Summary of ORD comments on OPP's glyphosate cancer assessment
December 14, 2015

1. ORD scientists have reviewed OPP's glyphosate cancer analysis and selection of cancer descriptor. The reviewers included two epidemiologists, a pathologist, and several scientists with significant expertise in cancer risk assessment. With the exception of one reviewer who participated in the recent IARC review and two reviewers who participated in the CARC review, an in-depth review of the original literature was not undertaken.

2. The goal of this focused, expedited review was to consider the characterization of glyphosate as "not likely to be carcinogenic to humans," given IARC's recent decision and looking at the totality of the available cancer database.

3. There are several epidemiological studies that vary in quality and study design. For many of the epidemiological studies, it appears that the small sample sizes limit their power to detect an outcome other than the null hypothesis. There are some epidemiological studies that show non-statistically significant elevated risks. One meta-analysis brings together those studies to strengthen the analysis and finds slightly elevated risks. The overall conclusion from IARC is that there is limited evidence of an association between glyphosate and NHL. One major point is that a determination of causality is not what one would expect from most of the studies that are available given their design and power.

   ORD's epidemiologists agree with IARC that there is "limited evidence" of carcinogenicity in humans and understand IARC's definition of "limited evidence" as "a positive association has been observed" for which a causal association is "credible, but chance, bias, or confounding could not be ruled out with reasonable confidence [IARC Preamble, section B6]." OPP preferred to dichotomize the epidemiological evidence to be either "causal" or "not causal." This dichotomization appears to be the major factor in the different positions between OPP and IARC with regard to the epidemiological data.

   Frameworks for data analysis and causal determinations that are currently in use by EPA and the risk assessment community include gradations of causality. EPA's Cancer Guidelines utilizes these gradations to inform cancer descriptor choices. An example of situation where a less than causal determination is used is for the descriptor "likely to be carcinogenic to humans" – an agent demonstrating a plausible (but not causal) association between human exposure and cancer. The OPP draft risk assessment does not appear to follow these approaches. It would appear that OPP's use of a "yes/no" approach would only lead to cancer descriptors of "carcinogenic to humans" or "not likely to be carcinogenic to humans."

4. Glyphosate has been tested in a large number of 2-year rat and mice studies, including several studies conducted in the same strains. A wide range of tumors have been observed in these studies, including adenomas and some carcinomas. Tumors have been observed in thyroid, liver, skin, pancreas, hemangiosarcoma, lymph, testes, mammary glands, kidney and lung. However, the tumor incidences were generally not statistically significant in pair-wise comparisons and were generally within the range of historical controls. Most tumor types were only observed in one study despite repeat studies within the same strain and similar doses at or above the limit dose.

The tumors found in more than one study were in the pancreas and liver, and were observed in 2 of 4 studies in SD rats. A positive trend was found for male combined renal tubule adenomas and carcinomas in one CD-1 mouse study. This tumor is relatively rare in CD-1 mice. A positive trend was also found for hemangiosarcoma in males in another CD-1 mouse study. What makes the database so unusual is the large number of animal bioassays that have been conducted and the variety of types of tumors that have been observed, albeit usually at very low incidences. The OPP evaluation concluded that all of the tumors found were not treatment-related.

OPP (and EFSA) focus on pairwise comparisons (which were generally not significant), while IARC also uses trend tests (which yielded several significant results). In a few cases, OPP reported trend test results that differed from those of IARC but did not report which test they used. EPA's cancer guidelines state that "Trend tests and pairwise comparison tests are the recommended tests for determining whether chance, rather than a treatment-related effect, is a plausible explanation for an apparent increase in tumor incidence. Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result."

5. The ORD reviewers noted that the analysis of the cancer data in the assessment was basically conducted on a study-by-study basis instead of using a more inclusive, systematic approach to provide an integrated analysis of the data. An integrated approach could provide better qualitative and quantitative evaluation of animal studies showing tumor responses and those without tumors. It is difficult to predict whether such an approach would yield a different outcome. It would likely be a large undertaking. An integrated, systematic approach could also include an evaluation of mutagenicity information on glyphosate and glyphosate-containing formulations. A thorough evaluation of the mutagenic potential of glyphosate was not included in the assessment and was not conducted as a part of this review. This aspect of the assessment is important because if there is evidence of mutagenic potential or if a mutagenic potential has not been adequately ruled out, then characterization of glyphosate as "not likely to be carcinogenic" could be problematic for this reason alone, given the lack of a high-quality negative epidemiological study.

6. The main issue is whether the characterization of cancer potential for glyphosate as "not likely to be carcinogenic to humans" represents the best evaluation of the data. There are five EPA cancer guideline categories:
   - Carcinogenic to humans
   - Likely to be carcinogenic to humans
   - Suggestive evidence of carcinogenic potential
   - Inadequate information to assess carcinogenic potential
   - Not likely to be carcinogenic to humans

   According to the cancer guidelines, characterizing a chemical as "carcinogenic to humans" or "not likely to be carcinogenic to humans" has a high bar with phrases such as "strong evidence" and "robust data" included in these descriptors. For glyphosate, nobody—including IARC, using their similar system—supports the top category (carcinogenic to humans). The descriptor "not likely to be carcinogenic to humans" is appropriate when "the available data are considered robust for deciding that there is no basis for human hazard concern." Examples include situations where there is "convincing evidence in both humans and animals that the agent is not carcinogenic" or animal evidence is available that "demonstrates a lack of carcinogenic effects in both sexes in well-designed and well-conducted studies in at least two appropriate animal

species (in the absence of other animal or human data suggesting a potential for cancer effects)."

"Likely to be carcinogenic" means that the "weight of the evidence is adequate to demonstrate carcinogenic potential to humans," giving as an example "an agent demonstrating a plausible (but not definitively causal) association between human exposure and cancer, in most cases with some supporting biological, experimental evidence, though not necessarily carcinogenicity data from animal experiments."

"Suggestive" evidence covers a spectrum of evidence ranging from "a positive cancer result in the only study on an agent to a single positive result in an extensive database that includes negative studies in other species." In ORD's experience, chemicals can fall into this category at the low end or the high end of the spectrum.

The descriptor "inadequate information to assess carcinogenic potential" is appropriate when "available data are judged inadequate for the other descriptors," and for which "additional studies would be expected to provide further insights." However, examples for when to use this descriptor range significantly from "little or no pertinent information," conflicting evidence (not to be confused with differing results, where "depending on the WOE, differing results can be considered either suggestive evidence or likely evidence)," to "negative results that are not sufficiently robust for not likely."

The ORD reviewers have not extensively discussed which descriptor might be most appropriate for glyphosate. In discussions to date, it appears that all of the descriptors have some scientific merit, with the exception of "carcinogenic to humans." For example, one might classify glyphosate as "likely" on the basis of experimental data alone, by accepting positive trend tests at two anatomical sites (despite differing results in other studies) or by viewing these tumors (which not everyone accepts) as rare. One level down on the continuum puts you at "suggestive evidence." For this descriptor, one could argue that the evidence is not strong enough for the "likely" descriptor but it cannot be dismissed. The positive association (i.e., limited evidence) of carcinogenicity in humans could arguably rule out the last cancer category ("not likely to be carcinogenic"). One could also argue that this unusual data set is best suited to the descriptor "inadequate information to assess carcinogenic potential" based on an argument that the results are not sufficiently robust for the descriptor "not likely."

ORD Recommendation: To strengthen OPP's human health assessment and address the differences in the potential cancer findings, we recommend the following:

- Expand the discussion of the cancer data and subsequent findings to include a detailed and thorough discussion of the issues that caused OPP to come to a different conclusion than IARC, if not directly noting the IARC findings themselves. Key controversies in how one could evaluate the data should be highlighted to provide transparency in how the Agency is making its determination. OPP could include a discussion of the strengths and weaknesses of choosing one cancer descriptor over the other.
- We understand that OPP plans to take the assessment to the SAP for external peer review. We recommend developing charge questions that will be specific to the cancer findings and ask the panel to address the specific scientific differences that exist between the IARC and OPP cancer determinations. ORD is willing to work with OPP to draft the charge questions, or review them

before they are finalized.