Reviewer’s report

Title: EFSA’s toxicological assessment of aspartame: was it even-handedly trying to identify possible unreliable positives and unreliable negatives?

Version: 0 Date: 07 Mar 2019

Reviewer: Lisa Lefferts

Reviewer’s report:

Review of "EFSA's toxicological assessment of aspartame: was it even-handedly trying to identify possible unreliable positives and unreliable negatives?" by Erik Paul Millstone & Elisabeth Dawson

This is an important, fascinating, and well-written paper that provides valuable insights into the safety of aspartame, and the evaluation practices of EFSA's Panel on Food Additives and Nutrient Sources Added to Food. It highlights critical flaws in EFSA's evaluation of aspartame specifically and raises questions on the transparency and conduct of reviews by EFSA's Additive Panel more generally. The results should stimulate a re-examination of the evidence on aspartame and reforms to the operations of EFSA.

The only other major comment of this reviewer is that it is theoretically possible that a well-conducted and unbiased review of a substance could objectively determine that studies that indicated possible toxicity were unreliable, and those that indicated safety were reliable (or vice-versa). In other words, a dissimilar proportion deemed reliable and unreliable when comparing studies with prima facie evidence of harm and those providing no such evidence is not de facto an indication of bias. For example, there could be a systematic flaw in studies of a certain type. The abstract only cites these proportions (i.e., 62 deemed reliable out of 81 studies that did not indicate any possible harm, vs. zero deemed reliable out of 73 studies indicating possible harm) as evidence for the conclusion that the EFSA review of aspartame was biased ("more alert to putative false positives than to possible false negatives"). More emphasis should be placed on more convincing evidence discussed elsewhere in the paper, for example including the practice of giving more weight to weak negative studies (e.g., that fail to meet generally accepted criteria for numbers of animals/dose group) compared to strong positive studies (e.g., with large numbers of animals/dose group), and the application of other, laxer criteria for finding negative studies reliable, compared to stricter criteria for finding positive studies reliable. It is the criteria used, and the application of the criteria, rather than the findings, which provide more compelling evidence of an uneven-handed approach.

I also offer the following minor comments, intended to further clarify/strengthen the paper.

Minor comments
Citations and explanation for statements are sometimes not provided initially, and only appear later in the manuscript. Although this is a matter of style, it would be helpful to the reader to either provide those cites/explanation initially or indicate that they are forthcoming.

For example, at the bottom of page 5 in Section 1, it is stated that evidence indicated that 15 toxicology tests conducted by Searle were improperly and incompetently conducted and that the reports of those tests were misleading, yet no cite or explanation of the basis for that statement is provided until page 8. Similarly, it is noted on page 7 that the UAREP was not provided with evidence of the incompetence of some of the laboratory staff or the fictional aspects of some of the documentation, and the reader has also not yet been provided with this evidence; it would be helpful to do so earlier, to put this in better context.

As another example, the authors' opinion of FDA's actions is given at the top of page 6 before the basis for that opinion is provided.

P. 6 first complete paragraph: is the implication that it would have been better for the 15 problematic studies to be reviewed all together? If so, explain why. If not, remove the word "but" in the phrase "...but in two separate sets and by different institutions." (It seems the main problem was the failure to provide evidence of improper/incompetent conduct, not whether the work was divided up).

P. 7: it is stated that a study was omitted from this analysis since it was about DKP, a breakdown product of aspartame, yet discussion of the DKP study appears on p. 8. Clarify.

It is stated that the FDA Bureau of Food Task Force reviewed 3 studies: E5, E-89, and E-77/78, and that Dr. Jacqueline Verrett found the three to be "...woefully inadequate." It would be helpful to clarify in the text (e.g., on p. 9 following Dr. Verrett's quote) what the EFSA Panel's conclusions were re: these three studies (I believe it found E5 unreliably positive, E-89 reliably negative, and E-77/78 reliably negative) and what types of studies they were (I believe the first two were reproductive/developmental toxicity studies and the last is already identified as a DKP toxicity study).

On page 7 it states the FDA Bureau of Food Task Force acknowledged that the 3 studies had not been properly conducted and that there were differences between raw data and the summaries of the data submitted in the petition to FDA, but that the differences were, in effect, minor/inconsequential ("would not significantly alter the conclusions of the studies.") Then on page 8 it states that one member of that Task Force (Dr. Verrett) found them "...woefully inadequate" and (on p. 9) that it was "unthinkable that any reputable toxicologist giving a completely objective evaluation of this data resulting from such a study could conclude anything other than that the study was uninterpretable and worthless and should be repeated." Please clarify these two seemingly contradicting conclusions. Did Dr. Verrett hold a minority opinion on the Task Force? Or had her views expressed in the quote changed from when the Task Force had done its review
based on new information regarding the conduct of the studies that was not available at the time of the Task Force review?

- Somewhere, also clarify whether the inadequacies described by Verrett and/or Gross only applied to the three studies reviewed by the FDA Bureau of Food Task Force, or whether they also applied (or were likely to apply) to additional studies as well (e.g., the other 12 studies)

- P. 11: Merrill is mentioned in the last paragraph but has not yet been introduced. Insert "Richard Merrill" after "FDA's Chief Counsel" in previous paragraph. Also, would be interesting to identify the two pharmaceutical products (perhaps in a footnote).

- P. 11 last paragraph: please provide cite(s). Astounding!

- P. 12 first paragraph: clarify if this was a public request for information or a private request from the secretariat. I assume this is what is described on p. 17 as a call for "interested parties" to provide "all necessary data." Could avoid repetition and clarify by just describing once.

- P. 14: as I understand it the human data was only correlational, so very weak. I think it is overstating it to describe this as "a full house".

- P. 15 first line: the typical duration for carcinogenicity studies in rodents is 104 consecutive weeks (two years according to FDA's Redbook (https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm078388.htm#ftnIV) and the US NTP (https://ntp.niehs.nih.gov/ntp/test_info/finalntp_toxcarspecsjan2011.pdf). OECD (http://www.oecd.org/env/test-no-451-carcinogenicity-studies-9789264071186-en.htm) also says the duration will normally be 24 months for rodents, although for specific strains of mice, duration of 18 months may be more appropriate. Also not clear whether "they" refers to mice, rats, or both. The last rat in the first Ramazzini study died at 159 weeks, and the mouse study ended at 130 weeks since <10% of the animals were still alive.

- P. 15 first para: replace "the ends of their average 'natural' lives" with "two years of age" for clarity

- P. 15 first para: suggest citing Gift et al (https://www.ncbi.nlm.nih.gov/pubmed/24045135): "The protocols characteristic of RI studies can cause interpretive challenges, but aspects of the RI design, including gestational exposure, life span observation, and larger numbers of animals and dose groups, may impart advantages that provide chemical risk assessors with valuable insights for the identification of chemical-related neoplasia not obtained from other bioassays."
- P. 15 first para: mention all 3 Ramazzini studies (first rat, mouse, second rat (included in utero exposure))
- P. 15 last para: delete "and specificity"
- P. 16 first complete para: could also cite Caldwell et al (https://www.ncbi.nlm.nih.gov/pubmed/18095346) who convincingly refutes the hypothesis that lymphomas/leukemias are induced by infection. For example, she notes that while respiratory infections frequently occur in old rats, and in most RI rat bioassays, leukemia/lymphoma are only reported in a few (8/112), and thus the link is unlikely.
- P. 17 last para: is this referring to the 15 studies mentioned previously? Clarify which studies.
- P. 21 (and elsewhere): isn't it likely that some data/results/conclusions could be considered reliable and others not? Would this explain why some studies are classified as C5?
- P. 23 (and elsewhere):
  o Explain how the NOAEL approach is not generally used for carcinogens
  o Clarify whether the EU ADI for aspartame was derived using a 100X safety factor
  o Explain that if aspartame is carcinogenic, as three independent studies in rodents conclude, then doses below the ADI pose a risk
- P. 32 last para: host-mediated mutagenicity test? (Clarify)
- P. 33 last para: almost certainly this was due to a lack of palatability at high doses
- P. 34 top: this is a third developmental toxicity study in rabbits?
- P. 34 bottom para: it would be helpful to briefly describe the specific numbered studies here and throughout the text (e.g., "In relation to E90 (a developmental toxicity study in rabbits) the panel explained ...."). Similarly, on p. 36. This provides important context and clarifies the text.
- P. 35 top: was there a statistically significant increase in abortions at the low dose? If not, then wouldn't it be reasonable to conclude there was no effect at this dose?
- P. 36 re: criterion for statistical significance: generally, a p-value < 0.05 is used to determine statistical significance. Same point on p. 45.
- P. 37 top: clarify whether this is maternal deaths or deaths of pups.
P. 39-40: first it says the panel reported than tumors were random with respect to dose and gender, and then later it implies that tumors affected one sex more than the other. If the latter, then they would not be random. Clarify. Also, were there statistically significant increases in cancers at some doses, but not in a monotonic dose-response manner? Clarify.

P. 40, para on E20: I can't tell if the criticism re: "equivocal" is one of semantics or something substantive. Perhaps look at the study and clarify what actually happened. There might have just been some minor fluctuations in consumption/weight without any discernable pattern. Ditto on p. 44 re: "limited;" rather than picking on word choice (they seem to be using the word "limited" to mean "small") make the point that a study with a small sample size is insensitive and yet effects were still noted and should not be discounted.

P. 46 criteria 5: qualify. "Virtually free" (?)

Soffriti et al studies in table and text: please expand comments. There is much more that could be said here. For example, IARC states "It is generally not appropriate to discount a tumour response that is significantly increased compared with concurrent controls by arguing that it falls within the range of historical controls …." Yet this is exactly what EFSA did. For example, the draft report stated "The ANS Panel … and EFSA … concluded that the hepatic and pulmonary tumour incidences …all fall within their own historical control ranges … Based on these data, the Panel concluded that the results … do not provide evidence for a carcinogenic effect of aspartame in mice." Also, EFSA only used historical control data to argue that tumor findings were not significant/should be disregarded, but never to argue that they were significant. Specifically, it never cited how extremely rare transitional cell carcinomas of the renal pelvis/ureter are in historical controls. In our comments to EFSA we cited a review of 17 studies using 2,669 control Sprague-Dawley rats that found them in only 1 male and 1 female, yet these tumors were found in 21/1500 aspartame-treated animals, and in 0 of concurrent controls. Furthermore, the draft report states "The ANS Panel noted that the only consistent [emphasis added] findings reported by the authors in the two rat studies were an increased incidence of lymphomas/leukemias…" But consistency should not have been expected since the two studies had different designs, and lack of consistency is not a reason to discount the results. For example, the first study was nearly four times larger and used a wider range of doses than the second and was thus more capable of detecting rare tumors (e.g., kidney). Also, the second study included in utero exposure. In utero exposure to carcinogens can produce different cancers than postnatal exposures (e.g., in utero exposure to DES increases risk of cervical/vaginal cancer, whereas exposure in adulthood increases risk of breast cancer).

Typographical/clerical errors

- P. 8. "…a particular animal" (not animals)
P. 9 3rd line from bottom: should "fissure" be "tissue"?

Check cite numbering, at some point in the manuscript I noticed that they appeared to be off (e.g., cite 38)

Be consistent about using dashes or not with E numbers (e.g., E5 or E-5).

P. 21 first line: replace "?" after "provided" with a period.

P. 33 first complete para: developmental toxicity, not development toxicity

Level of interest
Please indicate how interesting you found the manuscript:

An exceptional article

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

6. I provided testimony & comments to EFSA re: flaws in its draft assessment of aspartame. My organization rates most additives as “safe” and rates aspartame as “avoid,” primarily due to the
cancer evidence. I have exchanged information and views on aspartame with Dr. Millstone in this context.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

No