment, and are subject to comprehensive and multi-step evaluations. These assessments will consider all scientific evidence available to FDA to establish robust scientific conclusions on food chemical safety.

Although chemical prioritization is only one part of the larger post-market systematic assessment process, it provides a framework that assists FDA in allocating limited resources to the food chemicals with the most urgent need for scientific assessment. FDA recognizes that the chemical prioritization process must be science-based, data-driven, systematic, and reproducible. However, the chemical prioritization process must also be rapid such that the annual post-market workplan reflects the most current list of food chemicals for scientific assessment.

Charge Questions

FDA seeks to develop a science-based, data-driven, systematic, and reproducible process for the prioritization of chemicals in food that are candidates for post-market assessments, including food additives; color additives; generally recognized as safe (GRAS) substances (including GRAS substances that have not been notified to FDA); food contact substances; and chemicals that are present as unintentional contaminants. The draft Post-market Assessment Prioritization

Tool focuses on potential risk to public health (risk ranking) and also includes other decisional considerations, using a Multi-Criteria Decision Analysis (MCDA) method.

The focus of this review is the draft Post-market Assessment Prioritization Tool, which is a critical part of FDA's overall systematic post-market assessment process. Considering the information in Sections 1-3 and the appendix of this document, please provide feedback on the following questions:

The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

1. The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

2. The draft Post-market Assessment Prioritization Tool currently includes four Public Health criteria and three Other Decisional criteria.

a. Are the four Public Health criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.
b. Are the three Other Decisional criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.
c. Are there additional criteria that should be considered? If so, please describe additional criteria that might be considered and why.

3. The draft scoring definitions for all criteria were developed to consider the expected variability in the types and extent of data available for the wide variety of food chemicals that may be considered for review.

a. Given this context, are the scoring definitions for the Public Health criteria appropriate for the purpose of the tool?
    i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why?
    ii. The toxicity criterion described in Section 3.1.1 considers data for seven different toxicity data types and the score assigned reflects the highest toxicity data type score from the toxicity rubric, which is described in Appendix A Table A1. Is this the most appropriate strategy for assigning a toxicity criterion score? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.
b. Are the scoring definitions for the Other Decisional criteria appropriate for the purpose of the tool?

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why.
ii. FDA is exploring quantitative and qualitative methods to help inform the scoring of the 'building public confidence' criterion (Section 3.2.3) such as conducting public sentiment analysis (e.g., utilizing natural language processing). How might such tools or the information they provide be incorporated into this criterion? What additional strategies and metrics could FDA consider?

4. The prioritization methodology includes weighting factors.

a. FDA is considering equal weighting among the Public Health criteria and (separately), among the Other Decisional criteria for the Post-market Assessment Prioritization Tool.
   i. Should different weights be applied to the Public Health criteria when determining the Total Public Health Criteria Score? If so, please specify the weighting scheme that might be considered and why.
   ii. Should different weights be applied to the Other Decisional Criteria when determining the Total Other Decisional Criteria Score? If so, please specify the weighting scheme that might be considered and why.
b. FDA is considering equal weighting among the Total Public Health Criteria Score and the Total Other Decisional Criteria Score to determine the overall Post-market Assessment Prioritization Score.
   i. Should different weights be applied when determining the overall Post-market Assessment Prioritization Score? If so, please specify the weighing scheme that might be considered and explain why it would be more appropriate than equal weighting.

5. The draft toxicity rubric uses traditional toxicity data *(in vivo*, as well as limited *in vitro* such as for genotoxicity*),* human health outcomes (e.g., adverse event reports), and epidemiological data for determination of the toxicity criterion score within the Public Health criteria. Considering that the prioritization process is not a comprehensive review, please address the following questions.

a. How might FDA incorporate information from new approach methodologies (NAMs) into the toxicity rubric?
   i. Are there specific NAMs (e.g., systems biology, engineered tissues, artificial intelligence, *in vitro*, microphysiological systems, or other alternative data or modeling tools) that would be most appropriate for use in the toxicity rubric? If so, please explain which NAM(s) would be most appropriate and why.
   ii. Given that a single NAM is not expected to be a one-to-one replacement for a traditional *in vivo* toxicity test, how can the strengths and limitations of each NAM be appropriately considered if it is incorporated into the toxicity rubric?
b. Threshold of Toxicological Concern (TTC) approaches can be used to assess the toxicity of chemicals that lack sufficient safety data and have low dietary exposures. Although the Cramer classification scheme has historically been used in TTC approaches, FDA has recently developed the Expanded Decision Tree (EDT) that assigns chemicals to one of six EDT classes. How might such tools or the information they provide be incorporated into the toxicity rubric?

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

III. INDIVIDUAL REVIEWER COMMENTS

A. Reviewer #1

Comments on FDA's *Tool for the Prioritization of Food Chemicals for Post-Market Assessment*

Reviewer #1

I. RESPONSE TO CHARGE QUESTIONS

1. The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

The overall approach seems appropriate and reasonable for both intentionally added substances and unintentional contaminants in food and follows the traditional risk assessment structure (hazard, exposure, susceptible subpopulations, new data as trigger for re-assessments) which is applicable to all substances. Also the Other Decisional criteria which refer to risk perception are relevant for all substances.

However, there are major differences between intentionally added substances and contaminants: Substances that have undergone pre-market assessment (or acceptance, e.g. GRAS) by the FDA are supposed to be safe and it should be ensured that they remain safe also in light of new relevant toxicological data or increased exposure exceeding a Health-based Guidance value (HBGV). Genotoxic carcinogens and substances with other severe endpoints are expected to be excluded or highly restricted if unavoidable (e.g. in case of naturally occurring substances).
In contrast, contaminants and other unintentionally added substances may not have this "safe history". They may not have been evaluated by the FDA, may have been assessed as unsafe, a certain risk may be tolerated as they may be unavoidable and management measures are different (e.g. technical feasibility, mitigation measures instead of influencing conditions of approval or authorization).
As a consequence, as food safety and public trust in the food safety system are primary objectives, FDA might consider maintaining two distinct lists. The prioritization tool used would still be the same.

Combining the risk-/science-based Public Health criteria and the subjective Other Decision criteria into a single score appears worth reconsidering. The information is so different in nature that the immediate question arises as to how the two parts affected the overall score. One solution could be to derive a numerical score only for Public Health criteria. The risk-based indicators of "other government decisions" (e.g. new HBGV, BMDL/MOE, toxicological point of departure (POD) by a relevant body) could potentially be added. The Other Decision criteria could be used to adjust the score, e.g. by adding a rating of a, b, c. This would mean that a 9a would immediately be recognized as a high priority substance for (re-)assessment due to potentially high risk and high public/political interest.

Need for clarifications regarding the integration of the prioritization tool (step 3) in the overall post-market assessment process: I assume that all steps will be described in detail and some

information is already covered there. For example:

Step 1. signal detection: which sources are used for toxicity data (proprietary data by applicants, publications, databases, etc.), literature search strategy, quality of studies.

Step 2. triage: what exactly happens in this step? It would be important to describe if there are cases that directly lead to step 4 risk assessment, bypassing the prioritization tool (e.g. new evidence of genotoxic carcinogenicity in approved food colors). There may as well be cases where no prioritization is required because a low priority results from triage (e.g. if exposure is below the TTC of genotoxic carcinogens for substances within the scope of the TTC approach).

A flowchart illustrating the overall process would be beneficial.

2. The draft Post-market Assessment Prioritization Tool currently includes four Public Health criteria and three Other Decisional criteria.

a. Are the four Public Health criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

It seems appropriate to include the four proposed criteria toxicity (hazard), (change in) exposure, susceptible subpopulation, and new scientific information and potential impact. Comments on details and definitions will be provided below.

It is also beneficial that the chosen criteria are similar to the US EPA TSCA prioritization tool from 2012 which allows for comparisons and experiences gained over the years.

b. Are the three Other Decisional criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

The Other Decisional Criteria are mostly subjective, with exception of the Other Governmental Decisions category. The latter could be further defined in order to focus on comprehensive science-based risk assessments by other regulatory authorities and scientific bodies with potential impact on the FDA's previous assessments. For example, the US EPA, Joint FAO/WHO Expert Committee on Food Additives (JECFA), European Food Safety Authority (EFSA), Health Canada, Food Standards Australia New Zealand (FSANZ), etc. use basically the same internationally accepted principles for the risk assessment of chemicals in food when deriving Health-based guidance values (HBGV) or a benchmark dose (BMDL) combined with a margin of exposure (MOE) (applicable approach for substances with and without a toxicological threshold) based on a toxicological point of departure. FDA could specify which "governmental agencies" qualify and weigh this factor more than the other potentially highly subjective Other Decisional criteria.

"Building Public confidence" should be built through transparent and science/risk-based work of the FDA. It should particularly be ensured that substances requiring FDA approval/acceptance are closely monitored. "Building public confidence" should be the result of the post-market assessment process rather than being a category of the score. As

mentioned above, two separate lists for intentionally used substances with pre-market assessment by FDA and for contaminants could be useful and facilitate communication.

c. Are there additional criteria that should be considered? If so, please describe additional criteria that might be considered and why.

No. Additional criteria would add complexity, use more resources and potentially slow down the process.

3. The draft scoring definitions for all criteria were developed to consider the expected variability in the types and extent of data available for the wide variety of food chemicals that may be considered for review.

a. Given this context, are the scoring definitions for the Public Health criteria appropriate for the purpose of the tool?

In principle, yes. Table A1 addresses the different data sets and covers diversity with the term "weight of evidence." However, this leaves some room for interpretation.
In general, the focus is on the traditional toxicological approach (mainly animal/human data) and previously assessed substances ("new data", "higher exposure" - which makes sense for post-market surveillance; however, it leads to lower overall scores for all contaminants and non-intentionally added substance without a previous assessment by FDA).

For substances with no or limited toxicity data, other approaches and tools might be integrated (e.g. TTC/EDT, grouping of chemicals, QSARs).

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why?

Toxicity rubric: refer to ii.

Change in exposure (maybe add "and potential impact"): It could be clarified if the definitions relate to exposure from food only and/or also to overall exposure. Considerably, highly, moderately, somewhat higher, etc. are not very clear and data sources could be further explained. However, it will remain difficult to describe those terms in a more quantifiable way. It seems to be a "expert judgement" whether the change potentially impacts / leads to the exceedance of a HBGV or regulatory limit. "and potential impact" could be added to the title, consistently with the "3.1.4 new scientific information and potential impact" category.
Evidence from human biomonitoring could be a relevant source of information too.

Susceptible subpopulations: it should be clearly defined which subpopulations are meant. Infants only or also e.g. children (up to 3 years, more?), pregnant and breastfeeding women, others?

New scientific information and potential impact: The definitions are not applicable for substances which were not previously assessed by the FDA. Would they all be rated as zero? It should be clarified as it lowers the overall score for all such substances.

ii. The toxicity criterion described in Section 3.1.1 considers data for seven different toxicity data types and the score assigned reflects the highest toxicity data type score from the toxicity rubric, which is described in Appendix A Table A1. Is this the most appropriate strategy for assigning a toxicity criterion score? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

According to Table A1, FDA has chosen a structure with (1) acute and repeated-dose toxicity, addressing different endpoints and (2) species/study type (animal, human, in vitro/in silico) which corresponds to the information covered later in the risk assessment (step 4) and makes therefore sense.
Different information of hazard identification/characterization are used to assign a low, medium or high toxicity score; including purely hazard-based GHS classification, fixed dosages in animal studies (e.g. <250 mg/kg bw per day; which are different to substance-specific point of departures as e.g. a NOAEL or BMDL), or vague descriptions as "evidence". If TTC values were added, generic thresholds for humans (incorporating a uncertainty factor for extrapolation to human) would also be introduced. This seems complex and some data-collection/pre-evaluation may be potentially time-consuming, also depending on the number of substances that are prioritized.

In addition, it is unclear and should be clarified whether the FDA intends to always address all endpoints and information sources described in Table A1, or whether it is sufficient to specifically identify the most relevant information which will lead to the toxicity score of the substance, with all other endpoints being ignored, as they would not impact the existing highest score. This would safe resources particularly for endpoints where data is not readily available (e.g. immunotoxicity compared to e.g. GHS classification). This would create data gaps and should be explained.

These considerations could be further clarified in the example of table A2.
If always all toxicity rubrics are assessed (as now in the example of table A2), this should also be explicitly stated.

Comments to table A1:
Footnote 5 could be modified. As oral is the most relevant route for food, the focus should be on oral studies. Other routes could be included if relevant in the weight of evidence approach.

Footnote 6: dose levels discriminating against moderate/low scores should be further explained, especially if they differ from EPA's TSCA (e.g. developmental toxicity, neurotoxicity; compared to 2012

work_plan_methods_document_web_final.pdf unless there exists an updated version which could be referenced?). The statement "… and verified using food chemicals" should be referenced.

Acute toxicity, animals: Animal LD50 toxicity data is of limited value as mortality in laboratory animals is highly likely never used as sensitive endpoint for quantitative risk assessment of chemical substances in food. It could be removed.
Acute toxicity endpoints of animal studies as e.g. teratogenicity, developmental neurotoxicity will be covered under those specific endpoints.
Acute toxicity, human: "adverse events" could be clarified regarding to severity and causality, as usually done e.g. for food supplements.

Carcinogenicity/mutagenicity/genotoxicity: Carcinogenicity and genotoxicity/mutagenicity are not the same. To identify DNA-reactive, directly genotoxic carcinogens is currently a very impactful step in regulatory risk assessment internationally. Genotoxic carcinogens are generally excluded or highly restricted in approval processes. Such substances should directly undergo re-evaluation e.g. at step 2 triage unless exposure is very low (e.g. below TTC of genotoxic carcinogens).
Genotoxic carcinogens are currently considered as not having a toxicological threshold (paradigm of regulatory toxicology) which leads to a significantly higher margin of exposure compared to default uncertainty factors of substances with toxicological thresholds.
For prioritization, it is therefore important to identify those substances (high score 9). In footnote 7, GHS classes should be more precisely referenced (carcinogenicity AND?/OR? germ cell mutagenicity? definitions of the classes 1A, 1B, 2 could be added). IARC cancer categories could be used as well.
If such hazard classification is lacking, the approach of assessing genotoxicity used at EFSA and JECFA could be used in the proposed weight of evidence approach (in vitro OECD bacterial reverse mutation and in vitro micronucleus test, overruled by in vivo animal studies; EFSA 2023 Harmonised approach for reporting reliability and relevance of genotoxicity studies).
Substances with exposure below the TTC of 0.0025 µg/kg body weight (bw) per day which was derived for substances that have the potential to be DNA-reactive mutagens and/or carcinogens based on the weight of evidence could be rated as "low" (exceptions: the TTC approach is not applicable for "the cohort of concern" and some substance classes).
QSARs might be used for genotoxicity assessments in absence of experimental data.

Developmental and reproductive toxicity: GHS reproductive toxicity classes could be used in consistency to carcinogenicity/genotoxicity/mutagenicity.
Endocrine Disruptors could be added here or as a separate toxicity rubric, with explanations of study types and assignment of toxicity scores.

Immunotoxicity or immune activity: study types and assignment of toxicity scores should be further specified as they are less clear compared to the other toxicity categories.

In general, the term "activity" should be explained or deleted.

Bioaccumulation/biopersistence: a rational for the limits (half-life 12-24 h in mammals; BCF 1000) should be added and maybe a comparison to US EPA TCSA could be made.

As mentioned in the first comment, if not already explained for the overall process or signal detection, information on data sources and quality would be beneficial.

In the proposed toxicity score 1-5-9, substances without data are given an advantage over substances with (some) data. It could be considered splitting the score 5 into e.g. 5 (no data and no evidence for toxicity from e.g. QSARs) and 7 (no data and evidence for toxicity from QSARs).

b. Are the scoring definitions for the Other Decisional criteria appropriate for the purpose of the tool?

See comment to question 2b.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why.

The definitions are clearly explained using illustrating examples. These types of criteria can be strongly influenced by unrelated events occurring at the same time that may receive more attention. This applies in particular to reporting in the news/social media.

ii. FDA is exploring quantitative and qualitative methods to help inform the scoring of the 'building public confidence' criterion (Section 3.2.3) such as conducting public sentiment analysis (e.g., utilizing natural language processing). How might such tools or the information they provide be incorporated into this criterion? What additional strategies and metrics could FDA consider?

Building public confidence should be the outcome of FDA's risk-based work and transparent communication, rather than being a separate category. It is a highly subjective and rapidly changing.

4. The prioritization methodology includes weighting factors.

a. FDA is considering equal weighting among the Public Health criteria and (separately), among the Other Decisional criteria for the Post-market Assessment Prioritization Tool.

See comment to question 1.

i. Should different weights be applied to the Public Health criteria when determining the Total Public Health Criteria Score? If so, please specify the weighting scheme that might be considered and why.

Equal weighing seems reasonable for the prioritization tool.
If any criteria should get a different weight, toxicity should weigh more as the overall objective is that the general population shall not to be exposed to toxic substances.

It is reasonable to focus on susceptible populations, but their impact on the Public Health criteria appears to be quite high. This factor should therefore be clearly defined (it is not clear if it refers only to infant formula or also children, pregnant and breastfeeding women) and maybe given less weight in the overall score. Early stages in human life are also covered in the toxicity rubric of developmental and reproductive toxicity, and endocrine disruption could potentially be added.

ii. Should different weights be applied to the Other Decisional Criteria when determining the Total Other Decisional Criteria Score? If so, please specify the weighting scheme that might be considered and why.

As discussed in question 2b, risk-based "other government decision" (e.g. relevant HBGV, BMDL/MOEs since FDA's assessment) should be given greater weight (80-100%).

b. FDA is considering equal weighting among the Total Public Health Criteria Score and the Total Other Decisional Criteria Score to determine the overall Post-market Assessment Prioritization Score.

i. Should different weights be applied when determining the overall Post-market Assessment Prioritization Score? If so, please specify the weighing scheme that might be considered and explain why it would be more appropriate than equal weighting.

It seems favorable not to combine the two distinct criteria groups as they capture different rationales to prioritize substances in food. The FDA could consider basing its risk-rank tool exclusively on the Public Health criteria, maybe including the Other Governmental Decisions criteria which should as well be risk-based. The Other Decision criteria are subjective and could be used to adjust ranking, for example by using letters a, b, c for external perception/urgency. The highest

priority would be 9a, which would allow a quick overview of the two different sets of criteria.
The FDA policy could/should be used to justify the weighting of the two criteria groups.

To avoid many substances receiving the same rating, one could consider giving greater weight to the most important factors and multiplying them, for example, toxicity x exposure.

5. The draft toxicity rubric uses traditional toxicity data (in vivo, as well as limited in vitro such as for genotoxicity), human health outcomes (e.g., adverse event reports), and epidemiological data for determination of the toxicity criterion score within the Public Health criteria. Considering that the prioritization process is not a comprehensive review, please address the following questions.

a. How might FDA incorporate information from new approach methodologies (NAMs) into the toxicity rubric?

There is currently a lot of activity in developing, validation, and potentially implementation of NAMs in a regulatory context. NAMs should be considered in risk assessment (step 4) as part of the weight-of-evidence approach of hazard identification, potentially also quantitative using the "in vitro to in vivo extrapolation (IVIVE)" approach (e.g. NTP (2025) In Vitro to In Vivo Extrapolation).

With regards to the prioritization tool, the "weight of evidence approach" of assessing genotoxicity as included in the toxicity rubric (Table A1) could be considered as an internationally well-established NAM, e.g. by JECFA, EFSA, national food safety agencies in Europe; EFSA 2023 Harmonised approach for reporting reliability and relevance of genotoxicity studies. The term NAMs would not be used here in the sense of "New" approach methodologies, but rather in its meaning as "No-Animals" methodologies. In general, two negative in vitro tests covering all relevant genotoxicity endpoints are sufficient, with in vivo tests only required to follow up positive or equivocal results. QSAR toolboxes may be used where data are lacking (see also comments on Table A1, question 3).

i. Are there specific NAMs (e.g., systems biology, engineered tissues, artificial intelligence, in vitro, microphysiological systems, or other alternative data or modeling tools) that would be most appropriate for use in the toxicity rubric? If so, please explain which NAM(s) would be most appropriate and why.

See comment above on genotoxicity.

ii. Given that a single NAM is not expected to be a one-to-one replacement for a traditional in vivo toxicity test, how can the strengths and limitations of each NAM be appropriately considered if it is incorporated into the toxicity rubric?

If NAMs are used in the prioritization step, they should be used in the weight of evidence approach of the specific toxicity rubrics as foreseen in Table A1.
NAMs that are validated or in validation phase e.g. at the OECD, or whose implementation in the regulatory context is discussed (e.g. In vitro assays for developmental neurotoxicity | OECD) should be preferred to those of early experimental stages.
With reference to this question ii: NAMs should not be considered as replacement of traditional in vivo toxicity studies and rather directly be targeted at humans due to issues of animal-human extrapolation, limited evidence of the reproducibility of animal studies (for example, often only one oral subchronic animal study available), etc.

b. Threshold of Toxicological Concern (TTC) approaches can be used to assess the toxicity of chemicals that lack sufficient safety data and have low dietary exposures. Although the Cramer classification scheme has historically been used in TTC approaches, FDA has recently developed the Expanded Decision Tree (EDT) that assigns chemicals to one of six EDT classes. How might such tools or the information they provide be incorporated into the toxicity rubric?

The TTC approach is well-established and already used in different areas of food safety and might be implemented in the prioritization tool for substances with no or limited data (e.g. EFSA 2019 Guidance on the use of the Threshold of Toxicological Concern approach in food safety assessment). Exposure <TTC for genotoxic carcinogens (0.0025 µg/kg bw per day) could be incorporated into the "low" score of the carcinogenicity/mutagenicity/ genotoxicity category. An alternative option would be to use the TTC for genotoxic carcinogens already at step 2 of triage to exclude organic substances with exposure below this TTC from the post-market assessment process, provided the substance is within the scope of the TTC approach (exceptions are the "cohort of concern" as e.g. aflatoxins, N-nitrosamines, and specific chemical classes as metals, proteins, etc.).
As exposure might change in future, TTC-based decisions should be transparently documented.

The other Cramer Classes and their TTC values could potentially be used in the "other organ-specific toxicity" category, as the underlying NOAELs are from oral repeated-dose animal studies and do not differentiate in endpoints. Mapping high, medium, or low severity scores could be challenging and depend on the TTC or the Expanded Decision Tree (EDT) which offers 6 distinct classes. FDA's further development of the concept is highly welcome, and the EDT could potentially be used instead of the TTC. Given that the EDT has only recently been published, it will likely undergo further scientific discourse and consensus finding. I am reluctant to recommend either the EDT or TTC approach.

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

I would like to congratulate the FDA for establishing this post-market assessment process and its prioritization tool. I assume detailed outcome of the prioritization step (more than just a numerical score) will be published for transparency. General additional comments are already shared in question 1.

B. Reviewer #2

Comments on FDA's *Tool for the Prioritization of Food Chemicals for Post-Market Assessment*

Reviewer #2

I. RESPONSE TO CHARGE QUESTIONS

1. The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

Given the agency's limited resources, the FDA should be commended for its efforts to develop a quantitative tool designed to prioritize chemicals - both intentionally and unintentionally introduced into food - for post-market assessment.

A critical element in the implementation of such a tool will be the establishment of clearly defined criteria for the ranking process and ensuring that the process is conducted with full transparency. Equally important is the adoption of a weight-of-evidence approach when evaluating peer-reviewed scientific literature to determine whether "a chemical may produce severe health effects."

The application of a weight-of-evidence approach should include:

- An evaluation of the quality of each study (e.g., use of validated methodologies and instrumentation; demonstration of causality through multiple doses or concentrations; study design incorporating appropriate controls; application of suitable statistical analyses; use of doses or concentrations relevant to human exposure; relevance of the exposure route to human conditions; and adequacy of sample size per treatment group, etc.), with such factors informing both the quality rating of each study and the weight assigned to it in determining the review priority of a chemical.
- Recognition that individual studies reporting novel or previously unobserved effects should not disproportionately influence prioritization decisions and viewed with caution.

2. The draft Post-market Assessment Prioritization Tool currently includes four Public Health criteria and three Other Decisional criteria.

a. Are the four Public Health criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

1. "The chemical may produce severe health effects": This criterion is of critical importance; however, as discussed above, more than one study must demonstrate the effect(s), and those studies must be of sufficient quality to establish a robust weight-of-evidence for the observed adverse effect(s).

2. "The statement 'Dietary exposure to the chemical has increased' does not, in itself, constitute an important criterion. What is critical is the actual level of human exposure to that chemical and whether the recent increase approaches or exceeds the Threshold of Toxicological Concern (TTC). One can imagine a scenario in which either an intentional chemical additive or, more likely, an unintended contaminant is present in extremely low quantities and has recently increased. Even if the concentration of such a chemical rises substantially, it may still remain several orders of magnitude below the TTC and therefore pose minimal concern. Conversely, it is possible that a widely used chemical additive present in considerable quantities could, with only a modest increase in exposure, approach or exceed the TTC. Accordingly, the key consideration is the total extent of exposure and its relationship to the TTC, rather than the mere fact that an increase has occurred." The relationship to the TTC is most important and not whether there simply has been an increase in exposure.

3. "The chemical is present in food intended for vulnerable subpopulations." This is an important consideration; however, the term "vulnerable subpopulations" should be clearly defined to avoid any ambiguity and ensure that its meaning is consistently understood in the context of exposure assessment.

4. "Newly available information, data, or scientific findings indicate a potentially significant impact on the conclusions of the previous assessment." This is an important consideration; however, as noted above, it should not rely on a single paper or study. A weight-of-evidence approach must support the new information or data before it can justify prioritizing a chemical for reassessment.

b. Are the three Other Decisional criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

The "Other Decisional" criteria listed and as described are focused on whether public sentiment, changes by other governments and consumer group advocacy are not appropriate for the purpose of the tool.

The FDA should consider using complementary strategies and mechanisms to address the Other Decisional Criteria but outside the tool itself. One effective approach could be the development of a robust communications program that provides the public with accurate, science-based information on topics of concern. Such communication should be timely, accurate, easily accessible, and conveyed in language understandable to individuals with limited scientific backgrounds. It should also be clear that safety issues based on the best available science - whether or not they are of significant public interest - will be addressed through the Public Health Criteria within the tool. Therefore, while the Other Decisional Criteria remain important, they are not appropriate for inclusion in the tool's primary decision framework.

A final point relates to the criterion: "Not assessing this chemical could result in the public losing confidence…". One of the most significant contributors to a loss of public confidence is the absence of clear communication or transparency from a government

agency. The FDA must remain proactive, honest, and responsive in its communication with the public. On several occasions, the agency's lack of timely response to concerns raised by the public, consumer advocacy groups, or state and local governments has been noted. The FDA has performed exceptionally well in maintaining one of the safest food supplies in the world; however, much of the criticism directed toward the agency has stemmed from challenges in communicating effectively with the public.

c. Are there additional criteria that should be considered? If so, please describe additional criteria that might be considered and why.

No.

3. The draft scoring definitions for all criteria were developed to consider the expected variability in the types and extent of data available for the wide variety of food chemicals that may be considered for review.

a. Given this context, are the scoring definitions for the Public Health criteria appropriate for the purpose of the tool?

Please see below.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why?

As currently described, the scoring and definitions used to assess "Toxicity" appear to be hazard-based. There is no discussion of how toxicity scoring will be linked to actual human exposure levels resulting from the consumption of foods containing chemical additives. For acute and organ-specific toxicities, the criteria appear to rely on toxicological effects observed in animals at mg/kg/day dosing levels, or on reports of adverse effects in humans under any exposure scenario, without attempting to relate those effects to typical dietary exposure levels. A risk-based approach should instead be applied, founded on estimated daily exposure to the chemical additive through food or other relevant sources.

As previously discussed concerning the "Changes in Exposure" criterion, the critical factor is the actual exposure level in relation to the Threshold of Toxicological Concern (TTC). Substantial increases in exposure to a chemical present only at trace background levels may still result in negligible overall exposure. Conversely, modest increases in exposure to a chemical already present in foods at relatively high concentrations could have a meaningful toxicological impact.

The criterion "Susceptible Populations" also requires explicit definition, which is currently lacking. Without clearly defining which groups fall under this category, the term could be applied too broadly. Extending this logic further, virtually all

chemicals could be said to affect some subpopulation, which would render the term "susceptible populations" effectively meaningless.

ii. The toxicity criterion described in Section 3.1.1 considers data for seven different toxicity data types and the score assigned reflects the highest toxicity data type score from the toxicity rubric, which is described in Appendix A Table A1. Is this the most appropriate strategy for assigning a toxicity criterion score? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

Assigning the highest score derived from the seven different types of toxicity for the toxicity rubric is appropriate.

b. Are the scoring definitions for the Other Decisional criteria appropriate for the purpose of the tool?

As previously discussed, the Other Decisional Criteria are not appropriate for inclusion within the purpose of the tool. The introduction to this document states that "The Post-Market Assessment Prioritization Tool is intended to prioritize chemicals in food for post-market assessments and will be used to support Agency resource allocation for that purpose." Because the agency's primary responsibility is to protect public health, resource allocation should be driven by a risk-based approach grounded in the best available science and supported by a robust weight-of-evidence framework.

Incorporating factors such as public sentiment, intergovernmental decisions, political pressure, or consumer advocacy - elements that are often not based on scientific evidence - would weaken the integrity of science-based prioritization. Doing so would be inappropriate and could result in the inefficient use of limited resources, ultimately failing to serve the best interests of public safety. The considerations encompassed by the Other Decisional Criteria, while valid in their own right, should therefore be addressed outside the prioritization tool.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why.

The definitions are clear.

ii. FDA is exploring quantitative and qualitative methods to help inform the scoring of the 'building public confidence' criterion (Section 3.2.3) such as conducting public sentiment analysis (e.g., utilizing natural language processing). How might such tools or the information they provide be incorporated into this criterion? What additional strategies and metrics could FDA consider?

It is essential for the FDA to build and maintain public confidence. The agency should be encouraged to employ a range of strategies to achieve this goal,

supported by objective metrics to evaluate their effectiveness. The most effective way for the FDA to strengthen public trust is by consistently protecting public health while avoiding approaches that are overly restrictive or unnecessarily burdensome from a regulatory standpoint.

Equally important is ensuring transparency with all stakeholders, including the public, the private sector, and consumer advocacy groups. Timely and effective communication is also critical. Such communication must be accurate, clearly written, and easily accessible to the public. Remaining silent when there is public concern until certainty is achieved is no longer a viable option in today's information-rich environment. A more effective approach is to promptly reassure the public when available scientific evidence indicates that concerns are unfounded, or, when appropriate, to acknowledge awareness of the issue and communicate that the FDA is actively investigating or taking steps to address it. This should be followed by regular and timely updates detailing the agency's ongoing actions and findings.

4. The prioritization methodology includes weighting factors.

a. FDA is considering equal weighting among the Public Health criteria and (separately), among the Other Decisional criteria for the Post-market Assessment Prioritization Tool.

Please see below.

i. Should different weights be applied to the Public Health criteria when determining the Total Public Health Criteria Score? If so, please specify the weighting scheme that might be considered and why.

This question is challenging to address because several of the Public Health criteria require more explicit definition, particularly "Change in Exposure" and "Sensitive Populations." As currently described, the greatest weight within the Public Health category should be assigned to "Toxicity" and "New Scientific Evidence and Potential Impact." However, this weighting could change if the remaining two criteria were more clearly defined.

ii. Should different weights be applied to the Other Decisional Criteria when determining the Total Other Decisional Criteria Score? If so, please specify the weighting scheme that might be considered and why.

All three of the Other Decisional Criteria are inappropriate for inclusion in the tool and should not influence the overall weighting scheme. Even among the three criteria, it would be challenging to establish any prioritization, as none are grounded in scientific principles. This includes "Other Government Decisions," where the FDA has, on occasion and appropriately, taken positions that differ

from those of other regulatory authorities - most notably in jurisdictions that rely on a hazard-based rather than a risk-based approach, such as the European Union.

b. FDA is considering equal weighting among the Total Public Health Criteria Score and the Total Other Decisional Criteria Score to determine the overall Post-market Assessment Prioritization Score.

As previously stated, the Other Decisional Criteria are not appropriate for inclusion in the tool. Furthermore, if the prioritization framework is structured as proposed (i.e., "giving equal weighting among the total Public Health Criteria score and the total Other Decisional Criteria score to determine the overall post-market assessment prioritization score"), each of the Other Decisional Criteria would have a disproportionate influence on the overall prioritization compared to each of the Public Health Criteria.

This imbalance arises for several reasons. First, the FDA is already aware of factors such as public sentiment, political pressure, decisions by other governments, and advocacy from consumer groups - all of which are highly visible and continue to influence agency prioritization decisions even in the absence of the tool and if not formally included in the tool once it is developed. This influence from the above sources is clearly evident at present. Second, the Public Health component consists of four categories (toxicity, change in exposure, susceptible populations, and new scientific evidence), while the Other Decisional component consists of only three categories (external stakeholder activity, other government decisions, and building public confidence). Therefore, under an equal-weighting structure, each of the three Other Decisional categories would carry greater relative weight than each of the four Public Health categories, since the total scores for both components are proposed to contribute equally to the overall prioritization score.

i. Should different weights be applied when determining the overall Post-market Assessment Prioritization Score? If so, please specify the weighing scheme that might be considered and explain why it would be more appropriate than equal weighting.

As currently described, "Toxicity" and "New Scientific Evidence and Potential Impact" are the two most important criteria within the overall framework and should account for the majority of the total weighting. If more clearly defined, "Changes in Exposure"—which might be reconsidered to reflect a concept such as "Level of Exposure in Relation to the Threshold of Toxicological Concern (TTC)"—and "Sensitive Populations" could then warrant increased weighting, though not to a level exceeding that of "Toxicity" and "New Scientific Evidence and Potential Impact."

5. The draft toxicity rubric uses traditional toxicity data (in vivo, as well as limited in vitro such as for genotoxicity), human health outcomes (e.g., adverse event reports), and epidemiological data for determination of the toxicity criterion score within the Public Health

criteria. Considering that the prioritization process is not a comprehensive review, please address the following questions.

a. How might FDA incorporate information from new approach methodologies (NAMs) into the toxicity rubric

Toxicology as a discipline has historically employed in vivo, in vitro, and in silico methodologies that are reproducible across laboratories and fit for purpose. However, the utilization of NAMs has overwhelmingly been employed more so for mechanistic investigations rather than for regulatory purposes. A major obstacle to the broader application of NAMs for regulatory decision-making has been their qualification (i.e., determining if the NAM is fit for purpose) and/or validation (i.e., ensuring reproducibility across laboratories). These processes are lengthy, resource-intensive, and generally lack sufficient funding outside the private sector. Without qualification or validation, NAMs are unlikely to be utilized for making compelling safety arguments. Conversely, if a NAM indicates adverse effects for a chemical during safety assessments, it will require additional testing and explanation. Within this framework, there is little incentive for the private sector to use NAMs to establish safety. Furthermore, regulatory agencies still depend on in vivo guideline studies to make regulatory decisions. With no incentives for regulatory agencies or the private sector to adopt NAMs, it remains challenging to envision how NAMs will be integrated into toxicology beyond mechanistic studies in the foreseeable future. One specific area where NAMs may have near-term impact is in in silico approaches, which leverage existing data to make toxicity predictions.

i. Are there specific NAMs (e.g., systems biology, engineered tissues, artificial intelligence, in vitro, microphysiological systems, or other alternative data or modeling tools) that would be most appropriate for use in the toxicity rubric? If so, please explain which NAM(s) would be most appropriate and why.

A major advantage of certain NAMs is that they employ human cells and tissues. By doing so, NAMs have the potential to more accurately determine chemical toxicity in humans than animal-based assays, an important consideration given the differences across animal species in their biology. Conversely, a significant challenge with many, if not most, NAMs is their lack of drug-metabolizing activity. This limitation remains a major drawback in using NAMs to assess chemical toxicity.

Some microphysiological systems have made notable advances in developing metabolically active liver-on-a-chip (LoC) models, used alone or in combination with other organs-on-a-chip; however, challenges persist. Primary hepatocytes are considered the gold standard for LoC construction, yet they present difficulties such as their isolation, donor-to-donor variability, and de-differentiation - all of which negatively impact experiment-to-experiment reproducibility. Similarly, LoC models developed with immortalized cell lines exhibit reduced functionality and are less effective at replicating the susceptibility of human liver cells to

injury. Overall, metabolism remains a significant challenge in the application of NAMs for chemical safety assessment.

ii. Given that a single NAM is not expected to be a one-to-one replacement for a traditional in vivo toxicity test, how can the strengths and limitations of each NAM be appropriately considered if it is incorporated into the toxicity rubric?

As suggested by the question, no current single NAM can provide the breadth of information derived from animal studies. Individual NAMs are best suited to identify tissue- and cell type-specific changes induced by chemicals. Microphysiological systems that integrate multiple organs on a chip continue to progress toward the development of a "body-on-a-chip." However, significant technical challenges remain, and these systems are still far from replacing traditional in vivo toxicity testing. Likewise, the current state of the science would require a large battery of NAMs to even approximate traditional in vivo testing. In addition, metabolism poses an added challenge faced by many, if not most, NAMs, as discussed above.

b. Threshold of Toxicological Concern (TTC) approaches can be used to assess the toxicity of chemicals that lack sufficient safety data and have low dietary exposures. Although the Cramer classification scheme has historically been used in TTC approaches, FDA has recently developed the Expanded Decision Tree (EDT) that assigns chemicals to one of six EDT classes. How might such tools or the information they provide be incorporated into the toxicity rubric?

The EDT would be especially beneficial when limited toxicity data are available for a specific chemical additive, as one of its major strengths lies in its ability to make predictions based on chemical structures of related compounds with existing toxicity data. Moreover, the EDT predicts both the chronic oral toxic potential of a chemical and its safe level of exposure. It also refines and expands upon the original Cramer Decision Tree by incorporating a more detailed, fully structure-based set of questions, allowing for greater specificity and precision in classifying chemicals and predicting chronic oral toxicity compared to other tools. All of this information would be valuable for inclusion in the toxicity rubric. A current limitation, however, is that the EDT is not yet fully automated; thus, screening a chemical remains a manual process requiring expertise in organic chemistry and metabolism to conduct evaluations efficiently and reproducibly.

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

No.

C. Reviewer #3

Comments on FDA's *Tool for the Prioritization of Food Chemicals for Post-Market Assessment*

Reviewer #3

I. RESPONSE TO CHARGE QUESTIONS

1. The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

I very much welcome this initiative from the FDA to develop a Post-market Assessment Prioritization Tool, and I appreciate the opportunity to provide comments. The approach is appropriate from an overarching point of view in the sense that the types and subtypes of criteria that are part of the tool covers the type of issues that needs to be considered to fulfill the purpose. Also, the ability to perform weightings at the different levels of the approach is relevant. Depending on data availability and resources the approach may be further improved. I think it would benefit from including a specific "risk" criterion, or similar, thereby better aligning the approach towards how chemical risk is typically characterized. Using a more risk assessment-oriented framework was also a frequent comment in the open consultation. In relation to this, which is my most central comment, ideas for improvements are presented as part of related questions below.

The use of equal weighting across the various levels of the approach is an understandable starting point. However, it is not clear that this is reasonable in all instances. Also, there may be challenges to quantitatively derive an overall Prioritization score based on both Total Public Health and Total Other Decisional Factors. This is also discussed in more detail as part of my comments to related questions.

Also, my main points are summarized at the end of the document in association to question 6.

2. The draft Post-market Assessment Prioritization Tool currently includes four Public Health criteria and three Other Decisional criteria.

- ***a. Are the four Public Health criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.***
  Information related to all four criteria are relevant as part of a criterion for Total Public Health. It can, however, be debated how to best specify the four sub-criteria, scoring them, as well as combining them to a Total Public Health score. Broadly speaking, this may be done in numerous ways. To discriminate between potential alternatives the approach may, as far as possible, be aligned towards typical risk assessment principles. The toxicity criterion is a mix of toxicity/potency considerations, and general evidence considerations, and the exposure focuses on whether there is an increase. I think an improvement would be to amend this to a "Risk criterion", e.g., a point of departure or

Reference Dose compared to some exposure estimate/s. In the more detailed descriptions below I broadly use the term "tentative RfD" (tRfD) to describe the reference level that is part of the definition of such a "Risk criterion", i.e., a metric that considers dose-response (in this case the tRfD), severity of effect (ideally included), and exposure. However, it may be further discussed what terminology is most appropriate.

Risk criterion

It may be too resource intensive to derive a Risk criterion across the whole toxicity rubric. Therefore, an option could be a 2-step approach. As a first step, 1) the current design for assessing Toxicity may be used to decide the most relevant effect category for the assessed chemical, and 2) as a second step the Risk criterion is derived based on that effect category. The rubric entry "Bioaccumulation/biopersistence" would then be lifted out and could be considered as a separate criterion within Public Health (see more on this in association to question 4a). If building on the current approach, a new Risk criterion would be scored by 1, 5 or 9. Step 1 would not contribute to the Total Public Health score, but would only be used for effect category selection, while the result from step 2 (score 1, 5, or 9) would go into the equation for Total Public Health. For example, under such a method a Risk criterion score = 1 may imply exposures well below the tRfD for the considered compound with respect to the selected effect category; a score = 5 may imply exposures somewhat below the tRfD; and a score = 9 may imply exposures in the range of, or above, the tRfD. The exact definitions of the breaking points corresponding to scores of 1, 5, and 9 would need further consideration, but a coupling to concern levels, or similar, adds context and helps interpretation of the scoring.

Examples of straight-forward ranking approaches that include a Risk criterion are SFA (2015) and FSAI (2019). Both methods involve assessing the exposure relative to a tentative RfD. They differ in how the nature/severity of the effect is assessed as well as in the design of the scoring or classification systems. I will further describe my thinking using the margin of exposure metric in SFA (2015) as specific example, but the types of considerations made could also be applicable under some other risk-related metric. See SFA (2024, chapter 8.18) for a more recent (and simplified) application that uses tRfDs/RfDs in combination with per capita exposure estimates. The method includes a severity assessment of the *specific* health effect/endpoint (within the effect category considered) that the tRfD is based on resulting in a severity-adjusted tRfD, and thus a severity-adjusted margin of exposure (risk-related metric) that is classified in any of five Risk Classes. In the context of the current approach, the "Risk Classes" may be amended and considered in terms of scores of 1, 5, and 9. The determination of the severity factor is guided by a health effect classification scheme consisting of several broad health effect categories, somewhat like the broad health effect categories part of the rubric. The first version of this scheme in SFA (2015, Table 3) was developed using schemes in Burke et al. (1996) and Owen (2002) as starting points. As part of further development, a newer scheme can be found in Sand et al., (2018, Supplementary Table S2). In its most detailed form, it includes 9 severity levels within each broad effect category, but this type of scheme may also be simplified and divided into 5 or 3 levels within each broad effect category. In SFA (2024, chapter 8.18) the 5-graded version was used.

Derivation of the type of Risk criterion discussed above would require more data/information than the current set-up, e.g., more "dose classes" than 250 mg, 250-1000 mg, and > 1000 mg, or similar, in the rubric would be needed, covering different orders of magnitudes. For chemicals with limited data, use of the TTC concept (EFSA 2019) or associated refinements could help in the process of determining tRfDs. Regarding the exposure, a traditionally established exposure estimate is ideally of interest (combination of consumption and concentration data, or biomonitoring data). However, more general exposure estimates, e.g., per capita exposure estimates, may fit this type of tool better. Also, the exposure estimate could potentially represent something that is not a specific estimate of how much groups/individuals in the population are exposed but rather represent a "theoretically/practically available exposure", or similar, if this enables to cover more compounds. For example, the report speaks about data on production volume that might be used together with other information for deriving such exposure estimates.

In case data/information quality differs across compounds with respect to the variables behind the Risk criterion an uncertainty element is ideally attached to the applied risk metric. For example, a semi-quantitative approach for uncertainty analysis may be used (e.g., SFA 2015) instead of more data-driven resource intensive methods. The score may be determined by consideration how the distribution for the risk metric, rather than its point estimate, overlaps with associated critical ranges for the Risk criterion (associated with scores of 1, 5, and 9). As a more pragmatic alternative, the uncertainty discussed here may potentially be captured by the "New scientific information and potential impact" criterion (see also comments below and in responses to question 4a).

The suggestions above represent ideas that may or may not include a severity consideration. The feasibility of this may depend on the type of information available. Data availability may be a general challenge for use of a Risk criterion across a large set of compounds, but as noted initially, I think the possibility of this should at least be explored. It could be reasonable to design a clearly risk-based framework from the start that gradually may be subject to improvement (e.g., better basis for the parameter values used, broader applicability), including how the Risk criterion is defined, as relevant/new technology can support comparative quantifications on both the hazard and exposure sides. For example, Patlewicz et al., (2018) uses TTC in combination with high throughput exposure modelling for risk ranking.

Other Public Health criteria

Considering the other criteria part of Public Health, "change in exposure" may be generalized to "change in risk" under suggestion of a Risk criterion, which can imply a change in the tRfD and/or the exposure. However, in case of a Risk criterion, a change in exposure/risk seems not to be needed. If ranking is performed annually, this will illustrate if changes have occurred compared to the last assessment, e.g., a Risk criterion score may be modulated due to a change in exposure.

"Susceptible population" and "New scientific information and potential impact" are relevant additional criteria. In the case of a Risk criterion "New scientific information and potential impact" could be indirectly covered to some extent, e.g., "new information"

may be considered as part of the tRfD derivation. However, this type of criterion is still relevant to keep since it can also be seen as relating to uncertainty e.g., about the relevance of the data/information used for assessment. There are some additional points on this in comments to question 4a.

b. Are the three Other Decisional criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

"External stakeholder activity/attention", "Other government decisions", "Building public confidence" are appropriate for the purpose described. However, there may be challenges to quantitatively combine Other Decisional Criteria with Total Public Health. See comments related to this in association to question 4b.

c. Are there additional criteria that should be considered? If so, please describe additional criteria that might be considered and why.

As discussed above, I think the tool would benefit from the introduction of a Risk criterion as part of the Total Public Health criteria.

3. The draft scoring definitions for all criteria were developed to consider the expected variability in the types and extent of data available for the wide variety of food chemicals that may be considered for review.

a. Given this context, are the scoring definitions for the Public Health criteria appropriate for the purpose of the tool?

Overall, the definitions are appropriate for the purpose of the tool. See specific comments below.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why?

There is some level of ambiguity regarding all the scoring definitions, but expert judgment will ultimately be needed as be part of an approach like this. The "New scientific information and potential impact" criterion may be most difficult. Also, see comments under question 4 on weighting that overlaps a bit with considerations of the definitions.

ii. The toxicity criterion described in Section 3.1.1 considers data for seven different toxicity data types and the score assigned reflects the highest toxicity data type score from the toxicity rubric, which is described in Appendix A Table A1. Is this the most appropriate strategy for assigning a toxicity criterion score? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

Ideally, I think it is relevant to capture the domain of effects potentially caused by the assessed compound by integrating scores across the rubric in some way. However, the proposed approach is more practical, and I think it is adequate (i.e., highest toxicity data type score). As noted earlier, I think the most important improvement would be to modify the approach to include a Risk criterion, and under this suggestion the current Toxicity criterion may be used as an initial step to decide the effect category for which the Risk criteria is derived. As noted earlier, the rubric entry "Bioaccumulation/biopersistence" would then be lifted out and may be regarded as an additional criterion within the Public Health domain.

b. Are the scoring definitions for the Other Decisional criteria appropriate for the purpose of the tool?

Overall, the definitions are appropriate for the purpose of the tool. See specific comment below.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why.

While the definitions are logical it may be challenging to assess what category/score to select, i.e., scoring "building public confidence" appears most difficult given the definitions, scoring "external stakeholder activity/attention" appears medium difficult, and scoring "other government decisions" appears to be comparatively straight-forward. More clarification related to what is meant by public confidence would be useful. A clarification between score 9 and 5 for "external stakeholder activity/attention" would also be helpful. For example, would score = 5 more or less be the same as score = 9 without e.g., Congress, GAO, HHS Secretary, FDA Commissioner call for action; national news/social media coverage?

ii. FDA is exploring quantitative and qualitative methods to help inform the scoring of the 'building public confidence' criterion (Section 3.2.3) such as conducting public sentiment analysis (e.g., utilizing natural language processing). How might such tools or the information they provide be incorporated into this criterion? What additional strategies and metrics could FDA consider?

I do not have experience of conducting public sentiment analysis or other type methods for assessing public confidence. While broader analysis in the public is required as input for assigning this type of score, and while text data from the internet is a relevant domain, it may be important to also perform analysis outside this domain.

4. The prioritization methodology includes weighting factors.

a. FDA is considering equal weighting among the Public Health criteria and (separately), among the Other Decisional criteria for the Post-market Assessment Prioritization Tool.

I think the case can be made that unequal weighting may, or should, be considered within at least one of the two domains. This is further addressed in the comments below.

i. Should different weights be applied to the Public Health criteria when determining the Total Public Health Criteria Score? If so, please specify the weighting scheme that might be considered and why.

I think parameters like "Change in exposure or risk" (if at all needed, see earlier comment on this), "Susceptible population", and "New scientific information and potential impact" should modulate the priority, but such criteria seem not as important compared the proposed Risk criterion. This may also apply under the current approach. For example, a Toxicity score = 9 in combination with Susceptible population score = 1 gives the same weighted sum (using equal weighting) as a chemical with a Toxicity score = 1 and a Susceptible population score = 9, considering all else equal. I lean towards a set-up where additional factors modulate a primary risk-related factor within the Public Health domain, rather than a situation where all these factors are given the same weight. How to differentiate the weights quantitatively is a value-based consideration, with the consequence that a single weight (point estimate) may not be fully representative. However, use of distributions may be overly complex in this round of iteration of the method. Therefore, a starting point could be equal weighting between a main Risk criterion and the sum of remaining criteria, i.e., a weight of 0.5 for the Risk criterion, and weights of 0.5/3 for "Susceptible population", "New scientific information and potential impact" and "Bioaccumulation/biopersistence" (if regarded as a criterion outside the rubric), respectively. For example, chemical A with a risk score = 9, and remaining scores = 1, resulting in a weighted sum of 5.0 [(9×0.5) + (1×0.5/3) + (1×0.5/3) + (1×0.5/3)], would then compare to chemical B with scores = 5 across all Public Health criteria. This may be more balanced compared to using equal weighting (across all four criteria) where chemical A would obtain a Total Public Health score = 3 and chemical B would obtain a Total Public Health score = 5 even though the Risk criterion score is lower for chemical B. It may be noted that under the weighting scheme proposed, a Risk criterion score = 1 (i.e., exposure well below the tRfD) in combination with the remaining three criteria can provide a maximum Total Public Health score = 5. The system may be calibrated by such and related types of considerations.

Considering the additional criteria as modulators of the Risk criterion score may also be a way to address uncertainty, e.g., is there any new information that might suggest that a tRfD, or similar, is even lower? Could a sensitive population be affected? Is the compound potentially bio-persistent? Increased uncertainty about

such issues would then contribute to an increase in the Total Public Health score. The way the scoring definitions are currently written may be more framed in this manner, if regarded useful.

The rubric also covers acute effects, and it may be argued that such effects should be weighted differently compared to long-term effects, i.e., given the same level of severity/impairment an acute effect may be worse than a long-term effect since it is more immediate. If a severity consideration is part of a Risk criterion it may be set systematically higher for acute vs. chronic effects. If overly complex at this stage, however, such added level of detail may possibly be considered in future iterations of the approach.

ii. Should different weights be applied to the Other Decisional Criteria when determining the Total Other Decisional Criteria Score? If so, please specify the weighting scheme that might be considered and why.

According to the report these criteria should cover 1) the degree of concern about the chemical by the public generally and 2) whether any regulatory partners have taken action to restrict or expand uses of the chemical in food. As a default, it seems reasonable to weight 1) and 2) equally. If equal weighting of the three criteria reflects this, I think it is a relevant starting point.

b. FDA is considering equal weighting among the Total Public Health Criteria Score and the Total Other Decisional Criteria Score to determine the overall Post-market Assessment Prioritization Score.

Intuitively, it seems that the Total Public Health Criteria score should be at least as important as the Total Other Decisional Criteria score, if not more important. However, I have mainly chemical risk assessment perspective on these issues. See more detailed comments below.

i. Should different weights be applied when determining the overall Post-market Assessment Prioritization Score? If so, please specify the weighing scheme that might be considered and explain why it would be more appropriate than equal weighting.

It might be more likely that the scores across different compounds become more differentiated with respect to Total Public Health compared to Total Other Decisional Criteria, i.e., for Total Other Decisional Criteria perhaps only some compound, or a group of compounds, is the current focus during a given year. If so, it appears to me that Total Other Decisional Criteria could modulate the overall result/score in a reasonable manner, e.g., using equal weighting. This relates to my earlier point on the consideration of a risk-related criterion as the central element of the approach, which is modulated by other factors (related to Public Health as well as Other Decisional Criteria). However, if Total Other Decisional Criteria in practice becomes just as differentiated as Total Public

Heath across compounds it could be relevant to look at rankings based on the two domains separately, besides a joint ranking that uses equal weighting. An extra check point may then be to consider if e.g., the combination high Total Public Health and low Total Other Decisional Criteria excludes compound/s from further prioritization based on the overall Prioritization score and reaffirm/discharge the validity of this. In any case it might be helpful to present a separate ranking for at least Total Public Health to help interpretation.

5. The draft toxicity rubric uses traditional toxicity data (in vivo, as well as limited in vitro such as for genotoxicity), human health outcomes (e.g., adverse event reports), and epidemiological data for determination of the toxicity criterion score within the Public Health criteria. Considering that the prioritization process is not a comprehensive review, please address the following questions.

a. How might FDA incorporate information from new approach methodologies (NAMs) into the toxicity rubric?

I will address NAMs more generally since I do not have the expertise to make specific suggestions related to the sub-questions. While the consideration of NAMs could improve and/or substantiate the toxicity rubric I think the main opportunity of using NAMs, instead of traditional toxicity studies/data, would relate to improving the quantification of dose-response across compounds, rather than just supporting derivation of the current Toxicity score. Therefore, I believe NAMs, in the context of this type of tool, could mainly have role to facilitate more efficient derivation of point of departures, or similar, needed as part of derivation of a Risk criterion. As noted earlier, to me it seems reasonable to design a risk-based framework (as far as possible) already from the start that can then be subject to gradual improvement as new methods and technology advances or becomes reasonable established, e.g., using short-term in vivo-studies, NAMs, or AI in the process of derivation of individual parameter values needed for derivation of a Risk criterion.

i. Are there specific NAMs (e.g., systems biology, engineered tissues, artificial intelligence, in vitro, microphysiological systems, or other alternative data or modeling tools) that would be most appropriate for use in the toxicity rubric? If so, please explain which NAM(s) would be most appropriate and why.

See general comment above.

ii. Given that a single NAM is not expected to be a one-to-one replacement for a traditional in vivo toxicity test, how can the strengths and limitations of each NAM be appropriately considered if it is incorporated into the toxicity rubric?

See general comment above.

b. Threshold of Toxicological Concern (TTC) approaches can be used to assess the toxicity of chemicals that lack sufficient safety data and have low dietary exposures.

Although the Cramer classification scheme has historically been used in TTC approaches, FDA has recently developed the Expanded Decision Tree (EDT) that assigns chemicals to one of six EDT classes. How might such tools or the information they provide be incorporated into the toxicity rubric?

TTC or a more refined version of this approach could help to enable the implementation of a risk-related criterion as part of Public Health, so that such an approach can also be applied for compounds with limited toxicity data.

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

Within each criterion scoring is generally performed in terms of 1, 5, or 9, and there are some examples where also a score of 3 can be used. I think the approach could benefit from standardization so that 1, 5, and 9 is used across the board, or 1, 3, 5, 7, and 9 is used across the board in case of a 5-graded system. In the future, a more refined version of the tool might even make use of all 9 categories and maybe still be a somewhat compatible with what has been done in the past. So overall I think this set-up is fine.

Summary of main comments:

- I think the method would benefit from including a Risk criterion within the Public Health domain. This aligns the approach to how chemical risk is normally assessed, which adds context and could facilitate interpretation of the scoring.
- Unequal weighting within Public Health domain is proposed. If building the approach around a risk-related criteria, additional criteria may be considered as "modulating factors". The consideration of unequal weighting may also be relevant under the current set-up.
- The "modulating criteria" may be more/additionally framed in terms of uncertainty, where an increasing uncertainty would increase the Total Public Health Score.
- The domain for assessing a potential loss of public confidence may need to be broader than the internet.
- To improve interpretation and transparency it is suggested that rankings based on Total Public Health, and maybe also Total Decisional Criteria, is presented separately alongside the ranking based on the overall Prioritization score.

References

Burke TA, Doull J, McKone T, Paustenbach DJ, Scheuplein R, Udo de Haes HA, Young JS. 1996. Human health impact assessment. In Life cycle assessment: Analysis by an expert panel. Washington, DC: ILSI Health and Environmental Sciences Institute.

EFSA. 2019. EFSA Scientific Committee, More SJ, Bampidis V, Benford D, Bragard C, Halldorsson TI, Hernandez-Jerez AF, Hougaard BS, Koutsoumanis KP, Machera K, Naegeli H, Nielsen SS, Schlatter JR, Schrenk D, Silano V, Turck D, Younes M, Gundert-Remy U, Kass GEN, Kleiner J, Rossi AM, Serafimova R, Reilly L and Wallace HM, 2019. Guidance on the use

of the threshold of toxicological concern approach in food safety assessment. EFSA J 17(6):5708, 17 pp.
FSAI (Food Safety Authority of Ireland). 2019. FSAI risk ranking model for chemical contaminants in food. Report of the Scientific Committee of the Food Safety Authority of Ireland. ISBN: 978-1-910348-25-3.

Owens JW. 2002. Chemical toxicity indicators for human health: Case study for classification of chronic non-cancer chemical hazards in life cycle assessment. Environ Toxicol Chem 21:207-225.

Patlewicz G, Wambaugh JF, Felter SP, Simon TW, Becker RA. 2018. Utilizing threshold of toxicological concern (TTC) with high throughput exposure predictions (HTE) as a risk-based prioritization approach for thousands of chemicals. Comput Toxicol 7:58-67, doi: 10.1016/j.comtox.2018.07.002.

SFA (Swedish Food Agency). 2024. L 2024 nr 08: The Swedish Market Basket Study 2022. Per capita-based analyses of nutrients and toxic compounds in market baskets and assessment of benefit or risk. Swedish Food Agency report series. Uppsala.

SFA (Swedish Food Agency). 2015. The risk thermometer - a tool for risk comparison. National Food Agency Report No. 8.

Sand S, Lindqvist R, von Rosen D, Ilbäck N-G. 2018. Dose-related severity sequence, and risk-based integration, of chemically induced health effect. Toxicol Sci 165:74-89, doi:10.1093/toxsci/kfy124.

D. Reviewer #4

Comments on FDA's *Tool for the Prioritization of Food Chemicals for Post-Market Assessment*

Reviewer #4

I. RESPONSE TO CHARGE QUESTIONS

1. The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

The goal of a prioritization tool should be the inclusion of important factors or variables, making sure they are independent and not double counting some attribute, and the appropriate weighting of the factors to appropriately prioritize based on the endpoint of concern. The FDA *Post-market Assessment Prioritization Tool* (PART) is intended as a risk ranking but focuses primarily on qualitative measures rather than quantitative risk measures. The risks posed by chemicals in food are a function of both the doses required to elicit adverse effects and the levels of exposure from consuming food(s) containing the chemical. Strictly qualitative ranking systems have the potential to scramble the ranking of chemicals that would arise from a quantitative focus on risk. Examples are provided in the answers to subsequent questions.

An alternative method to prioritize efforts to assess the risks of chemicals in food would be to estimate the exposure necessary to be of concern (an Acceptable Daily Intake (ADI) or similar) based on available data and then to examine how high levels of the chemical in food would have to be to raise concerns. ADIs can be estimated using existing data and known relationships from well tested chemicals or short-term *in vivo* data (*e.g.*, Layton *et al*. (1987), Kvasnicka *et al.* (2024), Harrill et al. (2025). Similar methods can be used if the chemical is a potential carcinogen (Gaylor and Gold, 1998) although these do not distinguish between genotoxic and nongenotoxic carcinogens. These can then be compared to estimates of intake from food either from previous FDA assessments or from new information about changes in exposure. It is recognized that estimating total exposure to a chemical from all sources in the food supply could be a significant challenge. These estimates can reflect the uncertainty resulting from incomplete information (Paoli, Hartnett and Price (2025). A ranking of Predicted Exposure/Estimated ADI can then be used to set priorities.

2. The draft Post-market Assessment Prioritization Tool currently includes four Public Health criteria and three Other Decisional criteria.

a. Are the four Public Health criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

1. The chemical may produce severe health effects (e.g., cancer, cardiovascular toxicity)

This criterion would almost certainly rely heavily upon animal toxicity data since epidemiologic evidence for an emerging food safety concern would likely be highly preliminary and uncertain. Yet it is known that concordance of toxicologic endpoints, which is a necessary assumption for this criterion, is quite poor (e.g., Lin, Gold and Freedman (1995), Weitekamp, C.A. *et al.* (2025)). In addition, the human health risk from different potential carcinogens may vary widely depending on the mode of action for carcinogenicity (*e.g*., genotoxic *vs* non-genotoxic). This highly influential toxicity criterion could misorder priorities without reflecting the underlying scientific knowledge. These challenges raise the potential for putting an under- or over-emphasis on a specific chemical and response which may not be indicative of human risk.

2. Dietary exposure to the chemical has increased

   The key question for risk ranking is the exposure level. An increase may be significant for risk (if it lowers the exposure/ADI ratio to a level of concern) or insignificant if it changes it a very small amount or not at all. The direction of change holds only partial information for an accurate ranking

3. The chemical is found in or could potentially be present in food intended for vulnerable subpopulations (*e.g*., infants)

   Only question here is whether this is a general assumption or is chemical specific. There are known examples of toxicologic situations in which young animals are less sensitive to an adverse effect of a chemical than are adults. Also, what if the prior assessment (if it exists) already considered vulnerable subpopulations – would this be double counting?

4. Newly available information, data, or science indicates a potentially high impact on the conclusions of the previous assessment of the chemical.

   Seems OK although it may be redundant (double counting) with 1 – 3 for a previously assessed chemical.

b. Are the three Other Decisional criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

This is where I have a significant concern about double counting criteria. It is my understanding that the three Other Decisional criteria are essentially the same as the Signal Monitoring portion of the Enhanced Systematic Process for the FDA's Post-Market Assessment of Chemicals in Food that would identify chemicals for prioritization. Since both are reflecting non-scientific outside drivers (often the same – *e.g*., news coverage) this may give greater weight to some chemicals based on the same factors that put them on the priority list.

I would advocate for the Other Decisional criteria to be part of Signal Monitoring process but not combined with the risk-related scientific factors in setting priorities.

c. Are there additional criteria that should be considered? If so, please describe additional criteria that might be considered and why.

As discussed above, the use of a quantitative risk-based criterion (like Predicted Exposure/Estimated ADI) would be very valuable in understanding how likely it is that a chemical poses a risk.

3. The draft scoring definitions for all criteria were developed to consider the expected variability in the types and extent of data available for the wide variety of food chemicals that may be considered for review.

a. Given this context, are the scoring definitions for the Public Health criteria appropriate for the purpose of the tool?

No comments provided.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why?

Definitions are clearly explained.

ii. The toxicity criterion described in Section 3.1.1 considers data for seven different toxicity data types and the score assigned reflects the highest toxicity data type score from the toxicity rubric, which is described in Appendix A Table A1. Is this the most appropriate strategy for assigning a toxicity criterion score? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

For reasons noted elsewhere, the focus on qualitative attributes may lead to misleading scoring (*e.g.*, a high score from a focus on carcinogenicity for saccharin despite knowledge of mechanism of action). The prediction of a POD or ADI, as discussed above, would be a more useful output of the toxicity criterion.

For the other criteria, the use of ill-defined words to distinguish categories (considerably increased, moderately increased; potential high impact, potential moderate impact) seems to invite dispute and questions about rigor and reproducibility of the judgements.

b. Are the scoring definitions for the Other Decisional criteria appropriate for the purpose of the tool?

As explained elsewhere, my concern is that many of these criteria have already been considered in the Signal Monitoring portion of this process leading to a selection bias and overemphasis on these attributes.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why.

Definitions are fine.

ii. FDA is exploring quantitative and qualitative methods to help inform the scoring of the 'building public confidence' criterion (Section 3.2.3) such as conducting public sentiment analysis (e.g., utilizing natural language processing). How might such tools or the information they provide be incorporated into this criterion? What additional strategies and metrics could FDA consider?

This is outside of my area of expertise.

4. The prioritization methodology includes weighting factors.

a. FDA is considering equal weighting among the Public Health criteria and (separately), among the Other Decisional criteria for the Post-market Assessment Prioritization Tool.

Weights should reflect the relative importance of each criterion to the FDA decision making process. If equal weights are to be used the FDA should clearly describe why they believe equal weights are appropriate.

i. Should different weights be applied to the Public Health criteria when determining the Total Public Health Criteria Score? If so, please specify the weighting scheme that might be considered and why.

FDA should clearly describe why the weighting scheme reflects their belief about the relative importance of the attributes.

ii. Should different weights be applied to the Other Decisional Criteria when determining the Total Other Decisional Criteria Score? If so, please specify the weighting scheme that might be considered and why.

It is my view that the Other Decisional Criteria should not be included in prioritization scores because they are already reflected in the Signal Monitoring process.

b. FDA is considering equal weighting among the Total Public Health Criteria Score and the Total Other Decisional Criteria Score to determine the overall Post-market Assessment Prioritization Score.

No comments provided.

i. Should different weights be applied when determining the overall Post-market Assessment Prioritization Score? If so, please specify the weighing scheme that might be considered and explain why it would be more appropriate than equal weighting.

See comments above.

5. The draft toxicity rubric uses traditional toxicity data (in vivo, as well as limited in vitro such as for genotoxicity), human health outcomes (e.g., adverse event reports), and epidemiological data for determination of the toxicity criterion score within the Public Health criteria. Considering that the prioritization process is not a comprehensive review, please address the following questions.

a. How might FDA incorporate information from new approach methodologies (NAMs) into the toxicity rubric?

As discussed elsewhere, *in silico* techniques may be of greater utility than NAM(s). These tools allow the estimation of quantitative toxicity values for comparison to existing or newly updated exposure estimates. The FDA EDT, discussed below, would be very useful here. The use of NAM(s) may continue the focus on hazard traits rather that identification of health reference values that can be used to assess risk.

i. Are there specific NAMs (e.g., systems biology, engineered tissues, artificial intelligence, in vitro, microphysiological systems, or other alternative data or modeling tools) that would be most appropriate for use in the toxicity rubric? If so, please explain which NAM(s) would be most appropriate and why.

No comments provided.

ii. Given that a single NAM is not expected to be a one-to-one replacement for a traditional in vivo toxicity test, how can the strengths and limitations of each NAM be appropriately considered if it is incorporated into the toxicity rubric?

No comments provided.

b. Threshold of Toxicological Concern (TTC) approaches can be used to assess the toxicity of chemicals that lack sufficient safety data and have low dietary exposures. Although the Cramer classification scheme has historically been used in TTC approaches, FDA has recently developed the Expanded Decision Tree (EDT) that assigns chemicals to one of six EDT classes. How might such tools or the information they provide be incorporated into the toxicity rubric?

The EDT is a very good tool and would be a great start toward a more quantitative approach. At the least it would allow FDA assessors to ask "how high would exposure need to be to make this a priority chemical"? In this way, prioritization would be more risk based than hazard based.

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

FDA is to be commended for seeking a very structured and analytic approach to priority setting.

References

Gaylor, D. W., and Gold, L. S. (1998). Regulatory cancer risk assessment based on a quick estimate of a benchmark dose derived from the maximum tolerated dose. *Regulatory Toxicology and Pharmacology* **28**: 222-225.

Harrill, A.H., Hagiwara, S., Weitekamp, C.A., et al. (2025) Database-calibrated toxicity values for human health assessment based on existing toxicology data for one thousand chemicals. *Journal of Toxicology and Environmental Health B* **17**:1-40

Kvasnicka, J., Aurisano, N., von Borries, K., Lu, E., Fantke, P., Jolliet, O., Wright, F. A., and Chiu, W. A. (2024) Two-Stage Machine Learning-Based Approach to Predict Points of Departure for Human Noncancer and Developmental/Reproductive Effects. *Environmental Science and Technology* **58**: 15638-15639

Layton, D.W., Mallon, B.J., Rosenblatt, D.H., and M.J. Small (1987) Deriving Allowable Daily Intakes for Systemic Toxicants Lacking Chronic Toxicity Data. *Regulatory Toxicology and Pharmacology* **7**:96-112

Lin, T, Gold, L.S. and Freedman, D. (1995) Carcinogenicity Tests and Interspecies Concordance. *Statistical Science* **10**: 337 – 353

Weitekamp, C.A. *et al.* (2025) Quantitative and qualitative concordance between clinical and nonclinical toxicity data. *Toxicological Sciences* **206**: 253–272

E. Reviewer #5

Comments on FDA's *Tool for the Prioritization of Food Chemicals for Post-Market Assessment*

Reviewer #5

I. RESPONSE TO CHARGE QUESTIONS

1. The purpose of the Post-market Assessment Prioritization Tool is to assist in making decisions about which chemicals, including both intentionally added substances and unintentional contaminants in food, are a priority to review. Is the modeling approach we proposed appropriate for this purpose? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

The MCDA model is an appropriate tool for ranking and prioritization. To assist in evaluating the tool, it would have been helpful to know more about the steps that precede it in the full process (signal detection and triage). It also would have been helpful to know more about the prior round of pilot testing and lessons learned from that.

Maintaining public trust in the food supply is an important goal of the Human Food Program (HFP). To do that having and using a model with a trust criterion is not enough. The modeling effort needs to be part of an ongoing stakeholder-engaged and transparent process.

It is my understanding that chemicals from all categories (intentionally and unintentionally added, GRAS substances, food contact substances, etc.) are evaluated together in the model. It seems to me that there are potentially important differences in the way these chemicals and substances get into food and different stakeholder interests in each category as well as different control measures that might be taken depending on the category. Does it make sense then to combine them all together in one large ranking exercise? The model can also be run on each category separately. This may present a more balanced approach that allows HFP to respond more holistically to this challenge of multiple chemicals of concern and multiple stakeholder interests.

2. The draft Post-market Assessment Prioritization Tool currently includes four Public Health criteria and three Other Decisional criteria.

a. Are the four Public Health criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

The current Public Health criteria are appropriate. It would have been helpful to have further information on some of the definitions, e.g., susceptible subpopulations.

Change in exposure:
FDA may want to consider whether chemicals not previously assessed should have a higher score or put another way, whether filling assessment gaps is a priority. To

implement this the 'not previously assessed' category can be scored at 5 and the moderate increase in exposure can be scored at 3.

It would be helpful to understand the rationale behind giving change in exposure greater weight (higher score) in the prioritization model. Perhaps it is mainly pragmatic and designed to reduce the universe of chemicals of concern. However, given that some concerning signal has been detected, defining the exposure criterion based on current level of exposure in the food supply is the most appropriate input to a risk-based score.

b. Are the three Other Decisional criteria appropriate for the purpose of the tool? If not, please explain what changes might be considered and why.

The tool is described as a way to develop and update a workplan on an annual basis. High attention in the public sphere may come and go quickly. That aspect of the Other Criteria may be better served by having a rapid response capability and ongoing communication and outreach activities rather than building it into a planning tool. 'Multiple stakeholder groups actively setting standards' is important to include in the tool.

c. Are there additional criteria that should be considered? If so, please describe additional criteria that might be considered and why.

Public Health Criteria: I recommend a criterion be added to indicate whether a chemical is part of a family of chemicals of similar toxicity. Approaches to address chemical mixtures, cumulative risk and cumulative impact continue to be developed and applied in many decision contexts (NASEM 2025; US EPA 2023).

References:

NASEM 2025 State of the Science and the Future of Cumulative Impact Assessment https://www.nationalacademies.org/our-work/state-of-the-science-and-the-future-of-cumulative-impact-assessment

U.S. EPA 2023. Draft Proposed Principles of Cumulative Risk Assessment under the Toxic Substances Control Act https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/cumulative-risk-assessment-under-toxic-substances

Aggregate exposure: Consider adding an aggregate exposure criterion.

Other Decision Criteria: Another criterion that could be considered is "value" to the food or the food supply. This could be defined in different ways, e.g., is the chemical essential to food production or to shelf stability, etc. A recent work by Senerth et al (2025) has summarized many evidence to decision-making frameworks within environmental and occupational health; economic value and societal benefit are common components of many of those frameworks.

Reference:

Senerth E, Whaley P, Akl E, Beverly B, Alonso-Coello P, Rooney A, Schünemann HJ, Thayer KA, Tsaioun K, Morgan RL. GRADE guidance 40: The

GRADE evidence-to-decision framework for environmental and occupational health. Environ Int. 2025 Mar;197:109314. doi: 10.1016/j.envint.2025.109314.

3. The draft scoring definitions for all criteria were developed to consider the expected variability in the types and extent of data available for the wide variety of food chemicals that may be considered for review.

a. Given this context, are the scoring definitions for the Public Health criteria appropriate for the purpose of the tool?

The score numbers have no quantitative meaning but indicate some spectrum of potential concern from low to high. Greater spacing between the score levels will create greater variability in the combined score which should help distinguish the ranked chemicals. A 3-tiered item could have scores of 1, 6, 12; a 4-tiered item could have scores of 1, 4, 8, 12; and a 5-tiered item could be scored as 1, 3, 6, 9, 12.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why?

New Scientific Information: Suggest grouping potential low impact information with no new information or expand the scoring on this item so uncertain impact can have its own score. E.g., 1 – no new information, 3 – low impact, 6 uncertain/unknown impact, 9 – moderate impact, 12 – high impact). I would recommend putting somewhat greater weight on uncertainty.

ii. The toxicity criterion described in Section 3.1.1 considers data for seven different toxicity data types and the score assigned reflects the highest toxicity data type score from the toxicity rubric, which is described in Appendix A Table A1. Is this the most appropriate strategy for assigning a toxicity criterion score? If not, please explain your reasoning and provide alternatives for FDA to consider. Please be specific and provide references, as appropriate.

This approach is appropriate.

b. Are the scoring definitions for the Other Decisional criteria appropriate for the purpose of the tool?

Consider expanding the scoring range to 1 – 12, as described under the response to question 3a.

i. Are the definitions appropriately defined? If not, please describe changes that might be considered and why.

External stakeholder attention: 3.2.1: Re-define score 3 to conflicting. Following the common definition of uncertainty in risk assessment as something unknown, conflicting activity is not unknown.

ii. FDA is exploring quantitative and qualitative methods to help inform the scoring of the 'building public confidence' criterion (Section 3.2.3) such as conducting public sentiment analysis (e.g., utilizing natural language processing). How might such tools or the information they provide be incorporated into this criterion? What additional strategies and metrics could FDA consider?

Building public confidence: How will this be measured?

Building or maintaining public confidence is a valid organization-level risk concern but not necessarily appropriate to build into an evidence-based prioritization tool that is deployed once a year for internal planning. Building/maintaining public confidence can accrue over time from applying the tool within the "systematic post-market assessment process" with an added evaluation step, as described below (question 6). Assessing public confidence should be an ongoing or at least periodic activity.

Re: public sentiment analysis:
Typical agency outreach (hearings, public comment periods, etc.) primarily gets input from interested parties and less from the general public. So, a better understanding of public sentiment would be valuable. But I wonder whether there would be public sentiment data on many individual ingredients or additives beyond those already widely known.

Public sentiment data as might be gathered through applying natural language processing on social media posts was flagged by some public commentors as easily "gamed" through use of technology or social media campaigns. I share this concern. Information from a 'gamed' public sentiment analysis could bias the priority scoring approach.

4. The prioritization methodology includes weighting factors.

a. FDA is considering equal weighting among the Public Health criteria and (separately), among the Other Decisional criteria for the Post-market Assessment Prioritization Tool.

In the overview the tool description states, "The Post-market Assessment Prioritization Tool **focuses on potential risk to public health** (risk ranking) and also includes other decisional considerations, using a Multi-Criteria Decision Analysis (MCDA) method." (Bold text added for emphasis). If the tool serves HFPs public health mission, I recommend giving greater weight to the Public Health criteria.

i. Should different weights be applied to the Public Health criteria when determining the Total Public Health Criteria Score? If so, please specify the weighting scheme that might be considered and why.

Equal weighting is appropriate.

ii. Should different weights be applied to the Other Decisional Criteria when determining the Total Other Decisional Criteria Score? If so, please specify the weighting scheme that might be considered and why.

Equal weighting is appropriate.

b. FDA is considering equal weighting among the Total Public Health Criteria Score and the Total Other Decisional Criteria Score to determine the overall Post-market Assessment Prioritization Score.

I recommend greater weight be given to the public health score when calculating the overall score.

i. Should different weights be applied when determining the overall Post-market Assessment Prioritization Score? If so, please specify the weighing scheme that might be considered and explain why it would be more appropriate than equal weighting.

A greater weight on public health reflects the focus on potential health risk, as described in the documentation (see response to 4a above). There are alternative ways to do that:

Calculate the total score directly rather than separating the two sub-criteria. There are four public health criteria and three 'other' so overall public health makes a greater contribution to the total score.

Or give the public health criteria score a two-thirds weight and the other decision criteria score a one-third weight.

5. The draft toxicity rubric uses traditional toxicity data (in vivo, as well as limited in vitro such as for genotoxicity), human health outcomes (e.g., adverse event reports), and epidemiological data for determination of the toxicity criterion score within the Public Health criteria. Considering that the prioritization process is not a comprehensive review, please address the following questions.

a. How might FDA incorporate information from new approach methodologies (NAMs) into the toxicity rubric?

I defer to other's expertise in this area.

i. Are there specific NAMs (e.g., systems biology, engineered tissues, artificial intelligence, in vitro, microphysiological systems, or other alternative data or modeling tools) that would be most appropriate for use in the toxicity rubric? If so, please explain which NAM(s) would be most appropriate and why.

I defer to other's expertise in this area.

ii. Given that a single NAM is not expected to be a one-to-one replacement for a traditional in vivo toxicity test, how can the strengths and limitations of each NAM be appropriately considered if it is incorporated into the toxicity rubric?

I defer to other's expertise in this area.

b. Threshold of Toxicological Concern (TTC) approaches can be used to assess the toxicity of chemicals that lack sufficient safety data and have low dietary exposures. Although the Cramer classification scheme has historically been used in TTC approaches, FDA has recently developed the Expanded Decision Tree (EDT) that assigns chemicals to one of six EDT classes. How might such tools or the information they provide be incorporated into the toxicity rubric?

If there is inadequate data to score the toxicity criterion, EDT could be substituted. A 6-tiered scoring system could be used: I=1; II=2; III=4; IV=6; V=8; VI = 10.

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

The document describes a further round of pilot testing and review. I recommend that this be done within a public- and stakeholder-engaged process. Given the many constructive public comments there seems to be a large audience ready to engage.

The "systematic post-market assessment process" can be enhanced by adding an evaluation step following risk management action. The evaluation approach(es) may already be in place through existing post-market surveillance.